Using Survival Analysis to Model Retention in a Master's Program

Tracy Mohr
Senior Research Associate
Office of Institutional Research and Market Analytics
DePaul University
tmohr3@depaul.edu

DEPAUL UNIVERSITY
ENROLLMENT MANAGEMENT AND MARKETING
Institutional Research & Market Analytics

Purpose

- DePaul University's School of Computing and Digital Media (CDM) has the lowest master's graduation rate of the university's eight colleges. This study will use survival analysis as a method to identify characteristics that contribute to student attrition and determine those that are most significant.
- There has been much less research into graduate retention than undergraduate. Furthermore, data on graduate retention rates is not readily available. This study is a small effort to close that gap.

Institutional Background

- DePaul University (Chicago, IL) is the largest Catholic university in the country and the largest private institution in Illinois, with 25,000 students.
- DePaul's School of Computing and Digital Media (CDM) is currently the university's largest graduate college, but since 1996 has had the lowest four- and six-year graduation rates of DePaul colleges.

General CDM Context

- CDM currently has two distinct schools:
  - Computing.
  - Cinema and Interactive Media.

    (These two schools were established in 2009-10)
- Compared to the university, CDM graduate students are:
  - More likely to be men.
  - More likely to be students of color.
  - Older.

CDM Retention

- As noted, CDM has historically had the lowest retention rate of all DePaul graduate colleges.
- The question arises - can we determine if CDM graduate students have certain characteristics that make them more likely to leave without a degree?

Using Survival Analysis

- A general definition of survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest (dependent variable) is the *time until an event occurs*. (Kleinbaum & Klein, 2005, italics in original).
- Advantages of survival analysis over regression:
  - The time to the event is part of the model.
  - Events can be censored; for some units the event of interest has occurred and therefore we know the exact waiting time, whereas for others it has not occurred, and all we know is that the waiting time exceeds the observation time.
- The Cox proportional hazards (PH) model is used in this analysis.

Survival Analysis – Some Definitions

- Hazard rate - the instantaneous rate of failure at a given time *t*.
- Hazard ratio – a ratio calculated from hazard rates, but cumulative over a period of time. This assumes proportionality.
- Survival rate - cumulative probability of surviving the most recent interval multiplied by the probabilities of surviving all of the prior intervals.

Methodology

- The 2006 master's cohort was chosen, in order to obtain six years of data. (A cohort runs summer to spring, so this was Summer 2005 to Spring 2006.) There were 2,816 students in this cohort.
- The dependent variable was the number of quarters a student completed.
- The censoring variable was their departure status; students leaving with a degree were right-censored (see Table 2).
- Throughout this study, "departure" refers to a student leaving without a degree.

Twelve independent variables were used in the survival model, in addition to the dependent and censoring variables. (see Table 3)

Variable Name	Variable Description	Notes
terms_enrolled	**Dependent variable**: Number of terms enrolled	
departure	**Censoring variable**: censored defined as student either obtaining a degree or still enrolled	0 = right-censored (obtained degree), 1 = not censored (departed without degree)
gpa_group	Cumulative GPA after last term enrolled	0.00, 0.01-1.99, 2.00-2.49, 2.50-2.99, 3.00-3.49, 3.50-3.99, 4.00
FT_PT_group	Full-Time or Part-time Status	Calculated as percentage of enrolled terms that student was full-time (8 or more credit hours in a term)
retro_acadgrp	DePaul College	
F_aid_used	Federal Financial Aid Used	Binary: 0 = no aid, 1 = aid
S_aid_used	State Financial Aid Used	Binary: 0 = no aid, 1 = aid
I_aid_used	Institutional Financial Aid Used	Binary: 0 = no aid, 1 = aid
P_aid_used	Private Financial Aid Used	Binary: 0 = no aid, 1 = aid
race_collapsed	Race	Foreign students are classified separately
age_group	Age Group	Under 24, 24-29, 30-39, 40-49, 50 or older, unknown
SEX	Sex	
TERM	Term of Entry	Summer, Fall, Winter or Spring
GRAD_ASSISTANT	Graduate Assistant Status	Binary: 1 = was a graduate assistant at any time while enrolled

2006 CDM Demographics

- In 2006, compared to the overall cohort, CDM students were:
  - More likely to have a GPA of 0.00 or 4.00, less likely to have a GPA of 3.50-3.99.
  - More likely to be 100% part-time (more than 40% of all such students were in CDM).
  - Less likely to use any financial aid.
  - More likely to be men.
  - More likely to be students of color or foreign students.
  - Slightly older (average age 29.6 vs. 28.6 for cohort).

The overall model had five significant variables (at p <= 0.10); one of these was DePaul college (retro_acadgrp).

	Wald Chi-Square	Overall Model
Departure Rate	---	**24%**
2006 Cohort Enrolls	---	2,816
gpa_group	431.31	**<0.0001**
retro_acadgrp	321.39	**<0.0001**
FT_PT_group	74.61	**<0.0001**
F_aid_used	14.90	**0.0001**
P_aid_used	4.46	**0.0347**
race_collapsed	9.95	0.1267
age_group	6.63	0.2493
SEX	1.20	0.2726
TERM	2.33	0.5065
S_aid_used	0.04	0.8356
I_aid_used	0.02	0.8944
GRAD_ASSISTANT	0.002	0.9627

Comparing hazard ratios by college, using Business as the baseline (as it is the largest college and has the lowest departure rate), shows significant differences in departure rates for five of the seven other colleges. Despite its higher departure rate, CDM does have a lower hazard ratio.

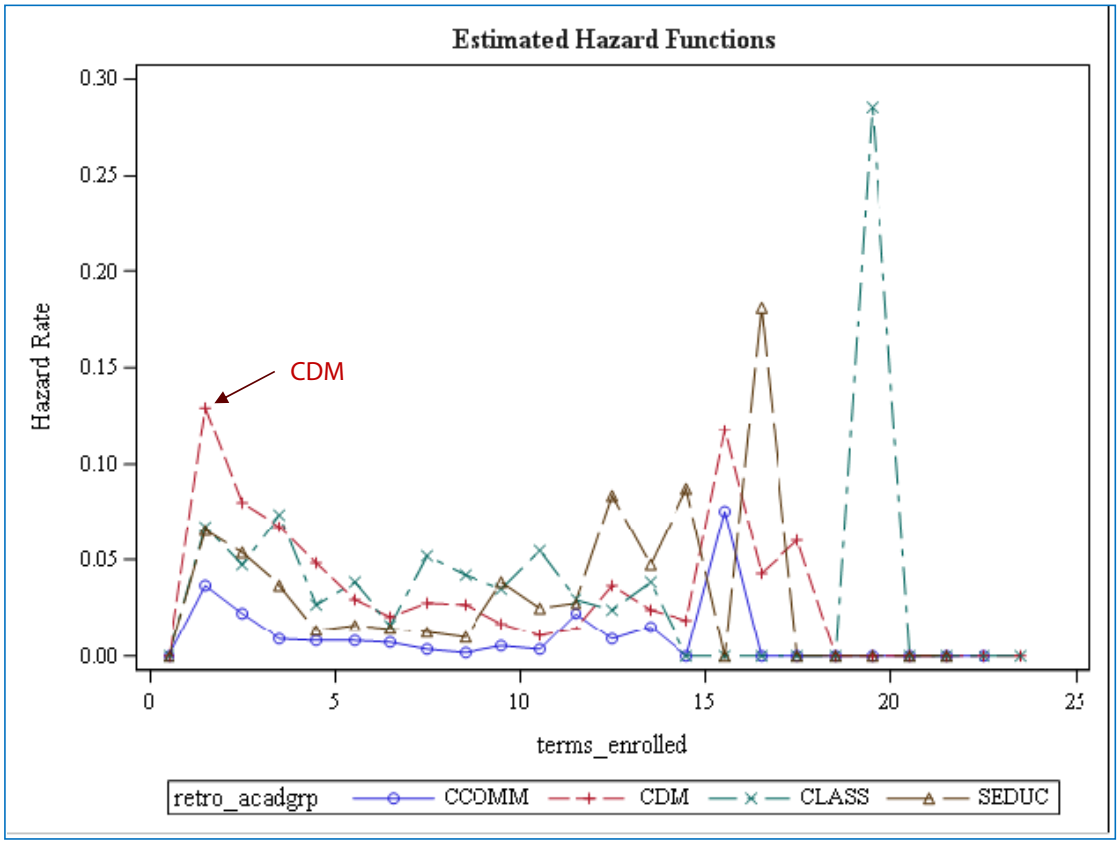
	N	Departure Rate	Chi-Square	P-value	Hazard Ratio
Business	839	10%	---	---	---
CDM	**781**	**36%**	**26.76**	**<0.0001**	**2.047**
Education	535	23%	57.58	<0.0001	3.390
Liberal Arts	368	32%	39.20	<0.0001	2.649
Science & Health	183	21%	24.13	<0.0001	2.766
Music	53	26%	25.75	<0.0001	5.296
Communication	44	14%	2.84	0.0920	2.074
Theatre	13	8%	0.15	0.6960	0.672

Through eight quarters (the equivalent of two years at DePaul), CDM's survival estimates are the poorest of the five largest colleges. Notably, 19% of CDM students left within two quarters, and 28% within four. None of the other four colleges saw more than 20% leave after four quarters.

Life Table Survival Estimates (by Quarter)

Quarter	CDM Survival	BUS Survival	LAS Survival	EDU Survival	CSH Survival
0	**1.0000**	1.0000	1.0000	1.0000	1.0000
1	**0.8784**	0.9641	0.9347	0.9360	0.9563
2	**0.8105**	0.9425	0.8910	0.8870	0.9344
3	**0.7578**	0.9341	0.8280	0.8549	0.9010
4	**0.7215**	0.9268	0.8059	0.8436	0.8727
5	**0.7002**	0.9192	0.7749	0.8303	0.8496
6	**0.6856**	0.9122	0.7625	0.8180	0.8371
7	**0.6665**	0.9090	0.7240	0.8074	0.8128
8	**0.6488**	0.9070	0.6942	0.7996	0.8128
9	0.6375	0.9016	0.6705	0.7690	0.8128
10	0.6303	0.8979	0.6348	0.7495	0.7062
11	0.6212	0.8777	0.6160	0.7288	0.6539
12	0.5987	0.8699	0.6014	0.6705	0.6539
13	0.5840	0.8569	0.5782	0.6393	0.5667
14	0.5729	0.8569	0.5782	0.5861	0.5667
15	0.5092	0.7946	0.5782	0.5861	0.5667
16	0.4875	0.7946	0.5782	0.4884	

The hazard rate is the instantaneous rate of failure at a given time – for example, as shown in this graph, the hazard rate for CDM students (red dashed line) in the first term is 13%, while for Business it is 4%. Graphing the hazard rates helps illustrate the likelihood of CDM students (red dashed line) departing sooner. This also provides a view of hazard rates by term.



Analysis for DePaul Colleges

- For the DePaul colleges in the 2006 cohort with enough students to allow it, Cox PH models were run for CDM and the other DePaul colleges with more than 100 new enrolls in the 2006 cohort. (The three colleges with fewer than 100 new enrolls, Communication, Music, and Theatre, were not included.)
- The models by college used the same variables as the overall model.

The overall model had five significant variables (in bold, at p <= 0.10); one of these was retro_acadgrp (DePaul college). The CDM model had six significant variables.

		College				
	Overall Model	**Business**	**CDM**	**Education**	**Liberal Arts**	**Science & Health**
Departure Rate	**24%**	**11%**	**36%**	**23%**	**32%**	**21%**
2006 Cohort Enrolls	2,816	839	781	535	368	183
gpa_group	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
retro_acadgrp	**<0.0001**	---	---	---	---	---
FT_PT_group	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**0.0022**
F_aid_used	**0.0001**	0.4881	**<0.0001**	0.1257	0.3059	0.8236
P_aid_used	**0.0347**	0.6797	**0.0664**	0.7769	**0.0266**	0.2507
race_collapsed	0.1267	**0.0597**	**0.0008**	0.1747	0.9084	0.9502
age_group	0.2493	0.6124	**0.0956**	**0.0883**	0.7549	0.3218
SEX	0.2726	0.6565	0.7429	0.5784	0.5900	0.5638
TERM	0.5065	**0.0118**	0.6822	0.7818	0.8269	**0.0393**
S_aid_used	0.8356	---	---	0.5173	---	0.9991
I_aid_used	0.8944	0.1129	0.9929	**0.0723**	**0.0716**	0.8644
GRAD_ASSISTANT	0.9627	0.5693	0.6051	0.2323	0.6198	0.3957

Findings (see Table 7)

- For both CDM and overall:
  - Students enrolled full-time in all quarters had a significantly higher departure rate than those enrolled part-time in at least one term.
  - Students with a 4.00 GPA had a significantly higher departure rate than those with GPA between 3.50 and 3.99, and were also more likely to depart than those in the 3.00-3.49 group.
  - Students receiving federal aid were significantly less likely to depart.

Findings (see Table 7)

- For both CDM and overall:
  - Students enrolled full-time in all quarters had a significantly higher departure rate than those enrolled part-time in at least one term.
  - Students with a 4.00 GPA had a significantly higher departure rate than those with GPA between 3.50 and 3.99, and were also more likely to depart than those in the 3.00-3.49 group.
  - Students receiving federal aid were significantly less likely to depart.
- CDM students enrolled 100% full-time were more likely to depart than the cohort as a whole.
- Overall, students receiving private aid were significantly less likely to depart, but CDM students were significantly more likely to depart (CDM n=18, so interpret with caution).
- Race is a significant variable in the CDM model, but there were no significant differences by race. The only significant difference by age was between those under 24 and age 24-29.

Discussion

- CDM students leave earlier – 19% in first two quarters. This is mostly those with GPA of 0.00, of which CDM has the highest percentage.
- However, academically successful CDM students do depart without degrees at a higher rate.

Discussion

- CDM students leave earlier – 19% in first two quarters. This is mostly those with GPA of 0.00, of which CDM has the highest percentage.
- However, academically successful CDM students do depart without degrees at a higher rate.
- It has been theorized that CDM students leave without degrees because that is not always necessary for employment in their field. This study was unable to confirm that.
- Interaction effects were tested, but resulted in a model with questionable convergence.

Future Actions and Recommendations

- Further investigate why high-performing students are leaving, including qualitative research.
- Use National Clearinghouse data to determine if students who leave are enrolling elsewhere.
- Apply models and findings to other DePaul colleges.
- Develop models using a shorter period (two or three years).
- Continue to develop models using more recent cohorts to validate results.

Suggested Actions for CDM

- Target retention efforts toward full-time students and those with high GPAs.
- Develop focused retention strategies based on most likely groups to depart.
- Develop separate models for the two CDM schools, using more recent cohorts.

Further questions?

tmohr3@depaul.edu

Using Survival Analysis to Model Retention in a Master's Program

Abstract – Retention and graduation from graduate programs is increasingly important to institutions for financial as well as social reasons. There is considerable literature using survival analysis and other event analysis methods to study undergraduate retention (e.g., Murtaugh, Burns, & Schuster, 1999). However, little has been published for graduate retention. A rare example is Haughton et al. (2011). Survival analysis is particularly well-suited to modeling student flow. In survival analysis, the variable of interest is the time until an event occurs (here, departure or graduation), while also accounting for those who continue. Those who leave but could return are also accounted for as well (censoring). This study applies survival analysis to a master's program, using it to identify factors that may significantly influence student retention for master's students. Suggestions will be made for further research.

Purpose

DePaul's School of Computing and Digital Media has the lowest six-year graduation rate of the university's colleges. In order to improve this graduation rate in an effective manner, it is important to find those characteristics that contribute to student attrition. This study will use survival analysis as one method to identify those characteristics and determine those that are most significant.

Survival Analysis

A general definition of survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest (dependent variable) is the *time until an event occurs*. (Kleinbaum & Klein, 2005, italics in original). Also referred to as event history analysis or reliability analysis, survival analysis differs from other statistical methods in that it takes into account observations where the event may not occur before the time the study is concluded. These incomplete observations are referred to as censored.

Table 3: *Variables Used for Survival Analysis*

Variable Name	Variable Description	Notes
terms_enrolled	Dependent variable: Number of terms enrolled	
departure	Censoring variable: censored defined as student either obtaining a degree or still enrolled	0 = censored (obtained degree or still enrolled), 1 = not censored (departed without degree)
gpa_group	Cumulative GPA after last term enrolled	
FT_PT_group	Full-Time or Part-time Status	Calculated as percentage of enrolled terms that student was full-time (8 or more credit hours in a term)
retro_acadgrp	DePaul College	
F_aid_used	Federal Financial Aid Used	Binary: 0 = no aid, 1 = aid
P_aid_used	Private Financial Aid Used	Binary: 0 = no aid, 1 = aid
race_collapsed	Race	Foreign students are classified separately
age_group	Age Group	
SEX	Sex	
TERM	Term of Entry	Summer, Fall, Winter or Spring
S_aid_used	State Financial Aid Used	Binary: 0 = no aid, 1 = aid
I_aid_used	Institutional Financial Aid Used	Binary: 0 = no aid, 1 = aid
GRAD_ASSISTANT	Graduate Assistant Status	Binary: 1 = was a graduate assistant at any time while enrolled

The largest colleges at DePaul had a sufficient number of students in the 2006 cohort to allow Cox PH models to be run for CDM and the other colleges with more than 100 new enrolls in the 2006 cohort. Significance is at p <= 0.10.

P-Values from Cox PH Models for Overall Model and by Colleges, 2006 Cohort

		College				
	Overall Model	**Business**	**CDM**	**Education**	**Liberal Arts**	**Science & Health**
Departure Rate	**24%**	**11%**	**36%**	**23%**	**32%**	**21%**
2006 Cohort Enrolls	2,816	839	781	535	368	183
gpa_group	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
retro_acadgrp	**<0.0001**	---	---	---	---	---
FT_PT_group	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**0.0022**
F_aid_used	**0.0001**	0.4881	**<0.0001**	0.1257	0.3059	0.8236
P_aid_used	**0.0347**	0.6797	**0.0664**	0.7769	**0.0266**	0.2507
race_collapsed	0.1267	**0.0597**	**0.0008**	0.1747	0.9084	0.9502
age_group	0.2493	0.6124	**0.0956**	**0.0883**	0.7549	0.3218
SEX	0.2726	0.6565	0.7429	0.5784	0.5900	0.5638
TERM	0.5065	**0.0118**	0.6822	0.7818	0.8269	**0.0393**
S_aid_used	0.8356	---	---	0.5173	---	0.9991
I_aid_used	0.8944	0.1129	0.9929	**0.0723**	**0.0716**	0.8644
GRAD_ASSISTANT	0.9627	0.5693	0.6051	0.2323	0.6198	0.3957

For each of the significant variables in both the overall and CDM models, the departure rates and hazard ratios were analyzed. An asterisk (*) after a parameter indicates that the other values of that parameter were compared to it to determine the hazard ratios.

Table 7: *Hazard Ratios for Significant Variables, 2006 Cohort and CDM*

		Cohort				CDM			
Variable		**N**	**Departure Rate**	**P-value**	**Hazard Ratio**	**N**	**Departure Rate**	**P-value**	**Hazard Ratio**
FT_PT_group	**100% Full-time ***	**566**	**31%**			**146**	**49%**		
	100% Part-Time	358	72%	<.0001	1.700	166	80%	0.2132	1.265
	1-25% Full-Time	248	20%	<.0001	0.330	72	29%	<.0001	0.318
	26-50% Full-Time	434	23%	<.0001	0.453	102	33%	<.0001	0.382
	51-75% Full-Time	698	10%	<.0001	0.244	153	13%	<.0001	0.160
	76-99% Full-Time	507	3%	<.0001	0.079	142	3%	<.0001	0.051
gpa_group	**4.00 ***	**246**	**31%**			**81**	**37%**		
	0.00	154	94%	<.0001	6.159	113	97%	<.0001	7.431
	0.01 - 1.99	38	95%	<.0001	4.172	16	100%	<.0001	4.083
	2.00 - 2.49	45	87%	<.0001	3.381	23	91%	<.0001	3.363
	2.50 - 2.99	154	32%	0.1222	1.355	57	30%	0.3806	1.337
	3.00 - 3.49	528	23%	0.2247	0.826	154	28%	0.3881	0.806
	3.50 - 3.99	1,650	12%	<.0001	0.441	337	14%	0.0002	0.398
F_aid_used	0 = no aid *	1,618	25%			548	37%		
	1 = aid	1,198	22%	0.0001	0.696	233	35%	<.0001	0.512
P_aid_used	0 = no aid *	2,714	24%			763	36%		
	1 = aid	102	15%	0.0347	0.542	18	39%	0.0664	2.206
race_collapsed (not significant in overall model	White *					246	40%		
	Asian					66	33%	0.1459	1.432
	African American					71	52%	0.4226	1.188
	Foreign					161	20%	0.3283	0.786
	Hispanic					37	46%	0.4726	0.813
	Missing					198	37%	0.1204	1.313
	Other					2	100%	<.0001	21.198
age_group (not significant in overall model	Under 24 *					158	22%		
	24-29					284	36%	0.0406	1.541
	30-39					202	42%	0.8591	1.042
	40-49					61	52%	0.5409	1.188
	50 or older					18	33%	0.4748	0.712
	Unknown					45	33%	0.8214	1.209

Findings:

- For both CDM and overall:
  - Students enrolled full-time in all quarters had a significantly higher departure rate than those enrolled part-time in at least one term.
  - Students with a 4.00 GPA had a significantly higher departure rate than those with GPA between 3.50 and 3.99, and were also more likely to depart than those in the 3.00-3.49 group.
  - Students receiving federal aid were significantly less likely to depart.
- CDM students enrolled 100% full-time were more likely to depart than the cohort as a whole.
- Overall, students receiving private aid were significantly less likely to depart, but CDM students were significantly more likely to depart (CDM n=18, so interpret with caution).
- Race is a significant variable in the CDM model, but there were no significant differences by race. The only significant difference by age was between those under 24 and age 24-29.

General Conclusions and Recommendations

- Further investigate why high-performing students are leaving, including qualitative research.
- Use National Clearinghouse data to determine if students who leave are enrolling elsewhere.
- Apply models and findings to other DePaul colleges.
- Develop models using a shorter period (two or three years).
- Continue to develop models using more recent cohorts to validate results.

Suggested Actions for CDM

- Target retention efforts toward full-time students and those with high GPAs.
- Develop focused retention strategies based on most likely groups to depart.
- Develop separate models for the two CDM schools, using more recent cohorts.

Using Survival Analysis to Model Retention in a Master's Program

Tracy Mohr
Senior Research Associate
Office of Institutional Research and Market Analytics
DePaul University
tmohr3@depaul.edu

Abstract – Retention and graduation from graduate programs is increasingly important to institutions for financial as well as social reasons. There is considerable literature using survival analysis and other event analysis methods to study undergraduate retention (e.g., Murtaugh, Burns, & Schuster, 1999). However, little has been published for graduate retention. A rare example is Haughton et al. (2011). Survival analysis is particularly well-suited to modeling student flow. In survival analysis, the variable of interest is the time until an event occurs (here, departure or graduation), while also accounting for those who continue. Those who leave but could return are also accounted for as well (censoring). This study applies survival analysis to a master's program, using it to identify factors that may significantly influence student retention for master's students. Suggestions will be made for further research.

Introduction

The benefits of a university-wide retention program are well known and can be demonstrated for both the student and the institution. The benefit of completing a college degree to the student is well documented, as census data continues to show a strong positive correlation between level of educational attainment and household income (Bureau of Labor Statistics, 2013). The benefits to the institution of retaining students are profound as well. Obviously, the more students that remain enrolled and complete their degrees, the more income flows to the institution. One study estimated that a 4% increase in new student retention at a large public institution would increase tuition revenue by $2.25 million at a cost of $345,000 (Mager, as cited in Simpson, 2005). There are also public relations benefits, albeit more difficult to quantify, as an institution may see its national rankings improve as its graduation rates rise.

Much of the study of retention in higher education has and continues to focus on undergraduates, particularly that of new freshmen. Comparative data is readily available: using the Integrated Postsecondary Education Data System (IPEDS), users can find retention, as well as graduation rates for first-time bachelor's degree-seeking students for most any institution in the United States. The literature addressing undergraduate retention is extensive and dates back the 1970s.

Less has been published on retention among graduate students. Alexander et al. (2009) summarize much of the extant resources and literature about graduate retention. They observe that much of the work in graduate retention addresses doctoral retention, not master's retention. For institutions with a majority of their graduate students at the master's level, the research literature may be found wanting. Data for graduate retention is also less readily available; the author knows of no central source similar to IPEDS for obtaining graduate retention data for other institutions.

Institutional background

DePaul University, in Chicago, is the largest private not-for-profit institution in the state of Illinois, and one of the largest in the United States. In recent years, between 25% and 30% of DePaul's students pursue master's degrees. DePaul enrolled 6,586 master's students in Fall 2013, with 1,611 being new enrolls.

DePaul has nine colleges, all of which enroll master's students. One of DePaul's colleges, the School for New Learning (SNL), is specifically oriented toward adult students: it does not enroll students younger than age 24, does not offer traditional programs, and offers students credit toward degrees for life experience. Because of this unique approach, SNL students are not included in this analysis.

Of the remaining eight colleges, the School of Computing and Digital Media (CDM) is the second largest (after the business school), with 1,903 master's students enrolled in Fall 2013, 408 of whom were new students. While university retention and graduation rates have steadily increased at both the undergraduate and graduate levels, CDM's six-year master's graduation rate remains the lowest among all DePaul colleges (excluding the aforementioned SNL). This has been the case for every cohort since 1996.

Survival Analysis

An acceptable retention model can be developed using logistic regression, with graduation as a binary dependent variable. While this is satisfactory for predicting if a student will graduate, as well as the factors that may have the greatest impact on success, it accounts for neither the length of time nor the path the student may take.

Survival analysis is a more appropriate method to use in this case. Several have used it to evaluate undergraduate retention, including Murtaugh, Burns, and Schuster (1999) and Singer and Willett (1993). However, there have been fewer examples of its use in graduate retention studies. A rare example is Haughton et al. (2011)

A general definition of survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest (dependent variable) is the *time until an event occurs*. (Kleinbaum & Klein, 2005, italics in original). Also referred to as event history analysis or reliability analysis, survival analysis differs from other statistical methods in that it takes into account observations where the event may not occur before the time the study is concluded. These incomplete observations are referred to as censored.

In survival analysis, the most common model used is the Cox proportional hazards (PH) model. Also known as Cox regression, this implements the proportional hazards model, and is designed for analysis of time until an event or time between events. One or more predictor variables, called covariates, are used to predict a status (event) variable. The central statistical output is the hazard ratio. A hazard ratio greater than 1 indicates a greater likelihood of failure (or departure, in this case). The proportional hazards model assumes that the ratio of hazards for any two observations is the same across time periods. This is the proportionality assumption.

If the proportionality assumption is tested and found to be violated, we assume that the independent variables are significantly related to survival times. To account for this, one option is to run the Cox model, including time-dependent interaction covariates with the variables violating the assumption. A second option is to run an accelerated failure time (AFT), or parametric regression, model. This model assumes that survival time changes by a constant factor when comparing different levels of covariates.

Purpose

DePaul's School of Computing and Digital Media has the lowest six-year graduation rate of the university's colleges. In order to improve this graduation rate in an effective manner, it is important to find those characteristics that contribute to student attrition. This study will use survival analysis as one method to identify those characteristics and determine those that are most significant.

Methodology and Results

At DePaul, a master's student cohort is considered to begin in the summer quarter and end in the spring quarter, with the cohort year being the year of the summer quarter. Retention research at DePaul has shown that graduation rates for a cohort increase until the sixth year, then begin to level off. Therefore, as the most recent cohort with six complete years of data, the 2006 cohort was chosen for analysis. As DePaul uses the quarter system, this cohort covers 24 quarters.

There were 2,816 new master's students enrolled in the 2006 cohort. The six-year graduation rate of the 2006 cohort was 74%, and 24% of all students in this cohort had departed without a degree by Fall 2013.

The following table shows selected statistics for the 2006 cohort and for each DePaul college. As already noted, CDM had the highest rate of students leaving without degrees; CDM also had the highest percentage of male students, the highest average age at entry and the lowest average GPA at departure.

Table 1: *2006 Master's Cohort by DePaul College*

	2006 Cohort	Business	CDM	Education	Liberal Arts	Science & Health	Communication	Music	Theatre
Enrolls	2,816	839	781	535	368	183	44	53	13
Left without degree	24%	11%	36%	23%	32%	21%	14%	26%	8%
Female	50%	42%	22%	81%	69%	68%	82%	45%	62%
Male	50%	58%	78%	19%	30%	32%	16%	55%	39%
Asian	7%	7%	9%	5%	4%	12%	5%	0%	0%
Black	8%	4%	9%	7%	15%	11%	16%	4%	15%
Foreign	11%	12%	21%	1%	2%	3%	5%	25%	0%
Hispanic	5%	3%	5%	5%	8%	4%	7%	4%	8%
Missing	18%	17%	25%	13%	15%	9%	11%	26%	0%
Other	0%	0%	0%	0%	1%	1%	0%	0%	0%
White	52%	57%	32%	69%	55%	60%	57%	42%	77%
Used Federal Aid	43%	29%	30%	62%	57%	65%	48%	53%	92%
Used Institutional Aid	11%	4%	4%	6%	28%	24%	21%	83%	100%
Average Final GPA	3.39	3.51	2.98	3.67	3.43	3.52	3.60	3.74	3.68
Avg. Pct. of Terms FT	67%	61%	54%	70%	49%	66%	70%	55%	100%
Average Age	28.5	28.8	29.6	27.7	28.0	27.9	26.2	25.2	25.3

In all survival models run in this study, the dependent variable was the number of quarters the student had completed. The censoring variable was the student's departure status, where students who left with degrees were right-censored. The following table shows how a student is classified in terms of censoring, where "failure" is considered to having left without a degree. Since they did not necessarily fail in the traditional academic sense, these students will be referred to as "departures."

Table 2: *2006 Censoring Status*

		Degree status	
		Did not receive degree	Received degree
Enrollment Status	Not enrolled at end of 24 terms	Departed ("failure")	Censored (graduated)
	Enrolled at end of 24 terms	Currently enrolled	Currently enrolled

The first model will look at the overall 2006 cohort. Following models will look at the results by college, with specific focus on CDM, as it is the college of interest with the highest departure rate.

To create a model for analysis of the 2006 cohort, the following variables were used.

Table 3: *Variables Used for Survival Analysis*

Variable Name	**Variable Description**	**Notes**
terms_enrolled	Dependent variable: Number of terms enrolled	
departure	Censoring variable: censored defined as student either obtaining a degree or still enrolled	0 = censored (obtained degree or still enrolled), 1 = not censored (departed without degree)
gpa_group	Cumulative GPA after last term enrolled	
FT_PT_group	Full-Time or Part-time Status	Calculated as percentage of enrolled terms that student was full-time (8 or more credit hours in a term)
retro_acadgrp	DePaul College	
F_aid_used	Federal Financial Aid Used	Binary: 0 = no aid, 1 = aid
P_aid_used	Private Financial Aid Used	Binary: 0 = no aid, 1 = aid
race_collapsed	Race	Foreign students are classified separately
age_group	Age Group	
SEX	Sex	
TERM	Term of Entry	Summer, Fall, Winter or Spring
S_aid_used	State Financial Aid Used	Binary: 0 = no aid, 1 = aid
I_aid_used	Institutional Financial Aid Used	Binary: 0 = no aid, 1 = aid
GRAD_ASSISTANT	Graduate Assistant Status	Binary: 1 = was a graduate assistant at any time while enrolled

To establish a baseline for analysis, an initial Cox PH model was run on all observations in the cohort using SAS PROC PHREG, with no grouping or strata. The dependent variable was the number of terms enrolled, and the censoring value was the student's departure status. The results follow, listed by ascending p-value. In this model, five variables were significant at $p <= 0.05$.

Table 4: *Results of Cox PH Model on 2006 Cohort*

Variable	Degrees of freedom	Wald Chi-Square	P-value
gpa_group	6	431.31	<0.0001
FT_PT_group	5	321.39	<0.0001
retro_acadgrp	7	74.61	<0.0001
F_aid_used	1	14.90	0.0001
P_aid_used	1	4.46	0.0347
race_collapsed	6	9.95	0.1267
age_group	5	6.63	0.2493
SEX	1	1.20	0.2726
TERM	3	2.33	0.5065
S_aid_used	1	0.04	0.8356
I_aid_used	1	0.02	0.8944
GRAD_ASSISTANT	1	0.002	0.9627

As college was a significant variable in the initial Cox PH model, as well as the subject of analysis, the hazard ratios of the individual colleges were compared. The college chosen as the baseline was Business; as it is the largest college at DePaul, yet has the lowest departure rate (with the exception of the smallest college, Theatre), its retention success may be considered a benchmark for the other colleges. The following table shows how the other colleges compared to Business. The p-values show that five of the other seven colleges had significantly greater departure rates (at $p <= 0.05$). CDM's hazard ratio of 2.047 was significant, and means that CDM students were about twice as likely to leave without a degree at any given time as Business students. However, it is worth noting that CDM's hazard ratio is lower than the other colleges, despite having a higher departure rate.

Table 5: *Hazard Ratios Comparing Colleges, 2006 Cohort*

	N	Departure Rate	Chi-Square	P-value	Hazard Ratio
Business	839	10%			
CDM	781	36%	26.76	<0.0001	2.047
Education	535	23%	57.58	<0.0001	3.390
Liberal Arts	368	32%	39.20	<0.0001	2.649
Science & Health	183	21%	24.13	<0.0001	2.766
Music	53	26%	25.75	<0.0001	5.296
Communication	44	14%	2.84	0.0920	2.074
Theatre	13	8%	0.15	0.6960	0.672

As noted, the initial intent of this project was to investigate why CDM has a higher departure rate compared to the other colleges. Since the largest colleges at DePaul had a sufficient number of students in the 2006 cohort to allow it, Cox PH models were run for CDM and the other DePaul colleges with more than 100 new enrolls in the 2006 cohort. The three colleges with fewer than 100 new enrolls, Communication, Music, and Theatre, were not included. The models by college used the same variables as the overall model. The following table compares the model results for the five colleges: Business, CDM and Liberal Arts did not have any students who used state aid, so there is no p-value for that variable. (For brevity's sake, only the p-values are included; a full table is available from the author.)

Table 6: *P-Values from Cox PH Models by Colleges, 2006 Cohort*

	College				
Variable	**Business**	**CDM**	**Education**	**Liberal Arts**	**Science & Health**
2006 Cohort Enrolls	839	781	535	368	183
gpa_group	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
FT_PT_group	<0.0001	<0.0001	<0.0001	<0.0001	0.0022
F_aid_used	0.4881	<0.0001	0.1257	0.3059	0.8236
P_aid_used	0.6797	0.0664	0.7769	0.0266	0.2507
race_collapsed	0.0597	0.0008	0.1747	0.9084	0.9502
age_group	0.6124	0.0956	0.0883	0.7549	0.3218
SEX	0.6565	0.7429	0.5784	0.5900	0.5638
TERM	0.0118	0.6822	0.7818	0.8269	0.0393
S_aid_used	.	.	0.5173	.	0.9991
I_aid_used	0.1129	0.9929	0.0723	0.0716	0.8644
GRAD_ASSISTANT	0.5693	0.6051	0.2323	0.6198	0.3957

The CDM model shares the same significant variables as the overall model, but also introduces race as significant at $p < 0.05$, and private aid used and age group as significant at $p < 0.10$.

For each of the significant variables in both the overall and CDM models, the departure rates and hazard ratios were analyzed. The following table compares them. An asterisk (*) after a parameter indicates that the other values of that parameter were compared to it to determine the hazard ratios.

Table 7: *Hazard Ratios for Significant Variables, 2006 Cohort and CDM*

		Cohort				CDM			
Variable		**N**	**Departure Rate**	**P-value**	**Hazard Ratio**	**N**	**Departure Rate**	**P-value**	**Hazard Ratio**
FT_PT_group	**100% Full-time ***	**566**	**31%**			**146**	**49%**		
	100% Part-Time	358	72%	<.0001	1.700	166	80%	0.2132	1.265
	1-25% Full-Time	248	20%	<.0001	0.330	72	29%	<.0001	0.318
	26-50% Full-Time	434	23%	<.0001	0.453	102	33%	<.0001	0.382
	51-75% Full-Time	698	10%	<.0001	0.244	153	13%	<.0001	0.160
	76-99% Full-Time	507	3%	<.0001	0.079	142	3%	<.0001	0.051
gpa_group	**4.00 ***	**246**	**31%**			**81**	**37%**		
	0.00	154	94%	<.0001	6.159	113	97%	<.0001	7.431
	0.01 - 1.99	38	95%	<.0001	4.172	16	100%	<.0001	4.083
	2.00 - 2.49	45	87%	<.0001	3.381	23	91%	<.0001	3.363
	2.50 - 2.99	154	32%	0.1222	1.355	57	30%	0.3806	1.337
	3.00 - 3.49	528	23%	0.2247	0.826	154	28%	0.3881	0.806
	3.50 - 3.99	1,650	12%	<.0001	0.441	337	14%	0.0002	0.398
F_aid_used	0 = no aid *	1,618	25%			548	37%		
	1 = aid	1,198	22%	0.0001	0.696	233	35%	<.0001	0.512
P_aid_used	0 = no aid *	2,714	24%			763	36%		
	1 = aid	102	15%	0.0347	0.542	18	39%	0.0664	2.206
race_collapsed (not significant in overall model	White *					246	40%		
	Asian					66	33%	0.1459	1.432
	African American					71	52%	0.4226	1.188
	Foreign					161	20%	0.3283	0.786
	Hispanic					37	46%	0.4726	0.813
	Missing					198	37%	0.1204	1.313
	Other					2	100%	<.0001	21.198
age_group (not significant in overall model	Under 24 *					158	22%		
	24-29					284	36%	0.0406	1.541
	30-39					202	42%	0.8591	1.042
	40-49					61	52%	0.5409	1.188
	50 or older					18	33%	0.4748	0.712
	Unknown					45	33%	0.8214	1.209

An interesting finding was that students with a 4.00 GPA in both the overall cohort and in CDM were more likely to leave without a degree than those whose GPA was between 3.00 and 3.99, although this was statistically significant only for the 3.50-3.99 group. Furthermore, those in CDM with 4.00 GPAs were also more likely to leave than those whose GPA was between 2.50 and 2.99.

Another interesting finding was that students who were enrolled full-time in all terms in both the overall cohort and in CDM were significantly more likely to leave without a degree than all other groups except 100% part-time CDM students.

The results can be analyzed further by the following variables.

Full-time Status

Students who were 100% part-time were more likely to depart than 100% full-time students; this ratio was not significant for CDM students, but was for the entire cohort. All other groups, both for CDM and overall, were significantly *less* likely to depart than 100% full-time students.

Final GPA Group

Students with GPAs less than 2.5 were significantly more likely to depart, both in CDM and the overall cohort.

Federal Aid Used

Students using federal financial aid were significantly less likely to depart, both in CDM and the overall cohort. About 30% of CDM students used federal aid, compared to 43% of all students.

Private Aid Used

CDM students using private financial aid were significantly more likely to depart. However, as only 3% of CDM students used private aid, this should be interpreted with caution.

Race

Asian and African-American CDM students were more likely to depart, but this is not statistically significantly. Hispanic students were less likely to depart, but this is also not statistically significant. (The 'Other' group has only two students, so this should not be considered significant.)

Age Group

Among CDM students, all groups except 50+ were more likely to depart than the under-24 group, but this was significant only for the 24-29 group.

Discussion

As previously noted, students with 4.00 GPAs were more likely to leave without a degree than those with GPAs between 3.00 and 3.99. One theory advanced within the university is that CDM students may be more likely to leave to take jobs without finishing their degrees, although this is still under study. This may offer some evidence that high-performing CDM students have mastered certain skills though their coursework without having to complete a degree, although again, this warrants further study.

These results also offer evidence that federal financial aid does help a student remain enrolled, and that it is important to make graduate students aware of their aid options.

In this study, the survival analysis performed does help pinpoint groups that may benefit from institutional action to improve retention. While a certain amount of additional data cleaning and analysis is required, it is not difficult for the experienced analyst.

One area to explore in future analyses of this type is interaction effects. Possible interactions to be considered include those between financial aid used and full-time status, and those between international (foreign) student status and full-time status.

One model was run with the following interaction effects: gpa_group and FT_PT_group; gpa_group and retro_acadgrp; and FT_PT_group and retro_acadgrp (three variables significant in the original model). While the model completed, SAS warned that the information matrix was not positive definite, and thus the model's convergence was questionable. This is possibly due to the number of

groups in each interaction effects, some of which have no observations. Furthermore, while the three interaction effects were all significant, hazard ratios were not calculated, rendering interpretation of the results more difficult. A modified model may yield usable results.

An important caveat to bear in mind is that, since only the 2006 cohort was used, this analysis captures a single moment in time, so to speak. Given the ever-evolving demographic composition of the university, certain of those aspects may become more or less significant in more recent cohorts. Furthermore, the introduction of new programs, which has already had an impact on DePaul's enrollment, may in turn affect retention and departure. This is particularly true for CDM, as it is constantly evolving to remain abreast of computing technology and trends.

Another consideration is that, as survival analysis can account for individuals who have not seen the event in question occur before the end of the study, it would be statistically justifiable to run an analysis using a shorter study period. As DePaul's graduation rate for its most recent master's cohorts is greater than 50% after three years, analyses for that shorter period would be possible.

One more characteristic not included in this analysis, but worth exploring, is current employment status and tuition reimbursement status, particularly for part-time students.

Conclusions

The author hopes that this paper makes a contribution to the literature for master's student retention, and offers some guidance for others interested in applying these methods.

In addition to studying other cohorts, further extensions of this study may include investigating more thoroughly the possible impact of stopping out for one or more terms, and adding data on whether or not a student is continuing in the same field as their undergraduate degree.

As one possible approach to retention, the methods presented here show definite value. The use of statistical methods can be of great service in identifying possible areas of concern and providing avenues to improve retention. Given the financial stakes for the institution and the personal stakes for the student, rigorous statistical approaches can help align university strategies with enrollment and retention objectives.

Acknowledgements

The author wishes to thank Gerry McLaughlin, Susan Stachler, Joe Filkins and Liz Sanders, of DePaul's Office of Institutional Research and Market Analytics, for their editorial input.

References

Alexander, C., Kohnke, M. & Naginey, A. (2011). *Undergraduate and Graduate Retention Two Concepts, One Outcome*. Accessed May 23, 2014. http://www.callutheran.edu/fact_book/documents/UndergraduateandGraduateRetention_TwoConceptsOneOutcome.pdf

Haughton, D., Eshghi, A., Li, M., Senne, L., Skaletsky, M. & Woolford, S. (2011). Enrollment management in graduate business programs: Predicting student retention. *Journal of Institutional Research,* 16(2). Pp. 63-79.

Kleinbaum, D. and Klein, M. (2005) *Survival analysis: A self-learning text.* Second Edition. New York, NY: Springer Science+Business Media, Inc.

Murtaugh, P., Burns, L. & Schuster, J. (1999). Predicting the retention of university students. *Research in Higher Education, 40*(3). 355-371.

Simpson, O. (2005). The costs and benefits of student retention for students, institutions and governments. *Studies in Learning, Evaluation Innovation and Development*, 2(3), 34-43.

Singer, J. & Willett, J. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics, 18*(2), 155-195.

U.S. Bureau of Labor Statistics. (2013) "Employment by summary education and training assignment, 2012 and projected 2022.". Accessed June 1, 2014. http://www.bls.gov/emp/ep_table_education_summary.htm