The Pennsylvania State University

The Graduate School

College of Education

A STUDY OF COLLEGE STUDENT GRADUATION USING DISCRETE TIME SURVIVAL ANALYSIS

A Thesis in

Educational Psychology

by

Randall C. Deike

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2003

The thesis of Randall C. Deike has been approved* by the following:

Dennis M. Roberts
Professor of Educational Psychology
Thesis Adviser
Chair of Committee

Edmond Marks
Senior Research Associate

John D. Swisher
Professor Emeritus of Education

Robert L. Hale
Professor of Education
Head of the Department of Educational and School
Psychology and Special Education

* Signatures are on file in the Graduate School.

ABSTRACT

College student graduation rates are often used as a measure of institutional success, of institutional accountability, and are increasingly tied to resource allocation. Degree completion provides benefits to the student, the institution, and society. Understanding the graduation behaviors of baccalaureate degree seeking students has become critical to the management of institutional enrollments. The study of student graduation behaviors has been plagued with methodological concerns, especially when the inherent longitudinal nature of college student graduation is considered. The purpose of this study was to investigate college student graduation using a technique called discrete time survival analysis. Discrete time survival analysis addresses many of the concerns associated with the longitudinal study of graduation, such as the use of variables that change over time and censored cases, and allows for the investigation of timing of graduation. Preenrollment, enrollment, and financial aid variables were considered in modeling the timing of graduation for a cohort of freshman, baccalaureate students studied over a 12 year period.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am indebted to the members of my committee for their patience and support. Dr. Dennis Roberts has been my adviser since I began my graduate study. His support, insight, and perseverance have been critical to the completion of this project and are very much appreciated. I want to thank Dr. Edmond Marks for his support and willingness to continue as a committee member through a relatively long hiatus on my part; I appreciate his perseverance. Also, I want to thank Dr. John Swisher and Dr. Robert Hale for agreeing to serve as committee members late in the process. I appreciate their willingness to become familiar with the topic quickly.

I also owe a collective thanks to all those friends and colleagues who have pushed kindly, prodded, needled, applied huge levels of guilt, and threatened. Without their encouragement, in all of its forms, I would not have finished. Thank you.

Finally, to Andrew and Sarah, I am sorry for the missed softball games, soccer games, basketball games, dance recitals, and Saturdays. I promise to make up for this. Thanks for understanding.

Chapter I

INTRODUCTION

The study of college student retention has been of importance for researchers, administrators, and the higher education community for decades. Costs accrue to society, the institution, and the student when degree completion is not realized (Pascarella & Terenzini, 1991).

From a societal perspective, the attainment of a college degree helps to mediate background resources, such as family socioeconomic status, on subsequent occupational status, potential earnings, and social status attainment (Pascarella & Terenzini, 1991). This is evidenced by the differential social status attainment of individuals from the same socioeconomic status with different levels of educational attainment (Pascarella & Terenzini, 1991). College graduates earn more than non-graduates, resulting in an increase in the local and federal tax base. College graduates are more likely to vote. In the 1998 congressional election, college graduates between the ages of 25 and 44 were 77% more likely to vote than high school graduates. And, high school dropouts were 72% less likely to vote than high school graduates. In the 1996 presidential election, college graduates between the ages of 25 and 44 were 70% more likely to vote than high school graduates, and high school dropouts were 49% less likely to vote than high school graduates (U.S Department of Education, 2000).

From the institution's perspective, maintaining enrollments is important to economic stability. As the number of new students has fluctuated, so has the importance of understanding the characteristics of those students most likely to remain enrolled and graduate. Institutions realize that the initial resources required to recruit, enroll, register, advise and support a new student are the same whether that student remains and graduates or not. For example, marketing and publication expenses, costs associated with maintaining a staff of professional counselors, travel costs associated with conducting college fairs and informational meetings, costs associated with telecounseling, staff time contacting students for yield enhancement, and staff time conducting academic advising, are all examples of preenrollment costs incurred by the institution. Having a better understanding of which students are more likely to enroll and remain enrolled to graduation is important in maximizing those preenrollment resources and generating revenue.

Understanding student graduation behaviors can affect the development or enhancement of retention strategies, leading to a higher graduation rate, and providing attrition prone students with support and guidance. Also, graduation rates are increasingly used as a measure of an institution's performance and are sometimes linked to resource allocation. Graduation information is required as part of the Common Data Set, a consistent set of information provided by individual institutions as a way for prospective students and their families to make fair comparisons among institutions. Graduation information is also required by other publications, such as U.S. News and World Report, to rank institutions and provide an additional resource for comparison. The state in which the study institution is located has already created a $6,000,000 grant program used to reward colleges and universities that graduate at least 40% of their state resident students within four years. The current federal administration has proposed a similar program as part of the reauthorization of the Higher Education Act, the legislation driving the administration of federal student aid. As part of the proposal, institutional accountability, in the form of better persistence and graduation rates, would be tied to federal grant funds. The current Higher Education Act is scheduled to expire in 2003. Higher education administrators are concerned that the administration of federal student aid could be linked to how well an institution is doing at retaining and graduating its students.

Since 1980, tuition and fee costs have increased at rates that are 2 and 3 times greater than the CPI. Over the 10 year period ending in 2000-2001, and after adjusting for inflation, costs at public four year institutions increased an average of 51%, and costs at private institutions increased an average of 35% (The College Board, 2000a). With decreases in state funding for higher education becoming a national trend, the cost of a college education will most certainly continue to rise. Large public research universities are most susceptible to decreases in state funding and have experienced an average 14% decline in state spending from 1986 to 1996, compared to a doubling of the share for Medicaid and a 25% average increase for correctional facilities (Yudof, 2002). Managing enrollments relative to an institution's fiscal viability coupled with continued increases in cost to students and institutions requires a sound understanding of retention and graduation behaviors.

Students and families faced with financing higher education are finding that a more substantial portion of their aid package reflects loan as opposed to gift aid. From academic year 1989-1990 to academic year 1999-2000, total aid has increased by approximately 90% in constant dollars, with two-thirds of that increase being attributable to loans (The College Board, 2000b). The affordability of college is a central issue in higher education today and most certainly impacts student retention and graduation. A better understanding of the relationship between student aid and student retention is needed if institutions are going to manage their enrollments. As noted, the increasing cost of education and decreasing gift aid available to finance costs have resulted in an increase in loan debt for students. The ability to repay those loans is greatly enhanced if a degree is earned.

College graduates earn more than high school graduates. In 1998, men who had completed at least a bachelor's degree earned 56% more than men who completed high school, and for women, the difference was 100%. Men who did not complete a high school education earned 30% less than men completing a high school degree, and the difference for women was 31% (U.S. Department of Education, 2000). The benefit of completing a college degree has increased over time. In 1975, a male college graduate earned $15,000 more annually than a high school graduate; that difference has increased to $32,000 today after adjusting for inflation. In the course of a lifetime of earnings, a college graduate will earn on average $1,000,000 more than a high school graduate and an individual with a professional degree will earn on average $3,000,000 more than a high school graduate (Yudof, 2002). Degree completion can result in a substantial increase in future earning potential.

Understanding graduation behaviors becomes more important as the number of college bound students increases. Institutions are faced with the potential for significant enrollment increases. Total enrollments in degree-granting two year and four year institutions are projected to increase by approximately 1.3 million students from fall 2000 to fall 2009; at four year institutions by 858,000 and at two year institutions by

406,000 (U.S. Department of Education, 2000). Changing demographics require an institution to better understand retention and graduation behaviors to meet enrollment goals. For institutions with enrollment caps, identifying those students most likely to succeed, defined by graduation, is important in their admission policies and economic stability. Especially when focusing on full-time baccalaureate degree seeking students, graduation increasingly becomes a criterion of importance. Although there are always competing goals that dictate admission policies, such as diversity in all of its forms, athletics, band, and the administration of financial aid, institutions strive to increase retention and graduation rates. Many institutions face an overwhelming number of applications for a very limited number of spaces in a freshman class. Given the benefit of earning a degree to the student, the institution, and society at large, it is a worthwhile goal to consider what is known, and what can still be learned, about retention and graduation when making admission decisions.

Not every student, however, can be admitted to a single institution. The acceptance rates from one institution to another vary drastically. Some institutions operate under an open policy, where every applicant is admitted. Others manage their freshman enrollments by adjusting the number of applicants admitted based upon the retention and graduation rates of upper-class students to reach an overall enrollment target. Although not likely, if the graduation rate is perfect, the freshman class would be reduced to maintain the overall enrollment. Although fewer students would be admitted at a particular institution, there would be other opportunities for those students elsewhere. Although unfortunate, many students are not admitted to their first choice institution.

The degree to which freshman enrollment targets would be adjusted if perfect graduation rates occurred would also vary drastically by institution. Continuing to consider full-time baccalaureate degree seeking students, one large public university in the northeast reported admitting 78% of its applicants with 29% of the admitted students enrolling. The same institution reported a four year graduation rate of 19% and a six year graduation rate of 44% for the entering class of 1995. Another large public institution in the northeast reported admitting 57% of its applicants with 38% of the admitted students enrolling. The four year graduation rate for this institution was 45% and the six year graduation rate was 81%. One private institution in the northeast reported admitting 23% of its applicants with 51% of the admitted students enrolling. This institution reported a four year graduation rate of 43% and a six year graduation rate of 95%. So, while institutions strive to increase retention and graduation rates because of the benefits to society, the institution, and the student, disparity in success is great.

Researchers studying college student retention have considered pre and post-enrollment characteristics including demographic factors, socioeconomic status, size and location of high school, SAT/ACT scores, high school GPA, class rank, financial aid, college grades, motivational factors, parental influence, parental education level, personality factors, college environment, student-faculty relationships, and extracurricular activities to name a few. The results of determining the relationship between these factors and student retention in many of these studies have been mixed due in no small part to differences in operational definitions, measurement, and methodology.

Terenzini (1987) indicates that the mission of the university must be considered in defining retention and describing its effects. Different definitions of dropout, stopout, and graduate have contributed to the mixed results of many retention studies. Summerskill (1962) indicated that the term attrition rate has been defined from a department or division level, from the college level and from a higher education or system level. Each of these stakeholders has a different perspective on success, retention, and graduation, leading to different approaches to determining the correlates of those outcomes.

The methodological approaches used to study student retention have also varied drastically. Although described as a longitudinal process, most studies of student retention have employed what equates to a cross-sectional methodology, or at best, a pre-post design. A cohort of students is selected. After a certain period of time has elapsed those students still enrolled are compared, on factors of interest, to those students who are not enrolled. This dichotomization, however, completely ignores the temporal nature of student retention by collapsing information across time, which potentially masks important behaviors. When students withdraw, or more importantly, the points at which students are at most risk of withdrawing, is a more meaningful question. Studying the temporal nature of student retention has been identified as lacking in the student retention literature. Tinto (1988) stated:

> …researchers have in fact done very little to explore the temporal dimensions of that process. Rather than pursue that possibility, past research has implicitly assumed that the process of student departure is essentially invariant over the course of the student's career. (p. 438)

Discrete time survival analysis is a statistical technique that allows timing of event questions to be framed appropriately. Willett and Singer (1991) describe the benefits of using survival analysis to study college student retention. This approach leads to a reframing of the retention question from whether students leave a particular college or university to when are they at most risk of leaving. Although well suited as a statistical technique for studying student retention, survival analysis has been used very infrequently in educational research. Willett and Singer (1991,1993) and Singer and Willett (1993) attribute this to a lack of statistical methods designed to answer the when question associated with event occurrence. With the extension of survival analysis from biostatistics to educational research, the methods now exist. Also, statistical software packages now make the application of discrete time survival analysis less cumbersome.

There is a small but growing body of retention literature where survival analysis has been used to study the timing of student departure (DesJardins, 1996; DesJardins, Ahlburg & McCall, 1999; DesJardins & Moye, 2000; Moore, 1994; Murtaugh, Burns & Shuster, 1999; Ronco, 1995). In most cases, time to first dropout has been the main criterion of interest. Adelman (1999), in a much touted recent study, considers the retention behaviors of the 1980 High School and Beyond/Sophomore cohort, and focuses on what contributes most to completing a bachelor's degree for students who attended four year colleges or universities. This cohort is tracked via transcript information over approximately 13 years. Adelman indicates that degree completion, and not retention rates per se, should be the criterion of interest:

> Degree completion is the true bottom line for college administrators, state legislators, parents, and most importantly, students—not retention to the second year, not persistence without a degree, but completion. (p. v)

Although graduation has been the outcome of interest in some of the retention literature, a cross-sectional approach with two times of measure has been used, and the time frame for analysis has been primarily four years. Time-to-degree has been increasing, with fewer students completing a degree in four years. Much of the reason for determining a specific end point for studying graduation comes from the lack of a consistent method for studying timing to graduation. Survival analysis lends itself well to this conceptualization of student retention, or more appropriately for the present study, graduation. The criterion of interest becomes timing to graduation and the outcome addresses when students are most likely to graduate and what are the characteristics of those students who graduate vs. those who do not.

The Focus of this Research

The purpose of the present research is to add to the growing literature using survival analysis as a methodology to study the temporal nature of student retention by focusing on graduation as the outcome of interest, and by incorporating the findings of Adelman (1999). The cohort under study will be first-time, full-time, freshman baccalaureate degree seeking students who enrolled at a large public university in the northeast in the fall of 1990. Their undergraduate enrollment histories will be studied over a 12 year period. Studying student graduation using survival analysis will provide researchers and practitioners alike with insights into the timing of graduation and the characteristics of those students most likely to complete a baccalaureate degree. Policy makers will have a better understanding of the characteristics of students at risk of not graduating, which could lead to better developed retention strategies and admission policies.

Chapter II

REVIEW OF THE LITERATURE

Introduction

The study of college student retention has been part of the literature in higher education for decades. Summerskill (1962) indicated that research on college student dropouts had at least a 40 year history at the point of his seminal review of the existing literature. Even in this early review of the literature, Summerskill identifies three origins of the interest in college student attrition research. First, colleges and universities faced accountability to students, parents, and legislators, and were criticized for doing a poor job if they did not retain and graduate their students. Second, as institutions grew in size, there was an increased interest in economic efficiency, with administrators being charged to manage large operational budgets. Summerskill indicates that student attrition results in the waste of time, energy, and money for both the student and the institution. Third, most institutions are primarily tuition driven, and loss of students, which can result in underenrollment, can cause severe budgetary issues. These reasons for interest in attrition research are just as relevant today.

Theories of College Student Departure

Tinto's Student Interaction Theory

Prior to the early 1970's, much of the student departure research was described as descriptive and atheoretical (Terenzini, Pascarella, Theophilides, & Lorang, 1985). Tinto (1975), building on the work of Spady (1970), developed a theoretical model of student departure based on Durkheim's (1951) theory of suicide. No review of retention literature would be complete without a discussion of Tinto's model, as it is the most cited and researched model of student departure.

Tinto (1975) describes dropout as a longitudinal process influenced by the interaction of preenrollment characteristics (family background, ethnicity, academic abilities, etc.) which have direct and indirect effects on goal commitment (graduation) and institutional commitment. Once enrolled (becoming part of the academic and social system of the institution) goal commitment leads to better grade performance and intellectual development, and institutional commitment leads to peer group and faculty interactions. The better a student performs academically, the more academically integrated that student becomes, and the more interactions with peers and faculty the student experiences, the more socially integrated that student becomes. The degree to which a student experiences academic and social integration affects the level of goal commitment and institutional commitment which then affects the dropout decision. Tinto asserts that the more socially and academically integrated into the specific institutions academic and social systems a student becomes, the less likely that student is to drop out. Tinto describes his model as an institutional as opposed to a systems model of dropout:

> In each instance, these different conceptual frameworks are applied to a model of dropout that seeks to explain dropout from institutions of higher education, not one that seeks to explain dropout in the system of higher educational institutions. It is, then, an institutional rather than a systems model of dropout. (p. 91)

Conceptualizing retention as a "longitudinal process of interactions that lead differing persons to varying forms of persistence and/or dropout behavior" (Tinto, 1975, p. 93) was one of the main contributions of Tinto's model. Tinto indicated that future research should use a longitudinal, as opposed to a cross-sectional methodology. Also, Tinto argued that the definition of dropout was critical in the study of student retention, making the distinction between voluntary (in good academic standing) and nonvoluntary (academic dismissal) dropouts.

Research conducted to test Tinto's (1975) model is vast and mixed. It is also incomplete as it is difficult to test the five major constructs and directions of influence comprising Tinto's model, simultaneously. Pascarella, Duby, and Iverson (1983), in a nonresidential setting, found results inconsistent with Tinto's model, with social integration having a direct negative effect on persistence. Pascarella and Terenzini (1983) found general support for Tinto's model, but found gender differences in how academic and social integration affected persistence.

In a recent comprehensive review of multi-institutional and single-institutional studies testing the influence of academic integration on subsequent institutional commitment and persistence, Braxton and Lien (2000) concluded that results varied by type of study. Multi-institutional studies provide support for the effect of academic integration on both persistence and subsequent institutional commitment while single-institutional studies provide, at best, modest support. Braxton and Lien offer two possible options for researchers studying student departure; abandon the construct of academic integration or rethink how academic integration is being measured. They indicate that "little consensus exists among scholars on the meaning of academic integration" (Braxton & Lien, 2000, p. 13) which leads to very different measures used as indicators. This is true for the ways in which other components of Tinto's (1975) model are measured as well.

Bean's Student Attrition Model

Often described as a competing model of student departure, Bean (1980) bases his model of student attrition on research in the area of worker turnover, and argues that students leave institutions for reasons similar to those causing workers to leave particular organizations. His model is an adaptation of Price's (1977) model of turnover in work organizations. Bean indicates that background variables in the form of past academic achievement, socioeconomic status, state residence, distance from home, and hometown size influence organizational determinants. Organizational determinants such as routinization, student development, practical value, and university GPA influence the student's satisfaction with the institution, which in turn influences institutional commitment. The higher the level of institutional commitment the student experiences, the lower the likelihood of the student dropping out.

Bean (1983) adjusted his model of student attrition by eliminating student background variables (such as past academic achievement and socioeconomic status) and including additional variables (such as intent-to-leave) that were consistent with a revised model of worker turnover developed by Price and Mueller (1981). Bean described his new model as the industrial model of student attrition. Bean found that intent-to-leave accounted for the largest proportion of explained variance in his model.

Bean (1982) has also offered a synthetic model of student attrition where components of the student integration and student attrition models are combined. In this model, background variables (such as high school grades, standardized test scores, and hometown size) directly influence objective interaction with the institution, the environment, outcomes and attitudes, and dropout. The objective interaction with the institution influences outcomes and attitudes, intent-to-leave, and dropout. Environmental factors (such as family approval of institution, family approval of major, and difficulty of financing education) also influence outcomes and attitudes, intent-to-leave, and dropout. Outcomes and attitudes influence intent-to-leave and dropout. And, intent-to-leave influences dropout.

Convergence of the Student Integration and Student Attrition Models

Cabrera, Castaneda, Nora and Hengstler (1992) also offer a model which combines aspects of both the student integration and student attrition models. Cabrera et al. found that a more comprehensive understanding of student attrition could be achieved if the student integration and student attrition models were combined. However, their results were based upon a sample of 466 students, which limited their use of confirmatory factor analysis, and subjects their conclusions to a cautious interpretation. In a follow-up study, Cabrera, Nora, and Castaneda (1993) attempted to move beyond identifying the overlap in the two models to testing an integrated model of student attrition using the same 466 subjects reported in the earlier study. They conclude that integrating the two models does provide a better understanding of student attrition.

The student integration and student attrition models have been the impetus for the vast majority of the research on student attrition. Empirical support for these models has been mixed, however. The underlying premise for both models is that the more a student is integrated into the social and academic culture of the institution, the more likely that student is to persist. Also, the influence of background, or preenrollment characteristics, on persistence is secondary to social and academic integration.

In a recent comprehensive review of attrition and persistence in higher education, Kennedy and Sheckley (1999) reviewed 125 articles from three general thematic areas in the literature. The first theme was described as demographic variables related to students and their families. The second theme was described as academic achievement, self-efficacy, and related personality characteristics. And, the third theme was described as the interactions between students and the college environment. They used percentage of variance explained as the outcome. Persisters were defined as individuals who remained enrolled in their original degree programs and completed their program of study within four or five years.

Kennedy and Scheckley (1999) indicate that demographic information accounts for a small percentage of explained variance associated with attrition, with parental support having the greatest impact. In reviewing articles related to academic achievement, self-efficacy and related personality characteristics, they suggest that students who persist are free from financial concerns, have living arrangements that connect them socially with other students, are well-adjusted psychologically, possess a strong belief in their academic abilities, and are academically successful. In terms of interactions between the student and the institution, Kennedy and Scheckley indicate that this relationship, or institutional fit, explains attrition behavior most consistently and also accounts for the most variance. They make their argument by pointing to widely accepted models of persistence. However, they also indicate that these models leave over 50% of the variance associated with retention unexplained.

Kennedy and Scheckley (1999) suggest that researchers might expand the methods used to study student attrition to include truly longitudinal designs, citing cross-sectional methodologies used to study what has been accepted as a longitudinal process, as problematic. They also suggest that the self-reported information used to assess perceptions, attitudes, or intentions is also problematic.

Although Kennedy and Scheckley (1999) indicate general support for academic and social integration, the extent to which the practical application of that research is possible for college and university administrators is extremely limited. Extensive surveys or data collection techniques are required to collect information associated with, for example, social integration (formal and informal contacts with faculty and peers), goal commitment (in the form of commitment to graduate), outcomes and attitudes (such as boredom, satisfaction, confidence, educational goals), and intent-to-leave. Although the research to date provides insight into the retention issue, its practical application at the institutional level is nearly impossible. Adelman (1999), in addressing this very issue concerning graduation, indicates:

> Both research traditions place an extraordinary emphasis on psychological variables: intentions, attitudes, influences, commitment, perceptions…These variables unfortunately refer to realities that lie beyond the control of those who can best steer students toward degree completion. (p. 27)

Much of the student retention research has focused on testing academic and social integration, or institutional fit, as defined in both Tinto's (1975) and Bean's (1980, 1982, 1983) models of student departure. Although there is moderate support for the benefit of institutional integration as a primary correlate of student departure, the results are mixed. Also mixed are the results assessing the benefit of background characteristics such as socioeconomic status, academic preparedness, distance from the institution, and ethnicity, to name a few.

Some of the discrepancies in the literature can be attributed to very different methodological approaches to studying student attrition. Much of the early research on college student attrition was criticized for its descriptive nature. Gekowski and Schwartz (1961) indicated, when studying dropouts, that many studies did not include a comparison group of students who remained enrolled, so there was often no control group. Gekowski and Schwartz also identified the univariate or bivariate nature of early studies as limiting, indicating that a more multivariate approach needed to be taken, as did Spady (1970). Jex and Merrill (1962) advocated for the use of longitudinal studies over cross-sectional studies. Longitudinal studies allow the researcher to study the effects of covariates on attrition at the time they occur. Also, a longitudinal study at a single institution, especially over a longer period of time, allows for the identification of stopouts. Pantages and Creedon (1978), in summarizing their review of research between 1950 and 1975, indicate that:

> The preponderance of the two-way analysis that distinguishes only between those students who graduate in four years and those who do not, combined with ex post facto methodology in attrition research, has obscured many important details of student withdrawal and has inflated estimates of attrition rates. (p. 92)

Pantages and Creedon also indicate that combining attrition data derived from different institutions is problematic, given the enormous variation in the attrition rates of individual institutions. The two most popular models of student attrition are both described as institutional models, as opposed to higher education system models, providing support to the notion that attrition should be studied at the institutional level. Terenzini (1982) states:

> Literature reviews provide a general and useful understanding of the complexity of the attrition phenomenon, but in no way do they substitute for the local research many institutional research offices are asked to undertake. (p. 55)

Also problematic in the study of student retention is the definition of the outcome. The definition of dropout is specific to the perspective of the researcher.

Dropout has been defined in multiple ways. It is difficult to generalize across studies with different definitions of dropout.

The methodological difficulties of student retention research and new methodologies that address these difficulties will be presented next.

Methodological Issues in the Study of Student Retention and Graduation

Defining the Criterion

In his review, Summerskill (1962) addresses issues associated with defining attrition and the potential covariates identified in the research. He indicates that the criterion measure varies from study to study, with attrition rate having been defined as students lost to a particular academic division, lost to the college or university as a whole, lost to the higher education system altogether, computed on a graduated-in-four-years basis, computed on a graduated-in-four-years-or-still-enrolled basis, and computed on a graduated-eventually basis. He also indicates that terms such as "holding power", "student mortality", "involuntary vs. voluntary withdrawal" and "net vs. gross attrition" added to the complexity.

In their comprehensive review of the college student attrition literature between 1950 and 1975, Pantages and Creedon (1978) indicate that a consistent definition of dropout does not exist. They state that the validity of summarizing or combining the results of multiple studies depends upon the degree to which consistent definitions have been used. They also indicate that many studies were cross-sectional in nature and defined dropout as any student previously enrolled at the institution who was not enrolled at the time of the study, leading to questionable results as distinctions were not made between permanent and temporary dropouts. If, as was often the case, studies were conducted at the end of the first semester or first year, a student no longer enrolled would be classified as a dropout, without considering their potential reenrollment at a later date. Some researchers make a distinction between dropouts, those students who leave the sample and never return, and stopouts, those students who leave the sample but return either before or after the study period. Some researchers defined dropout from a higher education systems approach, some from an institutional perspective, and some from a departmental perspective. Tinto (1982) in addressing the operational definition and differential interests of a system vs. institution perspective of dropout states:

> We will ask not only how dropout may be defined, but also how that definition may vary among different interested parties concerned with the character of dropout from higher education. (p. 3)

Prediger (1965), and Rose and Elton (1966), advocate a more specific definition of dropout based upon four categories of academically successful and unsuccessful persisters and dropouts. Panos and Astin (1968) define dropout vs. nondropout as those students who completed four years of college, whether or not they had graduated or were still enrolled at the institution of first enrollment. Others have indicated that dropouts should be categorized as voluntary and nonvoluntary, making distinctions between those who withdrew from an institution in good academic standing vs. those who were academically dismissed. Tinto (1975) indicates that it is important to make distinctions between the varying forms of dropout, especially between academic dismissal and voluntary withdrawal. Pantages and Creedon (1978) state, however:

> We believe that basing the distinction between "voluntary" and "nonvoluntary" withdrawal on scholastic achievement is inappropriate, since it ignores the factors that have caused poor academic performance. It is these factors that influence the decision to drop out, not the end result of these factors (i.e., poor grades). (p. 52)

Pantages and Creedon recommend that this approach be abandoned. Simpson, Baker and Mellinger (1980) argue that their findings provide support for Tinto's distinction between voluntary and nonvoluntary withdrawal.

More recently, Adelman (1999), in studying the persistence and bachelor's degree completion of students enrolled in four year institutions, suggests that graduation, and not persistence rates, be the outcome of interest when studying college student attrition:

> Degree completion is the true bottom line for college administrators, state legislators, parents, and most importantly, students—not retention to the second year, not persistence without a degree, but completion. (p. v)

Adelman's definition of the outcome of interest, graduation, eliminates much of the concern with inconsistencies in the definition of dropout. Adelman studied the 1980 High School and Beyond/Sophomore cohort by reviewing student records over an approximately 13 year period. Although using graduation as the outcome of interest is not completely novel, few studies have studied graduation behaviors for such an extended period of time.

Defining dropout has not been, nor does it continue to be, an easy matter. Differences of opinion concerning whether distinctions should be made between voluntary vs. nonvoluntary withdrawal still exist, as do concerns about the length of time under study.

Adelman (1999) has suggested that graduation, and not dropout or retention, be the outcome under study. This conceptualization eliminates many of the issues associated with defining dropout. There are those who would argue that some students do not intend to graduate and that this definition is just as problematic. Adelman, however, studied students who enrolled in four year colleges and universities as a measure of their intent to earn a bachelor's degree. As described in the introduction of the present study, there are benefits to society, the institution, and especially the student when a degree is realized. Graduation will be the outcome of interest in the present study.

Methodologies

Almost without exception, researchers studying retention and graduation have defined a particular cohort of interest, determined, or had determined by circumstances, a time period of study, and then compared those who have persisted or graduated with those who have not. Often, that time period is one semester or one year. The timing of retention or graduation has been virtually ignored. Studying the temporal nature of student retention has been identified as lacking in the student retention literature. Tinto

(1982) indicates that dropout is not invariable across student populations or across time. He states:

> While much work has been done to reveal the multiple associations between rates of dropout and the attributes of individuals (for example, sex, race, ability, and social origins), virtually no attention has been given to the possibility that the longitudinal process of dropping out may itself vary among different groups of students (for example, blacks and whites). (p. 7)

Pantages and Creedon (1978), after reviewing 25 years of attrition research, suggest that:

> The most meaningful research on attrition is provided by studies that cover a period of more than four years and that use precise operational definitions of dropout and nondropout. (p. 92)

The need to incorporate timing of withdrawal in retention research was identified as early as 1965. Barger and Hall (1965) used the week during which a student withdrew in the winter trimester of 1964 to dichotomize dropouts as early or late and compared the two groups on a number of characteristics. This early study illustrates several of the problems with most retention research where the timing of dropout has been considered.

This two-wave prospective study was conducted using only those students who withdrew, which potentially biases parameter estimates as a result of excluding those students who were still enrolled. The difficulty with including students still enrolled comes as a function of determining the time to dropout. Since those students had not yet experienced the event of interest, dropping out, they could not be categorized as early or late. These students had censored event times, creating a methodological dilemma when trying to study time as an outcome, which led to their exclusion in the Barger and Hall (1965) study. Some researchers have attempted to correct this issue by imputing time-to-event as the total time of the study. Barger and Hall (1965) studied dropout behavior during one trimester. Certainly the time to dropout would be greater than one trimester for those students still enrolled. The difficulty is in determining how long. Also, the one trimester study period could have been two trimesters, two years, or ten years.

Barger and Hall (1965) also studied students from several cohorts over one semester, confounding cohort and grade effects (see Baltes, 1968; Schaie, 1965). Willett and Singer (1991) recommend a multiwave or longitudinal methodology, where one cohort of students is followed over multiple years. Terenzini (1987) states:

> Longitudinal designs provide extensive, planned control of confounding background variables, as well as more precise estimates of the institutional influences of attendance behavior. Such designs are the most internally valid available for studying attrition and afford a measure of confidence in findings and associated conclusions that is not available with other designs. (p. 29)

Following a cohort of students over time presents several methodological issues, especially when the temporal nature of dropout behavior is of interest. The timing of event occurrence becomes the criterion, but defining time can be problematic. If a cohort of students is followed semester by semester and the outcome is timing of dropout for example, what measure of time is assigned to those subjects who do not experience the event? As mentioned above, these cases are censored, and must be treated in a way that allows for the inclusion of information about them but that also allows for the appropriate measure of time. Discrete time survival analysis is a statistical technique that allows timing of event questions to be framed appropriately.

Willett and Singer (1991) describe the benefits of using discrete time survival analysis to study college student retention. Survival analysis allows for the inclusion of the entire cohort under study, both those students who experience the event and those who do not, which addresses censoring. This approach leads to a reframing of the outcome under study from whether students leave a particular college or university to when are they at most risk of leaving. Although well suited as a statistical technique for studying student retention, survival analysis has been used very infrequently in educational research. Willett and Singer (1991, 1993) and Singer and Willett (1993) attribute this to a lack of statistical methods designed to answer the when question associated with event occurrence. With the extension of survival analysis from biostatistics to educational research, the methods now exist. Also, statistical software packages now make the application of discrete time survival analysis less cumbersome. There is a small but growing body of retention literature where survival analysis has been used to study the timing of student departure (DesJardins, 1996; DesJardins, Ahlburg & McCall, 1999; DesJardins & Moye, 2000; Moore, 1994; Murtaugh, Burns & Shuster, 1999; Ronco, 1995)).

Although much of the student retention research is described as longitudinal, it is barely more than cross-sectional. Many of these studies, incorporating two times of measure, benefit from the fact that a cohort of students is identified at time 1 and compared at time 2, but lose important information about the timing of dropout. With two times of measure, even if, for example, the time period is six years, it is impossible to determine if the student left the institution in the first week, the first semester, the first year or the sixth year. The most that can be determined is that the student was no longer enrolled at the end of the period of study. It is reasonable to assume that retention is not time invariant, especially across groups. A methodology now exists to incorporate timing of the event, graduation in the present study, as part of the outcome. Discrete time survival analysis will be used to study the timing of graduation behaviors, and a more complete description of this method will be provided in Chapter III. The study of student retention using survival analysis will be presented next.

Studying Student Retention and Graduation Using Survival Analysis

Willett and Singer (1991) describe the benefits of using survival analysis to study the occurrence of life events such as the entry into and exit from daycare, the entry into and exit from special education, the entry into and exit from the teaching profession, dropout and graduation from high school, and college student retention and graduation. Willett and Singer offer four reasons why the use of survival analysis is justified over more traditional approaches to studying event occurrence.

First, they indicate that the outcome of using traditional methods is inextricably linked to the time frame chosen for analysis, and that the time frame is rarely substantively motivated. If for example a five year time frame for graduation is chosen, the resulting graduation rates indicate the cumulative difference in graduation without accounting for the timing of graduation during that five year period. All other variation over the five year time period is lost.

Second, contradictory conclusions can result as a function of nothing more than the chosen time frame. If for example the same covariates are used to study graduation behaviors in two different studies, one using a one-semester time frame and the other using a five year time frame, the conclusions could be contradictory as a function of the potential variability in timing of graduation. In the research on student retention and graduation, this is clearly problematic.

Third, traditional methods are unable to address censored observations, or those subjects who do not experience the event of interest during the study period. When timing of the event is of interest, properly incorporating censored observations into the analysis is critical. Survival analysis allows for the inclusion of censored cases.

Fourth, traditional methods offer few mechanisms for the inclusion of time-varying covariates, or variables whose values vary from one time period to another. In studying student retention, variables such as GPA, and amounts of student aid vary over time. Researchers have tried to address time-varying covariates by imputing the average value over the time period. Much information is lost taking this approach.

Survival analysis clearly addresses many of the methodological concerns related to the longitudinal study of student retention and graduation. As noted above, the varying time frames and two times of measure associated with much of the existing research masks the timing and temporal nature of covariates related to dropout and graduation.

Although limited, survival analysis has been applied to the study of retention and graduation (DesJardins, 1996; DesJardins, Ahlburg & McCall, 1999; DesJardins & Moye, 2000; Moore, 1994; Murtaugh, Burns & Shuster, 1999; Ronco, 1995). Ronco (1994) studied the effects of ethnicity and full vs. part-time status on student departure over a 20 semester time period, finding that the greatest risk of departure was after the second semester and that Hispanic students were more likely to return after their first stopout than their White non-Hispanic counterparts. Ronco did not, however, include preenrollment demographic or academic variables, or measures of academic success such as GPA, so potentially important covariates were excluded from the model. Ronco (1995) implemented a competing risks model with dropout, transfer, and graduation as the outcomes of interest. In this study, ethnicity, gender, provisional vs. regular admission, age, semester GPA, major, and full vs. part-time enrollment were considered as covariates. Ronco found that the risk of transfer was almost as high as the risk of dropout. Provisional admits and students with low semester GPAs were at greatest risk of dropout and were less likely to graduate. Students with higher GPAs were more likely to graduate and less likely to dropout or transfer. However, the semester GPA was converted to an ordinal variable with three levels, which could mask important effects had the variable remained continuous. Hispanic students were less likely to transfer to a four year school, to graduate, or to drop out, but were more likely than non-Hispanic students to transfer to a two year institution. Full-time students were less likely to drop out or transfer, and over twice as likely to graduate. It is important to note that the transfer behaviors identified in this study are restricted to same state transfers, which limits the interpretation.

Murtaugh, Burns, and Schuster (1999) found that students with higher first quarter GPAs were less likely to withdraw, found significant college effects, found that African American students were less likely to withdraw than White students, and that graduation rates decrease with age. However, this study combined data from five fall semester cohorts (fall 1991, fall 1992 fall 1993, fall 1994, and fall 1995) and used the same end-of-study period, fall 1996, to determine the outcome. So, the number of years enrolled becomes the outcome. This is extremely problematic in that the risk set for each time period is inflated by the subsequent cohort. For example, the time period for the fall 1991 cohort could be five years. However, the maximum possible time period for the fall 1995 cohort is one year. A serious bias could result due to inflating the risk set for each time period when an artificial limit on the outcome, number of years enrolled, is imposed. In other words, it is not possible for the fall 1995 cohort to reach two years, for the fall 1994 cohort to reach three years for the fall 1993 cohort to reach four years, and for the fall 1992 cohort to reach five years. Also, collapsing across several cohorts confounds cohort and grade effects (see Baltes, 1968; Schaie, 1965).

DesJardins (1996) and Desjardins, Ahlburg, and McCall (1999) used survival analysis to study 3,975 new students beginning fall term 1986 at the University of Minnesota, Minneapolis campus. The time period under study was 22 semesters. Data for the study were collected from the University's data bases. In comparing their work to other research on college student departure, DesJardins et. al. indicate that the use of data readily available to institutional researchers may be advantageous compared to attitudinal data used by other researchers, citing non-response to surveys, resulting in reduced sample sizes, and costs, as deterrents. This study represents one of the most comprehensive studies of student departure using survival analysis to date, based upon the large sample size and the inclusion of variables identified in the existing literature.

DesJardins, Ahlburg, and McCall (1999) also include aid related variables, a factor missing in other studies using survival analysis. Their inclusion of aid variables is also disaggregated, allowing for the study of individual aid amounts by type. They indicate this makes an important contribution to the research on student departure, as many studies have used attitudinal data, usually captured via survey upon matriculation, to address aid related effects. They indicate that using disaggregated aid amounts over time allows for the study of the temporal nature of the effects of financial aid on student departure.

In terms of dropout behavior, DesJardins, Ahlburg, and McCall (1999) found that African American students were more likely to dropout than Whites in year one, and that there were no significant differences between Hispanics and White students. Asian students were less likely to drop out than their White student counterparts in the first two years of enrollment only. They found no gender differences. High school rank shows no significant effect on dropout, but the higher a students ACT score the less likely the student is to drop out in year two. Students from the local area were less likely to drop out in year one. Age was positively associated with dropout in year one. There were differential dropout effects by major. Semester GPA had a negative effect with students earning higher GPAs being less likely to drop out. However, the GPA effect diminished over time. Financial factors were also found. Loans were most likely to reduce the risk of dropping out in year three, and year five and beyond. The effect of student employment was quite constant and reduced dropout risk at year two and years four through seven. Work study reduced the risk of dropping out in year one only, and scholarships helped reduce the risk of dropping out in year three. The effects of grants were not significant.

Although the research using survival analysis to study student departure has at some minimal level addressed graduation, the real focus to date has been on dropout and stopout behaviors. As noted previously, the focus of the current study will be on graduation. Adelman (1999), in one the most comprehensive studies of graduation to date, indicates that degree completion, and not retention rates per se, should be the criterion of interest:

> Degree completion is the true bottom line for college administrators, state legislators, parents, and most importantly, students—not retention to the second year, not persistence without a degree, but completion. (p. v)

Adelman (1999) also echoes the sentiments of DesJardins, Ahlburg, and McCall (1999) concerning the use of variables readily available to institutional researchers and administrators when discussing the student integration and student attrition models of student departure by stating:

> Both research traditions place an extraordinary emphasis on psychological variables: intentions, attitudes, influences, commitment, perceptions…These variables unfortunately refer to realities that lie beyond the control of those who can best steer students toward degree completion. (p. 27)

Adelman describes his study as one directed at identifying what contributes most to bachelor’s degree completion for students who attend four year colleges at any point in their college careers. Adelman follows the 1980 High School and Beyond/Sophomore cohort for 13 years, through 1993. Adelman's study is unique in several ways. First, long-term graduation is the outcome of the study. Adelman also developed an "Academic Resources Index" variable composed of high school curriculum, test scores, and class rank. This resources variable accounted for most of the variance in graduation in Adelman's model. In addition, measures of continuous enrollment, community college to four year college transfer, and the trend in college grades account for most of the additional variance. Adelman's model accounts for 43% of the variance, with the Academic Resources Index accounting for 17%, continuous enrollment accounting for 10%, and the "DWI Index" (ratio of drops/withdrawals/incompletes to total courses attempted) accounting for an additional 6%, or 33% of the total 43%. The additional 10% of explained variance can be accounted for by low-credits (fewer than 20 credits in the first year), having children, not returning to the original institution of enrollment if stopping out or transferring, transfer from the original institution of enrollment, SES quintile, whether students worked to finance their education, the ratio of credits earned to credits attempted for the first year of enrollment, and a variable addressing remedial course work. Although Adelman's study makes a major contribution to the literature on student graduation, it does not employ a longitudinal methodology.

Adelman (1999) develops six progressive models based on sets of variables, but does not consider the timing of graduation. Extending Adelman's findings with a longitudinal methodology, like survival analysis, might make a significant contribution to the literature. Additionally, the aid variables included in Adelman's model are very limiting, as discussed by Adelman. Self reported aid information available only through 1986 was used to construct the aid variables. Including a more complete measure of financial aid could influence the model's outcome. Also, follow-up data collection occurred only four times over a 13 year period, with follow-up surveys conducted in 1982, 1984, 1986, and 1992. Additional transcript information was collected in 1993, but survey information was unavailable.

Adelman (1999) also uses linear regression to address the proportion of variance accounted for and then logistic regression to add credibility to his findings. Ordinary least squares regression is not appropriate for a dichotomous outcome. Important information about the timing of graduation is lost due to the statistical techniques used and the follow-up data collection protocol. Survival analysis is a more robust and appropriate technique for studying the timing and occurrence of graduation.

DesJardins and Moye (2000) apply survival analysis to Adelman's (1999) data. They found that the Academic Resources Index positively affects graduation, but that its effects are mitigated by GPA. They also report that the negative effects of being associated with a minority group become less when aid and GPA are included in the model. When aid effects are allowed to vary over time, loans enhance and work/study inhibits timely graduation, a finding contrary to Adelman (1999). Desjardins and Moye find that GPA is positively related to graduation.

DesJardins and Moye (2000) face the same limitations as Adelman (1999) in terms of the frequency of data collection and extremely limited aid variables. These facts result in a cautious interpretation of the outcome.

Summary

The temporal dimension of survival analysis should be beneficial to the study of college student retention. The timing of student departure, or the when, should be more informative than the whether. The definition of the outcome, timing of dropout, stopout, transfer, or graduation, is a function of the perspective of the stakeholder. The appropriateness of system wide research has been called into question, given the vast disparity in retention and graduation rates, mission, location, culture, and resources from one institution to another. This is evidenced by the philosophical underpinnings of the two most popular models of student departure, the student integration model and the student attrition model, both described as institutional models. System wide research will continue to be limited until a more systematic methodology is adopted and the effects of individual institutions are treated as random and not fixed covariates. Potential options could include meta-analysis and/or hierarchical mixed effects models.

For the current study, the findings of Adelman (1999) will be incorporated into a single institution discrete time survival model with graduation as the outcome. Variable selection will be driven by the existing literature, especially Adelman and studies using survival analysis as the method, in an attempt to extend the study of student departure using survival analysis. A more complete description of the methods will be provided in Chapter III.

Chapter III

RESEARCH METHODS

Survival Analysis Concepts

Attempting to explain the graduation patterns of college students is no easy task. Researchers have been studying retention and graduation for decades, and the results have been mixed due to differences in operational definitions, methodology, and design. Most often, studies are cross-sectional in nature, with at most a pre-post design being employed. College student graduation is a longitudinal process, which implies the requirement of a truly longitudinal methodology for its study. Willett and Singer (1991) describe the benefits of using survival analysis to study college student retention. This approach leads to a reframing of the retention question from whether students leave a particular college or university to when are they at most risk of leaving. Although well suited as a statistical technique for studying student retention, survival analysis has been used very infrequently in educational research. Willett and Singer (1991, 1993) and Singer and Willett (1993) attribute this to a lack of statistical methods designed to answer the when question associated with event occurrence. With the extension of survival analysis from biostatistics to educational research, the methods now exist. Concepts central to an understanding of survival analysis will be presented next.

Censoring

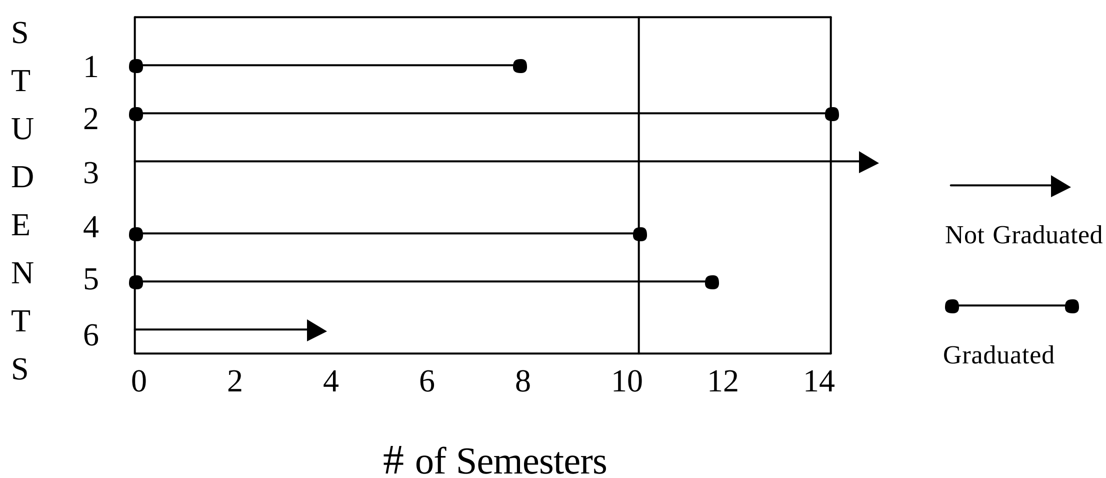
With survival analysis, answering the when question requires a metric for time to event occurrence. Some researchers have computed the actual time to event as the metric, but it is impossible to compute time to event for those subjects who do not experience the event but are still part of the study at the end of the study period. To address this issue, some researchers have imputed the study duration as the time to event for those subjects. This approach seriously biases parameter estimates. If a 10 semester time frame is used to study graduation behaviors, for example, and 30% of the students have not graduated during the 10 semesters but still remain in the sample, the measure of time to graduation would be 10 semesters. This clearly underestimates the time to graduation for the 30% of the sample who had not graduated by the end of the tenth semester. Subjects who do not experience the event of interest by the end of the study period, but remain in the sample, are described as censored. In this example, they are right censored. Figure 1 provides an example.

As can be seen from Figure 1, student 1 graduated at the eighth semester and student 4 graduated at the tenth semester, both within the 10 semester study period. So, time to graduation for these students is available. However, student 3 has not graduated by semester 14, so the time to graduation for student 3 is unknown. Even though student 2 and student 5 graduated, it was at semester 14 and semester 12 respectively. Information would not have been available at semester 10, the end of the study period, for those students. Using the 10 semester time of study as the time to graduation for students still in the sample but not having graduated by semester 10 would have resulted in an underestimation of 4 semesters for student 2, an underestimation of 2 semesters for student 5, and an unknown underestimation for student 3, assuming that student 3 eventually graduated. Survival analysis addresses censoring by the way in which time to graduation is developed, which will be discussed in subsequent sections.

The records for students 2, 3, and 5 would be described as right-censored. They remained in the sample through the end of the study period but did not experience the event of interest during the study period. Student 6 left the study at semester 4, possibly as a dropout, transfer, or stopout. Although this case could be described as an example of random censoring, or censoring that occurs as a function of some random event, it is more likely that the student left for some systematic reason. The student most likely left the institution as a dropout, transfer, or stopout, all events which are said to be "competing risks" with graduation as reasons for exiting the sample. The concept of competing risks will be addressed in more detail in a later section.

Figure 1

Example of Right Censoring

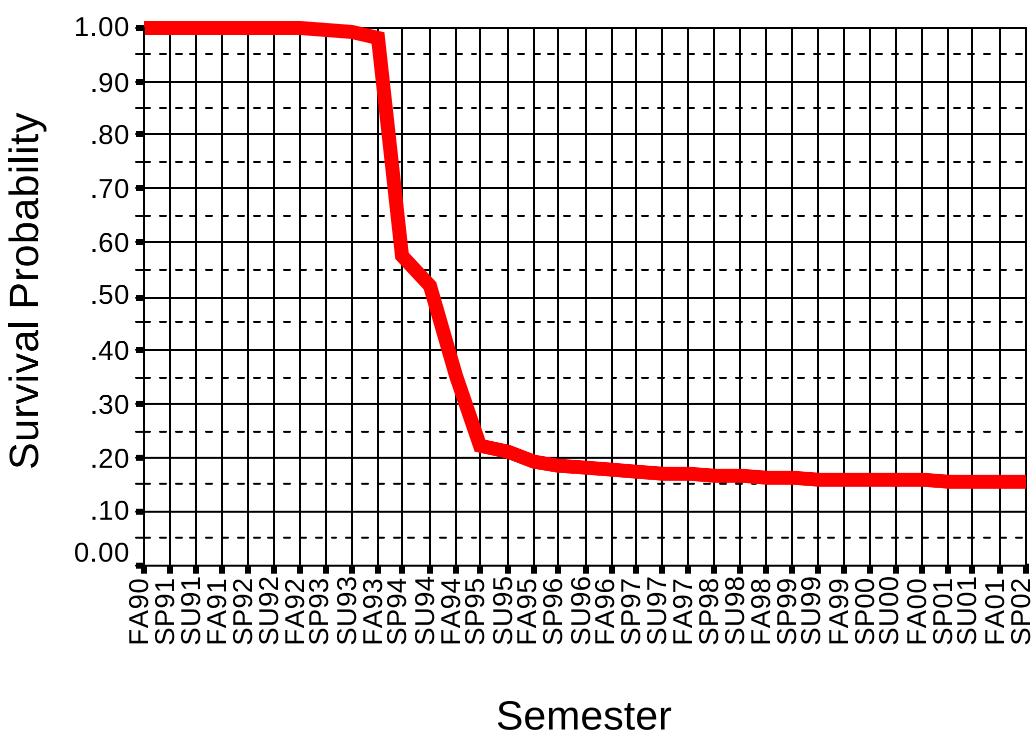


Survivor Function

The survivor function is defined as the plot of survival probabilities over time. Survival probabilities represent the proportion of the original sample that has not experienced the event of interest by a particular time point. Data for the present study are available on a semester by semester basis, so the numerator of the survival probabilities would be computed by summing the number of students who had not experienced the event of graduating at the end of a particular semester and dividing it by the denominator, the total number of students in the study. For example, the survival probability for the first semester would be 1.0, as no students had graduated. The survivor function would then be the plot of the survival probabilities for each semester over time. Figure 2 represents the baseline survivor function for the present study.

Figure 2

Baseline Survivor Function



The survivor function provides a chronological listing of survival probabilities and is a monotonically decreasing function of time. The survivor function eliminates the detrimental pooling effects that occur when a single time point is used to characterize graduation behaviors by parsing the graduation information into a semester by semester function. So, for example, 100% of the sample had not graduated by spring semester 1992 and approximately 15% of the sample did not graduate within the timeframe of the study.

A meaningful measure of the “center” of the survivor function is the median lifetime value, or the point in time at which 50% of the sample has experienced the event. From Figure 2, 50% of the sample graduated by summer semester 1994.

Hazard Function

Although an important function in summarizing the occurrence of events over time by providing a cumulative longitudinal summary of the proportion of students who have not graduated, the survivor function confounds information about graduation for each semester with cumulative information from the prior semesters. If graduation is an event and the “risk” of experiencing the event in each semester is the outcome of interest, the survivor function masks critical information about risk. The hazard function provides a better summary.

Central to an understanding of the hazard function is the “risk set”. The risk set for the present study is defined as all those students still enrolled at the beginning of a particular semester. These students represent those who have not left the institution for reasons other than graduation or have not graduated, so they are “at risk” of graduating at the end of that semester, and represent the denominator of the hazard probability.

The numerator of the hazard probability is the number of students out of the risk set who graduate. The hazard probability is the proportion of students in the risk set for each semester who graduate during that semester. These hazard probabilities represent the conditional probability that a student will graduate during the current semester given that they have not graduated in any prior semester. The conditional nature of hazard probabilities is the feature that makes the hazard function a better metric of risk/event occurrence. Students remain part of the risk set, hence the analysis, until they either leave the institution, graduate, or are censored. For example, all students are part of the risk set in the first semester but no one graduates during the first semester, so the hazard probability is 0. A plot of hazard probabilities over time determines the hazard function, a chronological summary of the risk of graduating.

Figure 3

Baseline Hazard Function

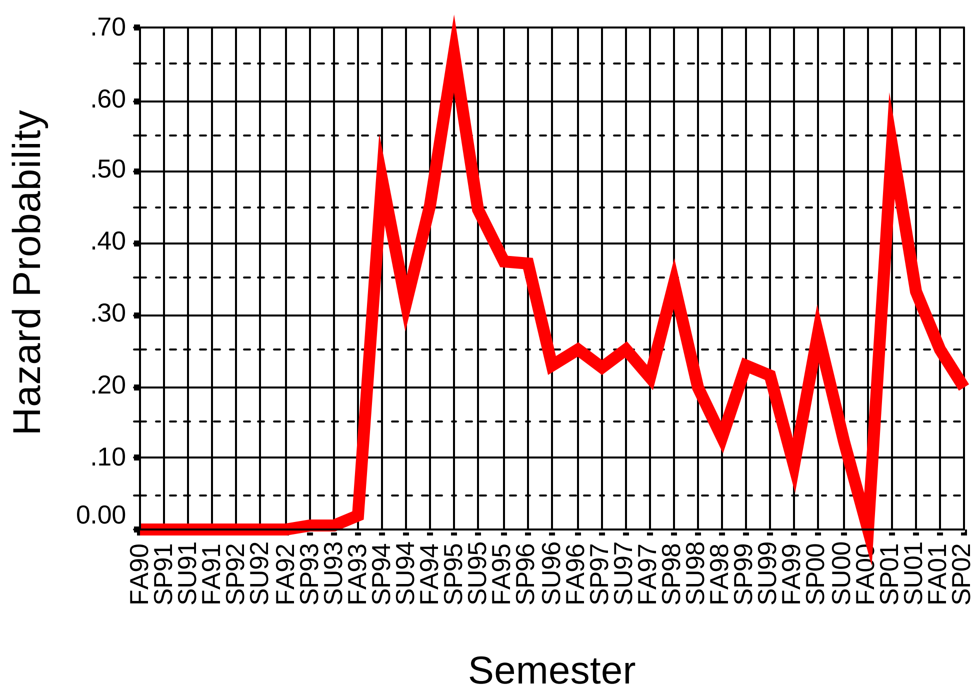


Figure 3 represents the baseline hazard function for the present study. As can be seen from Figure 3, the greatest risk of graduation occurs at spring semester 1994, spring semester 1995, and spring semester 2001. For spring semester1994 the risk set is composed of 2,660 enrolled students, 1,297 of whom graduated, for a hazard probability of .49. For spring semester 1995 the risk set is composed of 628 students, 418 of whom graduated, for a hazard probability of .67. And, for spring semester 2001 the risk set is composed of 13 students, 7 of whom graduated, for a hazard probability of .54. It is important to note that the stability of the hazard probabilities is a function of the size of the numerator and denominator. Hazard probabilities computed with large numbers of students in the numerator and denominator provide more predictive utility than hazard probabilities computed with smaller numbers of students, such as the hazard probability for spring semester 2001. Even though the probability of graduating is 0.54 for spring semester 2001, the small number of students left in the risk set and the small number graduating makes this probability less stable for predictive purposes.

This result exemplifies the flexibility of the hazard function in describing graduation behaviors over time. Even though the number of enrolled students declines, the hazard probabilities capture the semester behavior and the plot of the hazard probabilities captures the “risk” of graduation over time. It is evident from Figure 3 that spring semesters are points at which students tend to be more likely to graduate in general.

Time-dependent Variables

Researchers studying college student retention and graduation have had difficulty accommodating variables that change over time. Examples include semester grade point average, semester by semester aid amounts, and number of credits enrolled on a semester basis to name a few. Most researchers have taken an average across all time points and imputed that average as the measure for time-dependent variables. This approach completely masks the effects of fluctuations in grades, financial aid, and enrollment histories over time. Survival analysis allows for the inclusion of time-dependent variables directly into the analysis with semester by semester values. In this way the temporal effects of predictors are captured.

The terms time-dependent and time-independent will be used generously. Time-dependent will be used to describe variables that are allowed to change over time in the analysis, such as semester GPA or semester by semester financial aid amount. Time-independent variables will refer to variables that do not change over time in the present analysis, such as SAT scores and ethnicity.

Building the Data Set for use with Survival Analysis

Addressing censoring and time-dependent variables in survival analysis is accomplished through the manipulation of the longitudinal record developed for each subject. For the current study, a longitudinal data set was created by extracting data from several sources at the institution under study. First, preenrollment characteristics were extracted from the institution's mainframe and data warehouse admissions files. Enrollment data and financial aid data were also extracted from the mainframe, the

Table 1

Example of Longitudinal Data Set

Student	Gender	Age	Semester 1 GPA	Semester 2 GPA	Semester 3 GPA	Semester 4 GPA	Semester 5 GPA
1	Male	18	2.62	2.82	.	.	.
2	Female	18	3.10	3.81	3.13		
3	Female	18	2.83	3.48	3.93	3.94	4.00

institution's data warehouse, and a separate data warehouse unique to the student aid office. Each student was represented as a single row in this longitudinal file. Table 1 represents a subset of the data record for 3 students in the study.

As can be seen from Table 1, student 1 is a male 18 years of age who was enrolled for two semesters, as is evidence by the SEMESTER 1 and SEMESTER 2 GPA values. Student 2 is a female 18 years of age who was enrolled for three semesters, and student 3 is an 18 year old female who was enrolled for five semesters. GENDER and AGE are time-independent variables and SEMESTER GPA is a time-dependent variable. In order to prepare these data for analysis using survival analysis, we must transform this longitudinal data set to a person-period data set where each semester a student is enrolled is represented by a separate row in the data set. Table 2 provides the person-period data set for the data in Table 1.

Table 2

Person Period Data Set Example

Student		Gender	Age	Semester GPA
1		Male	18	2.62
1		Male	18	2.82
2		Female	18	3.10
2		Female	18	3.81
2		Female	18	3.13
3		Female	18	2.83
3		Female	18	3.48
3		Female	18	3.93
3		Female	18	3.94
3		Female	18	4.00
Total	N	10	10	10

As can be seen from Table 2, the three cases in the longitudinal data set represented in Table 1 have been expanded into 10 records in the person-period data set represented in Table 2. The ten cases represent two semesters for student 1, three semesters for student 2, and five semesters for student 3. The time-independent variables GENDER and AGE remain the same for each record for each individual student. The time-dependent variable, SEMESTER GPA, is permitted to change for each semester. Through this transformation, time-independent variables are able to change for each semester a student is enrolled and the information from censored students is part of the analysis until they depart from the study frame or are right censored. In order to address the issue of event occurrence, or graduation, additional variables must be computed.

In order to fit a survival analysis model two variables must be part of the person-period data set. The first, DURATION, provides an indication of how many semesters a student was enrolled before they graduated or were censored, and STATUS, an indicator identifying whether the student graduated or not. Table 3 provides an example of an expanded person-period data set with DURATION and STATUS included.

As can be seen from Table 3, the DURATION variable for student 1 is 2, indicating that student 1 was enrolled for 2 semesters. The EVENT variable for student 1 is equal to 0, indicating that student 1 did not graduate. Student 2 was enrolled for 10 semesters and did graduate, and student 3 was enrolled for 8 semesters and did not graduate. Also, the variable GENDER is time-independent as it remains the same for each student by semester, and the variables SEMESTER GPA and SEMESTER TOTAL AID are time-dependent as the values for each of these variables change from semester to semester.

An additional variable must be computed to identify the semester at which those students who graduated did graduate. The variable GRADUATE in Table 3 is an indicator variable with 1 indicating the semester during which a student graduated. For those students who did graduate, the value of GRADUATE will be 0 for all semesters except the semester during which the student graduated, and the value for GRADUATE for that semester will be 1. For those students who did not graduate, the value of GRADUATE will always be 0. The GRADUATE variable includes information about event occurrence and timing of the event. The GRADUATE variable will be the criterion in the survival analysis model.

Table 3

Person Period Table with Duration and Status

Student	Semester	Duration	Event	Graduate	Gender	Semester GPA	Semester Total Aid
1	1	2	0	0	Female	3.47	5500.00
1	2	2	0	0	Female	3.25	500.00
2	1	10	1	0	Female	2.54	3681.00
2	2	10	1	0	Female	2.95	1981.00
2	3	10	1	0	Female	3.22	2620.00
2	4	10	1	0	Female	2.00	2781.00
2	5	10	1	0	Female	1.97	500.00
2	6	10	1	0	Female	2.50	500.00
2	7	10	1	0	Female	1.44	.00
2	8	10	1	0	Female	.93	.00
2	9	10	1	0	Female	2.50	2500.00
2	10	10	1	1	Female	2.58	2500.00
3	1	8	0	0	Male	3.10	1200.00
3	2	8	0	0	Male	3.23	1200.00
3	3	8	0	0	Male	3.40	5697.00
3	4	8	0	0	Male	3.29	5698.00
3	5	8	0	0	Male	3.15	5697.00
3	6	8	0	0	Male	3.33	5698.00
3	7	8	0	0	Male	3.84	6642.00
3	8	8	0	0	Male	3.65	3128.00

As was described in a previous section, the hazard function is a probability function with values bounded by 0 and 1 and is the function modeled in survival analysis. When developing a multivariate survival analysis model “risk” is determined as a weighted linear combination of predictors. This gives rise to mathematical difficulties when the outcome is a probability, namely the possibility of deriving fitted values < 0 or > 1. A transformation is in order. When a probability is the outcome, the logit function is mathematically and conceptually appealing as the appropriate link function (Willett & Singer, 1991). To develop the logit function, the probabilities of the hazard function will be transformed to odds through the formula odds=hazard/(1-

hazard). Next, the natural log of the odds will be computed, resulting in the log odds or logit-hazard. This transformed function is the one modeled in survival analysis and is accomplished in this example by regressing covariates on the outcome variable GRADUATE.

Conceptually, and following the notation of Willett and Singer (1999), the entire hazard function is the outcome being modeled. If $h_{(t)}$ represents the entire hazard function $b_{0(t)}$ represents a baseline hazard function. As predictors are added to the model, the baseline hazard function shifts based upon the values of the predictor. The time-independent variable GENDER will be used as an example. If the value of GENDER is 1 for females and 0 for males the logit-hazard function could be represented as $\text{logit}\ h_{(t)} = b_{0(t)} + b_1 \text{GENDER}$.

When GENDER is equal to male (0), the logit-hazard function is equal to the baseline function. When GENDER is equal to female (1), the logit-hazard function shifts by b_1. This shift in the entire profile is what we describe through logistic regression. Beta-coefficients will be computed and evaluated to determine if the profiles for men and women differ statistically. This approach can be extended from the simple regression to the multiple regression case to address multiple time-independent variables.

Adding time-dependent variables, given the preparation of the person-period data set and the use of logistic regression, is straightforward. The time-dependent variable SEMGPA will be added to the model to include semester GPA values. The description of how the survival model works is motivated by the following set of equations:

$$\text{logit } h_{(1)} = b_{0(1)} + b_1 \text{GENDER} + b_2 \text{SEMGPA}_{(1)}$$
$$\text{logit } h_{(2)} = b_{0(2)} + b_1 \text{GENDER} + b_2 \text{SEMGPA}_{(2)}$$
$$\text{logit } h_{(3)} = b_{0(3)} + b_1 \text{GENDER} + b_2 \text{SEMGPA}_{(3)}$$
$$\text{logit } h_{(4)} = b_{0(4)} + b_1 \text{GENDER} + b_2 \text{SEMGPA}_{(4)}$$
$$\text{logit } h_{(T)} = b_{0(T)} + b_1 \text{GENDER} + b_2 \text{SEMGPA}_{(T)}$$

Conceptually, there is a separate equation for each time period in the data set, represented by the subscripts in the above equations. Each time period has a unique intercept represented by the $b_{0(t)}$ values. Because GENDER is time-independent (fixed), it shifts the logit-hazard function the same amount for each period. If gender is equal to 1, the logit-hazard is shifted by b_1 regardless of the time period. SEMGPA, however, is time-dependent. The value of SEMGPA can change from one time-period (semester) to another. The shift in the logit-hazard comes as a function of SEMGPA multiplied by b_2, which can vary from one semester to another because SEMGPA is not fixed. Beta-coefficients will be computed and evaluated to determine if the profiles for men and women differ statistically, controlling for SEMGPA, and if SEMGPA is a significant contributor to the risk of graduation, controlling for GENDER. This approach can be extended to include multiple time-independent and time-dependent variables in a logistic regression model. A fuller description of fitting and interpreting the survival model will be presented in Chapter IV.

Competing Risks

Dropping out, stopping out, or graduating can be identified as competing reasons for students to leave the sample prior to the end of the study period. For the present study, the perspective is that of the institution, which eliminates the distinction between transfer and dropout. Also, there is no distinction made between dropouts, students who are academically dismissed, and students who are dismissed for disciplinary reasons. Some distinction has been made in the literature concerning these voluntary vs. nonvoluntary modes of withdrawal, however, the position on this is mixed. Pantages and Creedon (1978) indicate:

> We believe that basing the distinction between "voluntary" and "nonvoluntary" withdrawal on scholastic achievement is inappropriate, since it ignores the factors that have caused poor academic performance. It is these factors that influence the decision to drop out, not the end result of these factors (i.e., poor grades). (p. 52)

Although an argument could be made for separating disciplinary dismissals into a separate category, information which would allow that to be accomplished in the present study is unavailable. Hence, the decision is a practical one. In the present study dropout will include students who withdraw in good academic standing, those who are academically dismissed, and those who are dismissed for disciplinary reasons.

Dropping out is a competing risk with graduation. Discrete time survival analysis allows for competing risk models, where a polytomous rather than a dichotomous outcome is predicted. However, the estimation and interpretation is more complex. Allison (1995) provides an alternative which allows separate analyses for each event without biasing parameter estimates and with only slight, at least in principle, loss of precision. Allison identifies the benefits of conducting separate as opposed to simultaneous analyses as being able to focus on only those events of interest and the ability to specify quite different models for different events. The approach to conducting separate analyses follows.

Table 4

Competing Risk Person Period Table

Student	Semester	Duration	Event	Graduate	Gender	Semester GPA	Semester Total Aid
1	1	2	2	0	Female	3.47	5500.00
1	2	2	2	2	Female	3.25	500.00
2	1	10	1	0	Female	2.54	3681.00
2	2	10	1	0	Female	2.95	1981.00
2	3	10	1	0	Female	3.22	2620.00
2	4	10	1	0	Female	2.00	2781.00
2	5	10	1	0	Female	1.97	500.00
2	6	10	1	0	Female	2.50	500.00
2	7	10	1	0	Female	1.44	.00
2	8	10	1	0	Female	.93	.00
2	9	10	1	0	Female	2.50	2500.00
2	10	10	1	1	Female	2.58	2500.00
3	1	8	2	0	Male	3.10	1200.00
3	2	8	2	0	Male	3.23	1200.00
3	3	8	2	0	Male	3.40	5697.00
3	4	8	2	0	Male	3.29	5698.00
3	5	8	2	0	Male	3.15	5697.00
3	6	8	2	0	Male	3.33	5698.00
3	7	8	2	0	Male	3.84	6642.00
3	8	8	2	2	Male	3.65	3128.00

Table 4 provides an altered person-period data set found in Table 3 where the event and outcome variables are altered to reflect a second event, dropout. For EVENT, 1 represents graduation and 2 represents dropout. The criterion, GRADUATE, is also altered so that a value of 1 in a person-period represents graduation and a value of 2 represents dropout. Following Allison, the appropriate analysis of graduation in this example would be accomplished by deleting the single records where GRADUATE is equal to 2. This would result in the elimination of the semester 2 record for student 1 and the semester 8 record for student 3. A logistic regression analysis would then be conducted with GRADUATE as the criterion.

For the present study, students who stopout will be included in the model and will be defined as any student who has returned to the institution after at least a 1 semester absence.

Variable Selection

Variables included in the present study are selected based upon a review of the literature and their availability to institutional researchers. The underlying premise of the present study is that the retention and graduation rates of institutions vary drastically and that research must be conducted at the institutional level using variables that are readily available to institutional researchers for the vast majority of students. This premise also guided variable selection.

Preenrollment Characteristics

Initial Age

The initial age (AGE) of students during the fall semester of 1990 will be included in the analysis. Although age increases with each passing year, only initial age will be included here and will be treated as time-independent.

Gender

The research results concerning the effects of gender on retention and graduation have been mixed. Gender (GENDER) is included in the model and is time-independent.

High School GPA

High school GPA (HSGPA) is included as a measure of a student's preenrollment academic abilities. The high school GPA used in the present study comes directly from the student's high school transcript and is then converted to a common metric. This conversion considers whether a high school weights the GPA and whether the student has had advanced placement courses. High school GPA is time-independent.

SAT Scores

SAT verbal (SATV) and math scores (SATM) will also be included in the model as a second measure of a student's preenrollment academic abilities. A very small percentage of students submit ACT scores to the institution under study. Where this is the case, ACT scores will be converted to SAT scores using concordance tables (Schneider & Dorans, 1999). SAT scores are time-independent.

Ethnicity

The ethnic background (ETHNIC) of students will also be included in the model. The White American category will be coded as the reference group. Ethnic background is a time-independent variable.

Residency

The residency (RESID) of the student will also be considered, with a dichotomous variable being computed representing in-state vs. out-of-state residency. Residency is a time-independent variable.

Legacy Status

A student's legacy status (LEGACY) will be included as a measure of possible institutional commitment. Students will be categorized based upon whether their parents, siblings, or grandparents were alumni of the institution under study. Legacy status is time-independent.

Adelman's (1999) Academic Resources Index

Generally following Adelman (1999), a modified version of the components of his Academic Resources Index will be included in the present study. Adelman combined three components to develop the composite index.

First, a measure of high school rank was used in Adelman's (1999) study. However, high school rank was unavailable for approximately 30% of the sample, which led to an imputation process based on the correlation between high school rank and an academic grade point average developed by Adelman. This correlation was computed using cases where both measures were available. Once the correlation was computed Adelman developed a specific percentile of academic grade point average and substituted it for the class rank percentile for students where class rank was unavailable. The resultant percentile ranks were converted to quintiles. Adelman indicates that "A larger unit of measurement was necessary to reduce the statistical noise, and quintiles were selected for the task" (p. 12).

Adelman (1999) also used the percentile rank scores of a test administered to the 1980 High School and Beyond/Sophomore cohort during their senior year (Senior Year Test), and described by Adelman as a “mini, enhanced SAT” (p.11). There were also subjects missing Senior Year Test score data. Where possible, Adelman derived a substitute for the Senior Year Test score by computing a percentile rank score form the SAT or ACT scores of those students missing the Senior Year Test score. The Senior Year Test scores for all students were then converted to quintiles.

The most complex component of Adelman’s (1999) index combines information about the number of Carnegie units students have completed in five subject areas; English, math, science, language, and history/social studies, with information about the highest level of mathematics a student has completed, whether they have any remedial mathematics or English courses, whether they have completed any advance placement courses, whether they have completed any computer science courses, and total academic units. The outcome is not total number of Carnegie units in any particular area, but rather a graduated 40 point index that combines information about the above listed factors. Adelman’s index was developed by observing patterns in the 1980 High School and Beyond/Sophomore cohort. In addressing whether this index is generalizable, Adelman states:

> These gradations of academic intensity and quality are based on the history of one national high school class that was scheduled to graduate in 1982. The next graduating class for which we possess similar national data is that of 1992. While the specific numbers of Carnegie units, APs, and remedial indicators

> might change, the basic form and principles of the gradations will probably not change. (p. 114)

Once the 40 gradations were determined, they were set out on a scale of equal intervals from 100 to 2.5, and converted to quintiles.

Finally, Adelman (1999) combined the three components by developing a logistic regression model with bachelor's degree attainment as the criterion and each of the three components of the index as the predictors. Each component in the model was weighted by its comparative contribution to the model based upon odds ratios and the Academic Resources Index was the sum of those weighted components.

There are several concerns relative to the development of the Academic Resources Index of Adelman (1999). Due in large part to the missing class rank data, Adelman determined that quintiles would be an appropriate method of dealing with the "statistical noise" associated with a process whereby missing percentile rank values were imputed for approximately 30% of his sample based upon a correlation between existing percentile rank scores and an academic grade point average he developed. Whether the imputation process was accurate or not, the conversion of the final percentile ranks to quintiles, which equates to collapsing across a continuous variable, creates a situation in which important information could be lost. This is true of every component of the Academic Resources Index. The continuous nature of the Senior Year Test score and the curriculum measure are limited by conversion to quintiles. Adelman at some level speaks to this issue when he states, relative to the class rank variable, "It was this variable more than any other that determined the used [sic] of quintile presentations for otherwise continuous variables in the analysis" (p.12).

Also, the way in which the weights were determined for each component in the development of the composite index is questionable. Odds ratios indicate the relationship between individual predictor variables and the criterion while controlling for the other predictors in the model. There is no cumulative effect of odds ratios. Computing the "comparative contribution" of each component as the percentage of the summed odds ratios is suspect.

Academic Resources in the Present Study

For the present study SAT verbal (SATV) and SAT math (SATM) scores, or converted ACT scores, will be used in place of the Senior Year Test scores used by Adelman (1999). The actual test scores, and not quintiles, will be used. This variable is time-independent.

The student's rank in high school class (RANK) will be included as a measure of how well the student performed in comparison to his/her peers and will be expressed as a percentile rank. Class rank is time-independent.

Finally, an academic intensity variable (ACAD_INT) was developed following Adelman's (1999) graduated 40 point index. Adelman used completion of Algebra 2 and number of advanced placement courses as two of the criteria for scale development. All students in the present sample have completed Algebra 2 and information is only available for completion of advanced placement courses, not number of advanced placement courses completed. As a result, the scale developed for the present study has 22 gradations and approximates Adelman's scale as closely as possible. The scale value, and not quintiles, will be used in the analysis. The academic intensity variable is time-independent.

Enrollment Variables

The following variables represent information available after students have enrolled at a particular institution. Some of these enrollment variables are time-dependent.

Semester GPA

The semester grade point average (SEMGPA) for each semester will be included in the current model as a measure of how well the student is progressing academically. GPA has been identified as a measure of academic integration in Tinto's (1975) model. GPA is time-dependent.

Cumulative GPA by Semester

Where semester GPA provides an indication of how well a student has performed that semester, cumulative GPA (CUMGPA) provides an overview of how well the student is performing overall. Cumulative GPA by semester will also be included in the model. Cumulative GPA is time-dependent.

Initial Academic College of Enrollment

Initial academic college of enrollment (COLL) will also be included in the model to test for college effects. College is a time-independent variable.

Number of Academic College Changes

A variable will be computed to indicate the number of times a student changes academic college during their tenure at the institution under study to determine if students who make changes in their major field which results in a change in college are more or less likely to graduate. Number of changes in academic college (COLLCNG) is time-independent.

Number of Academic Major Changes

A variable will be computed to capture the number of times a student changes their major during their tenure at the institution under study to determine if students who change their major are more or less likely to graduate. Number of major changes (MAJCNG) is a time-independent variable.

Continuous Enrollment

Following Adelman (1999), a dichotomous variable will be developed to indicate whether a student was enrolled continuously prior to their separation from the institution under study (CONTENR). Adelman found that students who remained continuously enrolled were more likely to graduate than those students who had gaps in their enrollment histories, identifying continuous enrollment as second only to the Academic Resources Index in its explanatory power. Basically, continuous enrollment identifies stopouts. If a student remained enrolled in each subsequent semester, without interruption, until they were no longer enrolled due to dropping out or graduating, for example, they were coded with a 1. A student's enrollment might only be for a few semesters; however, as long as they were enrolled continuously until their separation from the university under study, they were identified as continuously enrolled. If a student missed even one semester during his/her enrollment, but reenrolled thereafter, the continuous enrollment variable was coded as 0. The continuous enrollment variable is time-independent.

Percentage of Credits Earned to Credits Attempted

The percentage of credits earned to credits attempted (EARNATT) will be computed on a semester basis as a variation of Adelman's DWI Index, or the percentage of dropped courses, withdrawals, and incompletes to total courses attempted. Adelman dichotomized his DWI index by setting a threshold at 0.2 and assigning subjects accordingly. This could potentially mask important information. The percentage computed in the present study will be used without aggregation.

Financial Aid Variables

Total Aid Amount

The total amount of aid (TOTAID) from all aid sources will be included on a semester basis. Total aid amount is time-dependent.

Total Aid by Aid Source

Several variables will be created to represent the total aid amount by aid source. Variables representing the total amount of grant (TOTGRNT), scholarship (TOTSCHL), loan (TOTLOAN), and work aid (TOTWORK) will be created by semester. Each of these variables will be time-dependent.

Ratio of Total Aid to Cost of Attendance

A ratio of total aid amount to cost of attendance (COARAT) at the institution under study will be computed. This is done in an attempt to standardize the aid component over time by addressing increases in cost through the use of cost of attendance. The cost of attendance includes tuition and fees, and will vary by student based upon, for example, state residency, commuter status, and academic program. This ratio will be computed for each semester, and then converted to a percentage for analysis. COARAT is a time-dependent variable.

Ratio of Total Aid by Aid Source and Cost of Attendance

Variables will be computed to represent the ratio of total aid in each aid source mentioned above to cost of attendance. These variables will be included to determine if there is a differential graduation rate for students with a larger portion of their total cost being accounted for by gift versus loan aid. These variables will be computed on a semester basis and then converted to a percentage. These variables are time-dependent.

Sample

Subjects were selected from a large comprehensive doctoral degree granting institution in the northeastern region of the United States. Approximately 8,768 first-time, full-time, freshman, baccalaureate students were enrolled for fall semester 1990 at 19 campus locations. For the present study, only full-time, freshman, students admitted to baccalaureate programs at the central campus, who were United States citizens, were selected as subjects. Only those students attending the central campus were selected to eliminate effects which might be attributable to individual campus locations. Of the 3,409 students enrolled at the central campus, 3,336 met the above criteria. An additional 134 students were eliminated due to missing data, for a final sample size of 3,202, or 96% of the 3,409 students enrolled at the central campus.

Data for the study were extracted from several institutional databases and were manipulated using SPSS and SAS in a mainframe and personal computer environment. A longitudinal data set was created by merging preenrollment characteristics, enrollment data, and financial aid data, semester by semester. Students were followed from fall semester 1990 through spring semester 2002, including summer sessions, for a total of 35 semesters.

Chapter IV

ANALYSIS AND RESULTS

Descriptive Statistics

Descriptive statistics will be provided for the criterion and predictor variables. First, the criterion will be more fully defined. Then, descriptive statistics for time-independent and time-dependent predictor variables will be provided.

Criterion

A person-period data set was developed by transforming the longitudinal records of the 3,202 students in the present study into 29,008 person-period records with the variable GRADUATE identifying semesters in which students were enrolled (GRADUATE = 0), semesters in which students dropped out (GRADUATE=2), or semesters in which students graduated (GRADUATE=1). Table 4 provides an example of how the person-period records are developed.

As was described in an earlier section, graduation is the outcome of interest and dropout has been described as a competing risk with graduation. It is possible to fit a competing risks model, where a polytomous, as opposed to a dichotomous, outcome is modeled. In this case, the outcome, GRADUATE, would have three possible values for each person-period record. GRADUATE would be equal to 0 if the student were enrolled, 2 if the student had dropped out, and 1 if the student had graduated in that semester. However, a polytomous model, where all competing risks of leaving the sample are modeled simultaneously, is more complex and difficult to interpret. Following Allison (1995), it is possible to perform separate analyses for each event without biasing parameter estimates and with only slight, at least in principle, loss of precision. Allison identifies the benefits of conducting separate, as opposed to simultaneous, analyses as being able to focus on only those events of interest and the ability to specify quite different models for different events.

In the present study, the competing risks are identified as dropout and graduation. Hence, dropouts will be deleted from the data set by deleting each person-period record with GRADUATE equal to 2 prior to conducting a logistic regression analysis. This does not eliminate all person-period records for students who were dropouts, it only eliminates the person-period record for that specific semester during which the student dropped out. The elimination of these 495 records reduces the final data set to 28,513 person-period records.

The outcome of interest in the present study is graduation. The following graduation rates are provided as reference points. For the present sample, the four year graduation rate is 41%, the five year graduation rate is 76%, the six year graduation rate is 81%, and the graduation rate at the end of the study is 83%.

Time-Independent Variables

Table 5 provides frequencies for the discrete time-independent variables in the analysis. As can be seen from Table 5, the sample is composed of 6.6% more men than

Table 5

Frequencies for Discrete Time-independent Variables

GENDER

	Frequency	Percent
Female	1494	46.7
Male	1708	53.3

ETHNIC

	Frequency	Percent
White American	2869	89.6
Hispanic American	66	2.1
Asian American	159	5.0
African American	85	2.7
American Indian/Alaskan Native	4	.1
Unknown	19	.6

RESID

	Frequency	Percent
Resident	2340	73.1
Non-Resident	862	26.9

LEGACY

	Frequency	Percent
Non-legacy	2247	70.2
Legacy	955	29.8

CONTENR

	Frequency	Percent
Not Continuous	139	4.3
Continuous Enrollment	3063	95.7

COLL

	Frequency	Percent
Arts And Architecture	163	5.1
Agricultural Sciences	218	6.8
Business Administration	474	14.8
Communications	146	4.6
Division Of Undergraduate Studies	406	12.7
Education	98	3.1
Earth And Mineral Sciences	58	1.8
Engineering	823	25.7
Health And Human Development	87	2.7
The Liberal Arts	329	10.3
Science	400	12.5

women. The students are predominately in-state students, with 73.1% coming from the institution's home state. The sample is also predominantly white American, (89.6%). Approximately 30% of the students in the sample have a parent or sibling who is an alumnus or alumna. And, 95.7% of the students remained enrolled continuously until their departure from the institution. The largest academic college of initial enrollment was engineering, with 25.7% of the entering freshmen.

Table 6 represents descriptive statistics for continuous time-independent variables. These summary statistics are computed at the student level and represent preenrollment and first semester enrollment data.

Table 6

Descriptive Statistics for Continuous Time-independent Variables

	N	Minimum	Maximum	Range	Median	Mean	Std. Deviation
AGE	3202	16	90	74	18.00	17.97	1.462
HSGPA	3202	1.41	4.00	2.59	3.4900	3.4548	.37952
SATV	3202	210	800	590	510.00	514.74	83.895
SATM	3202	280	800	520	600.00	596.43	86.294
RANK	3202	4	99	95	91.00	87.26	11.860
ACAD_INT	3202	1	22	21	4.00	6.66	4.270
COLLCHNG	3202	.00	6.00	6.00	.0000	.4441	.75373
MAJCHNG	3202	.00	5.00	5.00	.0000	.4369	.71179
EARNATT	3202	.000	73.684	73.684	4.56274	5.81224	6.147348

Time-dependent Variables

As has been presented, time-dependent variables are ones that change semester by semester. The person-period data set created for the present study represents one record for each semester a student is enrolled. There have been 28,513 person-period records created for 3,202 subjects. The descriptive statistics found in Table 7 are computed across those person-period records, and represent a summary at the person-period level. So, for example, the average semester GPA of 3.03 is computed by summing the GPAs for the entire 28,513 records and dividing by 28,513.

Table 7

Descriptive Statistics for Continuous Time-dependent Variables

	N	Minimum	Maximum	Range	Median	Mean	Std. Deviation
SEMGPA	28513	.00	4.00	4.00	3.1900	3.0339	.79203
CUMGPA	28513	.00	4.00	4.00	3.0400	2.9909	.54080
TOTAID	28513	.00	10915.00	10915.00	939.7500	1545.4619	1880.17230
TOTGRNT	28513	.00	6192.25	6192.25	.0000	385.7095	794.64666
TOTSCHL	28513	.00	8500.00	8500.00	.0000	352.4078	985.98238
TOTLOAN	28513	.00	10290.00	10290.00	.0000	760.7344	1205.75259
TOTWORK	28513	.00	2375.90	2375.90	.0000	41.3831	199.59963
COARAT	28513	.00	312.89	312.89	13.4576	25.8059	30.89002
GRNTCOA	28513	.00	124.31	124.31	.0000	6.7249	14.07417
LOANCOA	28513	.00	132.64	132.64	.0000	12.7877	19.63763
SCHLCOA	28513	.00	312.89	312.89	.0000	5.5128	14.95975
WORKCOA	28513	.00	87.19	87.19	.0000	.6987	3.34031

Baseline Survivor Function and Survival Probabilities

Figure 2 represents the baseline survivor plot for the present study. Table 8 provides the survival probabilities that comprise the survivor function. As is shown in Table 8, 57% of the sample had not graduated by the end of four years and 15% of the sample had not graduated by the end of the study period, spring semester 2002. Approximately 50% of the subjects had graduated by summer semester 1994.

Table 8

Baseline Survival Probabilities

		Non-graduates/ Survivors	Cumultive # of Graduates	Survival Probability
Semester	FA90	3202	0	1.00
	SP91	3202	0	1.00
	SU91	3202	0	1.00
	FA91	3202	0	1.00
	SP92	3201	1	1.00
	SU92	3201	1	1.00
	FA92	3201	1	1.00
	SP93	3189	13	1.00
	SU93	3184	18	.99
	FA93	3130	72	.98
	SP94	1833	1369	.57
	SU94	1664	1538	.52
	FA94	1137	2065	.36
	SP95	719	2483	.22
	SU95	669	2533	.21
	FA95	617	2585	.19
	SP96	585	2617	.18
	SU96	579	2623	.18
	FA96	563	2639	.18
	SP97	551	2651	.17
	SU97	547	2655	.17
	FA97	539	2663	.17
	SP98	528	2674	.16
	SU98	526	2676	.16
	FA98	522	2680	.16
	SP99	516	2686	.16
	SU99	513	2689	.16
	FA99	511	2691	.16
	SP00	505	2697	.16
	SU00	504	2698	.16
	FA00	504	2698	.16
	SP01	497	2705	.16
	SU01	496	2706	.15
	FA01	494	2708	.15
	SP02	493	2709	.15

Baseline Hazard Function and Hazard Probabilities

Figure 3 provides the baseline hazard function for the present study. Table 9 provides the hazard probabilities.

Table 9

Baseline Hazard Probabilities

		Semester Graduates	Risk Set/ Students Enrolled	Hazard Probablility
Semester	FA90	0	3202	.00
	SP91	0	3072	.00
	SU91	0	568	.00
	FA91	0	2971	.00
	SP92	1	2895	.00
	SU92	0	756	.00
	FA92	0	2827	.00
	SP93	12	2781	.00
	SU93	5	977	.01
	FA93	54	2738	.02
	SP94	1297	2660	.49
	SU94	169	535	.32
	FA94	527	1164	.45
	SP95	418	628	.67
	SU95	50	112	.45
	FA95	52	139	.37
	SP96	32	86	.37
	SU96	6	26	.23
	FA96	16	64	.25
	SP97	12	53	.23
	SU97	4	16	.25
	FA97	8	38	.21
	SP98	11	32	.34
	SU98	2	10	.20
	FA98	4	31	.13
	SP99	6	26	.23
	SU99	3	14	.21
	FA99	2	22	.09
	SP00	6	22	.27
	SU00	1	8	.13
	FA00	0	11	.00
	SP01	7	13	.54
	SU01	1	3	.33
	FA01	2	8	.25
	SP02	1	5	.20

The hazard probabilities shown in Table 9 provide a summary of the “risk” of graduation for the entire sample. As is shown in Table 9, the greatest risk of graduating occurs in spring semester 1995. Other periods of substantial risk are spring semester 1994, summer semester 1995, and spring semester 2001.

The baseline hazard function provides a plot of the semester by semester hazard probabilities over time. The modeling process will allow for the assessment of differential risk based upon a set of covariates. So, for example, if gender is entered as a covariate the hazard function for men and women will be compared to determine if statistically significant differences exist.

Modeling Time to Graduation

The statistical modeling process will be conducted in two blocks. First, preenrollment characteristics will be entered into the model, and the outcome will be discussed. Next, enrollment and financial aid variables will be added to the model and their relative contribution, if any, will be determined.

Preenrollment Model

With GRADUATE as the outcome, a logistic regression model was fit using the following preenrollment variables:

AGE	Age	Time-independent
HSGPA	High School GPA	Time-independent
SATV	SAT Verbal Score	Time-independent

SATM	SAT Math Score	Time-independent
ETHNIC	Ethnicity	Time-independent
RESID	In vs. Out-of-state Residency	Time-independent
LEGACY	Parents or Siblings Alumni	Time-independent
RANK	High School Class Rank	Time-independent
ACAD_INT	Academic Intensity	Time-independent

The preenrollment variables were fit simultaneously using the SPSS (2001) binary logistic regression procedure. The parameter coding for the categorical variables in the preenrollment model is shown in Table 10, where the referent category for ETHNIC is

Table 10

Categorical Variable Parameter Coding Preenrollment Model

			Parameter coding				
		Frequency	(1)	(2)	(3)	(4)	(5)
ETHNIC	White American	25660	.000	.000	.000	.000	.000
	Unknown	158	1.000	.000	.000	.000	.000
	Asian American	1371	.000	1.000	.000	.000	.000
	Hispanic American	551	.000	.000	1.000	.000	.000
	African American	731	.000	.000	.000	1.000	.000
	American Indian/Alaskan Native	42	.000	.000	.000	.000	1.000
CONTENR	Continuous Enrollment	27084	1.000				
	Not Continuous	1429	.000				
RESID	Non-PA Resident	7076	.000				
	PA Resident	21437	1.000				
LEGACY	Non-legacy	19699	.000				
	Legacy	8814	1.000				
GENDER	Male	15535	.000				
	Female	12978	1.000				

White American, for LEGACY the referent category is non-legacy, for RESID the referent category is non-resident, and for gender the referent category is male.

To determine how well the model fit the sample data, a Hosmer and Lemeshow (1989) goodness of fit test was computed. This approach to determining model fit is most appropriate when some or all of the covariates are continuous. The Hosmer and Lemeshow test places subjects into deciles based on predicted probabilities and computes a chi-square test based on the observed and predicted number of subjects in the deciles (Stokes, Davis, & Koch, 1995). The test statistic, $\hat{C}$, is then compared to a chi-square distribution with number of decile groups minus 2 degrees of freedom. The null hypothesis states that the observed and predicted values are equal, so not rejecting the null indicates a good model fit. For the preenrollment model, $\hat{C}$ (8, N=28,513) = 1.91, p = .984, which indicates that the model fits the data well.

The model summary and individual parameter estimates for the preenrollment model are shown in Table 11. The omnibus tests of model coefficients provide a test for the null hypothesis that all beta coefficients are equal to 0. For the preenrollment model, the first block in the analysis, the block chi-square and the model chi-square are equal. As can be seen from the omnibus tests, the null hypothesis is rejected at the p < .05 level with, χ^2 (14, N = 28,513) = 32.41, p = .004.

A review of the individual parameter estimates for variables in the preenrollment model reveals that only GENDER, χ^2 (1, N = 28,513) = 4.15, p = .04, and RESID, χ^2 (1, N = 28,513) = 5.04, p = .03, are statistically significant at the p < .05 level.

Table 11

Model Summary Preenrollment Model

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Block	32.411	14	.004
Model	32.411	14	.004

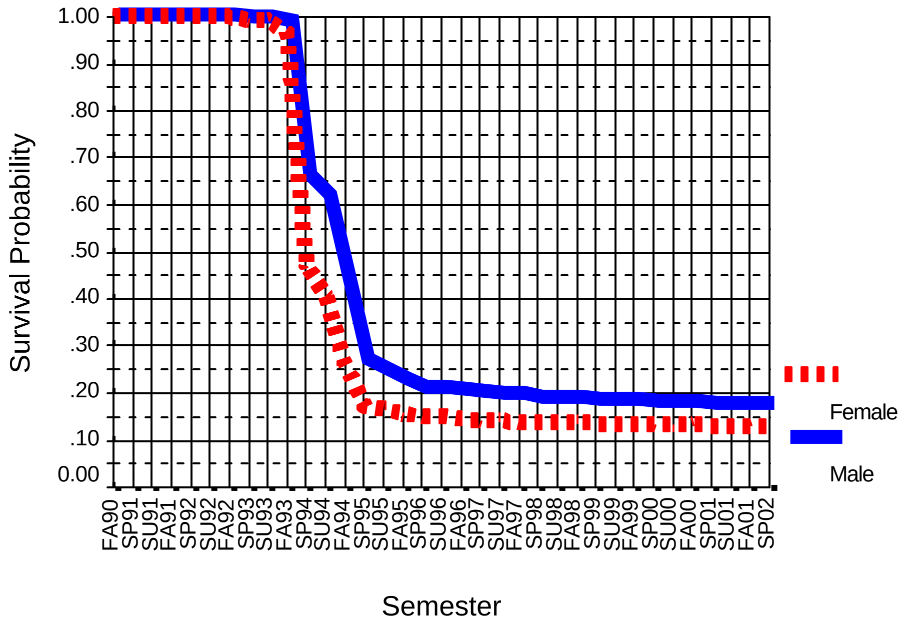
Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
AGE	-.004	.016	.061	1	.805	.996
GENDER(1)	.088	.043	4.148	1	.042	1.092
HSGPA	.141	.085	2.759	1	.097	1.151
SATV	.000	.000	.005	1	.945	1.000
SATM	.000	.000	.055	1	.815	1.000
ETHNIC			4.951	5	.422	
ETHNIC(1)	.080	.265	.090	1	.764	1.083
ETHNIC(2)	-.037	.098	.143	1	.705	.964
ETHNIC(3)	-.249	.166	2.237	1	.135	.780
ETHNIC(4)	-.247	.148	2.793	1	.095	.781
ETHNIC(5)	-.015	.526	.001	1	.977	.985
RESID(1)	-.107	.048	5.048	1	.025	.898
LEGACY(1)	.037	.044	.693	1	.405	1.038
RANK	.003	.003	1.000	1	.317	1.003
ACAD_INT	.000	.005	.000	1	.999	1.000
Constant	-2.849	.418	46.373	1	.000	.058

Based upon the odds ratio, Exp(B) in Table 11, females are 1.09 times more likely to graduate than males. The sample survivor function for GENDER is shown in Figure 4. The median lifetime value for females is found at spring semester 1994 and for males, fall semester 1994. Females are more likely to graduate, and graduate earlier than men.

Figure 4

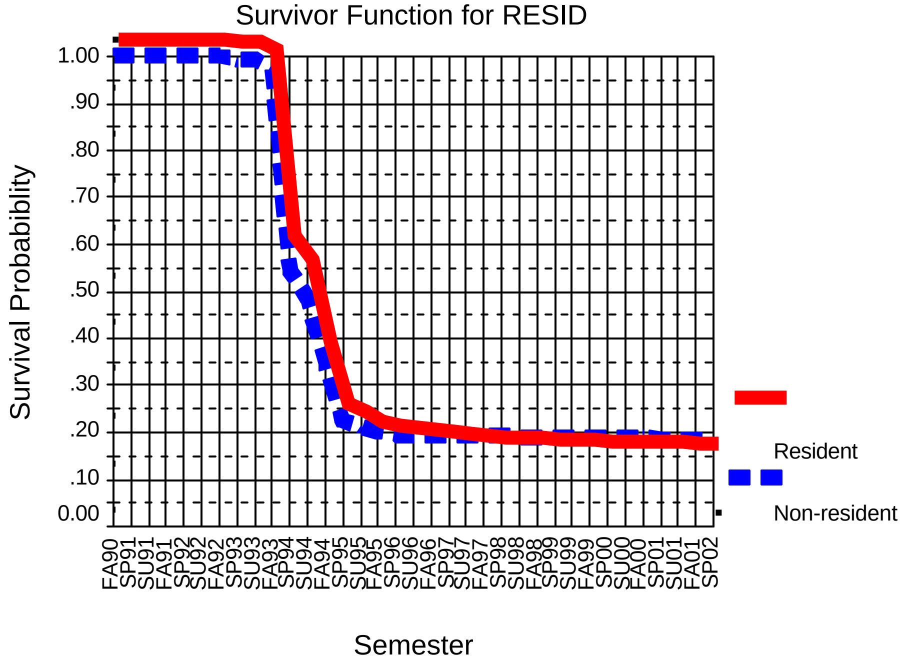
Survivor Function for Gender



Based upon the odds ratio for RESID, residents of the institution's home state are 90% as likely to graduate as non-residents, or non-residents are 1.11 times more likely to graduate than residents. Figure 5 provides the sample survivor function for

Figure 5

Survivor Function for RESID



RESID. Although the coefficient is significant, the difference in timing of graduation is mixed. The median lifetime value for non-residents is reached sooner than residents, at approximately summer semester 1994, and for residents at fall semester 1994. However, resident students are more likely to graduate after summer semester 1995, and ultimately graduate at a slightly higher rate.

Preenrollment, Enrollment and Aid Model

The following variables were added to develop the complete model:

SEMGPA	Semester GPA	Time-dependent
CUMGPA	Cumulative GPA by Semester	Time-dependent
COLL	First Semester Academic College	Time-independent
COLLCHNG	# Changes in Academic College	Time-independent
MAJCHNG	# Changes in Academic Major	Time-independent
CONTENR	Continuous Enrollment	Time-independent
EARNATT	% Credits Earned to Enrolled	Time-independent
TOTAID	Total Aid by Semester	Time-dependent
TOTGRNT	Total Grant Aid by Semester	Time-dependent
TOTSCHL	Total Scholarship Aid by Semester	Time-dependent
TOTLOAN	Total Loan by Semester	Time-dependent
TOTWORK	Total Work Study by Semester	Time-dependent
COARAT	% Total Aid by Cost of Attendance	Time-dependent
GRNTCOA	% Total Grant by Cost of Attendance	Time-dependent
SCHLCOA	% Total Scholarship by Cost of Attendance	Time-dependent
LOANCOA	% Total Loan by Cost of Attendance	Time-dependent
WORKCOA	% Total Work Study by Cost of Attendance	Time-dependent

The complete model was fit simultaneously using the SPSS (2001) binary logistic regression procedure. The parameter coding for the additional categorical variables in the complete model is shown in Table 12, where the referent category for COLL is Agricultural Sciences, and the referent category for CONTENR is not continuously enrolled. The output in Table 12 has been truncated in an effort to save space.

Table 12

Categorical Variable Parameter Coding Complete Model

			Parameter coding		
		Frequency	(1)	(2 - 9)	(10)
COLL	Science	3490	1.000	.000	.000
	Engineering	7874	.000	1.000	.000
	Earth And Mineral Sciences	488	.000	1.000	.000
	Business Administration	4086	.000	1.000	.000
	Arts And Architecture	1436	.000	1.000	.000
	Undecided	3716	.000	1.000	.000
	The Liberal Arts	2687	.000	1.000	.000
	Health And Human Development	782	.000	1.000	.000
	Education	854	.000	1.000	.000
	Communications	1256	.000	.000	1.000
	Agricultural Sciences	1844	.000	.000	.000
CONTENR	Continuous Enrollment	27084	1.000		
	Not Continuous	1429	.000		

To determine how well the model fit the sample data, a Hosmer and Lemeshow (1989) goodness of fit test was computed. For the complete model, $\hat{C}$ (8, N=28,513) = 9.99, p = .27, which indicates that the model fits the data well.

The model summary and individual parameter estimates for the complete model are shown in Table 13. The omnibus tests of model coefficients provide a test for the null hypothesis that all beta coefficients are equal to 0. As can be seen from the omnibus tests, the null hypothesis is rejected with, $c\,2$ (40, N = 28,513) = 952.36, $p < .001$. The block chi-square tests for any additional model improvement associated with the inclusion of the enrollment and financial aid variables.

Table 13

Model Summary Complete Model

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Block	919.946	26	.000
Model	952.356	40	.000

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
AGE	-.004	.017	.063	1	.802	.996
GENDER(1)	.015	.047	.103	1	.748	1.015
HSGPA	-.188	.089	4.418	1	.036	.829
SATV	-.001	.000	3.972	1	.046	.999
SATM	.000	.000	.060	1	.807	1.000
ETHNIC			.805	5	.977	
ETHNIC(1)	-.086	.269	.101	1	.751	.918
ETHNIC(2)	.061	.101	.363	1	.547	1.063
ETHNIC(3)	-.037	.171	.047	1	.828	.964
ETHNIC(4)	-.027	.161	.028	1	.867	.973
ETHNIC(5)	.262	.542	.234	1	.628	1.300
RESID(1)	.127	.057	4.983	1	.026	1.135
LEGACY(1)	.029	.046	.414	1	.520	1.030
RANK	-.002	.003	.658	1	.417	.998
ACAD_INT	.001	.005	.045	1	.832	1.001
SEMGPA	.614	.054	130.101	1	.000	1.848
CUMGPA	.384	.072	28.220	1	.000	1.469
COLL			9.344	10	.500	
COLLCHNG	-.004	.043	.007	1	.934	.996
MAJCHNG	-.055	.043	1.655	1	.198	.947
CONTENR(1)	.274	.120	5.199	1	.023	1.315
EARNATT	-.003	.005	.256	1	.613	.997
TOTAID	.001	.001	3.491	1	.062	1.001
TOTGRNT	-.001	.001	1.603	1	.205	.999
TOTLOAN	-.001	.001	1.272	1	.259	.999
TOTWORK	-.001	.001	1.789	1	.181	.999
TOTSCHL	-.001	.001	2.160	1	.142	.999
COARAT	-.105	.058	3.298	1	.069	.900
GRNTCOA	.073	.059	1.568	1	.210	1.076
LOANCOA	.084	.058	2.106	1	.147	1.088
SCHLCOA	.082	.058	2.015	1	.156	1.086
WORKCOA	.100	.066	2.335	1	.127	1.106
Constant	-4.510	.484	86.978	1	.000	.011

There is a significant improvement by adding the additional variables to the model, χ^2 (26, N = 28,513) = 919.95, $p < .001$.

A review of the individual variable parameters reveals that GENDER is no longer significant in the model, its effects having been mitigated by other variables.

HSGPA is significant, χ^2 (1, N = 28,513) = 4.42, $p = .04$, but its direction is counterintuitive. The interpretation of odds ratios when covariates are continuous is augmented by expressing the odds ratio as a percentage change in the risk of graduation (Allison, 1995). The conversion for HSGPA is accomplished by 100(.829-1) = -17.1%, where .829 is the odds ratio for HSGPA. For every unit increase in HSGPA there is a 17.1% decrease in the hazard for graduation.

SATV just reaches statistical significance, χ^2 (1, N = 28,513) = 3.97, $p = .05$. The coefficient for SATV is also negative, with a 0.10% decrease in the hazard of graduating for every unit increase in SATV. An increase of 100 points in the SATV score would result in a 10% decrease in the hazard of graduating.

The negative, and counterintuitive, coefficients for HSGPA and SATV can be explained in large part by collinearity. If a new model is fit using just SATV and SATM scores to predict graduation, neither variable is significant. If HSGPA is added, HSGPA is significant, χ^2 (1, N = 28,513) = 15.22, $p < .001$, and the coefficient is positive. Students are 25.4% more likely to graduate for every unit increase in HSGPA. SATV and SATM are still not significant. However, if SEMGPA and CUMGPA are added to the model, HSGPA remains significant, χ^2 (1, N = 28,513) = 6.14, $p = .01$, SATV becomes significant, χ^2 (1, N = 28,513) = 7.45, $p = .01$ but the signs are both negative. SEMGPA is highly significant, χ^2 (1, N = 28,513) = 149.75, $p < .001$, as is

CUMGPA, χ^2 (1, N = 28,513) = 15.60, $p < .001$. HSGPA and SATV are correlated with each other and with SEMGPA and CUMGPA. When all four variables are in the model simultaneously, collinearity affects the sign of the coefficients. This effect, namely some covariates having negative coefficients, when HSGPA, SATV, and SATM scores are included in models used to predict college students' first year grades, is well documented.

It is clear that GPA information is the most powerful predictor in the complete model. This makes intuitive sense. Information about how a student is progressing academically is more important in describing graduation than information relating to past performance. Adding GPA information in the form of SEMGPA and CUMGPA to a large degree mitigates the effects of preenrollment characteristics.

The approach taken in the present study is to fit sets of variables simultaneously, based on the institution's data sources and the graduation literature. Although somewhat controversial, stepwise procedures can be implemented to help with variable selection and deletion and could be part of future research. The stepwise procedure selects and deletes variables by maximizing the likelihood function.

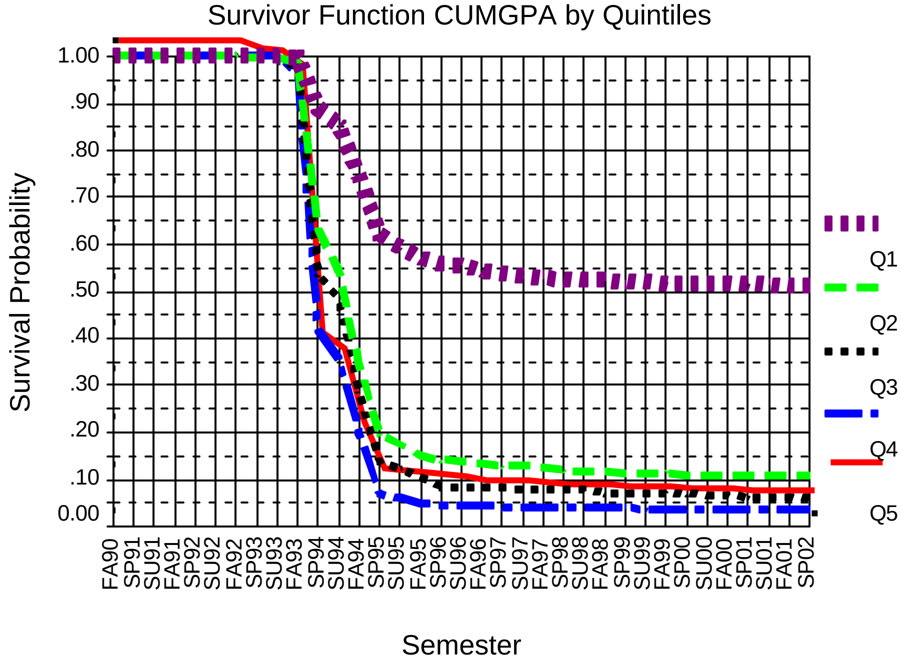
RESID remains statistically significant, χ^2 (1, N = 28,513) = 4.98, $p = .03$, but the sign has changed. State residents are now 1.14 times more likely to graduate than non-residents. It is likely that the change in sign comes as a function of the potential time effect observed in the sample survivor function (Figure 5), where the survival function for non-residents and residents crossed at approximately summer semester 1995. The inclusion of enrollment and financial aid variables could also affect the change in sign. Including a RESID by time interaction variable in the model could help illuminate the time effect.

SEMGPA is significant, χ^2 (1, N = 28,513) = 130.10, $p < .001$, and is time-dependent. The values of SEMGPA were permitted to change on a semester basis. The odds ratio for SEMGPA is interpreted as a one unit increase in SEMGPA results in an 84.8% increase in the hazard of graduating. SEMGPA provides the largest single contribution to the model.

CUMGPA is also significant in the model, χ^2 (1, N = 28,513) = 28.22, $p < .001$. A one unit increase in CUMGPA results in a 46.9% increase in the hazard of graduating. The survivor function for CUMGPA is represented in Figure 6 and is included to address the timing of graduation for students based upon their cumulative GPA. Students have been assigned to CUMGPA quintiles based upon their cumulative GPA during their last semester of enrollment. As can be seen from Figure 6, the overall

Figure 6

Survivor Function CUMGPA by Quintiles



graduation rates vary by quintiles, with the most drastic difference occurring within the first quintile.

Inclusion in the first quintile was based upon a cumulative GPA of 2.63 or lower. The overall graduation rate for the first quintile is 49%, much lower than the other quintile groups. Approximately 40% of the students in the first quintile who graduated did so by summer semester 1995.

Inclusion in the second quintile was based upon a cumulative GPA greater than 2.63 and less than or equal to 3.00. The overall graduation rate for the second quintile is 88.9%, with the median lifetime value occurring at approximately summer semester 1994. An additional 20%, or approximately 70%, graduate by fall semester 1994.

Inclusion in the third quintile was based upon a cumulative GPA greater than 3.00 and less than or equal to 3.26. The overall graduation rate for the third quintile is 94%, with the median lifetime value occurring at summer semester 1994. An additional 20%, or approximately 70%, graduate by fall semester 1994.

Inclusion in the fourth quintile was based upon a cumulative GPA greater than 3.26 and less than or equal to 3.51. The overall graduation rate for the fourth quintile is 96%, with the median lifetime value occurring between fall semester 1993 and spring semester 1994. Approximately 65% of the fourth quintile graduate by spring semester 1994. An additional 15%, or approximately 80%, graduate by fall semester 1994.

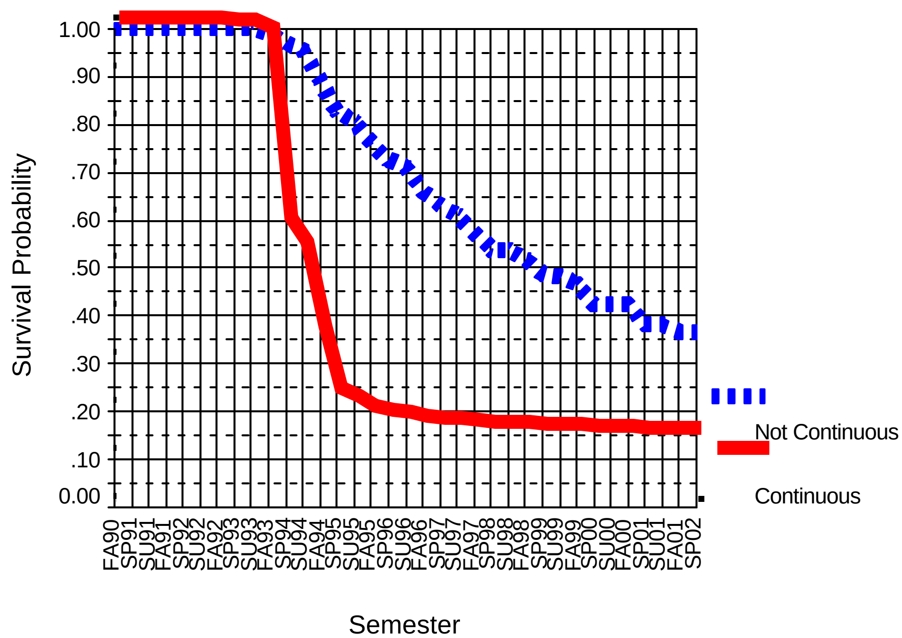
Inclusion in the fifth quintile was based upon a cumulative GPA greater than 3.51. The overall graduation rate for the fifth quintile is 96%, with the median lifetime value occurring between fall semester 1993 and spring semester 1994. Approximately

60% of the fifth quintile graduate by spring semester 1994. An additional 20%, or approximately 80%, graduate by fall semester 1994.

CONTENR is significant, χ^2 (1, N = 28,513) = 5.20, $p < .001$. Students continuously enrolled were 1.31 times more likely to graduate than those who were not continuously enrolled. The sample survivor function for CONTENR is shown in Figure 6. The median lifetime value for continuously enrolled students is summer semester 1994 and for students not continuously enrolled, the median lifetime value is spring semester 1999. However, the practical utility of the continuously enrolled variable is

Figure 7

Survivor Function for CONTENR



limited in the present study by the fact that 95.7% of the present study sample remained enrolled continuously prior to their departure from the institution under study. Any intervention based upon the current findings would be directed at a very small percentage of students.

Although not statistically significant at the p < .05 level, TOTAID, *c* 2 (1, N = 28,513) = 3.50, p = .06 and COARAT, *c* 2 (1, N = 28,513) = 3.29, p < .07 nearly reach significance and provide some support for the effect of student aid found in Adelman's study.

Summary

Two statistical models were fit in the present study. The first included just preenrollment characteristics in an effort to determine if graduation behaviors could be explained by background variables. The only significant factor in the preenrollment model was GENDER. Women graduated earlier and at a higher rate than men. The effect of GENDER, however, was mitigated by other variables in the model containing enrollment and financial aid data.

This outcome is consistent with the segment of the literature stating that background characteristics are minimally important in explaining retention and graduation when enrollment variables are included in the model. It is also supported by the student integration model of Tinto (1975) and the student attrition model of Bean (1980). In both cases, the importance of background or preenrollment characteristics is secondary to social and academic integration. Bean (1983) actually excludes background characteristics from one iteration of his student attrition model.

The substantial contribution SEMGPA and CUMGPA make to explaining the graduation behavior of students also supports the student integration model of Tinto (1975) and the student attrition model of Bean (1980). SEMGPA and CUMGPA serve as measures of academic integration, one of the central components of both models. For Tinto, they serve as proxies for academic institutional fit, and for Bean they serve as rewards in the worker turnover model. For every unit increase in SEMGPA, the hazard of graduation increases by approximately 85%, and for every unit increase in CUMGPA the hazard of graduation increases by approximately 47%.

Tinto (1975) and Bean (1980) would argue that students who perform at a higher level are more academically integrated, leading to better institutional fit. These students are motivated to persist and complete a degree. The importance of institutional fit over background characteristics is supported by a recent review of the literature by Kennedy and Scheckley (1999). Using proportion of variance as the determining criterion, they indicate that demographic information accounts for a small portion of explained variance. They also state that institutional fit explains most of the variance in modeling student retention and graduation.

A second model was fit by adding enrollment and financial aid variables to the preenrollment model. One of the primary goals of the present research was to add to the student retention literature using survival analysis as a methodology. Although the research has been limited, Ronco (1995) found that provisional admits and students with low semester GPA's were at greatest risk of dropout and were less likely to graduate. Students with higher GPA's were more likely to graduate and less likely to dropout or transfer. Hispanics were less likely to graduate. Full-time students were twice as likely to graduate as part-time students. Murtaugh, Burns, and Schuster (1999) found students with higher first quarter GPAs were less likely to withdraw, and found significant college effects. They also found that graduation rates decrease with age. DesJardins (1996) and DesJardins, Ahlburg, and McCall (1999) used survival analysis to study 3,975 new students beginning fall term 1986 at the University of Minnesota, Minneapolis campus. They found that the higher a student's ACT score the less likely the student was to drop out. Students with higher semester GPAs were also less likely to drop out. They found that students with loans, work study, and scholarships were less likely to drop out.

The present research supports the findings relative to GPA, and continuous enrollment. Although not significant, the total aid amount students receive by semester and the percentage of total aid to cost-of-attendance at the institution were approaching significance. In addition, state residents were more likely to graduate than non-residents, but not until after approximately six years. HSGPA and SATV were significant, but their coefficients were negative. As was explained previously, this effect is largely due to collinearity and the strong relationship between SEMGPA, CUMGPA, and graduation.

Another goal of the present research was to extend the work of Adelman (1999) by replicating Adelman's study, to the degree possible, with data from a single institution and using survival analysis as the statistical method. Adelman conducted one of the most comprehensive studies of graduation to date. He found that the Academic Resources Index he developed, continuous enrollment, and the ratio of credits attempted to credits earned, were the most significant contributors to his model of graduation behavior.

DesJardins and Moye (2000) did apply survival analysis to Adelman's (1999) data. They found that the Academic Resources Index positively affects graduation, but that the effects are substantially mitigated by GPA. When GPA is included in the model, the effects of the Academic Resources Index are reduced by approximately 41%. They also found that when aid is allowed to vary over time, loans enhance and work/study inhibits timely graduation, a finding contrary to Adelman. DesJardins and Moye did not include several important factors in Adelman's model, namely, a measure of continuous enrollment and the percentage of credits earned to credits attempted.

In an attempt to replicate Adelman's work, four variables were included in the present study to test the Academic Resources Index of Adelman (1999). SATV and SATM scores were included as a proxy for Adelman's senior test component of his index. High school class rank was entered as the actual rank score in place of Adelman's high school rank quintiles. Finally, ACAD_INT was developed and included in the model to represent Adelman's quintiles of the graduated 40 point academic intensity component of his index. CONTENR was included as an indicator of continuous enrollment and EARNATT was included as the percentage of credits earned to credits completed.

The only statistically significant variable representing Adelman's (1999) Academic Resources Index in the present study was SATV. The utility of the composite as a whole was not supported, and neither was the percentage of credits earned to credits attempted. However, the continuous enrollment variable was significant in the model, with students continuously enrolled being 1.31 times more likely to graduate than students who were not continuously enrolled. Continuously enrolled students were also likely to graduate much earlier. The median lifetime value for those continuously enrolled was summer semester 1994, and for students not continuously enrolled, the median lifetime value was spring semester 1999. However, 95.7% of students in the current study were continuously enrolled, limiting the practical import of this finding.

Chapter V

DISSCUSION, LIMITATIONS, AND IMPLICATIONS

Introduction

The retention and graduation of college students has been of critical importance to the higher education community for decades. Costs accrue to society, the institution, and the student when degree completion is not realized (Pascarella & Terenzini, 1991).

For the student, degree completion is linked to potential earnings. In the course of a lifetime, college graduates earn on average $1,000,000 more than high school graduates, and persons with a professional degree will earn on average $3,000,000 more than a high school graduate (Yudof, 2002). College costs continue to escalate, especially at large public institutions where state funding has dwindled. Students and their families are faced with financing an ever increasing total cost of attendance, with a substantial portion of the increase being covered by student loans as opposed to gift aid. The ability to repay those loans is enhanced by the additional earning potential a degree provides.

For the institution, maintaining enrollments is critical to economic stability. Initial resources in recruiting and enrolling students are maximized when those students remain enrolled and graduate. Also, graduation rates are often used by constituents as a measure of the institution's performance, and are increasingly linked to resource allocation. The state in which the study institution is located has created a $6,000,000 grant program used to reward colleges and universities that graduate at least 40% of their state resident students within four years. The current federal administration has proposed a similar program as part of the reauthorization of the Higher Education Act.

From a societal perspective, earning a college degree helps to mediate background characteristics, such as family socioeconomic status, on subsequent earnings, occupational status, and social status attainment (Pascarella and Terenzini, 1991). College graduates are more likely to vote. Also, the increased earnings of college graduates result in an increase in the local and federal tax base.

Focus of Present Study

The focus of the present study was twofold. First, advancing the use of survival analysis in studying college student graduation behaviors was considered. In addition, the major findings of Adelman (1999) were incorporated into the analysis.

The study of college student retention and graduation has been advanced by the use of survival analysis. Willett and Singer (1991) provide a compelling argument in support of the use of survival analysis in retention and graduation research. Much of the previous research has been conducted using a cross-sectional methodology, where typically a cohort of students is identified and an end-point of the study is determined. The end-point could represent one semester, one year, four years, or six years, and is often somewhat arbitrary. At the end of the study period, the characteristics of those students who graduated are compared to the characteristics of those students who did not. The temporal nature of graduation behaviors is ignored using a cross-sectional methodology.

Determining the duration of graduation studies and addressing the longitudinal nature of retention and graduation has been discussed in the literature. Pantages and Creedon (1978), after reviewing 25 years of attrition research, suggest that:

> The most meaningful research on attrition is provided by studies that cover a period of more than four years and that use precise operational definitions of dropout and nondropout. (p. 92)

Terenzini (1987) states:

> Longitudinal designs provide extensive, planned control of confounding background variables, as well as more precise estimates of the institutional influences of attendance behavior. Such designs are the most internally valid available for studying attrition and afford a measure of confidence in findings and associated conclusions that is not available with other designs. (p. 29)

Discrete time survival analysis allows for the study of graduation behaviors with a truly longitudinal design. In the present study, semester specific data were used over a 35 semester period to create a longitudinal profile of graduation, and to test the relationships among graduation and a set of covariates.

The outcome of the present study provides support for the strong relationship between the longitudinal effects of GPA and graduation, a consistent finding in the research using survival analysis as a statistical methodology for the study of retention and graduation. As semester or cumulative GPA increase, so does the likelihood of graduation. This makes intuitive and common sense, and adds support to the notion of academic integration in Tinto’s (1975) and Bean’s (1980) models of student retention and attrition. It is possible that GPA reflects not only a student's ability, but also the motivation that student has to do well and complete a degree.

The strong relationship GPA has with graduation mitigates other factors, such as background characteristics. Adelman (1999) found that the preenrollment variables in his Academic Resources Index accounted for 17% of the 43% of explained variance in his model. Adelman did not, however, include college GPA in his study. DesJardins and Moye (2000) analyzed Adelman's data using survival analysis and found that the inclusion of semester GPA reduced the contribution the Academic Resources Index made by 41%.

One goal of the present study was to replicate the work of Adelman (1999), especially relative to his Academic Resources Index. Variables included in the present study to replicate the Academic Resources Index included; ACAD_INT, representing the academic intensity of each student's high school curriculum; SATV and SATM, used as proxies for Adelman's senior test; RANK, representing Adelman's class rank quintiles; and HSGPA, representing Adelman's high school GPA quintiles. The ACAD_INT variable was not significant in the preenrollment or complete model, even without the inclusion of SEMGPA or CUMGPA in the model. SATM and RANK were not significant in either the preenrollment or complete models. SATV and HSGPA were significant in the complete model, but their coefficients were negative. As was discussed in Chapter IV, this is due in large part to the influence of SEMGPA and CUMGPA. The present study did not provide support for Adelman's Academic Resources Index as a predictor of graduation.

Adelman (1999) also found continuous enrollment as the second most important factor in his model. CONTENR, the continuous enrollment variable in the present study, was statistically significant, providing support for Adelman's findings. Not only was there a significant difference in GRADUATE, but those students who were continuously enrolled graduated much earlier than those who were not continuously enrolled. Again, this makes intuitive sense; those students who don't stop out should complete a degree and graduate sooner then those who do stop out. The added benefit, which is made possible by the use of survival analysis, is how much sooner, or when. The ability to address the temporal nature of retention and graduation is one of the added benefits of survival analysis.

Adelman (1999) also found the percentage of credits earned to credits attempted to be a significant factor in his model. This finding was not supported by the present research. The EARNATT variable was not significant in the preenrollment or the complete model.

RESID was also significant, but the results were mixed. Results from the preenrollment model showed non-resident students were more likely to graduate. Once enrollment and financial aid variables were added, the sign of the coefficient changed. The shift in sign could be a function of a time interaction effect with RESID, which could be explored by fitting a time by RESID interaction covariate in the model. Also, the inclusion of enrollment and financial aid variables introduces a new set of interrelationships among variables, which could also contribute to the change in sign.

Discussion

The benefits of survival analysis as a methodology for studying the timing of events are clear in many different areas of study. Extending survival analysis to the study of retention and graduation has provided an approach to studying the temporal nature of graduation behaviors. The question of whether students graduate is now implicit in the when do students graduate question. Even though survival analysis could be the most appropriate method for studying graduation behaviors, results from one study to another will almost certainly always be mixed.

One of the goals of the present study was to extend the work of Adelman (1999) using survival analysis. The degree to which Adelman's findings were supported was mixed. The differences in outcome could be explained by the different methodology, differences in the way variables were developed or collected, and differences in the two samples. With significant disparity in graduation rates from one institution to another, it should not be surprising that study results also vary drastically.

Adelman (1999) studied the graduation behaviors of a national sample of high school graduates enrolled at a multitude of institutions, with different entrance criteria, grading standards, cultures, and environments. There is certainly more variability in Adelman's sample than in the single institution sample used in the present study. Tinto (1982) in addressing the operational definition and differential interests of a system vs. institution perspective of dropout states:

> We will ask not only how dropout may be defined, but also how that definition may vary among different interested parties concerned with the character of dropout from higher education. (p. 3)

This statement can be extended to the study of graduation. The question becomes how generalizable to single institutions are the findings of studies conducted on national samples. The degree to which the findings are generalizable is a function of how much similarity there is between the study sample and the institution interested in effecting some change relative to retention and graduation. The single best place to influence graduation behaviors is at the institutional level. If institutions do not have information that is relevant to their unique situation, they will be ineffective in their efforts to increase graduation rates.

To illustrate the disparity in the samples of Adelman (1990) and the present study, a comparison of the variability for each factor will be made. Adelman developed a 40-point graduated scale, which was converted to quintiles, in an attempt to capture the academic intensity of a student's high school curriculum. Completion of Algebra 2 was one important criterion in the development of the index, and allowed Adelman to categorize students by whether or not they had completed Algebra 2. Adelman (1990) indicated:

> Of all the components of curriculum intensity and quality, none has such an obvious and powerful relationship to ultimate completion of degrees as the highest level of mathematics one studies in high school. (p. 16)

Adelman identified Algebra 2 as the real threshold for the influence of math on graduation. In the present study, all students had completed Algebra 2, thereby reducing the amount of variability in math compared to Adelman. It is possible that the ACAD_INT variable might have been significant if there had been more variability in the present sample.

HSGPA and SATV, both proxies for components of Adelman's (1999) Academic Resources Index, were significant in the complete model for the present study, but their signs were negative. This is mainly due to collinearity among these variables and other variables in the model. As has been discussed, the influence of SEMGPA and CUMGPA contributes significantly to the change in sign.

Also contributing to the differences in variability is the difference in graduation rates. Adelman (1999) indicated that 63% of his study sample completed a bachelor's degree sometime during the study period, where 83% of the sample for the present study earned a bachelor's degree. When modeling graduation behaviors, Adelman is comparing the 37% of his sample who did not graduate to the 63% who did graduate. In the present study, 17% of the sample who did not graduate are being compared to 83% who did graduate.

The generalizability of graduation studies is always suspect, in part for the reasons cited above. When research is done at the higher education system level across multiple institutions, the disparity in those institutions in terms of entrance criteria, cost, grading policies, competition, and mission, for example, is large. The degree to which findings are applicable to any one institution is diminished by the aggregation across multiple institutions. Policy makers develop plans that are based on averages across institutions, which do not fit any one institution well.

When graduation research is conducted at the institution level, criticisms are leveled relative to the generalizablility of results to the system. However, to realistically affect graduation rates, policy and practice must be implemented at the institution level. Institution specific research is critical to an understanding of graduation behaviors, and the subsequent development of strategies to influence students to remain enrolled and graduate. Terenzini (1982) states:

> Literature reviews provide a general and useful understanding of the complexity of the attrition phenomenon, but in no way do they substitute for the local research many institutional research offices are asked to undertake. (p. 55)

The two most researched and cited models of student retention, Tinto (1975) and Bean (1980), are described as institutional and not systems models. Retention and graduation studies must be conducted at the institutional level if they are to have any real practical import.

Limitations

Typically, research conducted at a single institution is described as limited in terms of its generalizability to the system as a whole. However, institution specific research has the greatest likelihood of effecting change in graduation behaviors.

The sample for the present study was homogeneous, with little variability in some of the covariates. Although this could affect the outcome, it is representative of the cohort under study, and provides an accurate representation of the relationships of covariates with graduation for the institution under study.

Only one cohort of students was studied. Cohort effects must be considered in future research. Additional research should include multiple cohorts to determine if the results of the present study are consistent from one cohort to the next.

When considering competing risks, dropout was identified as a competing risk for students leaving the sample. There was no distinction made between those students who were dismissed for disciplinary reasons, those students who were academically dismissed, or those students who withdrew in good academic standing. This was both a theoretical and practical decision. If data were available to disaggregate the above categories, a competing risks model of dropout would be helpful to determine if there were differences in the risk of dropout in these categories.

Implications and Recommendations

The present study has shown the benefits of survival analysis as a methodology for the study of college student graduation behaviors, and has extended the research using survival analysis to study graduation. As graduation rates are increasingly tied to resource allocation and institutional performance, community members must have a better understanding of graduation behaviors to develop interventions aimed at helping students complete a degree.

The implications of the present findings for national policy are limited. The study was conducted with data from a single institution. Possibly, the most important contribution of the present study at the higher education systems level is the advancement of survival analysis as a methodology for studying graduation behaviors.

The benefits of survival analysis to study time to event data have been shown in multiple disciplines over many years. Adopting a consistent methodology for studying retention and graduation behaviors would do much to move the research forward. Continuing to conduct studies at the higher education systems level across disparate institutions without considering the variability associated with those institutions is problematic. However, system level research must be conducted to inform policy development. If a consistent methodology, such as survival analysis, was used to conduct institution specific research, researchers could explore combining those studies in a hierarchical analysis where institutions were considered as random covariates. Using this approach would allow the variability of each institution to contribute to the variance component of the model, and is a recommendation for future research.

For the institution under study, students are not graduating in four years, the amount of time typically identified by legislators and policy makers as the appropriate interval for obtaining a bachelor's degree. As four year graduation becomes a more important indicator of success and access to resources, institutions need to better understand the impediments to completing a degree in four years.

Students enrolled continuously are more likely to graduate and graduate earlier. Exploring the causes of stopout behavior would also be of benefit to students and the institution, such as the influence of personal issues, and financing their degree. A better understanding of the causes of stopout behavior could facilitate the development of retention and graduation interventions. However, for the study institution, approximately 96% of the students in the sample were continuously enrolled. The potential impact realized by increasing that percentage is minimal.

The state residency of students was also a significant contributor to the understanding of graduation. Residents of the institution's home state were 1.14 times more likely to graduate than non-residents. However, the timing of graduation is mixed. Non-residents were more likely to graduate prior to summer semester 1995 and resident students were more likely to graduate after summer semester 1995. Both the difference in graduation rates and timing could come as a function of the tuition differential at the study institution. The tuition for non-resident students is double the tuition for resident students, which leads to a much greater financial burden for non-residents not completing a degree in a timely fashion. The influence of tuition, especially in regard to the resident/non-resident differential, must be considered in the development of retention and graduation interventions.

The findings of the present study support the intuitively appealing position that those students who have better semester GPAs are more likely to graduate. GPA is an indicator of ability and motivation. This supports the student retention model of Tinto (1975) and the student attrition model of Bean (1980). Academic integration, measured by GPA, has long been cited, and researched, as a critical component of retention and graduation. However, the more important contribution made in the present study, attributed to the use of survival analysis, is the addition of the temporal understanding of the effects of GPA.

To study the time effects of GPA, students were assigned to quintiles based upon the cumulative GPA of their last semester of enrollment. The survivor function based upon quintiles was computed and is presented in Table 6. The graduation patterns of students varied by quintile.

The four year graduation rate for students in the fourth and fifth quintiles (cumulative GPA greater than or equal to 3.51) was approximately 60%. The graduation rate increases to 80% with the addition of summer and fall semesters 1994. This compares to an overall four year graduation rate of 41%. The overall five year graduation rate is 76%, less than the 80% realized by the fourth and fifth quintiles through fall semester 1994. Students in the fourth and fifth quintiles are motivated and talented. The overall graduation rate for these students is 96%, which is excellent.

The four year graduation rate for students in the second and third quintiles (cumulative GPA greater than or equal to 3.00 and less than 3.51) was approximately 40%. The graduation rate increases to 70% with the addition of summer and fall semesters 1994. The overall graduation rate for students in the second quintile is 89%, and the overall graduation rate for students in the third quintile is 91%. Again these overall graduation rates are excellent.

For students in the second through fifth quintiles, the reasons for additional semesters of study are of concern, not the overall graduation rates. Determining whether program credit requirements, credit drop/add policies, course availability, or advising policies, for example, influence the need for an additional semester of study to realize degree completion is warranted. Administrators must consider the degree to which institutional policy might influence the need for students to take additional semesters of course work to complete a degree.

The situation is quite different for students in the first quintile (cumulative GPA 2.63 or less), where the four year graduation rate is approximately 10%. The graduation rate increases to approximately 30% with the addition of summer and fall semesters

1994, and only reaches 49% by the end of the study. Although identifying potential institutional impediments to timely graduation is important with this group as well, the most critical issue is that of overall graduation rates. Administrators must use poor performance, as measured by GPA, as an indicator of lack of institutional fit and must categorize students with low cumulative GPAs as an at risk population. Although this might seem obvious, it is clear that efforts to increase graduation rates will be most beneficial if they are targeted at students with cumulative GPAs in the 2.63 range and below. This might not be the optimal cut point, but it is clear that the graduation rates for students below this range are much lower and that some intervention is warranted. The use of survival analysis has provided important information regarding the timing of graduation and categorizing students based on cumulative GPA quintiles has identified where the most potential for improvement exists.

This study has addressed the two most important components of graduation. First, overall graduation rates are of importance for students, the institution, and society. Time to degree is important for those constituencies as well. Based on the findings of the present study, using semester GPA information as an early warning to intervene when students are doing poorly is critically important. This supports the student retention model of Tinto (1975) and the student attrition model of Bean (1980). Additionally, the findings of the present study indicate that interventions are most important for students with cumulative GPAs in the approximately 2.60 range and lower. The generalizability of these findings should be tested with additional cohorts for the study institution and at different institutions for the higher education system. The important finding is that interventions should be targeted based on cumulative GPA and have the most potential for success if they occur early. The graduation behaviors of students in the lowest cumulative GPA quintile are drastically different than their counterparts at the earliest time periods.

Administrators must also be cognizant of the differences in graduation rates and timing based upon residency. As tuition increases, the effect those increases have on the ability of non-residents to finance an education must be considered. As was shown in the present study, non-residents are less likely to graduate, but those who do graduate do so earlier than their resident counterparts. Pricing strategies must be considered carefully.

It is recommended that future research on graduation behaviors should be conducted using survival analysis at the institutional level, the level at which it is most likely that graduation behaviors can be influenced. For research at the higher education systems level, researchers should explore the possibility of combining the results of studies using survival analysis in a hierarchical model, where the individual study results would be treated as a random covariate, adding to the variance component of the model. In that way, the variability across institutions could be explored.

REFERENCES

Adelman, C. (1999). Answers in the tool box: Academic intensity, attendance patterns, and bachelor's degree attainment. Jessup, MD: U.S. Department of Education.

Allison, P. D. (1995). *Survival Analysis Using the SAS System: A Practical Guide*. Cary, NC: SAS Institute Inc.

Barger, B. & Hall, E. (1965). Time of dropout as a variable in the study of college attrition. *College and University, 41,* 84-88.

Baltes, P. B. (1968). Longitudinal and cross-sectional sequences in the study of age and generational effects. *Human Development, 11*, 145-171.

Bean, J. P. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in Higher Education, 12(2),* 155-187.

Bean, J. P. (1982). Student attrition, intentions, and confidence: Interaction effects in a path model. *Research in Higher Education*, *17(4),* 1982.

Bean, J. P. (1983). The application of a model of turnover in work organizations to the student attrition process. *The Review of Higher Education*, *6(2),* 129-148.

Braxton, J. M. & Lien, L. A. (2000). The viability of academic integration as a central construct in Tinto's interactionalist theory of college student departure. In J. M. Braxton (Ed.), *Reworking the Student Departure Puzzle* (pp. 11-28). Nashville, TN: Vanderbilt University Press.

Cabrera, A. F., Castaneda, M.B., Nora, A., & Hengstler, D. (1992). The convergence between two theories of college persistence. *Journal of Higher Education, 63(2)*, 143-164.

Cabrera, A. F., Nora, A., & Castaneda, M.B. (1993). College persistence: Structural equations modeling test of an integrated model of student retention. *Journal of Higher Education, 64(2),* 123-139.

Durkheim, E. (1951). *Suicide, a study of sociology*. (J. Spaulding and G. Simpson, trans.). Glencoe, Ill: Free Press.

DesJardins, S.L. (1996). Using event history modeling to study the temporal dimensions of student departure from college. *Dissertation Abstract International, 57(09)*, 3759. (UMI No. 9706199)

DesJardins, S.L., Ahlburg, D.A. & McCall, B. P. (1999). An event history model of student departure. *Economics of Education Review, 18*, 375-390.

DesJardins, S.L., & Moye, M. J. (2002). Studying the timing of student departure from college. Paper presented at the Annual Forum of the Association for Institutional Research.

Gekowski, N. & Schwartz, S. (1961). Student mortality and related factors. *Journal of Educational Research, 54*, 192-194.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied Logistic Regression*. New York: John Wiley and Sons.

Jex, F. B., & Merrill, R. M. (1962). A study in persistence. *Personnel and Guidance Journal, 40,* 762-769.

Moore, P. S. (1994). The role of discrete-time survival analysis in higher education enrollment management. *Dissertation Abstracts International, 56(01),* 111. (UMI No. 9517293)

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

Pantages, T. J. & Creedon, C. F. (1978). Studies of college attrition: 1950-1975. *Review of Educational Research, 48(1),* 49-101.

Panos, R. J. & Astin, A. W. (1968). Attrition among college students. *American Educational Research Journal, 5(1),* 57-72.

Pascarella, E., Duby, P., & Iverson, B. (1983). A test and reconceptualization of a theoretical model of college withdrawal in a commuter institution setting. *Sociology of Education, 56*, 88-100.

Pascarella, E. T., & Terenzini, P. T. (1983). Predicting voluntary freshman year persistence/withdrawal behavior in a residential university: a path analytic validation of Tinto's model. *Journal of Educational Psychology*, *75*, 215-226.

Pascarella, E. T., & Terenzini, P. T. (1991). How college affects students: Findings and insights from twenty years of research. San Francisco, CA: Jossey-Bass.

Prediger, D. J. (1965). Prediction of persistence in college. *Journal of Counseling Psychology*, *12*, 62-67.

Price, J. L. (1977). *The Study of Turnover*. Ames, IA: Iowa State University Press.

Price, J. L., & Mueller, C. W. (1981). A causal model of turnover for nurses. *Academy of Management Journal, 24,* 543-565.

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

Ronco, S. L. (1995). How enrollment ends: Analyzing the correlates of student graduation, transfer and dropout with competing risks model. Paper presented at the Annual Forum of the Association for Institutional Research.

Rose, H. A. & Elton, C. F. (1966). Another look at the college dropout. *Journal of Counseling Psychology, 13*, 242-245.

Schaie, K. W. (1965). A general model for the study of developmental processes. *Psychological Bulletin, 64*, 92-107.

Simpson, C., Baker, K., & Mellinger, G. (1980). Conventional failures and unconventional dropouts: Comparing different types of university withdrawals. *Sociology of Education, 53*, 203-214.

Singer J. D., & Willett, J. B. (1993). New methods for studying event occurrence: Using survival analysis in early intervention research. *Journal of Early Intervention, 17(3)*, 322-339.

Spady, W. (1970). Dropouts from higher education: An interdisciplinary review and synthesis. *Interchange, 1*, 64-85.

Stokes, M. E., Davis, C. S. & Koch, G. G. (1995). *Categorical Data Analysis Using the SAS System*. Cary, NC: SAS Institute Inc.

Summerskill, J. (1962). Dropouts from college. In N. Sanford (Ed.), *The American College* (Part 5, pp. 627-657). New York: Wiley.

Terenzini, P. T., Pascarella, E. T., Theophilides, C., & Lorang, W. G. (1985). A replication of a path analytic validation of Tinto's theory of college student attrition. *Review of Higher Education, 55*, 621-636.

Terenzini, P. T. (1987). Studying student attrition and retention. In J.A. Muffo and G. W. McLaughlin (Eds.) *A primer on institutional research*. (pp. 20-35). Washington, D.C.: Association for Institutional Research.

The College Board (2000a).*Trends in College Pricing*. New York City, NY: College Board.

The College Board (2000b).*Trends in Student Aid*. New York City, NY: College Board.

Schneider, D., & Dorans, N. (1999). Concordance Between SAT I and ACT Scores for Individual Students. *Research Notes* (RN-07). New York City, NY: The College Board.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research, 45(1)*, 89-125.

Tinto, V. (1982). Defining dropout: A matter of perspective. In E. Pascarella (Ed.) *New Directions for Institutional Research: Studying Student Attrition, No. 36*. San Francisco: Jossey-Bass.

Tinto, V. (1988). Stages of student departure: Reflections on the longitudinal character of student leaving. *Journal of Higher Education, 59*, 438-455.

U.S. Department of Education, National Center for Education Statistics, *The Condition of Education 2000*, NCES 2000-062, Washington, D.C.: U.S. Government Printing Office, 2000.

Willett, 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.

Willett, J. B., & Singer, J. D. (1993). Investigating onset, cessation, relapse, and recovery: Why you should, and how you can, use discrete-time survival analysis to examine event occurrence. *Journal of Counseling and Clinical Psychology, 61(6),* 952-965.

Yudof, M. G. (2002, January 11). Is the public university dead? *The Chronicle of Higher Education*, B24.

Vita

Randall C. Deike

WORK EXPERIENCE

June 1999-Present Director of Research, Enrollment Management and Administration, The Pennsylvania StateUniversity

June 1996-November 1999 Director of Undergraduate Admissions Data Analysis, Undergraduate Admissions, The Pennsylvania State University

July 1995-June 1996 Information Analyst, Office of Student Aid, The Pennsylvania State University

July 1991 - July 1992 Associate Dean of Students, Juniata College, Huntingdon, PA

July 1988 - July 1991 Assistant Dean of Students, Juniata College, Huntingdon, PA

PUBLICATIONS AND PRESENTATIONS

Deike, R.C., Drews, D.R., McLaughlin, R. K., Tilden, A. J., & Burroughs, W. J. (1989). Personal possessions and college student attrition. SEPA annual meeting, Washington, D.C.

Deike, R.C. (1999). Predictive modeling and geodemography: How representative are the data? *On Target*, 29, 4-5.

Blazina, M.K., Drews, D.R., and Deike, R.C. (1992). The process-product Relationship in student learning and critical thinking. EPA annual meeting, Boston, MA

Vicary, J.R., Doebler, M.K., Bridger, J.C., Gurgevich, E.A., & Deike, R.C. (1996). A community systems approach to substance abuse prevention in a rural setting. *Journal of Primary Prevention*, 16(3), 303-319.

Vicary, J.R., Swisher, J., Doebler, M.K., Yuan, J., Bridger, J.C., Gurgevich, E.A., & Deike, R.C. (1996). Rural community substance abuse prevention and intervention. *Journal of Family and Community Health*, 19(1), 59-72.

Deihl, L., Vicary, J., & Deike, R.C. (1997). Longitudinal trajectories of self-esteem from early to middle adolescence and related psychosocial variables among rural adolescents. *Journal of Research on Adolescence*, 7(4), 393-411.

Vicary, J.R., Swisher, J., VonEye, A., & Deike, R.C. (In Preparation). Impact of changes in psychosocial predictors in subsequent frequency of drunkenness among rural adolescents.