Paper presented at the 2006 Conference of AIRUM, Bloomington, MN November 2-3

Modeling the Incidence and Timing of Student Attrition: A Survival Analysis Approach to Retention Analysis

Peter M. Radcliffe
College of Liberal Arts
University of Minnesota
Minneapolis, MN 55455
radcl002@umn.edu

Ronald L. Huesman, Jr.
Office of Institutional Research
University of Minnesota
Minneapolis, MN 55455
huesm003@umn.edu

John P. Kellogg
Office of Institutional Research
University of Minnesota
Minneapolis, MN 55455
j-kell@umn.edu

Abstract - The goal of the study was to develop a practical application to help a large doctoral research extensive public university promote student success by identifying at-risk students. A probability model using logit and a longitudinal model using survival analysis were used to identify factors that impact a student's ability to persist and graduate.

1 Introduction

Student retention/graduation has become more important then ever to institutions in terms of accountability and rankings and recent national initiatives have focused on collegiate academic outcomes. It is clear from the final report from the Commission on the Future of Higher Education that accountability of institutions of higher education will be measured, in part, by graduation rates (Field, 2006). The Department of Education envisions a student record level national data base to track student's academic progress. If this system is implemented, this will invariably result in a better assessment of the nation's 6-year graduation rates at four-year institutions, which given current estimates of 56.4% (Knapp, Kelly-Reid, & Whitmore, 2006), are most certainly underestimating the true rate given the current methodology for calculating graduation rates (Adelman, 1999). Regardless of what may happen with the national tracking system and what is the "true" graduation rate, the importance of having a college degree is becoming more important and critical then ever. A current report by the Census Bureau highlights the economic advantages of having a college degree, the average earnings in 2004 for adults age 18 and over with a baccalaureate degree was $51,554 vs. $28,645 for those with only a high school diploma (U.S Census Bureau News, Oct. 26, 2006). College enrollment levels are at all time highs and education levels have reached an all-time high, with 28% of Americans aged 25 or older having completed at least a bachelors degree vs. 23% in 1995 (Snyder, Tan & Hoffman, 2006). But, it may still not be enough, "Nearly two-thirds of all high-growth, high-wage jobs created in the next decade will require a college degree" (Field, 2006, p.A23). It is imperative that each institution identify and attempt to understand the particular characteristics of its students and their progress to degree attainment within an institutional context. As part of an overriding strategic positioning process at the University of Minnesota-Twin Cities, student retention and graduation rates were identified as one of the metrics to be used to evaluate the institution's progress toward its strategic goal of becoming one of the top three public research universities in the world. The six-year graduation goal for students who start at the University in Fall 2008 has been set at 80%. Currently the University six-year graduation rates are about 19 percentage points short of the goal. This is an ambitious goal and will require the full commitment of all parts of the campus. The goals of the study were to identify student characteristics that are associated with an increased likelihood of dropout and gain information on when during a student's career they are most likely to depart. Questions to be answered by the study:

What characteristics of students help predict success or departure?
When is a student at greatest risk of not succeeding?
Does the profile of risk differ across subgroups (e.g. race, gender, etc.)?

2 Overview of the Research

There are a number of student retention models beyond Tinto's (1975) student integration model. But, Tinto was the first to lay out a detailed longitudinal model of student persistence that described the interconnections between the student and institution to the path of student success (i.e., graduation) over time. In a recent review of the state of student retention research and its potential future directions, Tinto argues what really is important is, "… not our theories per se, but how they help institutions address pressing practical issues of persistence." (Tinto, 2006, p. 6). It is widely understood that student's background/demographics and incoming academic ability (i.e., pre college measures) are important predictors of a student's ability to persist to graduation (Perkhounkova, Noble & McLaughlin, 2006; Ishitani & Snider, 2006; Ishitani, 2003; Tinto, 1975). The stated goal of much of the student retention literature is to identify at-risk students, using demographics and pre-college measures of academic ability (e.g. ACT/SAT scores) and measures of student/institutional fit. Generally speaking, Ishitani (2003) showed that pre-college attributes do vary over time, but what is not well understood is how these factors effect persistence over time (Ishitani & Snider, 2006). Previous retention research on student athletes at this institution found that not only background and pre-college characteristics were important, but also the "academic fit", as measured by first-term academic progress was a significant predictor of academic success and the effect of these variables also varied across time (Radcliffe, Huesman & Kellogg, 2006). Other factors, such as external support, job demands, social fit measures, etc. have been shown to be important, but often require special data sets and often reduce sample sizes due to missing data. Our goal was to select predictors that could be easily obtained prior to or early in the academic career of a student, have a high percentage of available data, and that were known to be accurate and reliable measures. Survival analysis is a regression method that allows us to identify important factors related to student graduation and allows us to model their impact over time. The use of a survival analysis approach to examine the longitudinal effect of these factors over time on student retention (Ishitani & Snider, 2006; Ishitani, 2003; DesJardins, Ahlburg, and McCall, 2002; DesJardins, McCall, Ahlburg, & Moye, 2002; DesJardins, & Moye, 2000; Murtaugh, Burns & Schuster, 1999) and student stopout behavior (Ishitani & DesJardins, 2002; DesJardins, Ahlburg & McCall, 1994; Ronco, 1994; Willett & Singer, 1991) is now well established.

3 Data

The data sample consisted of 9,890 students at the University of Minnesota-Twin Cities, a large, Midwestern Carnegie Doctoral-extensive university who entered as first-time full-time degree-seeking freshmen during the 1999 and 2000 fall semesters. The outcome observed was whether or not the student had graduated by the sixth year. The variable used to account for duration of time for the event under analysis was total credits completed at the university. Credits successfully completed corresponds more closely to progress toward a degree than simply the number of terms attended, and also allows for identifying important transitions in a student's career, such as major declaration and the shift from lower to upper division coursework. The model developed focused on: student background characteristics, demographics, financial need and end of first term performance indicators. It is not in particular, theory driven, though an attempt to use and operationalize readily available data from the institution's student record system loosely follows Tinto's model of student persistence.

A total of n=9,580 students composed the analysis sample (i.e., had complete data) or 96.8% of the original sample. As table 1 illustrates, a number of first semester academic progress variables were included in the development of the model. The ratio of credit hours earned toward a degree to credit hours attempted for the first semester (sans course withdrawals and remedial courses) was calculated as one measure of academic progress. The number of C's and D's earned during the first semester was also collected to assess academic progress, since these grades do count toward the progress to a degree but indicate borderline/marginal performance. The number of first-semester course withdrawals was also collected to augment academic progress. In addition, two indicator variables were developed to assess academic difficulty during the first semester; 1) did a student take a mathematics remedial course and 2) did they pass or fail this course. The sample was composed of 47.2% male students, 0.77% American Indian, 3.95% Black, 1.76% Hispanic, 10% Asian, and 0.8% international students. The reference group consisted of white and unknown race/ethnicity students (82.67%). In-state students comprise the reference group for tuition residency at 68.5% of the cohort, the remaining groups were from states with tuition reciprocity agreements (25.7%), and non-reciprocity states (5%). Only 3.5% of the students had student-athlete status during some point in their academic career. About 17.6% of the students were PELL eligible, this measure was used as a proxy measure of socioeconomic status. Approximately one-fourth of students lived off-campus their first year. The average ACT/SAT converted score for these two cohorts was 24.7. Approximately 98% of all students who enter the University of Minnesota have an ACT and/or SAT score. Of the 9,580 students in the analysis group, 62% had graduated within six years of entrance to the University. The proportion who graduated in the same amount of time in the original data sample was 60.9%, slightly lower but very similar to the analysis sample.

Table 1. Descriptive Statistics of the Sample (N=9,580)

Variable	Values	Mean	SD	Variable Description (type of variable)
Graduate*	0-1	0.620	0.485	If not successful at the U (response variable)
Credits**	0-232	106.3	46.91	Total credits earned while at the U (timing variable)
Ratio	0-1	0.941	0.177	Ratio of first term credits earned to attempted
D’s & C’s earned	0-5	0.881	1.010	Number of D & C grades earned first semester
W’s earned	0-4	0.124	0.373	Number of W grades earned first semester
ACT/SAT Composite	10-36	24.67	4.202	ACT composite score
Remedial Taken	0-1	0.113	0.316	If math remedial course taken first semester (dummy)
Remedial Failed	0-1	0.014	0.012	If Failed remedial course first semester (dummy)
Athlete	0-1	0.036	0.185	If Athlete (dummy)
Male	0-1	0.472	0.499	If Male (dummy)
American Indian	0-1	0.008	0.088	If American Indian (dummy)
Asian	0-1	0.101	0.301	If Asian (dummy)
Black	0-1	0.039	0.195	If Black (dummy)
Hispanic	0-1	0.018	0.132	If Hispanic (dummy)
International	0-1	0.008	0.087	If International student (dummy)
Reciprocity	0-1	0.257	0.437	If Tuition reciprocity state (dummy)
Non-Reciprocity	0-1	0.050	0.217	If Non-tuition reciprocity state (dummy)
Pell	0-1	0.175	0.381	If Pell grant eligible (dummy)
Off-campus Housing	0-1	0.250	0.433	If not living on-campus first term (dummy)

* Response variable

** Timing variable

4 Methodology

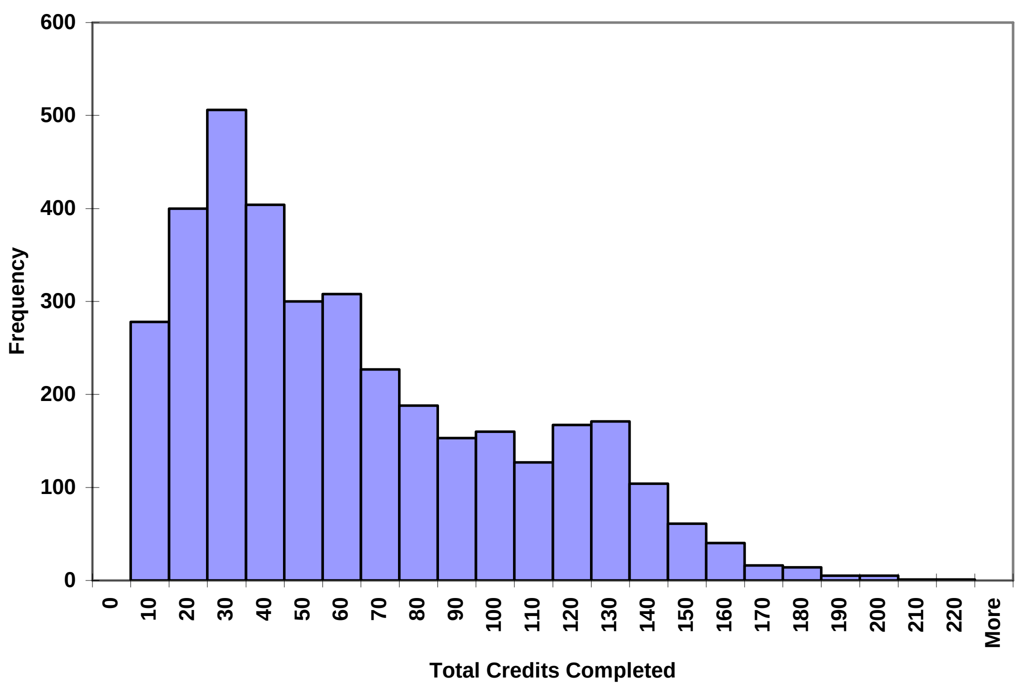
In attempting to improve student persistence and success, it is important to understand both whether a student will drop out, and when (Willet and Singer, 1991). A student’s experience in college is a dynamic process, involving the interaction of the student’s background and internal motivations with the institutional environment (Tinto, 1975). Understanding the points in that process at which student attrition takes place, and how those points differ between students, can help an institution to make personalized and effective interventions to improve student success.

Since our measure of student success, graduation within six years of initial admission, is inherently dichotomous, ordinary least squares regression techniques are inappropriate and would produce estimates with biased standard errors (Maddala, 1983). While there are a number of techniques that can be used for analysis of dichotomous dependent variables, logistic regression or “logit” models are the most widely used. The logit model estimates a latent variable representing the probability of a successful outcome (the non-zero value on the dependent variable) using a logistic distribution (StataCorp, 2005). The vector of coefficients and independent variables x_jB enters the equation through an exponential function, producing predicted probabilities that range from zero to one.

$$P\left(y_j \neq 0\right) = \frac{\exp(x_j \beta)}{1 + \exp\left(x_j \beta\right)}$$

In addition to understanding the incidence of graduation or failure to graduate, we are also interested in the factors that influence when during a student’s career they will depart the institution. To account for progress toward a degree for students despite their individual choices about semester credit loads and the use of summer enrollment, we assess the duration of enrollment by the total number of credit hours completed. Since this is a fundamentally time-based process, it is important to use a technique that accounts for the temporal patterns of enrollment. Survival models, also known as failure time models, duration models, or event history models, are ideally suited to the study of time-dependent processes (Box-Steffensmeier and Jones, 2004).

Figure 1. Histogram of Total Credits Completed by Non-Graduates



There are again a wide variety of different models that can be estimated. Discrete-time models have been extensively used in higher education research (see for example Desjardins, McCall, Ahlburg, & Moye 2002, Ishitani & Snider, 2006). These models are well-suited to term-based analyses of student retention. They are not, however, as appropriate for the analysis of continuous-time data, such as that used here. Parametric survival models allow for a representation of the underlying time-dependence of the process under investigation, as well as the influence of independent variables on the form of that dependence. In particular, this study uses a class of parametric survival models known as accelerated failure time models, where the independent variables influence the pace at which each individual experiences the underlying process, in this case either speeding up or slowing down a student's passage through the highest-risk period of their enrollment.

A variety of density functions can be used to model the time dependence of modeled process. A particular density can be selected either on a theoretical basis or by comparing models estimated with different distributions. Given that there is little history for parametric survival models in student persistence research, the most reasonable approach would appear to be comparing models. One method for comparing models is the Akaike information criterion (AIC) (StataCorp, 2005). Using this approach, which looks for the smallest AIC value, the generalized gamma distribution provides the preferred model (see Table 2). The gamma model can also be directly compared to the exponential, Weibull, and log-normal models as these distributions are special cases of the generalized gamma distribution. When the parameter κ is equal to zero, the gamma distribution simplifies to the log-normal. Where the parameter κ is equal to one, the gamma distribution simplifies to the Weibull, and when in addition the parameter s is equal to one, the Weibull (and therefore the generalized gamma) distribution simplifies to the exponential. Testing these parameter values, κ appears to be statistically significantly different from both zero and one, and σ is statistically distinguishable from one. This, therefore, would reject using the exponential, Weibull, or log-normal density in place of the generalized gamma in the estimation of the survival model for this data. Based on these diagnostics, the generalized gamma distribution appeared to be the most appropriate for this model.

Table 2. Akaike Information Criterion Measures for Alternative Survival Model Density Functions

Model Density	AIC Value
Log-Normal	11,292.04
Log-Logistic	10,279.51
Exponential	10,253.74
Weibull	9958.79
Generalized Gamma	9880.65

Survival models can be understood through the relationship between the instantaneous probability of a case failing given that it has not yet failed, known as the hazard rate, h(t), the statistical distribution specifying the general shape of the time-dependence, known as the density, f(t), and the proportion of all cases that have not yet failed at a given point in time, known as the survival function, S(t). These three components are mathematically related through the following generalized formula (Box-Steffensmeier & Jones, 2005):

$$h(t) = \frac{f(t)}{S(t)}$$

For the generalized gamma distribution, the form of the survival function is dependent on the value of the shape parameter κ. Where κ is positive, as it is in this analysis, the survival function takes on the following form, where I(*a*,*x*) represents the incomplete gamma function (StataCorp, 2005):

$$S(t) = 1 - I\left[|\kappa|^{-2}, \left\{|\kappa|^{-2} \exp\left(|\kappa| \cdot \frac{sign(\kappa)\{\ln(t_j) - x_j\beta\}}{\sigma}\right)\right\}\right]$$

This function can be evaluated at any value of *t*, representing total credits completed, and any set of observations and coefficients x_jB, to estimate the proportion of students with those characteristics who remain enrolled at that point.

5 Results

The logit model correctly identifies 71.8% of the freshmen studied as either graduates or non-graduates, performing better on the more common group of graduates. The model correctly identifies 87.5% of those who graduated, while only correctly identifying 46.1% of those who do not graduate in six years. The latter group is more difficult to isolate not only because it is less common, and therefore there is less information available for identification, but also because a number of those who do not earn their degree within the six-year window chosen for this analysis will, in fact, graduate in later years. Nevertheless, identifying nearly half of those students who will not graduate is of great practical utility for the targeting of outreach efforts. Also, since 69.3% of those predicted not to graduate were actual non-graduates, most of the resources devoted to an outreach effort would reach students in need of such services.

Table 3. Accuracy of Logit Model

		Actual		
Predicted		Graduated	Departed	Total
Graduated	Number	5,202	1,961	7,163
	Row %	72.6%	27.4%	100%
	Col %	87.5%	53.9%	74.8%
	Cell %	54.3%	20.5%	
Departed	Number	742	1,675	2,417
	Row %	30.7%	69.3%	100%
	Col %	12.5%	46.1%	25.2%
	Cell %	7.7%	17.5%	
Total	Number	5,944	3,636	9,580
	Row %	62.0%	38.0%	100%
	Col %	100%	100%	100%

Measures of first-term academic performance are among the most powerful predictors of graduation. All six measures of academic preparation or first-term performance are statistically significant at least the $p < .0001$ level. In addition, students who are from other states, whether or not they have a tuition reciprocity agreement with Minnesota, are less likely to graduate within six years. Students who live off-campus their first term are also less likely to graduate. Looking at demographics, most of the modeled factors are not statistically significant. However, student athletes were more likely to graduate than non-athletes, while male students, American Indian students, and students who were eligible for Pell grants were less likely to graduate within the six year window studied.

.

Table 4. Logit Model Parameter Estimates of Probability of Graduation

Logit (graduation)	Coef.	Std Error	z	Sig.	P>|z|
Ratio	3.3608	0.1844	18.23	***	0.0000
C’s & D’s earned	-0.4031	0.0238	-16.93	***	0.0000
W’s earned	-0.8338	0.0678	-12.29	***	0.0000
ACT/SAT Score	0.0536	0.0069	7.75	***	0.0000
Remedial Taken	-0.9846	0.0832	-11.83	***	0.0000
Remedial Failed	-1.3501	0.3844	-3.51	***	0.0000
Athlete	0.3468	0.1366	2.54	*	0.0110
Male	-0.1375	0.0484	-2.84	**	0.0040
American Indian	-0.7592	0.2972	-2.55	*	0.0110
Asian	0.1480	0.0849	1.74		0.0810
Black	0.1144	0.1300	0.88		0.3790
Hispanic	-0.0009	0.1822	-0.00		0.9960
International	0.3113	0.2884	1.08		0.2810
Reciprocity	-0.3740	0.0572	-6.54	***	0.0000
Non-Reciprocity	-0.2976	0.1141	-2.61	**	0.0090
Pell	-0.3701	0.0643	-5.76	***	0.0000
Off-Campus Housing	-0.5114	0.0588	-8.70	***	0.0000
Constant	-3.0417	0.2582	-11.78	***	0.0000

* = p < . 05   ** = p < .01   *** = p < .0001

Log-likelihood = -5339.294   p(chi-square) < .0001

Since the impact of any one variable in a logit model on the likelihood of graduation is conditional on the values of the other variables, one way to explore the practical significance of these variables is by establishing a “typical” or baseline case, and then observe what happens to the predicted probability of graduation when one of the variables in that case is changed to an alternative value. To construct this baseline case, each of the dichotomous variables in the model was set at the zero value, while the ratio of credits completed the first term was set at one, and the ACT/SAT score was set at its mean, rounded to the nearest whole number. Each of the possible alternative cases was then constructed by holding all the other variables at their baseline value, and changing each dichotomous variable individually from zero to one, the ratio variable from one to 0.8, representing failure to complete one of five courses taken, and the ACT/SAT score was lowered by one standard deviation, again rounded to the nearest whole number. Under each alternative specification, the predicted probability of graduation was calculated, and compared to the value from the baseline case. The results appear in table 5.

The two largest impacts in the model are from taking and from failing a remedial mathematics course. Enrolling in a remedial mathematics course during the first term lowers the predicted probability of graduation by eighteen percentage points, to 66%. Since a student can only fail a course if they enroll in it, the true impact of failing a remedial mathematics course must be estimated by setting both of these variables simultaneously to a value of one. Doing so produces a predicted graduation probability of 34%, a drop of a full fifty percentage points from the baseline prediction. Clearly, successful preparation for college-level mathematics and performance in remedial mathematics coursework is critical for student success at the University of Minnesota.

Table 5. Predicted Graduation Rates for Alternative Values of Each Variable Holding All Other Variables at Baseline Values

Logit (graduation)	Baseline	Alternative	Graduation	Change
Ratio	1.0	0.8	73%	-11%
C's & D's earned	0	1	78%	-6%
W's earned	0	1	70%	-14%
ACT/SAT Score	25	21	81%	-3%
Remedial Taken	0	1	66%	-18%
Remedial Failed[1]	0	1	34%	-50%
Athlete	0	1	88%	4%
Male	0	1	82%	-2%
American Indian	0	1	71%	-13%
Asian	0	1	86%	2%
Black	0	1	86%	1%
Hispanic	0	1	84%	0%
International	0	1	88%	4%
Reciprocity	0	1	78%	-6%
Non-Reciprocity	0	1	80%	-4%
Pell	0	1	78%	-6%
Off-Campus Housing	0	1	76%	-8%
		Baseline:	**84%**	**0%**
		Averages:	**63%**	**0%**
		Actual:	**62%**	**0%**

[1] Alternative for Remedial Failed also includes Remedial Taken set to 1.

Two other academic performance measures have large impacts. Withdrawing from a course after the second week of the term lowers the predicted probability of graduation within six years by fourteen percentage points, to 70%. In addition, failing to successfully complete a course (other than remedial mathematics) lowers the predicted probability of graduation by eleven percentage points, to 73%. These factors are also often cumulative, as a student who struggles in one course may well struggle in others. Earning a single C or D grade lowers the predicted probability of graduation by six percentage points, and in combination with other measures of first-term academic difficulties can contribute to a dramatic lowering of the expected odds of graduation.

A few other variables stand out as particularly powerful predictors. American Indian students have an expected probability of graduation thirteen percentage points lower than the baseline, even with controls for academic preparation and performance. None of the other variables identifying student race/ethnicity were statistically significant. It is not clear what factors outside the model account for this difference, but it is an area worthy of closer examination. In addition, living off campus during the first term of enrollment lowers the predicted probability of graduation by eight percentage points, suggesting this interferes with a student's social integration on campus. Students from neighboring states with tuition reciprocity agreements in place were six percentage points less likely to graduate within six years. Finally, students who were eligible for Pell grants have a predicted probability of graduation six percentage points less than the baseline, even without controlling for any additional financial aid they received.

The logit model, although it performs well at predicting the likelihood of graduation, does not provide information about when students will drop out. For those students who did not graduate within six years, therefore, a parametric survival model was estimated using total credits completed as a measure of the length of student persistence. As noted above, the model uses a generalized gamma distribution to model the time dependence of the underlying process, and the independent variables modify the shape of that distribution and therefore the proportion of cases that remain at any given time (see Table 6).

As with the logit model, academic preparation and first term academic performance measures were among the most clearly statistically significant. The ratio of credits completed to attempted, the number of W grades earned, and the student's ACT/SAT score were statistically significant predictors of credits completed before departure at an alpha level of less than .0001. Additionally, hailing from a bordering state with a tuition reciprocity agreement lowered the total number of credits completed before departure by a statistically significant amount at the same level of confidence. Relaxing the confidence level to a p value of .01 or lower, failing a remedial course (lowering total credits), or identifying as an Asian or Black student (increasing total credits) were statistically significant predictors. These last two may well reflect differences due to college of entry that are not otherwise accounted for in the model.

Table 6. Generalized Gamma Duration Model Parameter Estimates

Duration (credits)	**Coef.**	**Std Error**	**z**	**Sig.**	**P>|z|**
Ratio	0.7961	0.0698	11.41	***	0.0000
C's & D's earned	-0.0075	0.0109	-0.68		0.4940
W's earned	-0.0955	0.0257	-3.72	***	0.0000
ACT/SAT Score	0.0160	0.0033	4.80	***	0.0000
Remedial Taken	-0.0311	0.0351	-0.89		0.3760
Remedial Failed	-0.2278	0.0713	-3.20	**	0.0010
Athlete	0.0629	0.0721	0.87		0.3820
Male	-0.0218	0.0237	-0.92		0.3560
American Indian	-0.0882	0.0989	-0.89		0.3730
Asian	0.1245	0.0393	3.16	**	0.0020
Black	0.1464	0.0547	2.67	**	0.0080
Hispanic	0.1027	0.0783	1.31		0.1890
International	0.1052	0.1500	0.70		0.4830
Reciprocity	-0.1502	0.0298	-5.04	***	0.0000
Non-Reciprocity	-0.1184	0.0581	-2.04	*	0.0410
Pell	0.0612	0.0299	2.05	*	0.0410
Off-Campus Housing	-0.0510	0.0278	-1.84		0.0660
Constant	3.2547	0.1262	25.79	***	0.0000

Shape Parameters	**Coef.**	**Std Error**	**z**	**Sig.**	**P>|z|**
Ln(Sigma)	-0.3743	0.0236	-15.86	***	0.0000
Kappa	1.4746	0.0664	22.21	***	0.0000

* = p < . 05 ** = p < .01 *** = p < .0001

Log-likelihood = -4920.326 p(chi-square) < .0001 AIC = 9880.652

Using the same baseline versus alternative approach taken with the logit model to compare the impact of individual coefficients, the effect of differing values on these variables can be illustrated more clearly. Calculating the survival function, which represents the proportion of all cases that have not yet failed at a given time, after thirty, sixty, ninety, and one-hundred twenty credits completed, provides a snapshot of the proportion of students with different characteristics that are still enrolled after each year of progress toward their degree (see Table 7 below).

Table 7. Predicted Survivor Function for Alterative Values of Each Variable Holding All Other Variables at Baseline Values – Gamma Survival Model

			Survivor Function			
Duration (credits)	**Baseline**	**Alternative**	**30 credits**	**60 credits**	**90 credits**	**120 credits**
Ratio	1.0	0.8	79%	45%	17%	3%
C's & D's earned	0	1	83%	55%	27%	9%
W's earned	0	1	80%	49%	21%	5%
ACT/SAT Score	25	21	81%	51%	23%	6%
Remedial Taken	0	1	82%	53%	25%	8%
Remedial Failed[1]	0	1	75%	38%	12%	1%
Athlete	0	1	84%	59%	32%	13%
Male	0	1	82%	54%	26%	8%
American Indian	0	1	81%	50%	21%	5%
Asian	0	1	86%	62%	37%	16%
Black	0	1	86%	63%	38%	18%
Hispanic	0	1	85%	61%	35%	15%
International	0	1	85%	61%	35%	15%
Reciprocity	0	1	79%	46%	17%	3%
Non-Reciprocity	0	1	80%	48%	19%	4%
Pell	0	1	84%	59%	32%	12%
Off-Campus Housing	0	1	82%	52%	24%	7%
		Baseline:	83%	55%	28%	9%
		Averages:	77%	43%	14%	2%
		Actual:	69%	40%	24%	12%

[1] Alternative for Remedial Failed also includes Remedial Taken set to 1.

The differences are generally fairly small within the first year, but the cumulative impact of a higher probability of departure with each credit completed (the hazard rate) produces more dramatic impacts over time. As with the logit model, failure to successfully complete a remedial mathematics course in the student's first term of enrollment has a dramatically negative effect on the total number of credits completed before departure. Again, the true impact of this variable is seen by setting both it and the indicator for taking a remedial mathematics course to one. Of the total group of such students who did not graduate, 75% remained after 30 credits, 38% remained after 60 credits, 11% remained after 90 credits, and only 1% remained after 120 credits. The comparative numbers for the baseline case are 83% after 30 credits, 55% after 60 credits, 28% after 90 credits, and 9% after 120 credits.

The impacts of the other statistically significant academic preparation and performance measures are also clear from the survival function values. Failure to complete one course successfully (a ratio value of 0.8) lowers the proportion of the students who remain at 30 credits by four percentage points (79% versus 83% for the baseline), and that gap grows to ten percentage points at 60 credits, eleven percentage points at 90 credits, and six percentage points at 120 credits, where the surviving group is only one-third of that at the baseline (3% versus 9%). Earning a single "W" grade or scoring one standard deviation below the freshman mean on the ACT or SAT lowers the likelihood of persistence at each level of total credits.

Students from neighboring states with tuition reciprocity agreements were also more likely to drop out earlier in their careers, at a rate very similar to those who failed to successfully complete a first term course. Although the relationship is noisier, and therefore the confidence in the point estimate is lower, students from more distant states without tuition reciprocity agreements are predicted to have a very similar pattern of survival rates to those from reciprocity states, and noticeably lower than the rates for students from Minnesota.

Both Asian and Black students who fail to graduate within six years are predicted to complete a larger number of credits. This may well represent differences in the college of entry, which was not controlled for in this model to avoid confounds with levels of academic preparation. Asian and Black students are disproportionately represented in the entering classes of colleges at the University of Minnesota whose students generally complete more credits in the completion of their degrees.

6 Discussion

The results of the analysis provide guidance on the issues impacting the academic success of students and suggest possible routes for improvement. The most powerful predictors of both the likelihood of student departure and its timing for the cohorts in the study were levels of first-term academic performance and academic preparation. Worth noting was the significant impact of taking a mathematics remedial course first term or the unfortunate event of failing this course on a students' future success at the University. A disproportionate number of students enrolled in the University's General College were represented in this group. This college was designed to serve underserved and underrepresented students and it is likely these variables are reflecting more then poor academic preparation, but in addition, a mixture of socio-economic and racial/ethnic characteristics. Vigilance and support during the first term of enrollment are therefore of critical importance. The University of Minnesota has recently created a system of mid-term grade reports for all freshmen, and this data provides opportunities for targeted outreach to students who are struggling academically. Additional attention should also be paid to students living off campus and those students from out-of-state (reciprocity or not), given the evidence that they are more likely to leave before completing their degrees. This may in part, be a reflection of the lack of social integration into the campus community. Overall, being a student athlete also tends to be associated with student success, which may seem counter intuitive given all the recent negative publicity regarding student athlete graduation rates. The majority of student athletes are in the non-revenue generating sports and therefore may not have the have the additional stress/opportunities athletes in the major-revenue producing sports have. Perhaps, the opportunities athletics provides to develop relationships and identify with the campus is helping these students to succeed.

Worth noting was the overall lack of evidence that Asian, Black or Hispanic was significantly related to student departure after controlling for academic preparation and first-term academic performance (not true for Native American students). When modeled over time however, the picture changes, Asian and Black students were more likely to take longer, suggesting that they are likely to be enrolled beyond the six-year time frame examined, exhibiting stop-out behavior, or attending part-time. Finally, while the institution should be pleased with its retention of student of color, the fact that those who do not graduate within six years, on average, take more credits suggests there could still be benefit to targeted outreach during the transition into their major programs or during their senior year.

The study also has several implications for future research. Repeating the analysis with more recent cohorts will open up opportunities to connect data that has only recently been collected at the institution, such as the use of recreational sports facilities. Such measures could help with assessing the importance of the social integration of students. In addition, the integration of time-varying covariates could be used in future analyses. While working exclusively with first-term data is useful in identifying early warning indicators for interventions during the critical first year of enrollment, the ability to incorporate changing academic performance, financial need, and other measures over the student's career would enrich our understanding of the process of student retention and departure. Another approach, drawing on data from the National Student Loan Clearinghouse, would be to distinguish drop-outs from transfers to other institutions, using a competing-risks model to identify the factors that shape the enrollment paths of students (Ronco, 1996).

Bibliography

Adelman, C. (1999). Answers in the tool box: Academic intensity, attendance patterns, and bachelor's degree attainment. Washington, DC: U.S. Department of Education [online] Available: http://www.ed.gov/pubs/Toolbox/index.html.

Box-Steffensmeier, J. M., & Jones, B. S. (2004). Event history modeling: A guide for social scientists. Cambridge: Cambridge University Press.

DesJardins, S. L., Ahlburg, D. & McCall B. (1994). Studying the determinants of student stopout: Identifying "true" from spurious time-varying effects. Paper presented at the 34th Annual Forum of the Association for Institutional Research, New Orleans, LA.

DesJardins, S. L., Ahlburg, D., and McCall, B. (2002). Simulating the longitudinal effects of changes in financial aid on student departure from college. Journal of Human Resources 37(3), 653–79.

DesJardins, S. L., McCall, B., Ahlburg, D., & Moye, M. (2002). Adding a timing light to the "tool box". Research in Higher Education, 43(1), 83-114.

DesJardins, S. L. & Moye, M. (2000). Studying the timing of student departure from college (AIR invited paper). Paper presented at the annual meeting of the Association for Institutional Research, Cincinnati, OH.

Field, K. (2006, October 6). Spelling Lays Out 'Action Plan' for Colleges. The Chronicle of Higher Education, pp. A1, A23-A25.

Ishitani, T. (2003). A longitudinal approach to assessing attrition behavior among first-generation students: Time-varying effects of pre-college characteristics. Research in Higher Education, 44(4), 433-449.

Ishitani, T. & DesJardins, S. L. (2002). A longitudinal investigation of dropout from college in the United States. Journal of College Student Retention, 4(2), 173–201.

Ishitani, T. & Snider, K. (2006, May). Longitudinal effects of college preparation programs on college retention. IR Applications: Using Advanced Tools, Techniques and Methodologies, 9, 1-10.

Knapp, L., Kelly-Reid, J., & Whitmore, R.W. (2006). Enrollment in Postsecondary Institutions, Fall 2004; Graduation Rates, 1998 & 2001 Cohorts,; and Financial Statistics, Fiscal Year 2004 (NCES 2006-155). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006155

Maddala, G. S. (1983). Limited-dependent and qualitative variables in econometrics. Cambridge: Cambridge University Press.

Murtaugh, P., Burns, L., & Schuster, J. (1999). Predicting the retention of university students. Research in Higher Education, 40, 355-371.

Perkhounkova, Y., Noble, J., & McLaughlin, G. (2006, Spring). Factors related to persistence of freshmen, freshmen transfers, and nonfreshmen transfer students. AIR Professional File, No. 99, 1-9.

Radcliffe, P., Huesman, R., & Kellogg, J. (2006). Identifying Students at Risk: Utilizing survival analysis to study student athlete attrition. Paper presented at the National Symposium on Student Retention, Albuquerque, NM.

Ronco, S. (1994). Meandering ways: Studying student stopout with survival analysis. Paper presented at the Annual Forum of the Association for Institutional Research, New Orleans, LA.

Ronco, S. (1996, Summer). How enrollment ends: Analyzing the correlates of student graduation, transfer and dropout with a competing risks model. AIR Professional File, No. 61, 1-12.

StataCorp (2005). Stata Statistical Software: Release 9. College Station, TX: StataCorp LP.

Synder, T., Tan, A. & Hoffman, C. (2006). Digest of Education Statistics 2005 (NCES 2006-030). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006030

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. Review of Educational Research, 45, 89-125.

Tinto, V. (2006). Research and practice of student retention: What's next? Journal of College Student Retention, 8 (1), 1-19.

U.S. Census Bureau (2006, October 26). Census Bureau Data Underscore Value of College Degree (CB06-159). Washington, DC: U.S. Department of Commerce, U.S. Census Bureau. Retrieved from: http://www.census.gov/Press-Release/www/releases/archives/education/007660.html

Willet, J. B. & Singer, J. D. (1991). From whether to when: New methods for studying student dropout and teacher attrition. Review of Educational Research, 61(4), 407-450.