



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

**Peter M. Radcliffe**

Senior Analyst

College of Liberal Arts

University of Minnesota

Minneapolis, MN 55455

radcl002@umn.edu

**Ronald L. Huesman, Jr.**

Analyst

Office of Institutional Research

University of Minnesota

Minneapolis, MN 55455

huesm003@umn.edu

**John P. Kellogg**

Assistant Director

Office of Institutional Research

University of Minnesota

Minneapolis, MN 55455

j-kell@umn.edu

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

# Project Background



- University of Minnesota is going through a strategic positioning process
- University goal is to be one of the top three public research universities in the world
- As part of this process, all aspects of the University's functioning are being examined
- Retention and graduation rates have been identified as part of the set of measures that will be used to judge progress toward the strategic goal

# Research Questions



- Multivariate approach needed to answer the questions:
  - “What student characteristics help predict academic success or departure?”
  - “At what points in their careers are students with different characteristics likely to depart?”
  - Success defined as graduation within six years from entry for new freshmen

# Description of Data Set



- 9,580 students
- Entered as first-time, full-time freshmen
- Attempted at least one credit in first term of enrollment
- Enrolled at the University of Minnesota-Twin Cities a large, Midwestern, Doctoral-Extensive University
- Two cohorts, entering in 1999 and 2000

# Variables in Model



- Dependent variables
  - Graduation within six years of entry
  - Number of credits completed at departure
- Independent variables
  - First term academic performance
  - Academic preparation
  - Athletics status
  - Demographics
  - Student family income

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

Variable	Values	Mean	SD	Variable Description (type of variable)
Graduation	0-1	0.620	0.485	If graduated within six years (response variable)
Total Credits	0-232	106.268	46.907	Total credits earned while at the U (timing variable)
Ratio	0-1	0.941	0.177	Ratio of first term credits earned to attempted
C's & D's earned	0-4	0.881	1.010	Number of C grades earned first semester
W's earned	0-3	0.124	0.373	Number of W grades earned first semester
ACT/SAT Score	14-36	24.674	4.202	ACT composite score
Remedial Taken	0-1	0.113	0.316	Remedial math course taken first semester
Remedial Failed	0-1	0.014	0.012	Remedial math course failed first semester
Athlete	0-1	0.036	0.185	If a student-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 Tuition non-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)

# Logit Probability Model



- Since graduation is a dichotomous variable, OLS regression is not efficient and can produce estimated probabilities outside the acceptable range (0-1).
- A solution to this problem is to estimate a latent variable  $y^*$  that represents the probability of the non-zero outcome,  $y^* = xB + u$ , where  $u$  is a probability distribution such as the normal or logistic.
- Estimates can therefore be produced as points along the cumulative distribution function for the selected probability distribution.
- For the logistic distribution, the equation takes the form:

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

# Parametric Survival Models



- A variety of “event history” or “failure time” models
- Also used in biostatistics, economics, and political science
- Estimates the length of time an individual survives until they either fail, die, or otherwise experience the event of interest, or pass out of the window of observation
- In our case, the model estimates the number of credits a student completes before discontinuing enrollment or exceeds six years since their initial enrollment
- Hazard function, survival function, and density are linked by formula:

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





- Survival function:

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

- Represents the proportion of initial cohort remaining at a given time given that they are expected to eventually fail
- Follows a generalized gamma distribution
- Kappa ( $\kappa$ ) and sigma ( $\sigma$ ) determine the shape of the distribution
- $x_j \beta$  represents the vector of observations and coefficients

# Tables 2 & 3. Goodness of fit and model selection



Actual Outcome	Predicted Outcome		
	Graduated	Did not Graduate	Total
Graduated	5,202	1,961	7,163
Did not graduate	742	1,675	2,417
Total	5,944	3,636	9,580

Akaike Information Criterion (AIC)	
Log-Normal	11,292.04
Log-Logistic	10,279.51
Exponential	10,253.74
Weibull	9,958.79
Generalized Gamma	9,880.65

## Model Fit Statistics

- Percent correctly predicted: 71.8%
- Logit Log-likelihood = -5,339.29
- Logit p(chi-square) < .0001
- Gamma Log-likelihood = -4,920.33
- Gamma p(chi-square) < .0001



- Most powerful predictors are first-term performance and academic preparation
  - All six measures of first-term academic performance and academic preparation were significant
  - Taking a remedial math course and failing it lowers estimated likelihood of success by 50%
  - Earning a single W lowers estimated likelihood of success by 14%
  - Failure to complete one course successfully lowers estimated likelihood of graduating in six years by 11%
  - Earning a single C or D lowers estimated likelihood of success by 6%

# Logit Results Continued



- Some demographic indicators were also significant
  - Native Americans have an expected probability of graduation 13% lower than the baseline
  - Students who live off-campus their first semester decrease the estimated likelihood of success by 8%
  - Students from neighboring states were 6% less likely to graduate than the baseline
  - Student-athletes have an estimated likelihood of success 4% higher than the baseline

# Table 4. Logit Model Parameter Estimates



Logit (graduate)	Coif.	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

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



Logit (graduate)	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 <sup>1</sup>	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%
		<b>Baseline:</b>	<b>84%</b>	<b>0%</b>

# Duration Results



- First-term academic performance again has the strongest impact
  - Students who take and fail a remedial mathematics course in the first term take fewer credits, with 75% retained after 30 credits and 12% retained after 90 credits
  - Students who fail to successfully complete a one of five courses taken complete fewer credits in total, with 79% retained after 30 credits, and 17% retained after 90 credits
  - Students who earn a single W earned complete fewer credits, with 80% retained after 30 credits and 21% retained after 90 credits

# Duration Results Continued



- Academic preparation likewise has a significant impact
  - Scoring one standard deviation below the mean on the ACT (or converted SAT) lowers probability of retention after 30 credits to 81%, and after 90 credits to 23%
- Students from other states also complete fewer credits
  - 79% of students from reciprocity states remained after 30 credits, and 17% remained after 90 credits
  - 80% of students from non-reciprocity states remained after 30 credits, and 19% remained after 90 credits



# Table 6. Parametric Survival Model Parameter Estimate: Generalized Gamma Duration



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

Table 7. Predicted Survivor Function for Alternative Values of Each Variable Holding All Other Variables at Baseline Values



Duration (credits)	Baseline	Alternative	Survivor Function			
			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 <sup>1</sup>	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%
		<b>Baseline:</b>	<b>83%</b>	<b>55%</b>	<b>28%</b>	<b>9%</b>

# Policy Implications



- Academic performance in the first term is critical
- The University of Minnesota has in place a program to issue mid-term alerts to freshmen who are struggling in courses
- This program, which began after the cohorts in this study were admitted, affords the institution an opportunity to identify and reach out to students who are struggling before they fail or withdraw from classes

# Questions for future research



- Incorporate time-varying covariates: academic performance, financial measures over a student's career
- Results suggest that some departing students are in good academic standing, suggesting they may be transferring to another institution rather than dropping out – a competing risks model could be used to investigate this possibility
- Adding more extensive recent data may help in identifying issues related to social integration



# Questions?

**Peter M. Radcliffe**

Senior Analyst

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

**Ronald L. Huesman, Jr.**

Analyst

Office of Institutional Research  
University of Minnesota  
Minneapolis, MN 55455  
huesm003@umn.edu

**John P. Kellogg**

Assistant Director

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