DISSERTATION

GRADUATION HAZARDS AND SURVIVING COLLEGE: A DESCRIPTIVE STUDY OF THE LONGITUDINAL NATURE OF LOW-INCOME, FIRST GENERATION, AND MINORITY STUDENT ENROLLMENT AND GRADUATION

Submitted by

Heather Novak

School of Education

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2012

Doctoral Committee:

Advisor: David Most

Michael De Miranda
Michael Lacy
Paul Thayer

ABSTRACT

GRADUATION HAZARDS AND SURVIVING COLLEGE: A DESCRIPTIVE STUDY OF THE LONGITUDINAL NATURE OF LOW-INCOME, FIRST GENERATION, AND MINORITY STUDENT ENROLLMENT AND GRADUATION

There are ambitious institutional and national goals that aspire to improve the six year graduation rate for undergraduate students. An important element of increasing the overall rate lies in decreasing the educational attainment gaps for low-income, first generation, and other historically underserved students. Comprehensive theoretical approaches to student success show that campuses have the opportunity to influence these achievement gaps with intentional and integrated programming and policy; however, the first step of initiating campus changes is to understand how the longitudinal nature of enrollment varies for demographically different students. This study utilizes a competing risk event history analysis on six cohorts of Colorado State University (CSU) fall-start freshmen over eight academic years in order to describe their dropout and graduation trajectories across a variety of demographic and academic preparation variables. Results indicate that all students have the highest hazard of graduation at year five and the greatest dropout hazard at year one; however, the shapes of these hazards are different based on a student’s demographic characteristics. Students with high risk characteristics have much lower graduation hazards after year five and much higher dropout hazards after year one when compared to their low risk peers. Thus, findings from this analysis indicate that high risk students at CSU need to be directed on educational paths that keep them on track to graduate in five years and that these students may also need continued retention support during their second and third years.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION TO RESEARCH PROBLEM

There is a growing recognition among many higher education researchers that increasing the number of Americans who have four-year degrees is an issue of social justice (Baum, Ma, & Payea, 2010). In the 2009 State of the Union Address, President Obama stated that America would once again be the country with the largest proportion of adults with a college degree by the year 2020 (USA.gov, 2011). However, in meeting this ambitious goal, interventions are needed to help improve the graduation rates for certain groups of students. In particular, there is consensus in the research literature that low-income, minority, and first generation students attend and graduate from college at lower rates than their peers (Moore & Shulock, 2009). Institutions can use policy to influence student behavior because institutional environments can be designed to be an intervention for struggling students (Kuh, 2009). Therefore, institutions need to understand their own students' enrollment paths so that policies and curricula are effective for helping students who are falling behind (Offenstein, Moore, & Shulock, 2010). This study seeks to describe the longitudinal nature of enrollment and graduation for various subgroups of students at Colorado State University (CSU).

Increasing graduation rates for low-income, first generation and minority students is becoming an issue of national importance. International economic development data shows that the U.S. is still close to the top in college participation rates but ranks near the bottom in college completion rates (Moore & Shulock, 2009). The poor college completion rates in the U.S. are even more discouraging, considering that low-income students obtain degrees at considerably lower rates compared to upper income students- 25% compared to 56% (Tinto, 2004). Improving the rate that low-income, minority, and first generation students complete college is imperative because these are populations that colleges are increasingly going to have to serve, given demographic projections that low-income minority youth are the fastest growing population in the U.S. (Swail, 2002). Research shows that individuals and society benefit from increased education levels (Baum et al., 2010), but to fully recognize the benefit, these educational achievement gaps must be addressed.

Improving the graduation rates for subgroups of students is also a priority at the institutional level for CSU. In his fall 2011 presidential address, President Tony Frank set the ambitious goals of obtaining an 80% six-year graduation rate and eliminating the minority/nonminority graduation gap (Frank, 2011). This study helps inform student graduation and attrition at CSU by exploring the longitudinal nature of enrollment for various demographic groups. The analytical procedure presented serves as a model other institutions can utilize to explore the temporal dimensions of enrollment for subgroups of their students. For the purpose of this study, student success is defined as continuous enrollment through graduation. Student success can be defined to mean a variety of educational milestones or self-reported gains; however, the national and institutional priorities discussed revolve around improving graduation rates, so continuous enrollment and graduation are the outcomes of interest.

Low-Income Students

Low-income students are the least likely to enroll in and successfully complete a four-year degree. Demographically, low-income students tend to be ethnic minorities, first generation students, single parents, and have English as a second language (Engle & Tinto, 2008). Low-income students are the main subgroup of interest for this study because of their high risk of attrition. For the purpose of this study, low-income is defined as being Pell Grant eligible. Eligibility for the federal grant is commonly used as a proxy for low-income status, since nearly all Pell Grant recipients come from households in the bottom income quartile (Cook & King, 2007). Because first generation and minority status tend to be associated with low-income, these subgroups are also of primary interest. It is the author's intention to explore these enrollment patterns using these demographic variables not because these groups are inherently unable to complete postsecondary educations, but because their personal circumstances have made college graduation more difficult for them. As Colorado's land grant university, it is CSU's responsibility to serve all students, not just those who come from circumstances that enhance their ability to succeed in postsecondary education. This study will describe the enrollment and graduation trajectories of low-income, first generation, and minority students, with the hope that institutional practices can be informed to positively impact these lower performing student groups.

Most low-income students receive financial aid in the form of loans and grants to pay their college expenses. Despite receiving the majority of federal need-based grant aid, low-income students are more likely to borrow student loans and graduate with higher levels of personal debt. All low-income students have financial need (Advisory Committee on Student Financial Aid, 2010). Financial need is the difference between the cost of the institution and what the student is expected to pay based on their family's financial circumstances. Typically, 40% of college costs are not covered by financial aid for low-income students. It has been hypothesized that finances and unmet need play a major role in the graduation gap for low-income students (Cook & King, 2007). Financial aid is a possible institutional intervention that could improve the success rate for low-income, minority, and first generation students. Pell Grant eligible students who successfully graduate within six years are those who receive financial aid, are academically prepared for college, and enroll full time immediately after high school at a four-year institution (Wei & Horn, 2009). As a four-year institution with a land grant mission, CSU has the opportunity to help Pell Grant eligible students be successful.

The costs associated with higher education make success in postsecondary education difficult for low-income students. However, money isn't the only barrier to graduation. Student success is a complex issue and demographically, at-risk students (low-income, first generation, and minority) face a multitude of issues when navigating the college enrollment and completion process. When designing institutional interventions to help at-risk students, it is important to create policy designed specifically for these student subgroups. These policies have the added benefit of not only helping low-income, first generation, and minority students succeed, but in helping all students succeed. On the other hand, the opposite is not true. At-risk students often fall through the cracks if policy is aimed at the general population of students, rather than the groups described above (Thayer, 2000).

Graduation Rates of Low-income, First Generation, and Minority Students

Nationally, 57% of students who start at a four-year institution graduate within six years. This rate is considerably lower for specific groups of students; e.g., 40% for African Americans and 49% for Hispanics (Aud et al., 2010). CSU's graduation rates across ethnicity status follow this national trend.

Table 1.1 displays the six-year graduation rates for low-income, first generation, and minority students at CSU for the three most recent fall cohorts. Students are considered Pell Grant recipients if they received a Pell Grant their first fall semester. First generation status is determined by a student self-reporting that they will be the first in their family to obtain a bachelor's degree on their application. Minority status includes students who self-reported as Hispanic, African American, Asian, American Indian, Pacific Islander, or multiracial. Non-minority students are those who are international, white, or chose not to report their ethnicity.

Table 1.1

CSU Six Graduation Rates for Full-Time Freshmen

	Fall 2003	Fall 2004	Fall 2005
Pell	59.4	59.0	55.8
Non-Pell	65.7	65.6	66.2
First Generation	58.7	57.6	58.5
Non-First Generation	67.0	67.4	67.2
Minority	58.8	60.1	57.3
Nonminority	65.7	65.3	65.8
Total	64.8	64.7	64.6

As Table 1.1 displays, there are serious, substantial gaps in the rate that low-income, first-generation, and minority students graduate from CSU. Pell Grant recipients have lower graduation rates when compared to students who do not qualify for the Pell Grant. The fall 2003 cohort fell 6.3 percentage points behind non-Pell and this gap has increased to 10.4 percentage points for the most recent cohort, fall 2005. Over the last three cohorts, first generation students lagged behind their non-first generation peers by an average of nine percentage points. Minority students have the smallest gap compared to nonminority students when considering these three socioeconomic variables; however, minority students from the fall 2005 cohort still fell behind their nonminority peers by 8.5 percentage points. Gaps of this magnitude will not make President Frank's goal of closing the minority gap easily attainable. As Colorado's land grant institution, it is a part of the University's mission to serve low-income, first generation, and minority students. The graduation disparities presented in Table 1.1 show the need for continued assessment and efforts to improve the graduation rates of CSU's at-risk students.

Educating At-Risk Students

To increase the overall national bachelor degree attainment rate, our nation needs to address income-related inequalities in postsecondary education completion (Advisory Committee on Student Financial Aid, 2010). In a global economy, it behooves our nation to educate the majority of citizens. Beyond the global economic value of an educated workforce, there are national economic analyses showing that adults with a college education help local economies because they earn higher incomes and in turn generate higher tax revenues. This research also shows that human health, happiness and civic engagement is related positively to postsecondary degree completion (Baum et al., 2007). Educating a large proportion of our population is important from a social and economic perspective; thus, higher education needs to better serve the historically underrepresented to increase the percentage of educated people in our society.

It is important for institutions to assess their own data and track efficiencies for low-income and other historically underserved students. Students at one institution will tend to be more similar to each other than students across institutions; therefore, policy adjusted at the institutional level will be the most effective for an institution's own group of diverse students (Offenstein et al., 2010). The purpose of this study is to provide a description of the temporal dimensions of CSU student enrollment and graduation for diverse subgroups of students. A broader goal of this work is to provide a framework that CSU can use in the future while exploring various outcomes or research questions regarding the efficacy of institutional environment for at-risk students.

This chapter established that low graduation rates for low-income, minority, and first generation students are a serious problem at the institutional and national level, and that institutions need to address this issue individually. The next chapter reviews theories of student retention and common analytical approaches to answering research questions regarding enrollment. Chapter 3 will discuss the methodological considerations and design for the proposed study. Chapters 4 and 5 will present the findings and offer interpretations of the analysis. It is always important to ground empirical work in theoretical foundations and prior research; therefore, those will be discussed next.

CHAPTER 2: THEORETICAL FOUNDATIONS AND LITERATURE REVIEW

The purpose of this study is to describe the longitudinal nature of student enrollment and graduation at CSU. This chapter presents the theory and empirical research that is a basis for the research approach taken in this study. Five theoretical foundations for studying student retention are presented first. Then, common analytical approaches used in empirical work that explores student enrollment and graduation are discussed and critiqued. The chapter will conclude with presenting a theoretically grounded analytical approach for institutions to use in understanding the undergraduate enrollment path to graduation.

Conceptual Basis for Studying Postsecondary Student Success

This section discusses five different theoretical approaches to studying student success: human capital theory, student integration/attrition models, student engagement, financial nexus model, heterogeneous research approach. Each theory's premise is described, followed by a discussion of the factors each theory attributes as influencing student enrollment. The purpose of this section is to review the theoretical foundations for studying student postsecondary enrollment and to conclude with a list of factors that theory indicates as influential when conducting research on student enrollment.

Human Capital Theory

Human capital theory is a commonly used conceptual basis for studies that statistically model student enrollment. Human capital theory is used to describe the phenomena of people investing in their own education. Its basis is that decisions about college are a function of the direct costs, opportunity costs, and economic returns from schooling. Direct costs include the price paid for room, board, tuition, and fees. An opportunity cost is the cost associated with forgone earnings and time that could have been spent on something other than schooling. This theory unambiguously predicts that providing financial aid (decreasing the direct costs) will raise the likelihood of making a pro-schooling decision. However, the enrollment decisions will vary according to the student's characteristics, which are normally controlled for in the modeling as independent variables (DesJardins & Bell, 2006; DesJardins & Toutkoushian, 2005; Dynarski & Scott-Clayton, 2008).

Human capital theory uses empirical models to address research questions, such as how much the likelihood of enrollment changes for specified changes in costs and financial aid (Ehrenberg, 2004). This is a complex question because the net benefit needed to make a positive enrollment decision is based on the individual's experiences and expectations about the rewards of education. The models always try to infer and capture that variation by including covariates, but it is nearly impossible to include all of the variables that influence a person's behavior (DesJardins & Bell, 2006; DesJardins & Toutkoushian, 2005). Models utilizing human capital theory typically include predictors that control for student background and academic preparation. Typical control variables are students' gender, ethnicity, age, and family income, while financial aid or institutional costs tend to be the predictors that the research questions focus on (Dynarski & Scott-Clayton, 2008; Ehrenberg, 2004).

Human capital theory only indirectly acknowledges the individual and institutional variables that affect enrollment by recognizing that they can influence cost/benefit analysis. In response to this, student integration models developed to serve as a complementary conceptual basis for understanding student enrollment (Dai, 2008).

Student Integration Models

Two prominent theories that explain the college persistence process are Tinto's student integration model (1975; 1993; 2006) and Bean's student attrition model (1982). This section will review each theory and then discuss why some theorists group these two theories together.

The premise of Tinto's Student Integration Model is that the enrollment decision is a longitudinal process and that the factors that influence the departure decision come from the interaction of student and institutional characteristics. A major emphasis of Tinto's theory is that the decision to leave is dependent upon the student's perceptions of the interaction between the student characteristics and institutional characteristics. This interaction goes across social and academic categories. The decision to leave varies over time because daily interactions on campus influence whether or not a student stays. The enrollment decision is just as dependent upon what happens after the student enrolls as what preceded enrollment. Since the current environment influences attrition, student departure can serve as a barometer for the academic and social health of a university. An interpretation of this theory is that the university is partially responsible for student departure; therefore, universities have the ability to improve their graduation rates by self-improvement. Tinto's work offers practical applications of the theory by proposing retention programs, such as living-learning communities (Tinto, 1975; 1993; 2006). However, a weakness of this theory (with its focus on student institutional interactions) is the limited attention it gives to finances regarding the students' enrollment decisions (Cabrera, Nora, & Castaneda, 1993).

Bean's Student Attrition Model is very similar to Tinto's model except that the student departure process is perceived as similar to turnover in the workplace. This conceptualization stresses the importance of students' behavioral intentions (of staying or leaving) as the strongest predictor of persistence behavior. Students' intentions are shaped by how their personal characteristics interact with the institutional environment (Bean, 1982).

In practical application, both models both stress the interactions between the student and the institution as the primary determinant in the enrollment decision and their complementary use enhances the understanding of student departure and graduation (Cabrera et al., 1993; Hossler, 1984). There are several obvious similarities: both models regard persistence as a function of a complex set of personal and institutional variables, institutional fit or match is essential for persistence, and the enrollment process is longitudinal. Differences in the application of the theories in research tend to be limited to different names for constructs that include nearly identical variables (Cabrera, Castaneda, Nora, & Hengstler, 1992). Tinto's theory addresses academic integration and social integration, while Bean's theory focuses on the interaction between personal characteristics and the institutional environment and external factors. Tinto emphasizes institutional commitment while Bean uses language about students' intent to persist (Hossler, 1984). Both theories put relatively little emphasis on finances (Cabrera et al., 1993). Since both models regard persistence as a complex set of interactions over time, they will be grouped together in this study as student integration models.

Empirical research situated in student integration models focuses on questions regarding the students fit to a particular institution (Cabrera et al., 1992). Institutional fit has both social and academic elements. Typically variables related to student background characteristics (gender, ethnicity, age, parents' education, high school preparation) and institutional characteristics (public, private, four-year, two-year) are used in student integration studies (Cragg, 2009). Commitment and integration variables are usually acquired with survey research (Cabrera et al., 1993; Bean, 1980). Typical research questions ask about the institution/student fit (Cragg, 2009) or how academic and social factors influence the likelihood of retention (Wetzel & O'Toole, 1999). Studies grounded in student integration models tend to conclude that student retention is a complex issue but that higher levels of integration and commitment are associated with higher levels of retention or graduation (Cabrera et al., 1993; Wetzel & O'Toole, 1999; Bean, 1980).

Recently studies rooted in student integration have started to evaluate the longitudinal nature of student departure (Ishitani, 2006; 2008). Longitudinal analysis is particularly appropriate for the studies utilizing Tinto's theory because of its emphasis on the time-varying nature of student's social and academic integration with their university (Tinto, 2006).

Comprehensive Theories

The complexity of student retention has led to the development of theories that are more comprehensive in the constructs that they emphasize in order to understand student enrollment (Kuh, Kinzie, Buckley, Bridges, & Hayek, 2006). Three theories that take a holistic perspective are presented next.

Student engagement.

The premise of student engagement theory is that the activities students spend time on and the relationships that they develop are major components to their success (Kuh, 2009; Pascarella & Terenzini, 2005; Astin, 1993). Importantly, institutional practices can positively and significantly impact student development, which encourages students to engage in educationally purposeful activities (Chickering, 1993). Success is defined broadly and can be measured by a variety of variables beyond the standard persistence and graduation measures (Kuh et al., 2006). Empirically, studies have shown that greater levels of engagement result in higher retention and graduation rates (Carini, Kuh, & Klein, 2007; Kuh, Cruce, Shoup, Kinzie, & Gonyea, 2007). In addition to this finding, there is also research showing that engagement has a conditional effect for low-income, first generation, and minority students, which means that engagement has an even stronger positive effect for at-risk students (Kuh et al., 2007). Some authors even suggest that improving student engagement for students with high risk characteristics could be a possible intervention to improving institutional graduation rates (Kuh et al., 2007). Student engagement theory is a comprehensive approach to understanding how schools can improve their policy and curriculum to most help their students engage in behaviors that will help them succeed.

Generally, the theory of student engagement is associated with the National Survey of Student Engagement (NSSE). This is because the NSSE instrument was developed to measure constructs of student engagement. Unlike outcomes such as graduation or GPA, student engagement variables are not available from institutional databases and must be gathered by survey methodology. Research has concluded that the NSSE instrument measures the constructs it intends to measure (Kuh et al., 2007); however, there are scholars who critique the NSSE instrument's validity and reliability (Porter, Rumann, & Pontius, 2011). Further critique comes from empirical work that fails to show a positive correlation between NSSE constructs of student success and measures of graduation and GPA (Gordon, Ludlum, & Howey, 2008).

Acknowledging its limitations, NSSE data is still an important higher education data source. NSSE data has been used for analyzing differences in the educational experience of subgroups of students. Examples of this type of work include the comparison between first generation students and non-first generation students (Pike & Kuh, 2005), analysis of students who participate in learning communities (Zhao & Kuh, 2004), and between international and native students (Chun-Mei, Kuh, & Carini, 2004).

Financial nexus model.

The premise of the financial nexus model is that student enrollment is a function of a student's background characteristics, academic preparation, and academic expectations, all of which interact with college experiences and costs. A student's evaluation of his or her college enrollment is a decision that is made each semester and is based on financial and academic interactions with the institution. The financial nexus emphasizes that important interactions exist between student characteristics and college costs (St. John, 1992; Paulsen & St. John, 1997).

Studies grounded in the financial nexus model typically use an analytical approach referred to as the workable persistence model. The workable persistence model specifies that a sequential logistic regression be used to evaluate the effect of financial aid on student persistence after controlling for student demographic and college experience variables. The model's first construct (student demographics) should include variables such as gender, ethnicity, family income, and high school GPA or high school rank. The second construct (college experiences) typically contains variables such as college GPA, whether the student lives on or off campus, and what type of postsecondary institution the student attends. The final construct includes the variables related to the specific research questions of interest, usually about financial aid and cost. The outcome variable of interest is always persistence, although it varies from being within-year or between-year persistence (St. John, 1992). This technique was originally developed to offer institutions a straightforward approach to evaluate their financial aid policy (St. John, 1992), but its use has been expanded to national (St. John, Oescher, & Starkey, 1994; St. John, Kirshstein, & Noell, 1991) and state level data (St. John, Hu, & Weber, 2001; Hu & St.John, 2001).

Research utilizing the financial nexus model focuses on evaluating the impact of financial aid on students' persistence. Research under this approach has found that student aid does improve the likelihood of persistence (St. John et al., 2001; Hu, & St. John, 2011); however, the quantity of aid received is sometimes too small to show an effect (St. John, 2000). Another important finding from this line of inquiry is that multiple studies have identified an interaction between students' income and receiving financial aid; the negative association of being low-income and persisting is mediated by the receipt of financial aid (St. John, Musoba, & Simmons, 2003; St. John, Hu, & Weber, 2000).

Heterogeneous research approach.

The heterogeneous research approach combines a comprehensive theoretical framework with an analytical process to explore temporal patterns of student dropout. This approach proposes that important variables from the student integration models and human capital theory should be combined into eight constructs that are used as independent variables in a discrete time event history model. The dependent variable is the probability of dropping out. This approach specifies a three-step process to arrive at the final event history model. The first step is to run an initial model without any interactions to develop a baseline feel for the data. The second step is to run the initial model on subgroups of the data to explore if any interactions exist. The final step is to use the subgroup analysis to inform which interactions should be added to the event history model from the first step. A tenant of the heterogeneous research approach is that student groups are heterogeneous (very different in their cultural and personal characteristics); therefore, analysis needs to be done on subgroups so that possible interactions can be explored. Minority status and income level are two variables that should always be explored for interacting with financial aid and college cost (Chen, 2008).

Studies utilizing this approach need to include variables from the eight constructs. The first construct, student background characteristics, includes gender, age, race/ethnicity, family income, and parental education. The second and third constructs include variables that represent students' educational expectations and their precollege academic preparation. The fourth construct contains variables representing college experiences such as academic and social integration variables. The fifth construct is only applicable to studies utilizing multi-institutional databases and represents college characteristics, such as public or private. The sixth construct is the exploratory variable of interest and includes variables related to finances such as financial aid and cost. The seventh construct represents time which is measured discretely in academic years. The last construct is the interaction terms that are included based on the subgroup analysis (Chen, 2008). Research done using the heterogeneous research approach focuses on understanding how financial aid mediates the relationship between parental income and the likelihood of student dropout. Results showed that low-income students have higher probabilities of dropping out but that financial aid decreased the gap in probabilities across income status (Chen & DesJardins, 2008).

Important Factors for Student Enrollment

All of the theories presented here suggest that student characteristics need to be controlled for to attempt to make any causal assertions regarding the relationship between a variable of interest and a postsecondary education outcome (DesJardins & Bell, 2006; Tinto 1993; Kuh, 2009; St. John, 2000; Chen, 2008). Gender, ethnicity, age, parental education, and income are common demographic variables that models include as covariates. Academic preparation is also thought to be an important control, so high school GPA or test scores are often included (Ishitani, 2006; DesJardins, 2010; Dai, 2008; Carini et al., 2006; Hu & St. John, 2001).

Factors related to institutional environment are often the predictor variables research questions focus on after controlling for background characteristics. Student engagement theory is focused on how institutional policies and curriculum influence students to behave in ways that will increase their learning (Kun, 2009). Similarly, student integration models propose that students' interaction with their school's characteristics determines whether or not they will stay and graduate (Wetzel & O'Toole, 1999). Theories that are most focused on finances also ask questions about the institutional environment by using models to address questions regarding institutional financial aid policy (St. John, Hu, & Tuttle 2000). Institutional environment variables are typically the variables about which researchers attempt to explore causal relationships. As will be further elaborated on in the following section, research goals of causality need to be properly aligned with the appropriate methods (Murnane & Willet, 2011).

Possible interactions among covariate and predictors and the time-varying nature of enrollment are other recurring theoretical themes in the literature. The heterogeneous research approach explicitly states a modeling building step in which interactions between financial aid, time, income, and ethnicity are explored (Chen, 2008). Research utilizing the financial nexus model has demonstrated that financial aid mediates the positive association between higher income and persistence (St. John, Musoba, & Simmons, 2003; St. John, Hu, & Weber, 2000). As will be discussed below, regression studies in higher education often do not include important interactions in research models (Dowd, 2008). In order to address this critique, student retention research should explore the possibility of interactions (Tinto, 2006). Accounting for the time-varying nature of student enrollment is also another common theoretical theme and is a central tenant of the heterogeneous research approach (Chen, 2008) and student integration models (Tinto, 1993).

Empirical Studies of Student Graduation and Retention

With these prominent conceptual frameworks in mind, the next section synthesizes published research studies within these theoretical bases. The purpose of this section is to review the analytical techniques used in empirical studies that explore relationships between predictors and student enrollment/graduation. Although the theoretical frameworks identify factors that try to understand causal relationships with student enrollment, empirically, it is very difficult to show a cause and effect relationship (Dowd, 2008; Murnane & Willet, 2011). This section reviews and critiques the analytical techniques used in the studies reviewed, setting the stage for the analytical technique and goal of this study.

Bivariate Analysis

Bivariate analysis is used to make simple comparisons across a variable of interest. For instance, the graduation rates discussed in the first chapter show the bivariate graduation rates across Pell Grant Recipient status at CSU. Inferring from this gap that Pell is not effective at improving student graduation is not appropriate because that ignores all of the other factors that influence students' enrollment, which probably account for some of the graduation gap.

Initially, in the study of student retention, some researchers used bivariate comparisons before and after a policy change to infer about the efficacy of a program. Hansen (1983) compared enrollment rates before and after the start of the Pell Grant program across levels of achievement and socioeconomic status. He concluded that the Pell Grant program did not promote access and was in contradiction to economic growth because the taxpayer money could be spent on more productive programs (Hansen, 1983). In a similar methodological manner, Kane (1994) used bivariate comparisons of cost and parental education to explain the variation in college enrollment among black youth. Kane found that parental education increased enrollment and increased college costs decreased enrollment. Bivariate comparisons make useful descriptions; however, they fail to control for the multitude of other important variables when trying to understand what causes some students to graduate and some to drop out (Leslie and Brinkman, 1988).

Multiple Regression

Multiple regression techniques allow researchers the ability to explore the association of a predictor variable with an outcome while statistically controlling for other predictors in the model. Linear regression is used when the outcome is a continuous variable, such as years of schooling (Dynarski, 1999) or time to degree (Dai, 2008). Logistic regression is similar to linear regression in its ability to explore the association between a predictor and outcome while controlling for other factors, but it allows the outcome variable to be categorical. Logistic regression is particularly popular in higher education research because it can handle binary outcomes (graduate/did not graduate or persist/did not persist) and its coefficients are easily interpreted as odds ratios (Peng, Lee, & Ingersoll, 2002; Peng, So, Stage, & St. John, 2002).

Studies across various theoretical foundations utilize multiple regressions. The financial nexus model specifies a sequential logistic regression as its analytical technique (Paulsen & St. John, 1997). Regression techniques are common in studies using human capital theory. These studies tend to focus findings on the marginal effect of a price change on student enrollment or graduation. For instance, one study found that an additional $1,000 dollars in scholarship offers increases the likelihood of enrollment by almost eight percent. They also found that the impact of scholarship offered varied by income groups (Singell, 2002). Studies grounded in the student integration model also utilize logistic regression. Findings tend to focus on the factors that predict student's ability to integrate at their university. For instance, one study found that a student's academic match to the university had a stronger impact on retention than a student's economic match after controlling for student and institutional characteristics (Cragg, 2009). Despite the popularity of regression analysis, there are some critiques of the way that this tool has been used in higher education research.

Research design limitations often do not warrant causal interpretation of the results because data in higher education is often not collected as the result of intentional experiments. To assert causation, a researcher must be able to rule out any other possible explanations for the relationships; unfortunately, a regression equation that controls for coefficients does not meet this requirement. Data in higher education is typically observational. The problem with asserting causation with observational data is that people self-select into their groups. This self-selection is called endogenous assignment and makes it difficult to partition out the effect of a predictor variable from the unobserved variables that influenced the person to join one group over the other (Murnane & Willet, 2011; Dowd, 2008). An example of this issue is when researchers try to assert the effect of paid financial aid. The student had to decide to apply for aid to be a recipient, so it is difficult to truly control for all of the characteristics that played into a student's decision to file for financial aid and use the offered loans and/or grants.

Another critique of the use of multiple regressions in higher education is that interaction effects are often not properly explored. An interaction is when the level of the predictor's coefficient is dependent on the level of another variable. If an interaction exists and the researcher does not include it, then there can be bias in the main effects. Despite multiple theories stating the importance of interactions, empirical studies often do not include them (Dowd, 2008).

These limitations undermine the findings of decades of research on student retention; however, researchers are aware of these problems and are implementing techniques that are more appropriate for higher education datasets. The next section discusses a set of analytical approaches that attempt to account for the endogenous assignment and self-selection bias inherent in observational data.

Quasi-Experimental Techniques

To account for the lack of random assignment, there are some specific design approaches that can be used to help provide a rationale for interpreting explanatory variables in a causal manner when using observational data. This section discusses three quasi-experimental techniques and gives empirical examples of their use in higher education research.

Natural experiments.

A natural experiment is when a sudden, unexpected change assigns people into either the treatment or control group. The unexpected change acts as a forcing variable, which provides exogenous association. A fundamental assumption of this approach is that the only difference between the groups is that they are on either side of the forcing variable. If this assumption is valid, then any difference between the groups can be attributed to the treatment (Murnane & Willet, 2011). For instance, the start of Pell Grant program provided data for a natural experiment regarding its efficacy. In this case, time is the forcing variable, before and after the start of the Pell Grant program. Seftor and Turner (2002) asked if the Pell subsidy resulted in greater numbers of nontraditional students returning to school. The researchers compared the college going rate of nontraditional students shortly before and after the start of the Pell grant program to estimate the effect of the Pell Grant program. The authors concluded that the availability of federal aid had a significant effect on the enrollment behavior of older nontraditional students.

Similar to the Pell Grant natural experiment, other studies have used the start/demise of a financial aid program to test the causal relationship between funding and enrollment. For instance, one study utilized the impact of the start of the GI bill by exploring cohort variation in college completion from before and after WWII to determine if the subsidies affected the proportion of veterans with a college degree. The policy change of offering GI benefits to veterans is the forcing variable. They found that the effect of serving in WWII and gaining GI benefits resulted in a 10 percent higher chance of college completion. These authors cautioned that their research could not separate out the effect of WWII and the GI bill from each other. It is possible that the experience of serving in WWII influenced enrollment rather than (or in addition to) the GI benefits (Bound & Turner, 2002). Similarly, Dynarski (1999) compared students who would have been eligible for a specific social security benefit (SSB) with students who were eligible except that the program was canceled. In this example, time is the forcing variable. The number of years of completed schooling for the last three cohorts that received the SSB award was compared to the number of years of completed schooling for the first cohort that would have qualified but the program was canceled. The author concluded that receiving the SSB award resulted in a greater number of completed years of schooling.

Natural experiments can also utilize eligibility criteria as the forcing variable that creates exogenous assignment. For example, Bettinger (2004) used differences in the Pell Grant award amount to study whether the amount of funding influences retention. Regression discontinuity was used to explore if variation in the Pell Grant award amount based on family size related to the likelihood of attrition. In this natural experiment, family size is the forcing variable. The author concluded that a $1,000 increase in funding resulted in an almost nine percentage point decrease in the likelihood of withdrawal. Kane (2003) used the Cal Grant (a state-need based aid program in California) income and GPA qualification criteria to test if the aid had a positive impact on college going behavior. Comparing grant recipients with students who were just outside the eligibility requirements, the author concluded that grant eligibility had a large impact on the likelihood that a graduating senior would enroll in college the following fall.

Instrumental variables and propensity scores.

Instrumental variables and propensity scores are statistical techniques that can be used to account for the endogenous assignment in observational data. Instrumental variables are intended to carve out the exogenous variation in a predictor so that this exogenous variation is all that is used when estimating the causal impact of the predictor on an outcome (Murnane & Willet, 2011). Pike, Hansen and Lin (2011) utilized instrumental variables to explore the relationship between learning community participation and retention. Analyses of observational data often show that participating in learning communities has a positive effect on student success. However, this type of analysis ignores the fact that students self-select to participate in these communities and that the students who choose to participate may be different from the students who do not choose to participate. This study showed that once self-selection into the learning community is controlled for by an instrumental variable, the positive relationship between learning community and student persistence is eliminated.

Propensity scores allow the researcher to match the treatment and control groups by obtaining a likelihood score for their probability of experiencing the treatment. Once groups are matched, the assumption is that any differences between the groups can be attributed to the treatment (Murnane & Willet, 2011). Using propensity score matching can dramatically influence regression results; for instance, the direction of the effect of a freshman seminar class on first year GPA actually flipped once propensity scores were used to match the groups. Prior to propensity score matching, students who took a first year experience course had lower GPAs compared to all first year students who did not take the course. After controlling for selection bias with propensity score matching, students who took the course had higher GPAs (Clark & Cundiff, 2011).

Higher education researchers want to draw conclusions about causal relationships and quasi-experimental techniques lend some validity to causality with observational data (Murnane & Willet, 2011). There are, however, some limitations to the quasi-experimental approaches. Internal validity is not a guarantee with the adoption of a quasi-experimental technique. Data needs to be assessed to make sure that attempts to account for endogenous assignment really worked (Murnane & Willet, 2011). For instance, if a student held off enrolling because he/she knew the Pell Grant program was being implemented next year, time would not be an appropriate forcing variable. Second, the quasi-experimental methods discussed are cross-sectional in nature, but student enrollment is a longitudinal process. Students must evaluate whether or not to enroll each semester (Dowd, 2008; Tinto 1993; DesJardin, 2003). In higher education research, it is common to have predictors whose values change over time: financial aid and cost of attendance changes every year, GPA changes every semester, as does the student's credit load. Empirical research utilizing longitudinal methods allows researchers to address questions about the temporal dimensions of student enrollment.

Longitudinal Analysis

Longitudinal analysis utilizing an event history model is emerging as a popular way to study student retention and graduation issues. The benefit of using a longitudinal approach to studying enrollment in higher education is that the results show the probability of enrollment over time. The event of graduation is not something that occurs in a single point of time but a process that takes at least four, and sometimes eight or more, years (Tinto, 1975; 1993; 2006). Event history analysis allows researchers to ask questions about the timing of predictors' effects and include predictors that are not static but with values that change over time. Event history analysis is the preferred method to analyze how multifaceted factors influence student enrollment because it acknowledges the longitudinal nature of student attrition (Cheng, 2008).

Event history analysis is a technique that has been applied to a variety of research questions and data sets. It is commonly used in higher education research to show the probability of dropout (Ishitani, 2003; 2006; 2008; Chen & DesJardins, 2008; Ishitani & DesJardins 2002; Johnson 2006; Lott, Gardner, & Powers, 2009), graduation (DesJardins, Ahlburg, McCall, & Moye, 2002b; DesJardins, Ahlburg & McCall, 1999), or competing risks of graduation and stopout (DesJardins, Ahlburg, & McCall 2006; 2002; Ronco, 1996) for students over time. There have been studies with less common outcomes, such as the number of completed credits (Radcliffe, Huesman, & Kellog, 2006) and the rate of progress towards a degree (Bahr, 2009). An alternative application of the technique is a study where event history analysis was used to analyze which states would adopt a merit aid program, such as the Georgia Hope Scholarship (Doyle, 2006). Event history analysis has mostly been utilized with institutional (Scott & Kennedy, 2005;

DesJardins & McCall, 2010; DesJardins et al., 2006; 2002; 1999; Ishitani, 2003; 2008; Radcliffe et al., 2006; Johnson, 2006; Murtaugh, Burns, & Schuster, 1999; Ronco, 1996), but it also has been used with state (Bahr, 2009) and national (Chen & DesJardins, 2008; Ishitani & DesJardins 2002; Ishitani, 2006; DesJardins et al., 2002b) higher education data sets.

Event history analysis has been utilized in studies with a variety of research questions and is often used to test the causal nature of relationships. For instance, due to the ability to allow predictor values to change over time, event history analysis has been used to evaluate various financial aid packaging policies. The goal of this research is to simulate if different combinations of financial aid have differing effects at various points of students' undergraduate education (DesJardins & McCall, 2010).

Event history analysis has also been used as a descriptive tool. In an exploratory study of doctoral student attrition, the goal of the research was in quantifying the multivariate relationships between variables rather than finding causal relationships (Lott et al., 2009). In a separate study of graduate student completion rates, a descriptive approach is utilized by showing the cumulative completion rates across students' field of study, gender, ethnicity, and previous academic achievement. The goal of this research is to explore the patterns of completion across interesting variables (Most, 2008). Using analytical techniques that can account for the longitudinal nature of student enrollment is an improvement over cross-section regression analysis; however, like all analytical techniques, event history analysis has limitations.

Event history analysis requires a detailed amount of data; this is probably why it is most commonly used with institutional analysis where institutional databases can be mined to obtain the appropriate variables. National datasets often contain the majority of variables necessary but tend to lack the detail necessary for continuous time-varying predictors, such as financial aid (Ishitani, 2006; Chen & DesJardins 2008; DesJardins et al., 2002b), which is often a variable in institutional studies (DesJardins & McCall, 2010; DesJardins et al., 1999).

Interpretation of results using observational data generated from an event history model is subject to the same limitations as multiple regression. Self-selection bias has been addressed by using offered rather than paid financial aid (DesJardins & McCall, 2010; DesJardins et al., 1999; 2002); however, students still have to apply to get offers of most types of aid. A limitation of these studies is the presence of self-selection into the financial aid receiving group. The descriptive approaches that utilize event history analysis in exploratory manners are appealing because the analytical goal is not at odds with the type of data available in higher education research. For instance, Most (2008) used observation data to describe graduate degree completion across disciplines of study. The analytical goal does not involve causal relationships and is therefore not at odds with utilizing observational data.

Comprehensive Approach to Studying Low-Income Student Success

Thus far, discussion has focused on theoretical foundations and analytical approaches of studying student retention. This chapter will conclude with a presentation of the theoretical and analytical considerations that will guide the analysis in this study.

The theoretical premise of this study's analysis is based on the understanding that the set of characteristics students arrive with do impact their ability to succeed at college. Importantly, institutions have the opportunity to help students succeed by using policy and curriculum that foster student success. The institutional environment is important because it is an intervention source for students whose background characteristics might put them at risk of departure. The institutional environment includes elements of student engagement and satisfaction (as measured by NSSE) as well as financial aid packages. The institutional environment interacts with students based on their characteristics, which produce student behaviors that determine enrollment intensity, time allocation, engagement with the university, and GPA. All of these interactions and effects have the possibility of varying over time.

The comprehensive approach to studying low-income student enrollment is presented below in Figure 1. The arrows on the left side represent students' background characteristics and academic preparation. These variables are gender, ethnicity, age, parental education, household income, index score, and their residency status. Students enter the institutional environment and their characteristics interact with CSU policies and culture. This interaction results in student behaviors that can be measured by their enrollment intensity, GPA, time allocation, and other NSSE measures of engagement. The outcome of this student/institutional interaction results in students who stop out, drop out, transfer, or graduate. This interaction happens over time as represented by the arrow that goes through the circular representation of institutional environment.

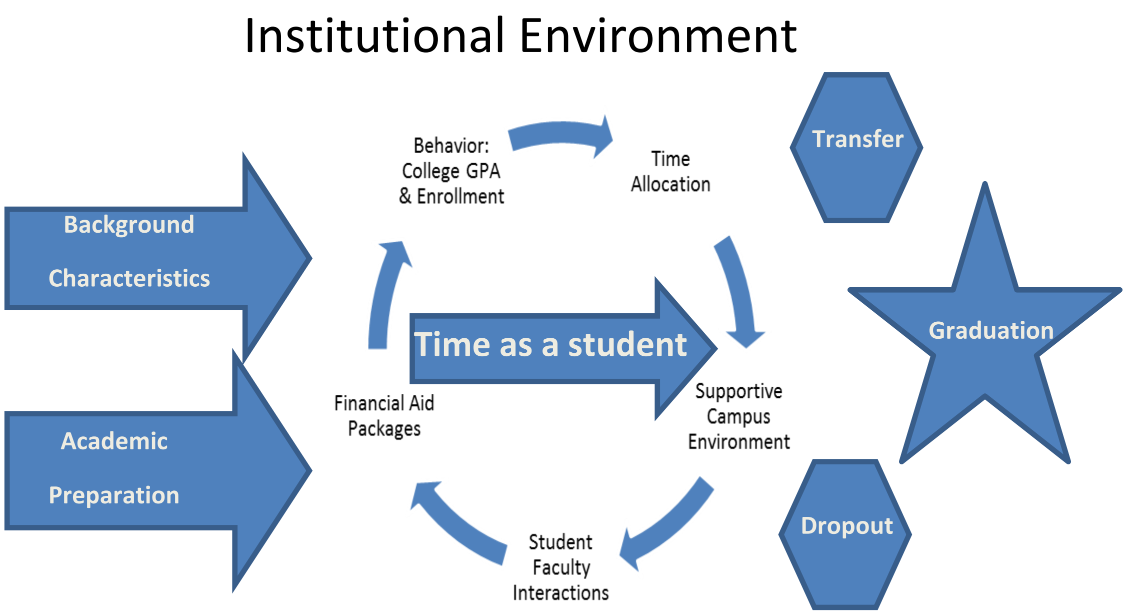

Figure 2.1

The purpose of this study is to understand enrollment patters ignoring environmental variables. As an institution interested in eliminating the graduation gaps an important first step is to explore the timing of these gaps. Future work can start to explore how the environment influences enrollment and graduation patterns for first generation, minority, and low-income students. This study is not intended to address how the institutional environment can influence students so institutional environmental variables are purposively not included as covariates in the model. Interesting future work should explore how students' major, GPA, and credit loads influence their enrollment trajectories. The next chapter of this dissertation will provide detailed information on the data and methodological procedures proposed.

CHAPTER 3: METHODOLOGY

Event history analysis is a useful tool for postsecondary student research because it allows researchers to address questions about when students will experience educational milestones not- just if they will experience them. This chapter focuses on the study's methodological considerations and starts by stating the research question and purpose. Basic conceptual foundations of discrete time event history analysis are reviewed next. The rest of the chapter will focus on describing the study's data, variables, exploratory analysis, and model fitting process.

Research Purpose and Guiding Question

The purpose of this study is to describe student dropout and graduation for various types of students using an event history model. Student types or subgroups of interest are based on demographic (gender, ethnicity, low-income status, and first generation status) and academic variables (index and residency). The following question guides this study: *what is the longitudinal nature of dropout and graduation for different types of students at CSU?* The relative levels and shapes of dropout and graduation hazard profiles will be explored by using a discrete time event history model.

This descriptive study will be based on both exploratory analysis and event history modeling. The exploratory analysis will be focused on understanding the observed hazards of each student group individually and in combination with each other. Then the discrete time event history model will be fit in order to describe the partial relationships that exist among the hazards and various student types. Results from the descriptive model will be displayed as odds ratios in a table and also graphically by fitting the descriptive model to certain student prototypes.

Conceptual Foundations of Event History Models

Event history analysis was first used in the biomedical field and it typically studied the timing of death. This tradition referred to the analytical technique as survival analysis. Event history analysis was first used to study education questions in the early 1990s (Singer & Willet, 2003). It is an emerging technique in the field of institutional research and is becoming more commonplace in studies of postsecondary student enrollment and graduation. Despite the commonality of the technique, the next section describes important concepts for longitudinal analysis with a discrete time event history model.

Events and States

An important methodological consideration when designing an event history analysis is to clearly define the event(s) of interest and the possible states that an individual can exist in (Singer & Willet, 2003; 1993; 1991). The target event occurs when an individual moves qualitatively from one state to another. Often in event history analysis there is a single event being studied that results in only two possible states. For example, in studies examining attrition, the event of interest is the first time a student drops out of higher education and the two states are enrollment and post dropout nonenrollment (Ishitani 2002; 2008; Johnson, 2006; Murtaugh et al., 1999; Chen & DesJardins, 2008). A competing risk is when one event removes the individual from the risk of a different event type (Allison, 1984). This study utilizes competing risks to jointly assess the longitudinal nature of graduation and dropout. Due to data limitations, the event of dropping out cannot distinguish if the student leaves CSU to enroll in another institution or if the student drops out of higher education entirely.

Measuring Time

Defining the start, duration, and measurement of time are important methodological considerations for longitudinal studies (Singer & Willet, 2003; 1993; 1991). Time must be clearly defined so that everyone in the sample is eligible for the event to occur, but has not yet experienced the event. The beginning of time is the moment when everyone in the population occupies the initial state. A common technique is to set the beginning of time at the occurrence of a precipitating event that places all individuals at risk of experiencing the event of interest. In studies of student departure, this precipitating event is often entry into college as cohort. In this study, the cohort approach is used and three cohorts are tracked from the start of their first fall semester.

The length of time that the analysis follows individuals should be based on conceptual understanding of the event of interest. At a minimum, the study should be long enough that at least half of the sample has experienced the event of interest. This study will follow the three cohorts of students for eight academic years. Studies of student graduation typically track students for at least six years (Ishitani, 2003; 2006; Ishitani & DesJardins, 2002; Chen & DesJardins, 2008; Johnson, 2006; Murtaugh et al., 1999), but some studies have tracked students for longer periods (Most 2008; Lott & Powers, 2009).

Studies of student retention are typically discrete time studies because academic years or semesters are meaningful units of time. The distinction between continuous time and discrete time is important because the analytical techniques differ substantially (Allison, 2010). When time is discrete, the analytical process is very similar to cross-sectional logistic regression and is therefore less analytically challenging than when time is measured at a continuous level (Singer, 1991). All reference in this study to event history analysis assumes that time is being measured discretely.

Censoring

Censoring is an essential concept to analyzing longitudinal data. It is when the event of interest does not occur in the study time period; in other words, censored individuals have unknown event times. Censoring can by minimized with good research design, but it is nearly impossible to eradicate. The number of censored cases in a study is related to the length of time that data are collected and also the rate that the event occurs. Censoring is minimized in studies which measure a frequently occurring event for longer periods of time. Censored cases in this study are students who continuously enrolled for eight years but did not graduate. They are censored because they have neither experienced graduation nor dropout in the eight years that they are followed.

There are different types of censoring: informative versus noninformative and right versus left. An assumption of event history analysis is that all censoring is noninformative and that censored cases are representative of people who experienced the event. Attrition can be an example of informative sampling. For instance, in a longitudinal study that is tracking first drinking episodes of people who have gone through alcohol abuse treatment, it is problematic if the people who stop coming to the follow-up meetings are the ones who relapsed into their prior drinking behavior (Singer & Willet, 1991). Censoring can also be right or left censored. Right censoring is when the event is unknown because event occurrence is not observed (due to attrition or because the event occurred after data collection ended). Left censoring is when the event time is unknown because the beginning of time is not observed. Left censoring should be eliminated through research design (Singer & Willet, 2003). In this study, censored cases are noninformative and right censored.

Event History Output

Life tables contain all of the raw data necessary to calculate and graph the observed hazard and graduation probabilities. These tables have a row for every discrete time period in the study. They display the counts of people who are either in the risk set, had the event occur, or are censored. Life tables also display the calculated hazard and survival proportions. In this study, a life table will be made for each subgroup of students. These tables are essential to the explorative analysis done in this study because they include the raw data that is needed to calculate the probabilities of dropout and graduation. Comparing the observed probabilities across various groups of students is the center of the exploratory analysis.

The hazard probabilities are the outcomes of interest in an event history analysis. The hazard probability is the conditional probability of the event occurring in a specific time period (Allison, 1984). Conditionality means that the probability of event occurrence is based on those who have not yet experienced the event. It is informative to graph the observed hazards of subgroups because comparing the shape and height of hazard functions indicates possible interactions that should be explored when fitting the model (Singer & Willet, 1991; 1993; 2003).

The cumulative survivor proportion is the proportion of participants who are still in the risk set but have not had the event occur to them (Allison, 1984). When using an event history analysis to study student success, the graduation function is more useful than the survivor function because it represents the cumulative proportion of students who have graduated. Comparing the graphed hazard probabilities, cumulative graduation proportions, and life tables among student subgroups is essential to the explorative analysis conducted in this study.

Assumptions of the Discrete Time Event History Model

The discrete time event history model has three assumptions that must be evaluated and, if necessary, corrected (Singer & Willet, 1993). The first assumption is that the discrete time event history model assumes that in the logit hazard scale, the predictors are linearly related to the hazard. Statistically, linearity can be tested for by adding a nonlinear predictor specification to the hazard model. The linearity assumption is met if the addition of a nonlinear term does not improve fit. When the linearity assumption is violated, it can be resolved by using a transformation on the predictors (Singer & Willet, 1993).

The second assumption is that there is no unobserved heterogeneity. This means that all variation in hazard profiles across participants is solely dependent on the observed variation in the predictors. When predictors that explain variation are excluded from the model, unobserved heterogeneity is present (Singer & Willet, 1993). Advanced models include an error term to account for unobserved heterogeneity. Specifying an error term is more advanced than the current study's analysis, but is a technique that has been used when studying student retention (DesJardins, 2003).

The proportionality assumption states that in the logit scale, the covariates' hazard profiles share a common shape and are parallel. Proportionality is assessed by examining graphs of the observed hazard probabilities across each predictor. Proportionality is violated if the risk of event appears to vary over time across levels of a predictor. A predictor time interaction term should be included to account for the non-proportionality (Singer & Willet, 2003).

Proposed Event History Analysis

This study uses an event history model to describe how enrollment and graduation probabilities vary across the covariates. The purpose of this section is to describe the dataset and variables used in this study.

The Data

The data selected for this study contains seven cohorts of new freshmen who start in a fall semester (fall 2001, fall 2002, fall 2003, fall 2004, fall 2005, fall 2006, and fall 2007). These cohorts have their enrollment data tracked for eight years (or as many as possible, i.e. four for 2007, five for 2006, six for 2005, and seven for 2004) after their initial term. The dataset contains 24,065 students. 16 students from these 7 cohorts died, these students are excluded from the dataset. There are 313 students who are missing an index score, which decreases the total sample size to 23,752 for the modeling process.

Compared to other published studies that use institutional level data to conduct event history analysis, seven cohorts and over 23,000 students is a large dataset. Several of the studies using institutional data only followed one cohort (Ronco, 1996; DesJardins & McCall, 2010, DesJardins et al., 2002). Studies using national data typically only have one cohort (Chen & DesJardins, 2008; Ishitani, 2006). A benefit of having a large dataset is that the numbers of students in interesting subgroups are large enough to be analyzed in a meaningful way.

The predictor variables.

The variables selected to control for student background and academic preparation are based on the variables that theory and previous research indicate influence student retention and graduation. Gender, ethnicity, first generation status, and Pell Grant recipient status are all variables that theory indicates research should control for to understand student retention (Chen 2008; Tinto, 2006; DesJardins & Bell, 2006). Residency is also included because institutional graduation and retention rates show a large discrepancy between resident and nonresident students (Institutional Research, 2010). The reasons for the large amount of variation across residency status could be due to substantially different tuition charges or could be related to students feeling too far from family and friends in their home state. Academic preparation is measured by the student's index score. Index is specific to the state of Colorado and is a composite score of both high school performance (GPA or high school rank) and test scores (ACT or SAT). All of the model's covariates are dichotomous except for index, which is a continuous variable.

An important element of event history analysis is that the model can incorporate both time variant and time invariant predictors. When a variable is time invariant, its value does not change over time.

Gender and ethnicity are common time invariant variables. A time-varying variable is a predictor whose value changes over time. Enrollment level (full-time/part-time) is a common time-varying predictor in longitudinal studies of student graduation. In this study, gender, ethnicity, first generation status, and index are all time invariant while Pell recipient and residency are permitted to change over time. Table 3.1 below displays and describes each predictor.

Table 3.1

Student Subgroup Variables and Covariates in the Descriptive Model

Construct	Variable	Time-varying	Variable description
Student background characteristics	Gender	No	Dichotomous
	Pell Recipient	Yes	Dichotomous
	Ethnicity	No	Dichotomous
	First generation	No	Dichotomous
	Residency	Yes	Dichotomous
Academic preparation	Index	No	Continuous
Time	Measured in academic years 1 through 8		Discrete

The outcome variables.

Since this analysis is accounting for the competing risk between dropping out and graduating both of the outcome variables must be analytically defined. For the purpose of this study dropping out is defined as a student not enrolling for three consecutive semesters (not including summer). Therefore, a person can stopout for one or two consecutive semesters and not be included as a dropout; however, if a student stops out for three semesters and returns the fourth semester that student is no longer in the dataset, since they have been counted as experiencing the event of dropping out. The event of graduation is defined as occurring the academic year a student earns their degree.

Time measured in academic years.

The unit of time used for analysis in this study is an academic year. Although, some sensitivity is lost using the bigger discrete measure of academic year rather than semester this decision was made based on the fact that policy and campus programming are put in place over academic years rather than semesters. The data was initially prepared in the semester format because the semester detail is needed to define the event of dropping out and then converted into a person period file with academic year as the time measure. For instance, if a student dropped out in the spring semester of their first academic year (enrolled their first fall and then do not enroll their first spring, second fall, or second spring) they are counted as dropping out in year one. If a student graduates in the fall, spring, or summer of their fifth academic year they are counted as graduating in year five.

Descriptive Explorative Analysis

Extensive exploratory analysis will be completed prior to attempting to fit the event history model. This is based on the recommendation that researchers should look at the observed hazard of subgroups prior to any model fitting attempts (Singer & Willet, 1991; 1993; 2003). Life tables and graphs of the dropout hazards, graduation hazards, and graduation proportions for each subgroup are the main outputs for this explorative process.

Graphing the observed conditional probabilities is important to help uncover relationships between subgroups of interest and their probabilities of graduation and dropout. This explorative process helps determine if the proportionality assumption is violated. If the shape of the observed hazard probabilities differ across a subgroup, it is possible that the covariate interacts with time. A covariate and time interaction term can be included to alleviate this violation. The interaction term allows the effect of the variable to vary across time. Time-varying effects can be included for time varying or invariant predictors. This exploratory analysis is also useful for exploring if any of the covariates interact with each other. Interactions between gender and minority status will be explored along with the possible interaction between first generation status and Pell Grant recipient status.

Graphs will be displayed in the raw probability scale. However, they will also be reviewed in the logit scale as it is important to look at the observed data in the logit scale, since that is the scale that the data will be modeled with (Singer & Willet, 1993; 2003). The benefit of looking at the raw probability is that this scale is intuitive. Comparisons can be made to explore if the gaps between subgroups change over time.

The purpose of this explorative analysis is to obtain an understanding of how hazards vary among subgroups of students. This knowledge is essential for the model fitting process described next.

Descriptive Modeling

A descriptive model will be used to summarize the multivariate associations among student characteristics and academic preparation. This section will first discuss some fundamental considerations of the model and then review the proposed four-step modeling fitting process.

The Generic Period Specific Hazard Model

To assess how background characteristics and academic preparation result in different hazard profiles for CSU students, the relationship has to be modeled as shown in equation 1 (Singer & Willet, 2003).

$$\text{Logit } P_{it} = [\alpha_1 D_{1it} + \alpha_2 D_{2it} + \cdots + \alpha_k D_{kit}] + [\beta_1 X_{1it} + \beta_2 X_{2it} + \cdots + \beta_k X_{kit}] \qquad \text{(Equation 1)}$$

P_{it} represents the conditional probability that the event occurs for person *i* and time *t*. When α is the constant coefficient, β is the coefficient for an explanatory variable, *D* represents a dummy variable for time, and x represents the value of the predictor, then P_{it} can be described as the linear function of the explanatory variables. The logit transformation of P_{it} needs to be taken so that the function's outcomes are no longer bounded by 0 and 1. This transformation changes the interpretation of the βs so that they are equal to the change in the log-odds for each one-unit increase in X (Singer & Willet, 2003). Equation 1 can include terms for the interactions between time and the predictors and terms for the interactions between the predictors. This equation is the generic equation used in this study's descriptive model.

Multivariate Logistic Regression

The descriptive model will estimate the maximum likelihood parameters using the *multinomial logit* command in Stata (StataCorp, 2010) by regressing the event indicator on the covariates and time. The *multinomial logit* command allows for two binary logistic regressions to be run simultaneously on the data. It is important to explore the covariates' relationship with both graduation and dropout to see how the associations vary with different outcomes. To properly estimate the odds of competing risks, both the probability of graduation and dropout need to be evaluated at the same time. A multinomial logistic regression will yield different results than if multiple binary logistic regressions are run because the multinomial model imposes appropriate constraints on the model parameters (Long, 2006). This study has three possible states, so the multinomial logistic regression will be run comparing two of the three possible alternatives. The base category in this study will be continued enrollment. The two alternative outcomes will be the probability of dropout compared with continued enrollment, as well as the probability of graduation compared with continued enrollment. The probability of dropout compared to the probability graduation can be obtained by subtracting alternative one from the second alternative.

An assumption of multinomial logistic regression is the independence of irrelevant alternatives (IIA). IIA assumes that adding or deleting alternative outcome categories does not affect the odds among the original outcomes (Long, 1997). Most likely there are unobserved variables in my model that are related to the error terms of both graduation and dropout. This could be biasing my parameter estimates.

Parameterizing Time

When conducting event history analysis, it is important to explore if time can be specified in a more parsimonious manner. When there are many different time periods or if hazard is very low at some time periods, it is a good idea to explore different specifications of time. Several event history analyses of student enrollment have used an alternative specification of time to improve model fit (DesJardins & McCall, 2010; DesJardins et al., 2002). When the sample size is large and the ratio of predictors to sample size is also large, there is flexibility to use the more general parameterization of time. In this analysis, the dataset is large, but alternative specifications of time will be evaluated when fitting the descriptive model. As recommended in Singer & Willet (2003), models will be compared that specify time in the general form to a set of polynomial specifications. Various specifications will be compared by the BIC statistic. The alternative specification will be based on the following general rule: “if a smooth specification works nearly as well as the completely general one, appreciably better than all simpler ones, and no worse than all more complex ones, consider adopting it” (Singer & Willet, 2003 p. 417).

Building the Descriptive Model

The descriptive event history model will be built following a four-step process.

Step 1.

The first model that will be estimated will only regress the event indicators on time (D_1 through D_8). This is the unconditional hazard model because it does not include any of the covariates. This model represents how hazard varies over time, ignoring student demographic background and academic preparation variables. The evaluations of alternative specifications of time will be done in this first step.

Step 2.

The next model that will be fit will build on the unconditional hazard model by including background and academic preparation covariates in addition to time. The second model can be referred to as the conditional hazard model (Singer & Willet, 2003). Theory and previous research indicate that all of the covariates should be controlled for in students' enrollment decisions (Chen, 2008; Tinto, 2006; DesJardins & Bell, 2006); therefore, the inclusion of each predictor will not be statistically evaluated. Regardless of statistical significance, each of the predictors will be retained. The linearity assumption of the predictor variables will be evaluated and transformations of the predictor variables will be made if necessary.

Exploring Interactions

The final two steps of the model building process evaluate which interactions to include in the final model. A result of the exploratory analysis is to inform which interaction evaluated and model fit statistics are compared across competing models to inform which of the interactions are included in the final model. The model fit statistics of focus are the BIC and Cragg-Uhler pseudo r-squared. The BIC is used because it accounts (penalizes) for the addition of parameters while the AIC does not. For instance, the AIC will always get smaller with the addition of a predictor where the BIC will not always decrease with additional predictors (Singer & Willet, 2003). The BIC statistics are all computed using the actual sample size (unique count of students included in the model) rather than the number of records in the person period data file (StataCorp, 2010). The Cragg-Uhler pseudo r-squared is utilized by comparing the relative change in the statistic among competing models rather than interpreting its absolute value (Lacy, 2006).

Step 3.

The third model will build on the conditional hazard model by testing for time varying effects of the covariates. Previous event history analysis done at CSU showed that predictors did not interact with time (Lacy & Long, 2007) and there is some disagreement in the literature about how often the effect of predictors is dependent on time. Singer and Willet (1991; 1993; 2003) state that finding interactions is more often the case, while Allison (1984) reported this as a rare situation. Nonetheless, ignoring a predictor's interaction with time can seriously bias the magnitude and direction of coefficients (DesJardins, 2003); therefore, they will be evaluated. The best possible combination of interactions will be determined based on substantive knowledge, pseudo r^2 evaluations and BIC evaluations. Previous studies have shown that gender and ethnicity both have a time varying effect on student graduation (DesJardins et al., 1999; DesJardins et al., 2002b). An interesting result of the descriptive model will be this exploration of which covariates' coefficients depend on time.

Step 4.

The fourth step finalizes this study's descriptive model. This step evaluates the addition of predictor by predictor interactions. The exact interactions that will be looked at will be guided by the exploratory analysis done in research approach 1. The researcher hypothesizes that there will be interactions between gender and ethnicity and Pell Grant eligible and first generation. Other possible predictor interactions may become apparent with the initial exploratory analysis. All of the interactions will be evaluated by the BIC criteria, pseudo r^2, and substantive knowledge (Singer & Willet, 2003). Therefore, an interaction term could be retained if the BIC criterion improves slightly with its retention and substantively it is important; similarly, an interaction term may be dropped if the BIC improves slightly but substantively it is less important.

Interpretation of the results

The covariates' coefficients in the descriptive model can be exponentiated so that they become odds ratios. The benefit of looking at the results in terms of the odds ratio is in its interpretability. In the absence of interactions, an odds ratio that is greater than one indicates that the group of students represented by the value of that predictor is more likely than the reference group to have the specified event happen. If the odds ratio is equal to one, then there is no difference in the likelihood of the group represented by the predictor to be any different than the reference group. Of course, if the odds ratio is less than one, then the group represented by the predictor is less likely than the reference group to experience the event (Long, 2006). The model results presented in chapter 4 will show the odds ratios rather than the actual coefficients.

Odds ratios can be discussed in terms of their percentage change of the event occurring (Ishitani, 2008). Equation 2 shows this relationship.

$$\%\Delta O = (\exp(\alpha_j))^{\Delta A} - 1)*100\% \qquad \text{(Equation 2)}$$

Equation 2 shows that the percentage change in odds of the event occurring, $\%\Delta O$, is equal to the exponentiated coefficient, $\exp(\alpha_j)$, at a specified change (ΔA subtracted from 1). In one example (Ishitani, 2008), a coefficient of -.286 was associated with students whose mothers had a four-year degree or higher compared to students whose mothers had not earned a four year degree with the event of interest being dropout. Using equation 2, $\%\Delta O = (\exp(-.286)-1)*100\% = (.75)-1*100\% = -25\%$, the coefficient of -.286 is equal to -25%. This can be interpreted as students whose mothers had a four-year degree had rates of dropping out that were 25 percent lower than students whose mothers had not earned a four-year degree (Ishitani, 2008). It is intuitively appealing to talk about the coefficients in terms of the percentage change in a group's odds; however, the descriptive model's results will also be displayed graphically.

Graphical Presentation of Descriptive Model

The descriptive model will be presented graphically by displaying the fitted hazard functions of two prototypical student characteristic sets (high-risk/low-risk)). It is often more informative to look graphically at fitted hazard profiles rather than just focus interpretation on coefficients. Fitting a graph to the mathematical model requires that the level of all the predictors be determined. Graphing fitted profiles is a commonly used presentation technique (Singer & Willet, 1991; 1993; 2003; Ishitani 2003; 2006; 2008; DesJardins et al., 1999; 2002).

The purpose of looking at a high-risk student versus a low-risk student is that those two profiles will represent the extremes present in the data (Ishitani, 2003; Ishitani & DesJardins, 2002). All of the CSU students in our three cohorts will fall between these extremes. Identifying the student characteristics associated with the highest and lowest risk profiles is important for understanding the differing risk profiles for students at CSU. Based on the theory and previous research presented in the previous chapter, it is expected that low-income and first generation students will be included in the high-risk prototype (Ishitani, 2003; St. John, 2000). Understanding the hazard profiles of low-income and first generation students is an important first step to studying how institutional environmental factors can positively impact these students' enrollment and graduation.

This chapter described the methodology that is used for this analysis. The next chapter displays and discusses the results.

CHAPTER 4: RESULTS

The purpose of this chapter is to display and discuss the results derived from the previous chapter's methodology. Results will be discussed in three sections: descriptive explorative analysis, building the descriptive model, and results of the descriptive model.

Descriptive Explorative Analysis

The first step of this analysis is to explore the observed hazard probabilities and graduation proportions. No modeling is done in this section. Life tables, graphs displaying the observed hazard probabilities of dropout and graduation, and graphs of the graduation proportions are shown and discussed. Due to the complex nature of student enrollment and graduation, it is essential to look at these observed probabilities and proportions across demographic and academic preparation subgroups. The benefit of exploring the enrollment and graduation trajectories for each subgroup is to obtain a sense of whether or not differences among the groups vary over time or interact with each other. This section will begin by looking at the life table and graphs for all students, then explore the same information across each of the demographic/academic preparation variables included as predictor variables in the descriptive model. Risk sets become very small in the later years across the underrepresented subgroups; therefore, caution should be used when interpreting observed relationships in the later years due to inherent volatility of hazards caused by small risk sets.

All Students

Table 4.1 is the life table for all new freshmen who started in the fall in the 2002-03 through 2007-08 academic years. Each row represents an academic year; students are followed through the fall and spring semesters for a total of eight years. The first column in the life tables indicates which academic year the headcounts, hazard probabilities, and graduation proportions are associated with. The second column shows the risk set, or the number of students who are enrolled at the start of the academic year. The number in the risk set is equal to the previous year's risk set minus the number of dropouts, graduates, and censored students in that previous year. For instance, the risk set in year number six

(1,560) is equal to the risk set in year five (7,765) minus the number of dropouts (265), graduates (4,471), and censored students (1,469) in year five. The number of students in the dropout column represents the number of students who dropped out that year (students who did not reenroll for three consecutive semesters, excluding summer). The number in the graduate column represents the number of students who graduated in the academic year (including summer graduates).

The censored column represents the count of students who have not experienced either event in the time frame of the analysis. There are no censored students in years one through four because all of the cohorts have had at least four academic years since their first fall semester. The large number of censored students in years five, six, and seven are all in cohorts that have not had eight years since their initial fall semester and respectively started in fall of 2007, 2008, and 2009. For instance, all of the 1,469 censored students in year five are students who started in the in the fall of 2007 and are still enrolled with unknown educational outcomes. The censored students in year eight are all from cohorts that started in 2004 or prior. To summarize, censored students are those who are still pursuing their undergraduate education (have not graduated or dropped out) and are in their fifth (2007 cohort), sixth (2006 cohort), seventh (2005 cohort), eighth (2004 cohort), ninth (2003 cohort), or tenth year of undergraduate education (2002 cohort).

The main descriptive statistics for this exploratory analysis are shown in the last three columns. The dropout hazard shows the conditional hazard of dropout that is equal to the number of dropouts in a specific year divided by the risk set for that year. The graduation hazard is equal to the number of graduates in a specific year divided by the risk set in that year. These probabilities will be the focus of the discussion since they are the outcomes from the competing risk event history model discussed in the remaining two sections of this chapter. The graduation proportion is the sum of all graduates through the current year divided by year one's risk set. At year five, censoring begins and the current year's graduation and dropout hazards are used to estimate how many of the censored students would graduate if they had been observed for that amount of time. Therefore, the cumulative graduation rate of .581 in year five is equal to 7,889 graduates in years one through five plus, 845 (the estimated amount of the 1,469 censored students who would graduate if we had been able to observe them for five years, i.e., 1,469 multiplied by .576) divided by 24,065. It is important to explore the cumulative rate of graduation because it is the result of the previous terms' dropout and graduation hazards.

Table 4.1
Life Table FA02-FA07 Cohorts of New Freshmen

Year	Risk Set	Dropout	Graduate	Censored	Hazard Probability		Graduation Proportion
					Dropout	Graduate	
1	24065	3815	2	0	0.159	0.000	0.000
2	20248	2434	7	0	0.120	0.000	0.000
3	17807	913	331	0	0.051	0.019	0.014
4	16563	462	8336	0	0.028	0.503	0.361
5	7765	265	4471	1469	0.034	0.576	0.581
6	1560	89	761	312	0.057	0.488	0.631
7	398	4	180	93	0.010	0.452	0.637
8	121	0	55	66	0.000	0.455	0.646

The conditional hazard of dropout from table 4.1 is displayed in figure 4.1. The hazard of dropout is highest at year one. Dropout out hazard decreases from year one to year four. Hazard increases slightly (2.9 percentage points) from year four to year six; however, a dwindling risk set in year six may cause spurious changes in hazard from year five to year six. For instance, 2.3 percentage points of the 2.9 percentage point change in year six's hazard rate is the result of only 89 students dropping out in the 6th year. Due to the analytical definition of dropout there are no dropouts in year eight since students still enrolled in their eighth year do not have enough time remaining to meet the definition of dropout (not enrolling for a consecutive year and a half). The small number of dropouts in year seven is also a result of the analytical definition of dropout. Only students who dropped out in the fall semester of their 7th year can be coded as a dropout since students who dropout in the spring of their seventh year are censored because the timeframe of the data cannot confirm if a student did not enroll for three consecutive fall/spring semesters. The graphical representation of dropout hazard is only displayed through year six since data limitations prevent obtaining a rate calculation for years seven and eight.

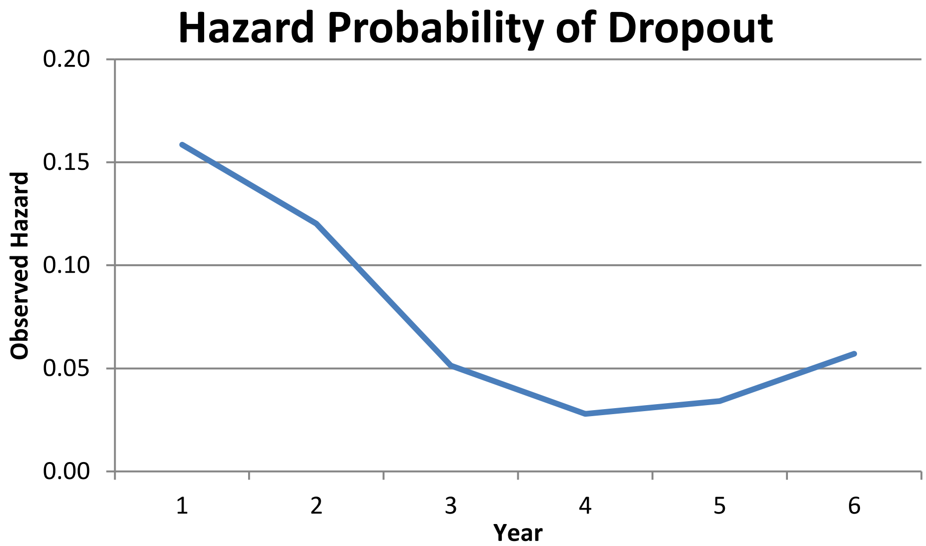

Figure 4.1

The conditional probability of graduation from table 4.1 is displayed in figure 4.2. The graduation hazard is near zero for the first three years and then increases steeply in year four and peaks at year five. The hazard probability drops nearly 9 percentage points in year six and then drops a bit more in year seven and flattens out to year eight. This indicates that after students have been enrolled for more than five years, they have lower probabilities of graduating than they did during their fifth year of enrollment.

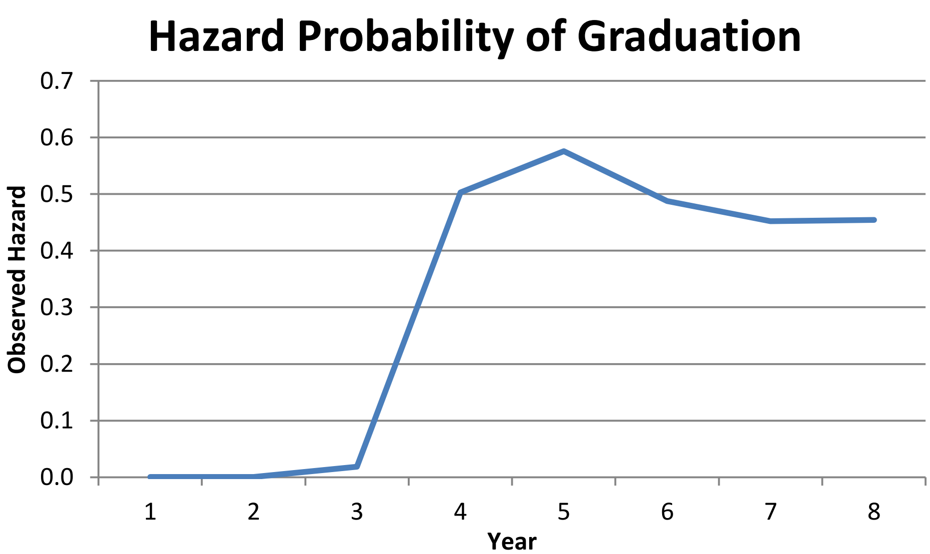

Figure 4.2

The graduation proportions from table 4.1 are displayed in figure 4.3. As expected from the hazard of graduation, the cumulative proportion remains near zero until the students' fourth year when the proportion of the risk set experiencing graduation increases to .581 at the fifth year. The cumulative graduation proportions will always increase through years six, seven, and eight; however, after year six the increase is at a very slow rate. There is a 4.9 percentage point change from year five to six and only a .07 percentage point change from year six to year eight.

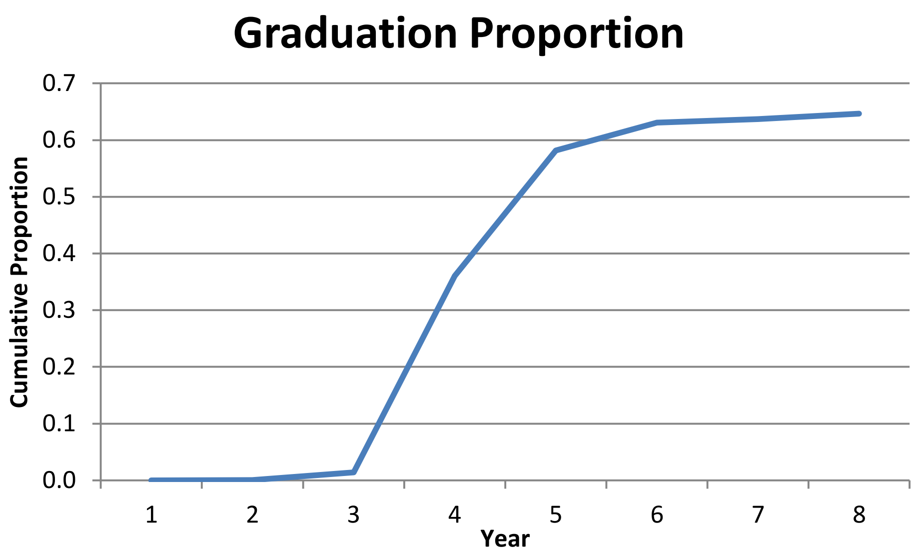

Figure 4.3

Pell Grant Status

The purpose of this section is to display and review the conditional hazards and cumulative graduation proportion of students by Pell Grant recipient status. In the descriptive model, Pell grant status is a time varying predictor, since a student's eligibility is determined each year based on their yearly application for financial aid. However, for the exploratory analysis, Pell Grant status is determined by whether or not the student qualified for a Pell grant during his/her first freshman year. Table 1.1 shows that the six year graduation rate of Pell Grant recipients lags behind non-Pell Grant recipients. This negative association between the graduation rate and Pell Grant eligibility is reflected longitudinally in the following tables and graphs. Table 4.2 displays the same information as table 4.1, except the information is provided at both possible levels of Pell Grant eligibility. There are 3,368 Pell Grant recipients in year one's risk set, which is 14% of the entire risk set.

Table 4.2

Life Table of FA02-FA07 Cohorts of New Freshmen by Pell Status

Group	Year	Risk Set	Dropout	Graduate	Censored	Hazard Probability		Graduation Proportion
						Dropout	Graduate	
Non-Pell	1	20697	3186	2	0	0.154	0.000	0.000
	2	17509	2004	6	0	0.115	0.000	0.000
	3	15499	769	289	0	0.050	0.019	0.014
	4	14441	381	7378	0	0.026	0.511	0.371
	5	6682	224	3881	1260	0.034	0.581	0.594
	6	1317	79	643	254	0.060	0.488	0.642
	7	341	4	148	80	0.012	0.434	0.647
	8	109	0	51	58	0.000	0.468	0.657
Pell	1	3368	629	0	0	0.187	0.000	0.000
	2	2739	430	1	0	0.157	0.000	0.000
	3	2308	144	42	0	0.062	0.018	0.013
	4	2122	81	958	0	0.038	0.452	0.297
	5	1083	41	590	209	0.038	0.545	0.506
	6	243	10	118	58	0.041	0.486	0.562
	7	57	0	32	13	0.000	0.561	0.574
	8	12	0	4	8	0.000	0.333	0.578

The conditional hazard of dropout for Pell Grant recipients starts out higher than it is for non-Pell recipients: .19 compared to .15. The non-Pell group sees a slightly larger decrease in hazard from year one to year two since the non-Pell group's hazard decreases 3.9 percentage points compared to the Pell recipient group, which only experiences a 3.0 percentage point decrease. However, the Pell group begins to catch up with the Non-Pell group in year three since they experience a 9.5 percentage point decrease and the non–Pell group only had a 2.3 percentage point decrease. In years five and six, the non-Pell group increases their dropout hazard and has a higher hazard rate than the Pell group, whose hazard remains flat. This flip of hazard rates is probably an artifact of a small risk set in the Pell group. For instance, out of the total 89 dropouts in the sixth year, 79 of them were non-Pell recipients compared to only 10 Pell recipients.

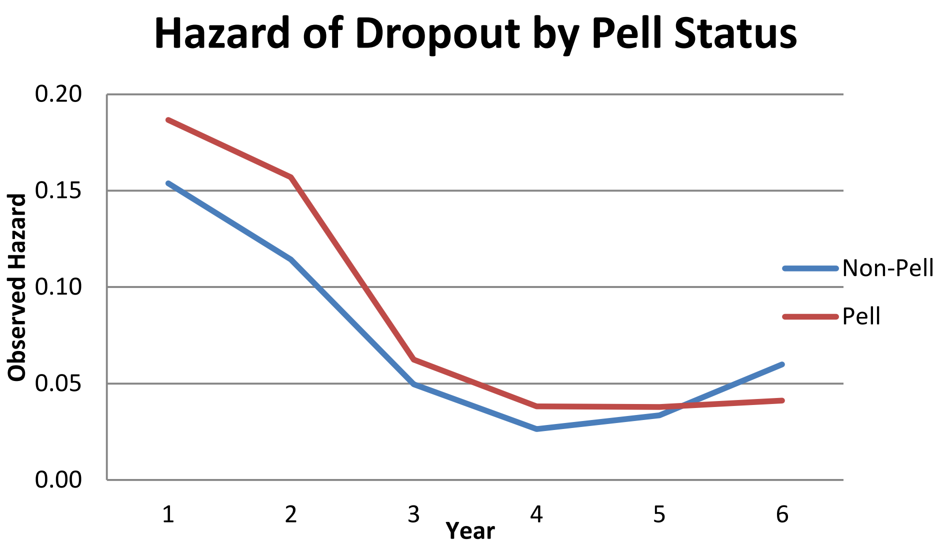

Figure 4.4

The conditional hazard of graduation for Pell Grant recipients and non-Pell grant recipients is displayed in Figure 4.5. Non-Pell Grant recipients see a 49.2 percentage point change in the graduation hazard from year three to year four. Pell Grant recipients only saw a 43.3 percentage point change in their fourth year's hazard rate. The 5.9 percentage point gap between the hazard rates of Pell compared to non-Pell in the fourth year is decreased to a 3.6 percentage point gap at the fifth year and a negligible gap in the sixth year. The Pell Grant recipients' graduation hazard peaks at year seven, which causes this group to have a graduation hazard that is 13.5 percentage points higher than the non-Pell group. The relationship flips back in the eighth year; however, the variation in graduation hazard over the seventh and eighth years is probably due to small risk sets since the Pell students only have 57 and 12 students in their seventh and eighth year risk sets. Interestingly, the Pell Grant recipients' graduation hazard peaks much later than the overall and non-Pell student graduation hazards.

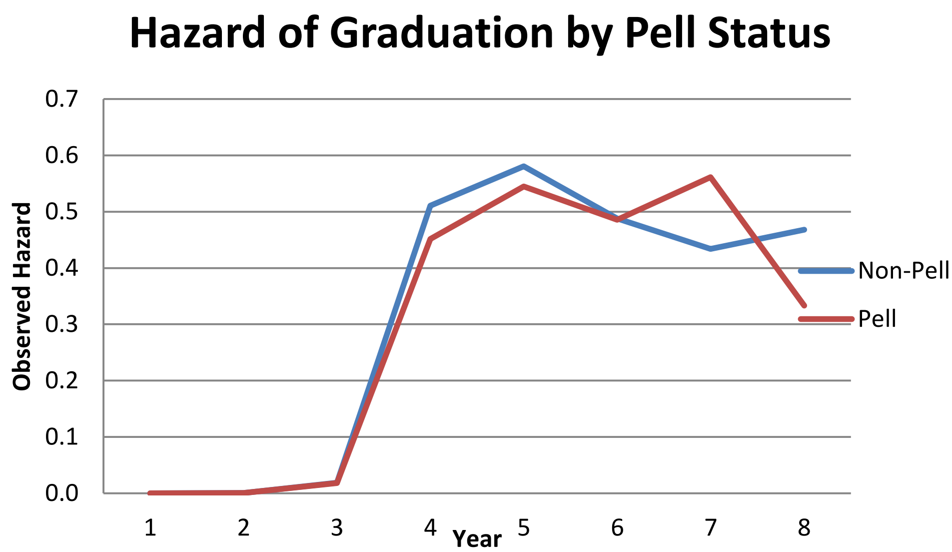

Figure 4.5

The cumulative graduation proportions are important to explore across independent variables because they display the cumulative disparities in graduation rates that are the result of different dropout and graduation hazards in the previous academic years. As displayed in figure 4.5, the Pell Grant recipients' hazard of graduation exceeded that of the non-Pell grant recipients in year seven; however, that peak in the conditional probability of graduation came too late for Pell Grant recipients to catch up in their total number of graduates compared to non-Pell. The cumulative gap in graduation proportions for non-Pell compared to Pell is 7.4 percentage points at year four, 8.8 percentage points at year five (when the non-Pell group experiences their peak in graduation hazard), and stabilizes at just over seven percentage points in the remaining years.

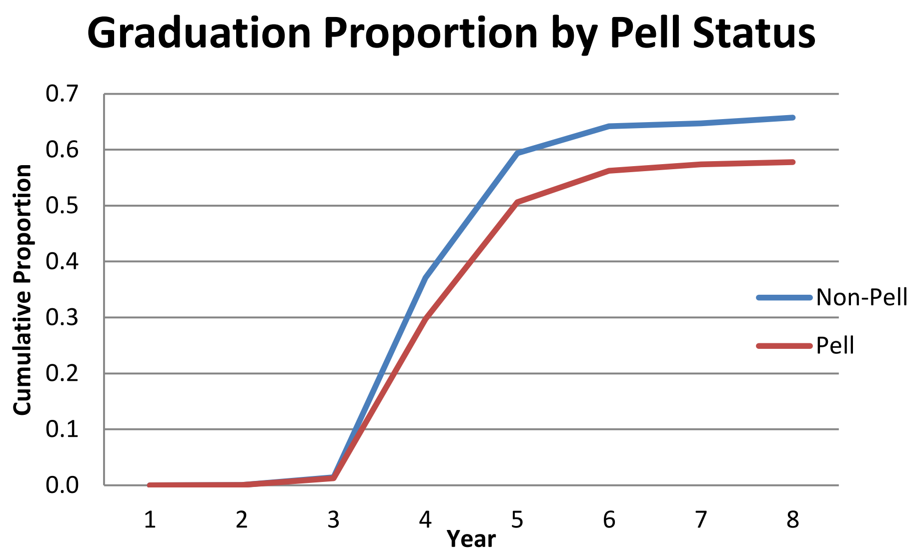

Figure 4.6

First Generation Status

The purpose of this section is to display and review the conditional hazards and cumulative graduation proportion of students by first generation status. Table 4.3 displays the life table for first generation and non-first generation students. 27% of the year one's risk set are first generation students.

Table 4.3

Life Table of FA02-FA07 Cohorts of New Freshmen by First Generation Status

Group	Year	Risk Set	Dropout	Graduate	Censored	Hazard Probability		Graduation Proportion
						Dropout	Graduate	
Non-First Generation	1	17495	2525	1	0	0.144	0.000	0.000
	2	14969	1627	5	0	0.109	0.000	0.000
	3	13337	642	237	0	0.048	0.018	0.014
	4	12458	304	6328	0	0.024	0.508	0.376
	5	5826	185	3349	1128	0.032	0.575	0.604
	6	1164	59	557	234	0.051	0.479	0.654
	7	314	4	141	72	0.013	0.449	0.661
	8	97	0	44	53	0.000	0.454	0.671
First Generation	1	6570	1290	1	0	0.196	0.000	0.000
	2	5279	807	2	0	0.153	0.000	0.000
	3	4470	271	94	0	0.061	0.021	0.015
	4	4105	158	2008	0	0.039	0.489	0.320

5	1939	80	1122	341	0.041	0.579	0.521
6	396	30	204	78	0.076	0.515	0.569
7	84	0	39	21	0.000	0.464	0.573
8	24	0	11	13	0.000	0.458	0.580

Figure 4.7 shows the dropout hazards in table 4.3 across first generation status. The conditional hazard of dropout for first generation students starts out higher than it is for non-first generation students: .196 compared to .144. The initial dropout hazard for the first generation students is one of the largest initial hazards across all of the demographic variables explored. The non-first generation students and the first generation students see a similar decrease in hazard from year one to year two: both groups decrease 4 percentage points. The first generation students have a larger decrease in year three's hazard than non-first generation students, but after year four, the changes in hazards for both groups are very similar. First generation students consistently have a higher dropout hazard compared to non-first generation students, since these students started with a higher hazard.

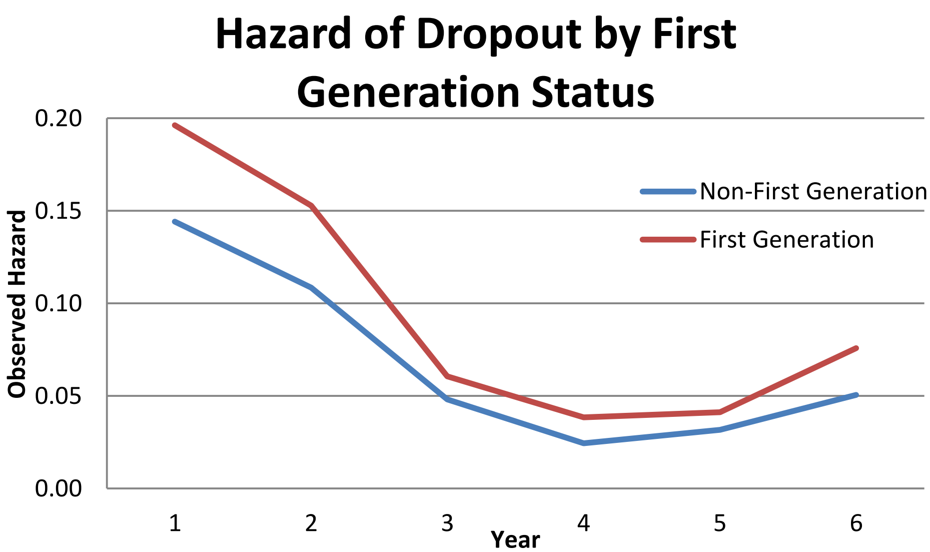

Figure 4.7

Figure 4.8 shows the hazards of graduation for first generation and non-first generation students. The graduation hazards of first generation students compared to non-first generation are very similar in both the shape and rates. At year four, the non-first generation students have a graduation hazard two percentage points greater than the first generation students' graduation hazard: .508 compared to .489. However, at year six, first generation students have a graduation hazard 3.7 percentage points higher than non-first generation.

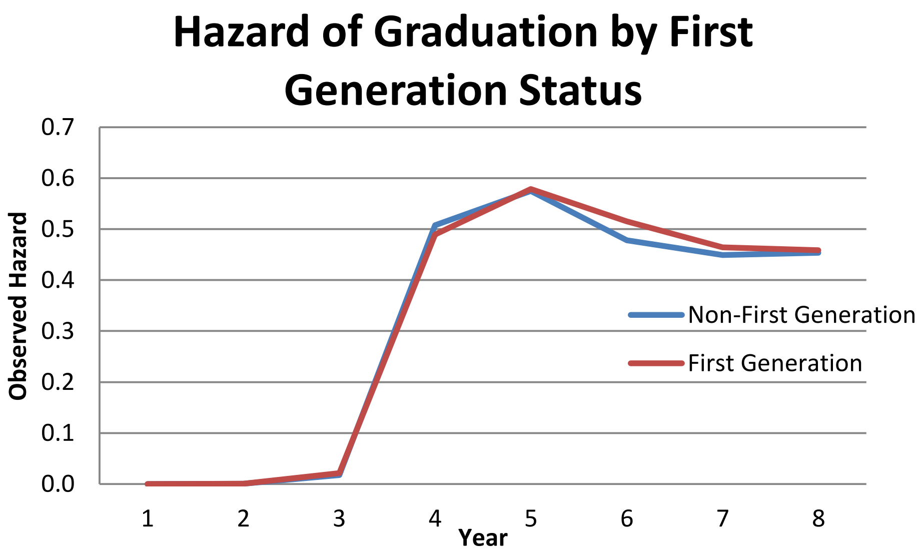

Figure 4.8

Figure 4.9 shows the cumulative graduation rates for first generation and non-first generation students. At year four, there is a 5.5 percentage point gap in the cumulative graduation proportion. This gap grows to a 9.1 percentage point difference by year eight. The gap in cumulative graduation for first generation students is most likely due to the higher dropout hazard of first generation students since the graduation hazards are similar. The first generation group loses more students in the risk set to dropout, so the cumulative impact of similar graduation hazards results in a smaller proportion of first generation graduates.

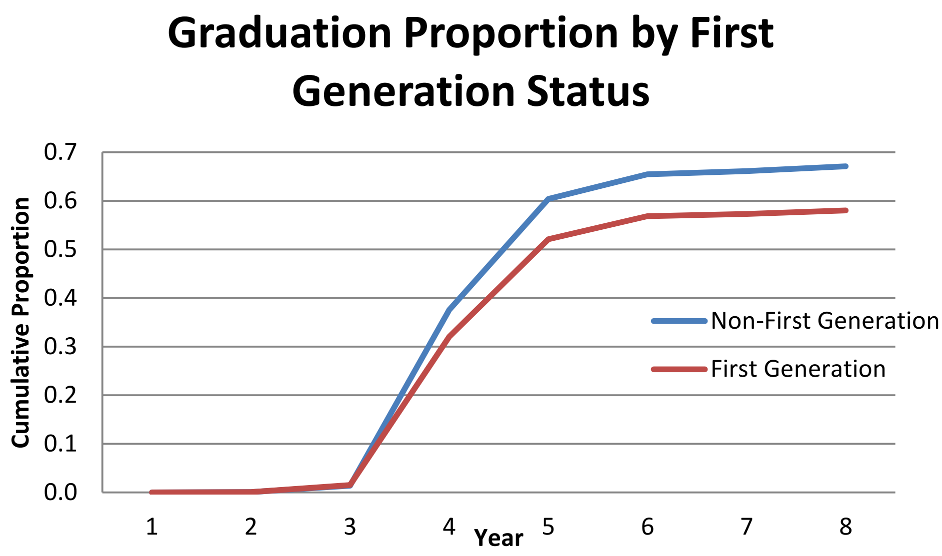

Figure 4.9

First Generation and Pell Grant Status

This section explores the dropout and graduations hazards along with cumulative graduation proportions across both Pell Grant recipients and first generation status. The purpose of looking across both variables is to explore if there is an interaction between Pell Grant recipients and first generation status. 66% of students are non-first generation and non-Pell Grant recipients, 20% of the risk set is first generation non-Pell Grant recipients, 7% of the risk set is non-first generation and a Pell Grant recipient, and 7% of the risk set are both first generation students and Pell Grant recipients.

Table 4.4

Life Table of FA02-FA07 Cohorts of New Freshmen by First Generation and Pell Status

Group	Year	Risk	Dropout	Graduate	Censored	Hazard Probability	Graduation

	Set					Dropout	Graduate	Proportion
Non-First Generation Non-Pell	1	15875	2239	1	0	0.141	0.000	0.000
	2	13635	1449	5	0	0.106	0.000	0.000
	3	12181	575	214	0	0.047	0.018	0.014
	4	11392	274	5838	0	0.024	0.513	0.382
	5	5280	168	3047	1025	0.032	0.577	0.611
	6	1040	54	501	203	0.052	0.482	0.661
	7	282	4	125	65	0.014	0.443	0.667
	8	88	0	41	47	0.000	0.466	0.677
First Generation Non-Pell	1	4822	947	1	0	0.196	0.000	0.000
	2	3874	555	1	0	0.143	0.000	0.000
	3	3318	194	75	0	0.059	0.023	0.016
	4	3049	107	1540	0	0.035	0.505	0.335
	5	1402	56	834	235	0.040	0.595	0.537
	6	277	25	142	51	0.090	0.513	0.581
	7	59	0	23	15	0.000	0.390	0.583
	8	21	0	10	11	0.000	0.476	0.593
Non-First Generation Pell	1	1620	286	0	0	0.177	0.000	0.000
	2	1334	178	0	0	0.133	0.000	0.000
	3	1156	67	23	0	0.058	0.020	0.014
	4	1066	30	490	0	0.028	0.460	0.317
	5	546	17	302	103	0.031	0.553	0.538
	6	124	5	56	31	0.040	0.452	0.593
	7	32	0	16	7	0.000	0.500	0.606
	8	9	0	3	6	0.000	0.333	0.613
First Generation Pell	1	1748	343	0	0	0.196	0.000	0.000
	2	1405	252	1	0	0.179	0.001	0.001
	3	1152	77	19	0	0.067	0.017	0.011
	4	1056	51	468	0	0.048	0.443	0.279
	5	537	24	288	106	0.045	0.536	0.476
	6	119	5	62	27	0.042	0.521	0.533
	7	25	0	16	6	0.000	0.640	0.544
	8	3	0	1	2	0.000	0.333	0.545

The initial hazard is highest for first generation students despite their Pell Grant recipient status. Figure 4.10 shows that first generation students (both Pell and non-Pell) have the highest initial dropout hazard of .196. However, the first generation Pell Grant students have the smallest decrease in their dropout hazard in the second year, since they only experience a 1.7 percentage point decrease. The first generation, non-Pell students experience a 3.5 percentage point decrease in their dropout hazard. In the third year, all of the groups experience a large decrease in their dropout hazard rates. The first generation Pell group begins to close the gap by dropping 11.3 percentage points from year two to year three. The non-first generation, non-Pell group experienced the smallest decrease in the third year (5.9 percentage points); however, this group had the lowest initial rate and maintains the lowest rate at the third year compared to the other groups. Pell Grant recipients (regardless of first generation status) experience a small percentage point change in year three. By the fourth year, all groups experience similar changes in dropout hazards- two to three percentage points. During the fifth and sixth years, all groups (besides the first generation Pell Grant recipients) start to see slight increases in the dropout hazard. At year six, the first generation non-Pell students have the highest dropout rate- .090- and non-first generation Pell students have the lowest rate- .040. This change in the ordering of dropout hazard could be due to smaller risk sets at year six (just over 100 for both of the Pell recipient groups). The biggest interaction between Pell and first generation status appears to be at year two when first generation Pell Grant recipients experience markedly reduced decrease in hazard compared to the other three groups.

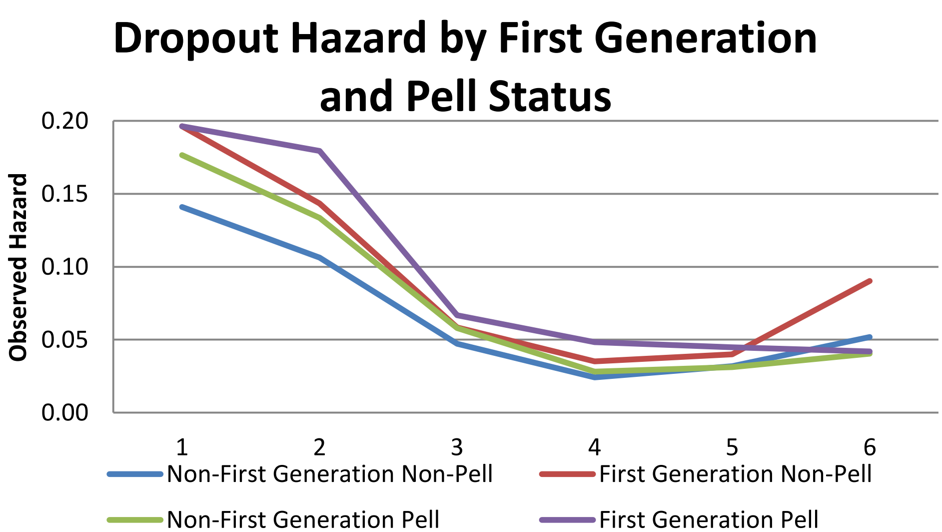

Figure 4.10

Figure 4.11 displays the graduation hazard across first generation and Pell Grant recipient status. The graduation hazard at year four is similar for Pell recipients (regardless of first generation status) both groups have a hazard rate less than .5. In contrast, non-Pell students have a graduation hazard above .5 at year four, since the non-first generation non-Pell group has a graduation hazard of .513 and the first generation non-Pell group has a graduation hazard of .505. The first generation Pell recipients have a different graduation hazard trajectory compared to the other groups because the highest hazard rate for this group occurs at year seven with the rate .640. For all of the other groups, the graduation hazard peaks at year five. The non-first generation Pell recipients have a second peak in graduation hazard at year seven, but this second peak is a lower rate than this group's hazard at year five: .553 compared to .500. Again (similar to the dropout hazard) the first generation Pell grant recipients appear to have a distinct graduation hazard trajectory compared to the other three groups.

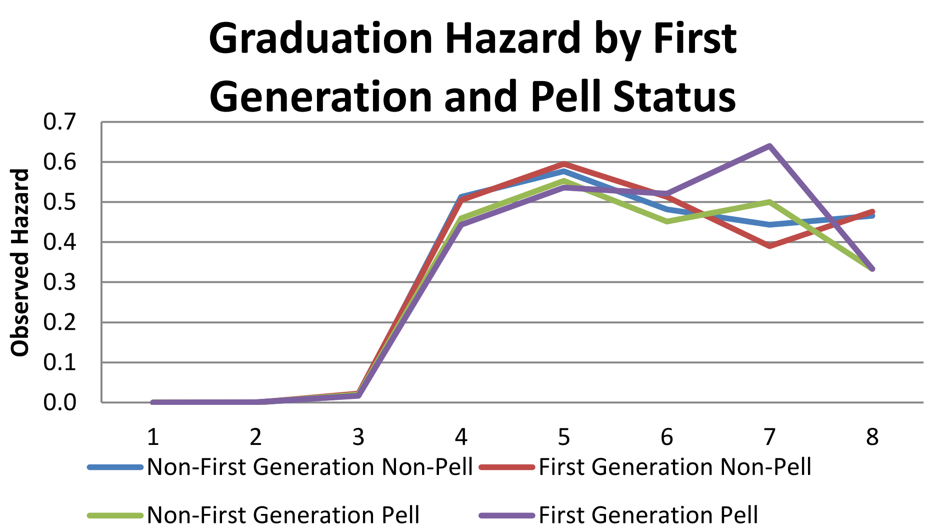

Figure 4.11

Figure 4.12 displays the cumulative graduation proportion across first generation and Pell Grant recipient status. The non-first generation non-Pell students have the highest graduation proportions and the first generation, Pell Grant recipients have the lowest graduation proportions. The gap in graduation proportions is 10.2 percentage points at year four. This gap is the greatest at year five when it grows to

13.4 percentage points. The gap shrinks slightly at years six and seven, but then increases in year eight to 13.2 percentage points. The first generation non-Pell students have very similar cumulative graduation proportions to the non-first generation Pell recipients. First generation non-Pell have a slightly higher graduation proportions through year four and then the non-first generation Pell students obtain a higher graduation proportion for years five through eight. This switch in proportions is probably due to the second peak in graduation hazard that the non-first generation, Pell recipients experience at year seven.

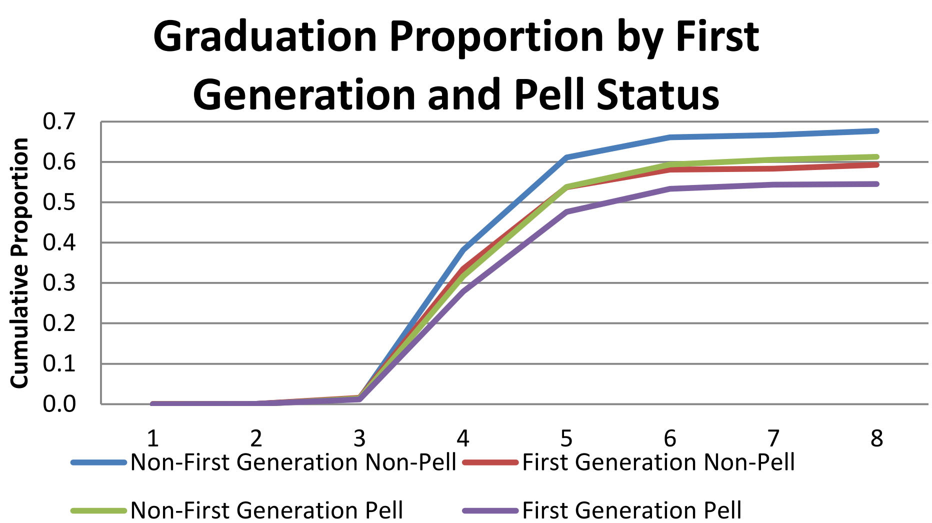

Figure 4.12

Minority Status

The purpose of this section is to explore the differences in graduation and dropout trajectories for students across minority status. Table 4.5 displays the dropout and graduation hazards along with the cumulative graduation proportions for minority and non-minority students. 13.6% of the risk set in year one are minority students.

Table 4.5

Life Table of FA02-FA07 Cohorts of New Freshmen by Minority Status

Group	Year	Risk Set	Dropout	Graduate	Censored	Hazard Probability		Graduation Proportion
						Dropout	Graduate	
Non-Minority	1	20800	3227	1	0	0.155	0.000	0.000
	2	17572	2034	7	0	0.116	0.000	0.000
	3	15531	756	304	0	0.049	0.020	0.020
	4	14471	393	7385	0	0.027	0.510	0.370
	5	6693	220	3920	1236	0.033	0.586	0.594
	6	1317	77	640	259	0.059	0.486	0.635
	7	341	4	154	78	0.012	0.452	0.645
	8	105	0	50	55	0.000	0.476	0.650
Minority	1	3265	588	1	0	0.180	0.000	0.000
	2	2676	400	0	0	0.150	0.000	0.000
	3	2276	157	27	0	0.069	0.012	0.012
	4	2092	69	951	0	0.033	0.455	0.300
	5	1072	45	551	233	0.042	0.514	0.505
	6	243	12	121	53	0.049	0.498	0.554
	7	57	0	26	15	0.000	0.456	0.566
	8	16	0	5	11	0.000	0.313	0.570

Figure 4.13 shows that minority students have a higher hazard of dropout in all of the years except year six. In year one, the non-minority students have a dropout hazard that is 2.5 percentage points lower than the minority students. Non-minority students decrease 3.9 percentage points from year one to year two and minority students decrease 3.1 percentage points over that same time period. In year two, this gap increases to a 3.4 percentage point gap, but by year three the gap has decreased to 2 percentage points. Minority students see larger decreases in dropout hazards in years three and four which allows them to get closer to the non-minority dropout hazard rate. In years five and six, both groups see a slight

increase in their dropout hazards. In years four and five the gap between the groups' hazard rate is very small and by year six the non-minority students have a slightly higher dropout hazard than minority students. Overall, the difference between dropout hazards for minority and non-minority students decreases slightly over time.

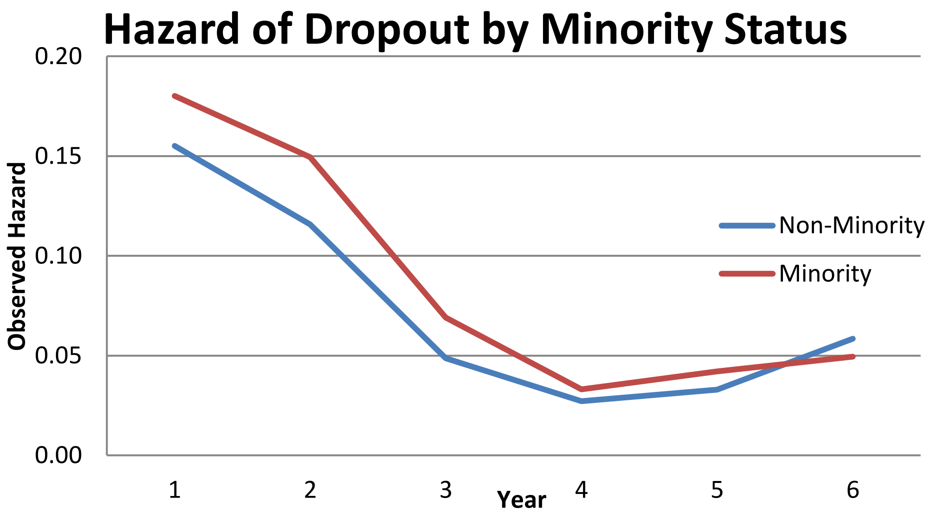

Figure 4.13

Figure 4.14 displays the graduation hazards for minority and non-minority students. Minority students have lower graduation hazards than non-minority students except for years six and seven where the rates for the two groups are very similar. The gap between the groups' graduation hazard is largest at year four (.056) and year five (.072). The non-minority group has a graduation hazard that peaks at year five, while the minority students' highest rate is also highest at year five but there is not much of peak-rather a gradual decrease. For instance, the non-minority students have a graduation hazard of .586 in year five and that decreases to .486 in year six, while the minority students have a graduation hazard of .514 in year five and that drops slightly to .498 in year six. The shape of the graduation hazards for minority and non-minority students are slightly different.

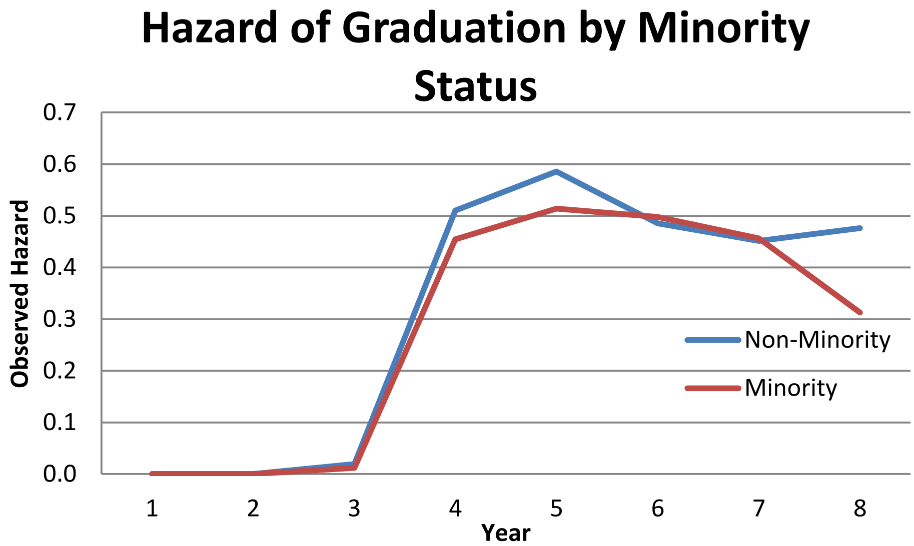

Figure 4.14

The cumulative graduation proportions across minority status are presented in figure 4.15. A larger proportion of non-minority students experience graduation compared to the proportion of minority students who graduate. At year eight, the gap between the groups is 7.8 percentage points because minority students have a cumulative graduation proportion of .579 and non-minority students have a proportion of .657. The gap between the two groups' cumulative graduation proportion is largest at year five (.088). This is most likely due to the pronounced peak in the graduation hazard for non-minority students.

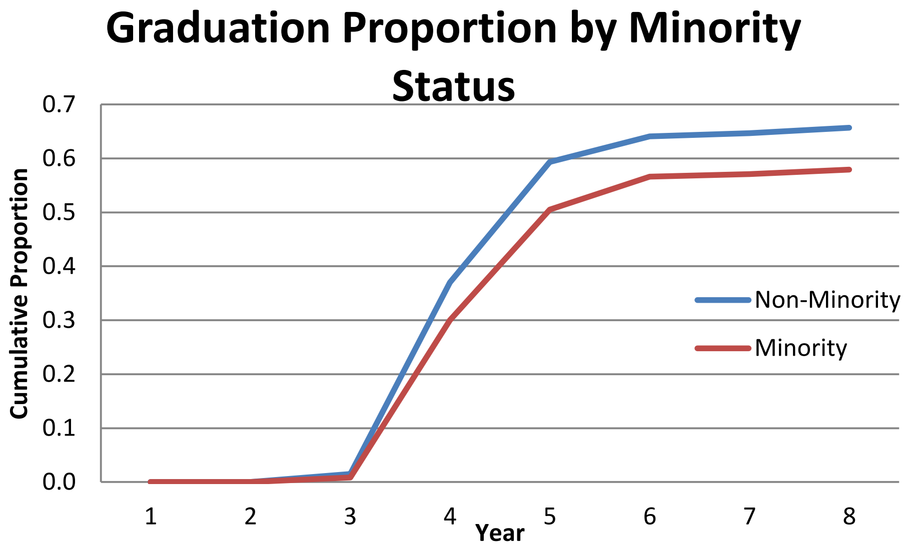

Figure 4.15

Gender

This section discusses differences in the dropout and graduation trajectories for female and male students. Table 4.6 displays the graduation hazards along with the graduation proportions across gender. 55.6% of the risk set in year one are female.

Table 4.6

Life Table of FA02-FA07 Cohorts of New Freshmen by Gender

Group	Year	Risk Set	Dropout	Graduate	Censored	Hazard Probability		Graduation Proportion
						Dropout	Graduate	
Female	1	13386	2164	0	0	0.162	0.000	0.000
	2	11222	1293	6	0	0.115	0.001	0.000
	3	9923	481	244	0	0.049	0.025	0.019
	4	9198	215	5519	0	0.023	0.600	0.431
	5	3464	125	2138	633	0.036	0.617	0.620
	6	568	29	275	123	0.051	0.484	0.653
	7	141	0	72	32	0.000	0.511	0.658
	8	37	0	14	23	0.000	0.378	0.662
Male	1	10679	1651	2	0	0.155	0.000	0.000
	2	9026	1141	1	0	0.126	0.000	0.000
	3	7884	432	87	0	0.055	0.011	0.008
	4	7365	247	2817	0	0.034	0.383	0.272
	5	4301	140	2333	836	0.033	0.542	0.533
	6	992	60	486	189	0.061	0.490	0.604
	7	257	4	108	61	0.016	0.420	0.610
	8	84	0	41	43	0.000	0.488	0.626

The hazard of dropout for female and male students is displayed in figure 4.16. Male students have a higher hazard of dropout in all years except one and five. Initially, females start with a slightly higher hazard- .162 compared to .155. However, females experience a larger decrease in dropout hazard in year two (.047) compared to the decrease males experience (.028). Also, in year four the male hazard continues to decrease while the female dropout hazard begins to increase; nonetheless, the shape of the hazard trajectories is very similar for both genders. The gap between the gender's dropout hazards is also very small.

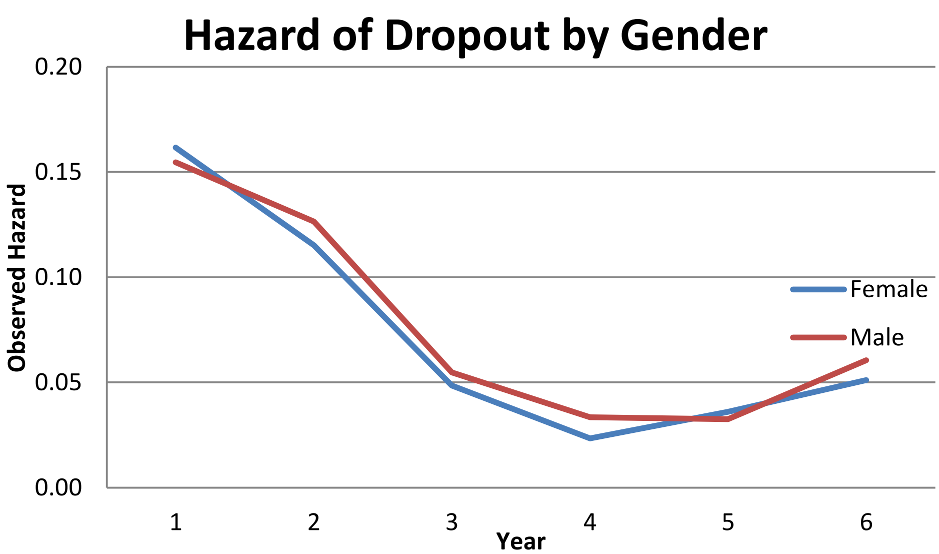

Figure 4.16

Figure 4.17 shows the graduation hazard for males and females. Unlike the observed dropout hazards across gender, the observed graduation hazards for males and females shows different shapes and large gaps. The graduation hazard at year four for females is 21.8 percentage points higher than the graduation hazard at year four for males. Male students have a graduation hazard that peaks at year five, while female students also have their highest graduation hazard at year five but year four's graduation hazard is nearly as high. Females have more of a plateau than a peak. The graduation hazards for males and females at year six are nearly identical, and in year eight, male students have a higher graduation hazard than female students.

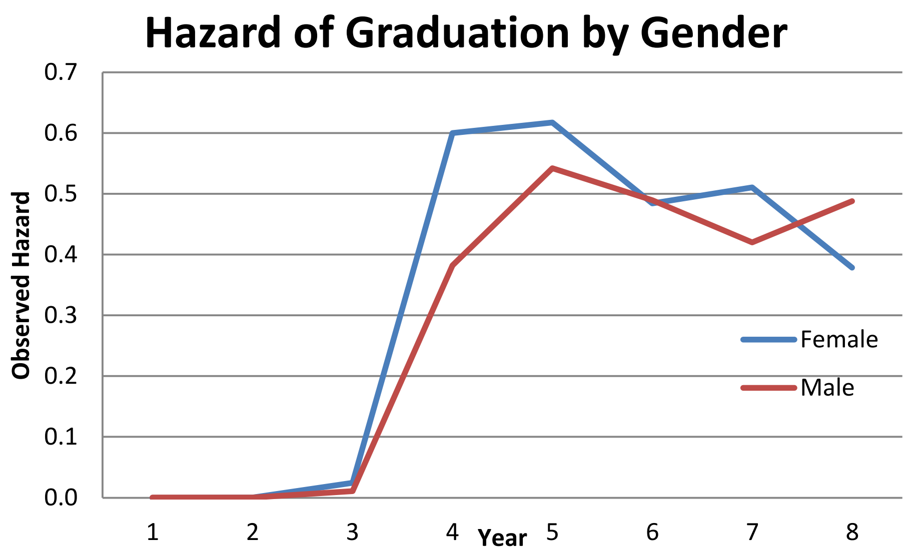

Figure 4.17

The cumulative graduation proportion for male and female students is displayed in figure 4.18. The largest gap in graduation proportions between males and females is at year four when .431 of females have graduated compared to only .272 of males. Despite this 15.9 percentage point gap at year four, the disparity between gender narrows with additional years. At year eight there is only a 3.6 percentage point difference between the cumulative proportion of females who have graduated and males.

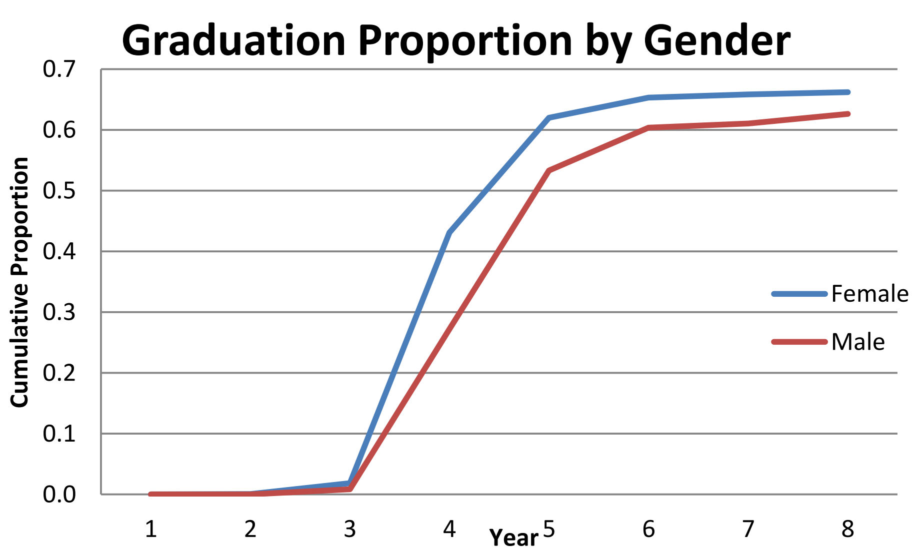

Figure 4.18

Minority Status and Gender

This section explores the dropout and graduation hazards along with cumulative graduation proportions across gender and minority status. The purpose of looking across both variables is to explore if there is an interaction between gender and minority status. 39% of the risk set in year one are male non-minority; 48% are female non-minority; 6% are male minority students; and 8% are female minority students.

Table 4.7

Life Table of FA02-FA07 Cohorts of New Freshmen by Gender and Minority Status

Group	Year	Risk Set	Dropout	Graduate	Censored	Hazard Probability		Graduation Proportion
						Dropout	Graduate	
Male Non-Minority	1	9307	1410	1	0	0.152	0.000	0.000
	2	7896	967	1	0	0.123	0.000	0.000
	3	6928	355	79	0	0.051	0.011	0.009
	4	6494	212	2519	0	0.033	0.388	0.279
	5	3763	118	2070	711	0.031	0.550	0.544
	6	864	52	421	161	0.060	0.487	0.613
	7	230	4	96	56	0.017	0.417	0.620
	8	74	0	39	35	0.000	0.527	0.637
Female Non-Minority	1	11493	1817	0	0	0.158	0.000	0.000
	2	9676	1067	6	0	0.110	0.001	0.001
	3	8603	401	225	0	0.047	0.026	0.020
	4	7977	181	4866	0	0.023	0.610	0.443
	5	2930	102	1850	525	0.035	0.631	0.633

	6	453	25	219	98	0.055	0.483	0.664
	7	111	0	58	22	0.000	0.523	0.669
	8	31	0	11	20	0.000	0.355	0.672
Male Minority	1	1372	241	1	0	0.176	0.001	0.001
	2	1130	174	0	0	0.154	0.000	0.001
	3	956	77	8	0	0.081	0.008	0.007
	4	871	35	298	0	0.040	0.342	0.224
	5	538	22	263	125	0.041	0.489	0.460
	6	128	8	65	28	0.063	0.508	0.539
	7	27	0	12	5	0.000	0.444	0.541
	8	10	0	2	8	0.000	0.200	0.549
Female Minority	1	1893	347	0	0	0.183	0.000	0.000
	2	1546	226	0	0	0.146	0.000	0.000
	3	1320	80	19	0	0.061	0.014	0.010
	4	1221	34	653	0	0.028	0.535	0.355
	5	534	23	288	108	0.043	0.539	0.538
	6	115	4	56	25	0.035	0.487	0.586
	7	30	0	14	10	0.000	0.467	0.592
	8	6	0	3	3	0.000	0.500	0.602

Figure 4.19 displays the dropout hazard across gender and minority status. For years one through three, the minority groups cluster together and the non-minority students cluster together. The female and male minority dropout hazards are less than a percentage point different from each another in years one and two. The hazard for these two groups is only about two percentage points different at year three. Similarly, for non-minority students, the dropout hazards only differ by a percentage point across all of the years. More substantial gaps exist between the minority and non-minority students; however, the gap between minority and non-minority students is similar for both males and females. Regardless of minority status, females have a slightly higher dropout hazard at year one and then a slightly lower dropout hazard in the remaining years compared to males.

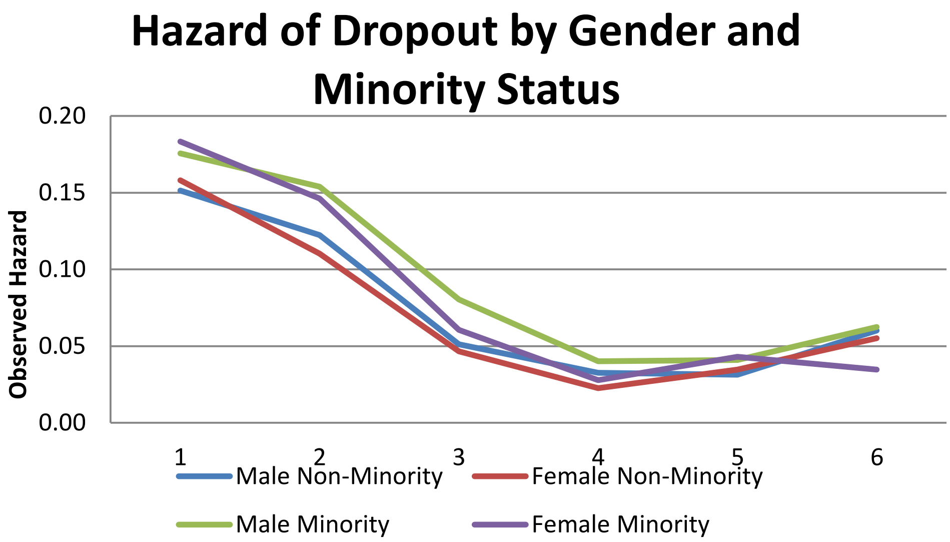

Figure 4.19

Figure 4.20 shows that the graduation hazards appear to cluster around gender in contrast to the dropout hazards which clustered around minority status. Similar to patterns in figure 4.17, female students (regardless of minority status) have a considerably higher graduation hazard at year four compared to the male students' graduation hazard at year four. All of the groups have the highest graduation hazard at year five. At year five, female non-minority students have a considerably higher graduation hazard than the other three groups. The five year graduation hazard for female non-minority students is .631, while female minority students and male non-minority students have similar graduation hazards of .539 and .550 respectively. At year six, all four of the groups have a graduation hazard within one percentage point. At year eight male minority students have a considerably lower graduation hazard (.200) than the other three groups. Although, male minority students have a lower graduation hazard at most time points this considerably lower year eight hazard is most likely due to a small risk set of only ten students.

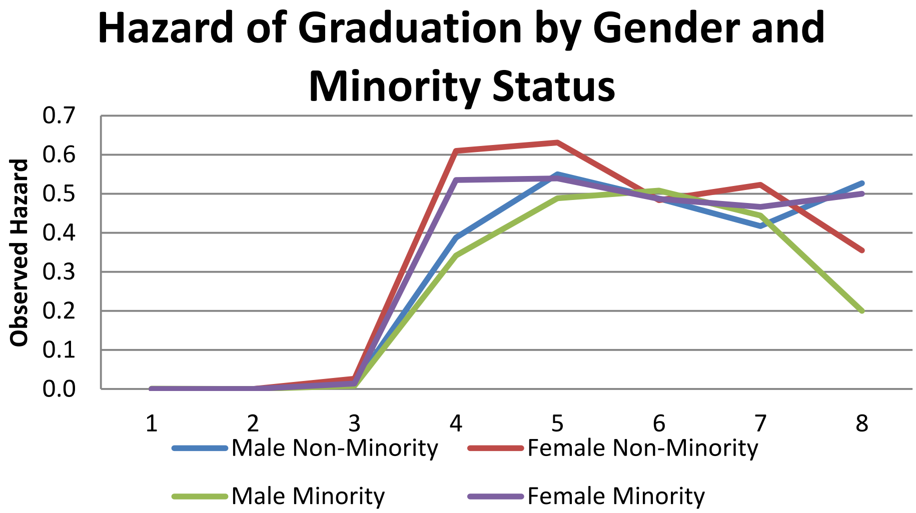

Figure 4.20

Figure 4.21 shows the cumulative graduation proportion across gender and minority status. Female non-minority students have the largest graduation proportion and male minority students have the lowest graduation proportion. The graduation proportions for male non-minority students and female minority students are very similar. The gap in cumulative graduation proportions between female non-minority students and male minority students is largest at year four when the gap is 22 percentage points (.443 compared to .224) and narrows to 12.3 percentage points by year eight.

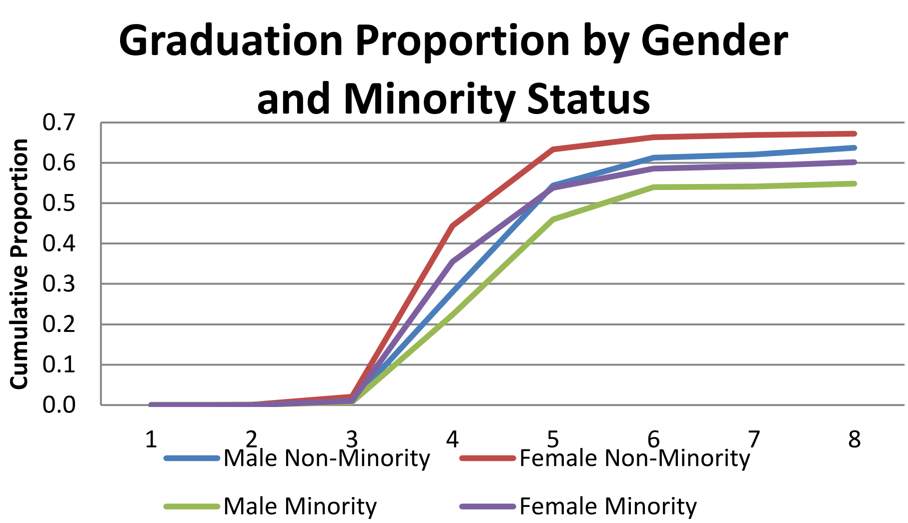

Figure 4.21

Residency

Table 4.8 displays the graduation hazards along with the graduation proportions across residency status. 21.4% of year one's risk set are non-resident students.

Table 4.8

Life Table of FA02-FA07 Cohorts of New Freshmen by Residency

Group	Year	Risk Set	Dropout	Graduate	Censored	Hazard Probability		Graduation Proportion
						Dropout	Graduate	
Non-Resident	1	5155	1069	0	0	0.207	0.000	0.000
	2	4086	596	1	0	0.146	0.000	0.000
	3	3489	199	62	0	0.057	0.018	0.012
	4	3228	91	1705	0	0.028	0.528	0.343
	5	1432	44	844	258	0.031	0.589	0.536
	6	286	18	143	57	0.063	0.500	0.579
	7	68	0	39	9	0.000	0.574	0.586
	8	20	0	12	8	0.000	0.600	0.593
Resident	1	18910	2746	2	0	0.145	0.000	0.000
	2	16162	1838	6	0	0.114	0.000	0.000
	3	14318	714	269	0	0.050	0.019	0.015
	4	13335	371	6631	0	0.028	0.497	0.365
	5	6333	221	3627	1211	0.035	0.573	0.594
	6	1274	71	618	255	0.056	0.485	0.645
	7	330	4	141	84	0.012	0.427	0.651
	8	101	0	43	58	0.000	0.426	0.661

Figure 4.22 displays the dropout hazard across residency status. Non-resident students have higher dropout hazards in years one through three and then have nearly identical dropout hazards for years four through six. At year one, non-resident students have a dropout hazard of .207 and resident students have a dropout hazard of .145- a .062 difference between the groups. This disparity decreases to .032 in year two and to less than one percentage point at year three. At year four, the dropout hazards are identical (.028) and in year five, the residents have a slightly higher dropout hazard. At year six, the non-resident dropout hazard is once again slightly higher than the resident hazard. Overall, the difference in the hazard of dropout between residents and non-residents decreases over time.

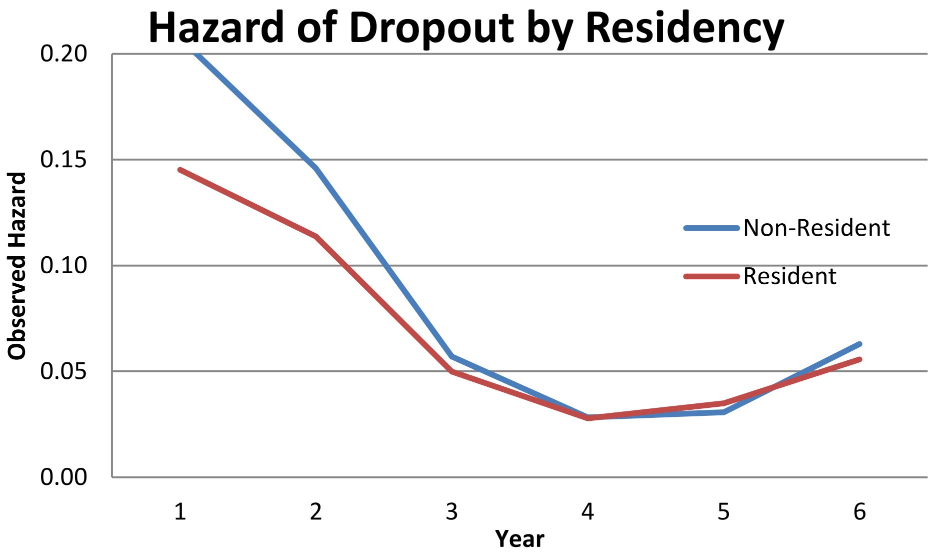

Figure 4.22

Figure 4.23 shows the graduation hazard across residency status. The non-resident group consistently has a higher graduation hazard than the resident group, but the changes in hazard between the years are very similar for both groups. Both groups have a peak in the graduation hazard at year five and then decrease in year six; however, the non-resident hazard increases in years seven and eight. Unlike any of the other subpopulations, the non-resident hazard is highest at year eight. The graduation trajectories are nearly identical for resident and non-resident students until year six.

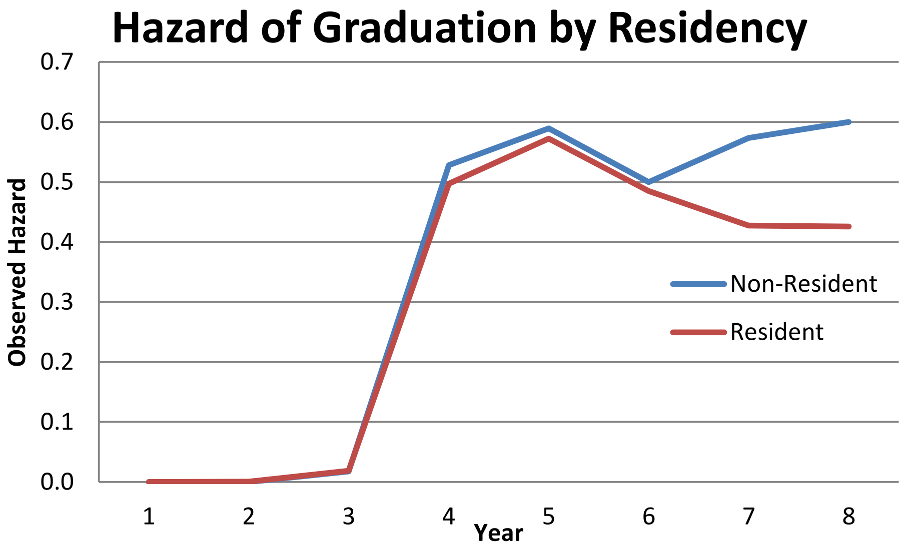

Figure 4.23

Figure 4.24 shows the cumulative graduation proportion by residency. A larger proportion of resident students graduate compared to non-residents. This gap in the graduation rate is due to the non-residents' higher dropout hazard, since they have a slightly higher graduation hazard at all years. The large disparity in graduation hazards for non-residents in years seven and eight does not allow the group to catch up to resident students' cumulative rate, since the higher graduation hazard is being applied on a smaller risk set. The gap in cumulative graduation proportions is largest at year eight and is equal to a 6.8 percentage point difference.

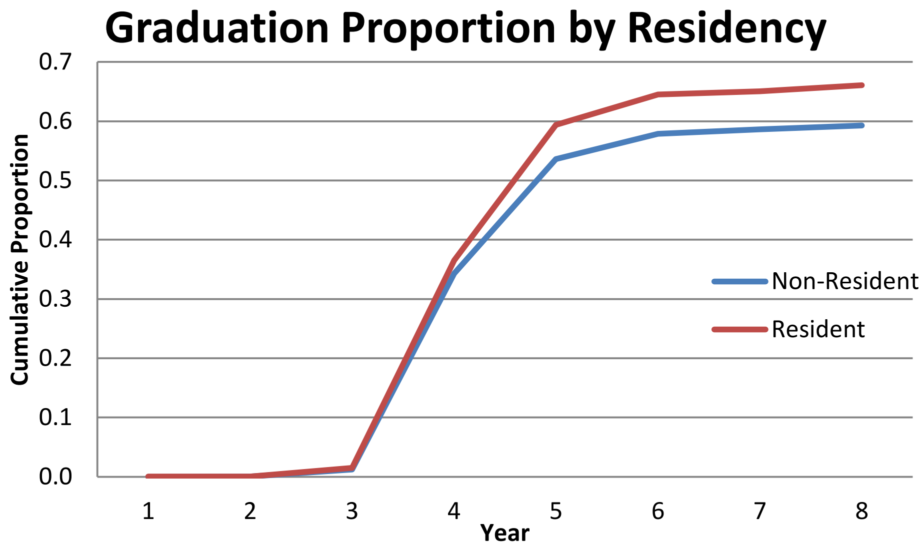

Figure 4.24

Academic Preparation (Index)

Table 4.9 shows dropout and graduation hazards along with the cumulative graduation proportion. Index is a continuous variable that was put into a dichotomous coding for this explorative analysis. It will be treated as continuous during the modeling section of the analysis. Index is also the only independent variable that has missing cases, so the subgroup's sum for each risk set will not equal the risk sets in table 4.1. For the purpose of the exploratory analysis low index was defined as 105 or less. A score of 105 was chosen as the cutoff point because it is the 25th percentile. As a continuous variable, index has a mean of 113.3 and a standard deviation of 110.6.

Table 4.9

Life Table of FA02-FA07 Cohorts of New Freshmen by High/Low Index

Group	Year	Risk Set	Dropout	Graduate	Censored	Hazard Probability Dropout	Hazard Probability Graduate	Graduation Proportion
Low Index	1	5391	1090	2	0	0.202	0.000	0.000
	2	4299	736	0	0	0.171	0.000	0.000
	3	3563	245	11	0	0.069	0.003	0.002
	4	3307	132	1233	0	0.040	0.373	0.231
	5	1942	71	1036	388	0.037	0.534	0.462
	6	447	27	210	101	0.060	0.470	0.524

	7	109	2	48	33	0.018	0.440	0.532
	8	26	0	11	15	0.000	0.423	0.543
High Index	1	18361	2671	0	0	0.146	0.000	0.000
	2	15690	1648	7	0	0.105	0.000	0.000
	3	14035	650	317	0	0.046	0.023	0.018
	4	13068	318	7023	0	0.024	0.537	0.400
	5	5727	191	3387	1063	0.033	0.591	0.619
	6	1086	62	541	198	0.057	0.498	0.665
	7	285	2	132	60	0.007	0.463	0.670
	8	91	0	42	49	0.000	0.462	0.678

Figure 4.25 shows the dropout hazard across the dichotomous coding of index. Students with an index less than 105 have a higher initial hazard rate compared to students who have an index score greater than 104: .202 compared to .146. The difference between the groups' dropout hazard is largest at year two when a 6.6 percentage point difference between the groups exists. The disparity in dropout hazard decreases over time, and by the sixth year, the lower index students only have a .3 percentage point difference in their dropout hazard compared to higher index students.

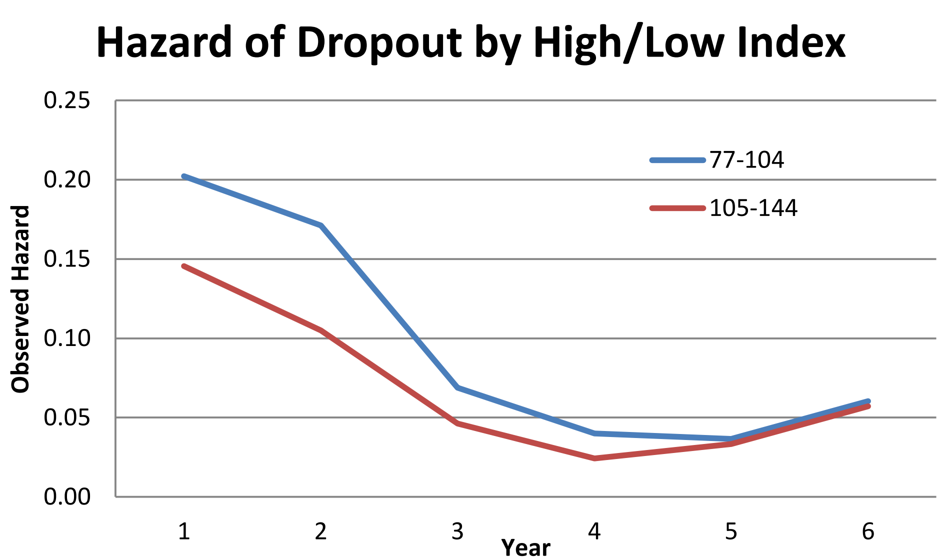

Figure 4.25

Figure 4.25 shows the graduation hazard for low and high index students. The high index students have the greatest advantage in their hazard rate at year four, since they have a graduation hazard equal to .537 compared to the low index students whose graduation hazard is only .373. The difference between the groups dwindles from the .165 difference in year four to .058 at year five. By year six, the gap is close to three percentage points, which is roughly where the gap remains through year eight. Generally, the graduation trajectories share a very similar shape for the two groups, even though the high index students have higher graduation hazards than the low index students.

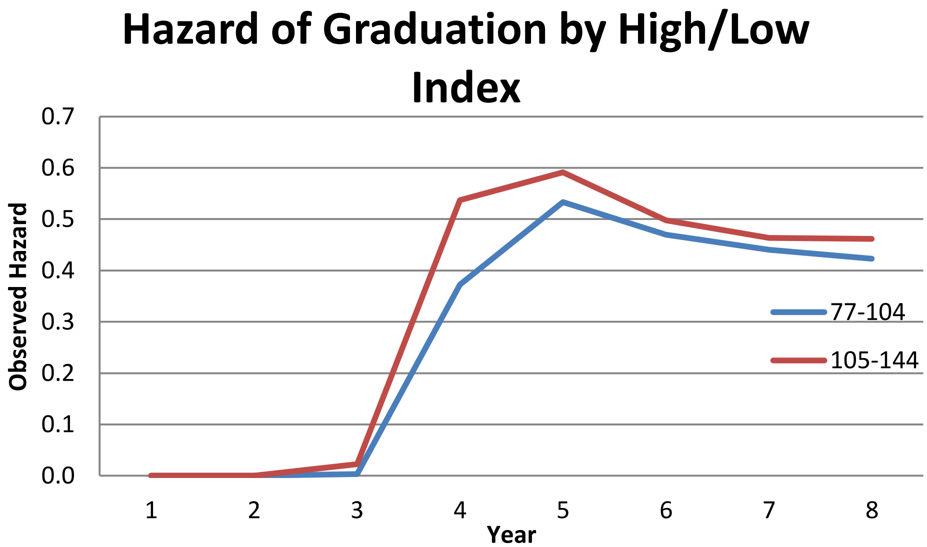

Figure 4.26

The cumulative graduation proportions for low and high index students are displayed in figure 4.26. High index students have a larger proportion of students that graduate compared to the low index students. The gap in cumulative graduation hazard is greatest at year four (16.9 percentage points) when the cumulative graduation proportion for high index students is .400 compared to the low index graduation proportion which is only .231. By year eight, the gap in the cumulative graduation proportions decreases to 13.5 percentage points.

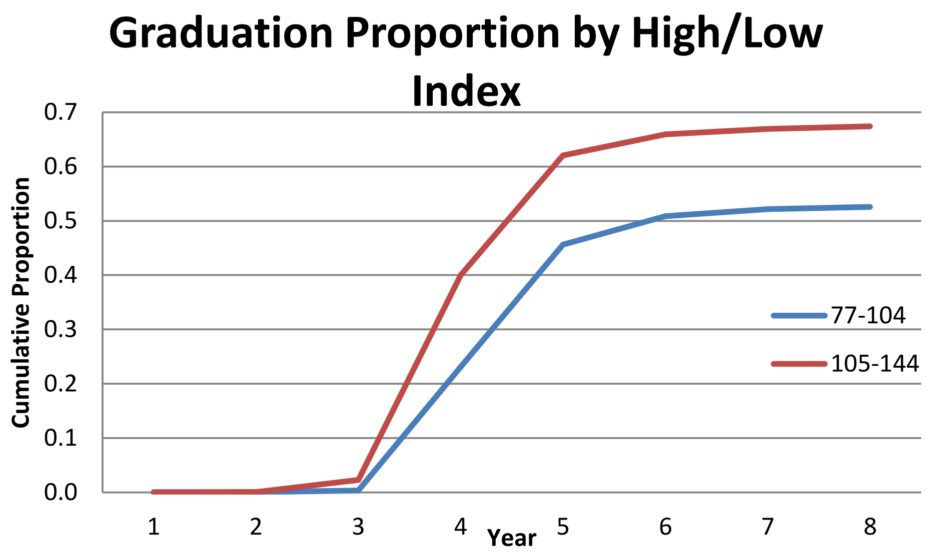

Figure 4.27

Exploratory Analysis Conclusions

This section summarizes the general trends in the observed dropout and graduation hazards. Since the hazards are the outcomes of the competing risk event history model, it is important to stay close to the data and examine how the model's predictor variables vary across each hazard. The observed dropout and graduation hazards for each of the demographic variables are examined in order to inform which interactions are included in the descriptive model. The trends of the demographic variables did not differ across the hazard rates in the same manner for dropout and graduation; therefore, the summary is organized by the two outcomes: dropout and graduation. Finally, this section concludes with a discussion of predictor by predictor interactions that will be tested in the modeling process.

Dropout Hazard.

A general trend when looking at the gaps in hazard across time over the demographic variables is that the difference in hazards decreased over time. For instance, this trend was present across Pell recipient status, minority status, residency, and index. The data was also graphed in the logit scale (since that is the scale modeling will be done) to help inform which of the model's inputs warrants further exploration in the modeling process. Based on the analysis above and evaluating the graphs in the log odds scale, a time by predictor interaction will be explored for Pell status and across residency.

Graduation Hazard.

Generally, the shapes of the graduation hazards varied due to different peaks in the hazard rate. Diverse students tended to have peaks later than the majority groups. For instance, Pell Grant recipients showed a much higher graduation rate at year seven when the majority of students peaked at year five. There was also a distinct difference in graduation hazard trajectories across gender. Female students peaked at year five, but their four year hazard was nearly as high. In contrast, male students had much more defined peak at year five, albeit their graduation hazard was much lower. The graduation hazards in the logit scale confirmed this trend of later peaks for Pell recipients and male students. Thus, an interaction between gender and time as well as Pell recipient status and time will be explored.

Predictor Interactions.

It is also important to explore if the effect of one independent variable depends on the level of another independent variable. First generation Pell year recipients appear to have slightly different graduation and dropout trajectories than their peers. Also, male minority students appear to have lower graduation hazards than the other groups. The interactions between gender and minority status will be explored during the modeling process along with the interaction between Pell and first generation.

Selecting a Descriptive Model

This section uses steps one through four described in chapter three to evaluate competing models that describe the longitudinal nature of dropout and graduation. Step one focuses on evaluating various time specifications. Step two takes the time specification determined in step one and introduces all the input variables into the model. The model presented in step two is the baseline model all other competing models are compared with. Step three introduces the time by predictor interactions that were decided in the exploratory analysis and compares models with all possible combinations to the step two model. Step four then compares models with predictor by predictor interactions to the baseline model. Step four also compares models that include the predictor by predictor interactions along with the predictor by time interactions that were determined in step three in order to determine this study's final model.

Step One

The first step of the descriptive modeling approach evaluates seven different specifications of time based on the recommendations for evaluating various time specifications in Singer and Willet (2003). The simplest model is the constant model, which removes time from the model. If the constant model is the best fitting approach, then that would indicate that the hazards of dropout and graduation do not vary across time. The next simplest specification enters time as a linear variable. If the linear specification is the best fitting model, then the graduation and dropout hazards would vary the same amount and in the same direction each academic year. The quadratic model squares the time variable so that this specification allows for one peak in the hazards, while the cubic specification cubes the time variable so that there are two peaks in the hazard. The fourth order model raises the time variable to the fourth power, while the fifth order model raises the time variable to the fifth power. The fourth and fifth order models allow more flexibility of hazards to change over time compared to the lower order models. The general specification enters time as eight dummy variables and allows the direction and magnitude of the hazards to change at all time points. The general specification is the most flexible of all time specifications. The goodness of fit statistics for each of these seven different models are displayed in table 4.10 below.

Table 4.10

Comparing Various Time Specifications

	Constant Model	Linear Model	Quadratic Model	Cubic Model	Fourth Order Model	Fifth Order Model	General Model
Log Likelihood	-64241.1	-49117.7	-45623.5	-44781.8	-44597.7	-44521.7	-44484.2
BIC	128502.3	98275.8	91307.5	89644.1	89296.1	89164.3	89109.5
AIC	128486.1	98243.5	91259.0	89579.5	89215.4	89067.4	-88996.5
Cragg-Uhler R2	0	0.378	0.448	0.465	0.468	0.469	0.470

Based solely on the goodness of fit statistics, the general specification for time is the best specification for this study. The simpler specifications have larger BIC statistics and smaller R^2 statistics. Intuitively, the general specification seems most appropriate because the observed graduation and hazards across each of the input variables were very different from one another, which could indicate that it is difficult to find a simpler specification that fits both hazards. Based on both the goodness of fit statistics as well as the exploratory analysis, the general specification of time will be used in this study's descriptive modeling. Table 4.11 below shows the model coefficients, odds ratios, and p-values for the general specification.

Table 4.11

Model Coefficients Only Including Time

	Graduated			Dropout		
	Odds Ratio	95% CIs *LL*	95% CIs *UL*	Odds Ratio	95% CIs *LL*	95% CIs *UL*
Year 1	0.000	0.000	0.000	0.188	0.182	0.195
Year 2	0.000	0.000	0.001	0.137	0.131	0.143
Year 3	0.020	0.018	0.022	0.055	0.052	0.059
Year 4	1.074	1.041	1.107	0.059	0.054	0.065
Year 5	1.476	1.410	1.546	0.087	0.077	0.099
Year 6	1.072	0.968	1.187	0.125	0.101	0.156
Year 7	0.841	0.690	1.026	0.019	0.007	0.050
Year 8	0.833	0.583	1.192	0.000	0.000	>999
Goodness of Fit						
Log Likelihood						-44484.235
Deviance						88968.469
n parameters						8
BIC						89109.53
AIC						88996.47
Cragg-Uhler R2						0.470

Step Two

The purpose of the second step is to introduce the predictor variables into the model. As stated earlier, the purpose of this step is not to statistically evaluate the input variables for inclusion or exclusion, since substantively, the literature and prior empirical research indicate their importance. The purpose of this step is to present the baseline model that will be used to compare more complicated models that include interactions in steps three and four. The odds ratios and 95% confidence intervals for both outcomes and the goodness of fit statistics for the baseline comparison model are presented below in table 4.12.

Table 4.12

Model Coefficients with Time and Predictor Variables

	Graduated			Dropout		
		95% CIs			95% CIs	
	Odds Ratio	*LL*	*UL*	Odds Ratio	*LL*	*UL*
Year 1	0.000	0.000	0.000	4.001	3.057	5.237
Year 2	0.000	0.000	0.000	2.925	2.231	3.834
Year 3	0.001	0.000	0.001	1.190	0.903	1.568
Year 4	0.033	0.025	0.043	1.222	0.923	1.618
Year 5	0.053	0.040	0.069	1.714	1.284	2.289
Year 6	0.043	0.032	0.057	2.411	1.715	3.391
Year 7	0.034	0.024	0.048	0.370	0.133	1.028
Year 8	0.033	0.021	0.052	0.000	0.000	>999
Female	1.860	1.770	1.955	1.038	0.989	1.089
First Generation	1.043	0.984	1.105	1.396	1.324	1.472
Pell Recipient	0.823	0.765	0.886	1.023	0.953	1.098
Minority	0.851	0.790	0.916	1.066	0.997	1.141
Non-Resident	0.859	0.808	0.914	1.284	1.214	1.359
CCHE Index	1.028	1.026	1.031	0.971	0.969	0.974
Goodness of Fit						
Log Likelihood						-42710.152
Deviance						85420.304
n parameters						14
BIC						85702.42
AIC						85476.3
Cragg-Uhler R2						0.493

The inclusion of the predictors improves the model fit compared to the model presented in step one. For instance, the R^2 statistic increases from .470 to .493. This is a 4.9% increase, which is the largest percentage increase in R^2 that will be seen in any of the later model comparisons. The BIC fit statistic also improves with the inclusion of the predictors; for instance, the BIC decreases to 85702.42 from 89109.5. This is a 3.8% decrease in the BIC. These goodness of fit statistics will be used as the baseline comparison for competing models that include various interactions in the next two steps.

Exploring Interactions

The purpose of the next two steps is to evaluate if the model presented in step two improves with the inclusion of interactions. Predictors can interact with time, and predictors can interact with each other.

First, competing models that include predictor by time interactions will be evaluated. Second, predictor by predictor interactions will be introduced and assessed.

Predictor by time interactions (step 3).

The purpose of the third step is to evaluate if the model presented in step two improves with the inclusion of time by predictor interactions. A time by predictor interaction is important to explore because it allows the effect of a predictor to vary by time. Based on conclusions from the exploratory analysis, time interactions will be evaluated for three (gender, residency, and Pell recipient status) of the six predictor variables. Table 4.13 displays the model fit statistics for five competing models compared to the model presented in step two. Model 3.1 includes the interaction between time and residency along with the remaining five predictors (gender, first generation status, Pell recipient status, minority status, and index). Model 3.2 includes the interaction between gender and time along with the remaining five predictors (first generation status, Pell recipient status, minority status, residency, and index). Model 3.3 includes the Pell recipient by time interaction along with the remaining predictors (first generation status, gender, minority status, residency, and index). Model 3.4 includes all of the interactions in the first three competing models along with the remaining three predictors (index, first generation status, and minority status). Finally, model 3.5 includes the gender by time as well as residency by time interactions along with the remaining four predictors (first generation status, Pell recipient status, minority status, and index).

Table 4.13

Model Fit Statistics for Competing Models With Different Time by Predictor Interactions

	Step 2	Model 3.1	Model 3.2	Model 3.3	Model 3.4	Model 3.5
Log Likelihood	-42710.15	-42641.50	-42619.25	-42683.22	-42527.90	-42554.56
Deviance	85420.30	85282.99	85238.50	85366.44	85055.81	85109.12
n parameters	14	21	21	21	35	28
BIC	85702.42	85645.71	85601.21	85739.23	85690.56	85602.82
AIC	85476.30	85354.99	85310.50	85440.44	85181.81	85207.12
Cragg-Uhler R2	0.493	0.494	0.494	0.493	0.496	0.496

Results from table 4.13 show that model 3.5 has the strongest goodness of fit statistics. The R^2 statistics are largest for models 3.4 and 3.5. Model 3.5 has nearly the smallest BIC (model 3.2 is slightly smaller), which is 1.2% lower than the baseline BIC. The decision to create model 3.5 was based on the results from models 3.1, 3.2, and 3.3, since these simpler models indicated that the Pell recipient status by time interaction produced least desirable goodness of fit statistics compared to the statistics produced by models that include gender by time or residency by time. For instance, the time by Pell recipient interaction model (model 3.3) has the least desirable goodness of fit statistics, since its BIC is larger than the baseline model by a small amount and its R^2 statistic is the smallest of all the competing models. Conclusions from step three indicate that the final descriptive model should include a gender by time as well as a residency by time interaction.

Predictor by predictor interactions (step 4).

Step four introduces two predictor by predictor (gender by minority status and first generation status by Pell Grant recipient status) interactions that the exploratory analysis indicated as possibly having a differential impact on graduation and dropout hazards. Table 4.14 displays the model fit statistics for four competing models all compared to the baseline model produced in step two. Model 4.1 builds on the baseline model by including the first generation and Pell Grant recipient status interaction. Three dummy variables are entered into the model (Pell and first generation, non-Pell and first generation, Pell and non-first generation all compared to non-Pell and non-first generation) rather than two dummy variables for Pell and first generation status in addition to the remaining four predictor variables (gender, residency, minority status, and index). Model 4.2 also builds on the baseline model by adding one predictor by predictor interaction. This model has three dummy variables for the gender by minority interaction (male and minority, female and minority, male and non-minority all compared to female and non-minority) rather than the two dummy variables for gender and minority status in addition to the remaining four predictor variables (first generation status, residency, Pell recipient status, and index). Model 4.3 includes both predictor by predictor interactions along with the remaining two predictors (index and residency).

Finally, model 4.4 integrates the first generation and Pell grant recipient status interaction into a model that also includes the recommendations from step three. This model includes three sets of interactions (residency by time, gender by time, and first generation and Pell Grant recipient status) along with the remaining two predictors (index and minority status).

Table 4.14

Model Fit Statistics for Competing Models With Different Predictor by Predictor Interactions

	Step 2	Model 4.1	Model 4.2	Model 4.3	Model 4.4
Log Likelihood	-42710.15	-42705.25	-42708.25	-42703.35	-42549.75
Deviance	85420.30	85410.49	85416.50	85406.69	85099.49
n parameters	14	15	15	16	29
BIC	85702.42	85712.75	85718.76	85729.10	85613.34
AIC	85476.30	85470.49	85476.50	85470.69	85201.49
Cragg-Uhler R2	0.493	0.493	0.493	0.493	0.496

Results from table 4.14 indicate that model 4.4 has very similar goodness of fit statistics compared to the best model from step three (model 3.5). For instance, the R^2 in model 4.4 is the same as the R^2 in model 3.5. The BIC statistic in model 4.4 is slightly larger than the BIC statistic in model 3.5; however, this difference only results in a .01% increase in BIC and is probably a meaningless difference. Interestingly, the inclusion of predictor by predictor interactions without the residency by time along with gender by time interactions did not improve the model fit statistics compared to the baseline model. The model fit statistics also indicate that the Pell by first generation status interaction improves the model fit more than the gender by minority interaction. This statement is based on model 4.1's BIC being the lowest (although essentially the same as 4.2's BIC) compared to model 4.2 and 4.3. The inclusion of both predictor by predictor interactions produces a worse fit compared to introducing either interaction independently. Therefore, the descriptive model used for this study should include three sets of interactions (time by residency, time by gender, and first generation by Pell Grant recipient) as well as the remaining predictors (index and minority status).

The purpose of exploring the goodness of fit statistics for various predictor by time and predictor by predictor interactions is to help inform the descriptive model that is described next. The predictors that are evaluated for possible interactions were chosen based on conclusions from the exploratory analysis in this first section of this chapter. Theory and empirical research indicate that all six predictors are influential for student success; thus, substantively, all of the interactions have influence on students' hazard of dropout and graduation. This section assessed the goodness of fit statistics to help inform which of these interactions will be included in the descriptive model used to answer the research question posed in this paper. The final section of this chapter reviews the statistical model that can be used to describe the dropout and graduation trajectories of CSU freshmen based on various student characteristics.

A Descriptive Model of the Longitudinal Nature of Dropout and Graduation

The purpose of this final section of chapter four is to discuss and graphically display this study's model that describes the longitudinal nature of dropout and graduation for new freshmen at CSU. The predictors included in the model were selected based on theory and previous empirical research. The observed dropout and graduation hazards as well as cumulative graduation proportions were evaluated across each predictor variable in the first section of this chapter. Conclusions from the exploratory section provided insight on which interactions should be statistically evaluated for inclusion. The second section of this chapter, selecting a descriptive model, evaluated various time specifications along with interactions. Conclusions from this second section indicated that the final model should include an interaction between first generation status and Pell Grant recipient status, residency by time, as well as an interaction between gender and time. This final section will begin with discussing the results from the model that are displayed below in table 4.15.

Table 4.15

Odds Ratios for the Descriptive Model

	Graduated			Dropout		
	Odds	95% CIs		Odds	95% CIs	
	Ratio	*LL*	*UL*	Ratio	*LL*	*UL*
Year 1	0.00	0.00	0.00	3.71	2.82	4.86
Year 2	0.00	0.00	0.00	3.11	2.37	4.10
Year 3	0.00	0.00	0.00	1.32	0.99	1.76
Year 4	0.03	0.02	0.04	1.29	0.96	1.74
Year 5	0.06	0.05	0.08	1.78	1.30	2.45
Year 6	0.06	0.05	0.09	3.63	2.46	5.35

Year 7	0.04	0.03	0.06	0.81	0.29	2.27
Year 8	0.07	0.04	0.14	0.00	0.00	>999
Pell & FG	0.87	0.78	0.96	1.33	1.21	1.45
Non-Pell & FG	1.03	0.97	1.10	1.45	1.37	1.54
Pell & Non-FG	0.79	0.72	0.87	1.13	1.02	1.24
Minority	0.86	0.80	0.93	1.07	1.00	1.15
CCHE Index	1.03	1.03	1.03	0.97	0.97	0.97
Non-Resident (year 1)	0.00	0.00	>999	1.59	1.47	1.72
Non-Resident (year 2)	0.63	0.08	5.25	1.23	1.11	1.37
Non-Resident (year 3)	0.78	0.58	1.05	1.04	0.88	1.24
Non-Resident (year 4)	0.97	0.90	1.06	0.90	0.70	1.16
Non-Resident (year 5)	0.73	0.65	0.81	0.57	0.40	0.81
Non-Resident (year 6)	0.54	0.43	0.68	0.33	0.18	0.60
Non-Resident (year 7)	0.67	0.43	1.05	0.00	0.00	>999
Non-Resident (year 8)	0.25	0.11	0.55	0.48	0.00	>999
Female (year 1)	0.00	0.00	>999	1.10	1.02	1.18
Female (year 2)	4.49	0.54	37.33	0.95	0.87	1.04
Female (year 3)	2.08	1.63	2.67	0.94	0.82	1.08
Female (year 4)	2.36	2.21	2.51	1.09	0.90	1.32
Female (year 5)	1.35	1.23	1.49	1.41	1.10	1.82
Female (year 6)	0.93	0.75	1.16	0.86	0.53	1.38
Female (year 7)	1.33	0.87	2.04	0.00	0.00	>999
Female (year 8)	0.64	0.28	1.50	0.89	0.00	>999

Table 4.15 displays the regression coefficients exponentiated as odds ratios and the 95% confidence intervals around those odds ratios. The life tables showed that there was very low event occurrence for graduation in the early years and very low event occurrence for dropout in the later years. A result of the low event occurrence and also dwindling risk sets in later years are large amounts of uncertainty in the parameter coefficients for years one through three for the graduation outcome and in years six through eight for the dropout outcome. Interpretations will focus on the years with more precise parameter estimates. The discussion of the results is organized by predictor variable.

Pell Recipient and First Generation Status

The exploratory analysis indicated that the gap in hazard across first generation status varied by whether or not the individual was a Pell Grant recipient. To represent the interaction, three dummy variables were entered into the model with non-Pell and non-first generation being the reference group. The Pell and first generation students have an odds ratio of .87 for graduation, which is interpreted as having 13% lower odds of graduation compared to students who are not first generation or Pell recipients.

Students who are not Pell recipients but are first generation have very similar odds of graduation compared to students who are neither. Students who are Pell recipients but are not first generation have 21% lower odds of graduating compared to the reference group after controlling for the other demographic and academic preparation variables in the model. First generation status has a negative association with graduation if the student is also a Pell recipient, but does not appear to be negatively associated with graduation if the student is first generation but not a Pell recipient.

Results indicate that first generation status has a different relationship with the hazard of dropout than it did with the hazard of graduation. For instance, a student who is first generation and a Pell recipient has 33% higher odds of dropping out compared to a student who has neither characteristic after controlling for the other predictors in the model. A student who is a not a Pell recipient but is first generation has 45% higher odds of dropout compared to the reference group. A student who is a Pell recipient, but is not first generation still has higher odds of dropout compared to a student who is not first generation or a Pell recipient; however, these higher odds are a smaller magnitude (13%) than the previous two comparisons. First generation status and Pell grant recipient status has a positive association with dropout; however, the magnitude of the association for Pell recipients is dependent on whether or not the student is first generation.

Minority Status

As shown in figure 4.13 and figure 4.14, minority students had an observed graduation hazard that is lower than the graduation hazard of non-minority students and a dropout hazard that is higher than the dropout hazard of non-minority students. This relationship was maintained in the descriptive model after controlling for the other predictors. For instance, minority students have 14% lower odds of graduating compared to non-minority students after controlling for the demographic and academic preparation variables also included in the model. Minority students have 7% higher odds of dropping out compared to non-minority students after controlling for other predictors in the model. Since a minority status and time interaction was not included in the model, the difference in odds between groups is forced to stay the same at each academic year.

Index

Theory and previous empirical work suggest that the better the academic preparation, the greater the likelihood of student success. This relationship was also evident in the exploratory analysis (figures 4.25 and 4.26), which showed that the observed graduation hazard is higher for students with a high index and the observed dropout hazard is lower for students with a high index. This relationship is maintained when considering the fitted hazards that can be obtained from the descriptive model's coefficients. For instance, with every one point increase in index score, a student has 3% higher odds of graduation, and for every one point increase in index score, a student has 3% lower odds of dropout after controlling for demographic variables.

Residency

Residency is a time varying variable that also has a time varying effect, since a residency by time interaction is included in the descriptive model. When graduation is the outcome in each time period, the coefficient for residency is negative, indicating that non-residents have lower odds of graduating compared to residents after controlling for the other variables in the model. However, once the uncertainty around the point estimates of the coefficients is accounted, for non-residents have lower odds of graduation in years five, six, and eight. In years one through four and in year seven, non-residents students have similar odds of graduating compared to resident students. The magnitude of the difference in odds is larger in year six when non-residents have 46% lower odds of graduating compared to residents and in year eight, when non-residents have 75% lower odds of graduating. Interestingly, figure 4.23 shows that non-residents have an observed graduation hazard that is slightly higher than the observed graduation hazard of residents. This possible contradiction is explained by the difference between fitted and observed hazards. Figure 4.23 is not accounting the other characteristics that non-residents have that also influence graduation, such as socioeconomic status and academic preparation. The results presented in table 4.14 account for all of the predictors in the model; therefore, all else equal, non-residents have lower odds of graduating compared to residents in years five, six, and eight.

The magnitude and direction of the association between residency and student dropout varies considerably by academic year. For instance, in years one and two, non-residents have considerably higher odds of dropping out compared to resident students after controlling for the other variables in the model. A non-resident has 59% higher odds of dropping out compared to a resident student in year one and 23% higher odds in year two. In years three and four, the odds of dropping out for residents and non-residents are very similar. In years five and six, the relationship flips because in year five, non-residents have 43% lower odds of dropping out compared to residents after controlling for the other predictors in the model and 66% lower odds in year six. This switch in the direction of the association is not surprising, since figure 4.22 showed that the observed hazard of dropout was higher for non-residents until year four.

Gender

A gender time by time interaction is included in the model so that the association between gender and the hazards of dropout and graduation can vary by each academic year. Females have considerably higher odds of graduation in years three, four, and five. For instance, in year three, female students have 108% higher odds of graduation compared to males, and in year four, the magnitude of this association is even stronger, since female students have 136% higher odds of graduation compared to males after controlling for the other predictors in the model. By year five, the magnitude of this association decreases to 35% higher odds of graduation and by years six, seven, and eight. Interestingly, female students also have higher odds of dropout in year one and in year five. In the first year, female students have 10% higher odds of dropping out compared to male students and 41% higher odds of dropping out in year five. The odds of dropping out in years two through four and in year six are very similar for males and females.

Graphical Representation of the Descriptive Model

A benefit of the multinomial logistic regression is that coefficients are easily interpreted as odds ratios; however, a weakness of evaluating the model's results solely in terms of odds ratios is that there is little indication of actual difference in the fitted hazards for students with different sets of characteristics. Figure 4.28 graphs the fitted graduation hazard and figure 4.29 graphs the fitted dropout hazard for two different student prototypes: a low risk and high risk student. The characteristics for low and high risk students are based on the results from table 4.15. A low risk student has the following characteristics: not first generation or a Pell recipient, non-minority, has an index score one half of a standard deviation higher than the mean (118), resident, and female. This student will have a dropout hazard that is lower than the comparison (high risk) student and a graduation hazard that is higher than the comparison student. A high risk student has the following characteristics: first generation and a Pell recipient, minority, has an index score one half of a standard deviation lower than the mean (108), non-resident, and male. This student will have a dropout hazard that is higher and a graduation hazard that is lower than the comparison (low risk) student.

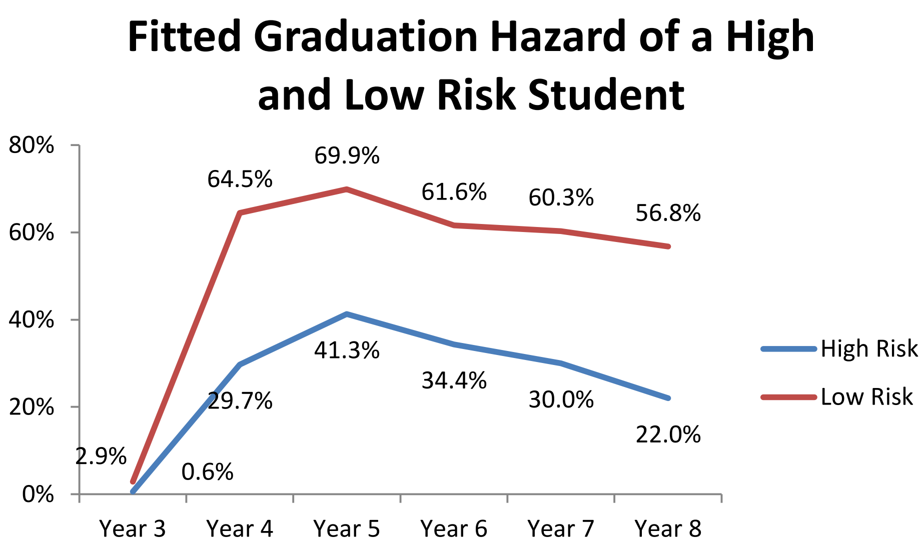

Figure 4.28

Figure 4.28 displays the graduation hazard probabilities obtained from the model. The fitted predictions can be interpreted as the probability of graduating in a given year given that the student hasn't already graduated. For instance, a low risk student has a 70% likelihood of graduating in year five if she has not already graduated, while a high risk student has only a 41% likelihood of graduating if he has not already graduated. The student characteristics that are all negatively associated with success (high risk) or positively associated with success (low risk) were purposefully grouped together so that these prototypes represent predictions at the extremes. The fitted hazard predictions for any other combination of demographic characteristics will fall somewhere in between the high and low risk hazards. The largest gap is at year four where there is nearly a 35 percentage point difference in the conditional probabilities. In year five, the hazard of graduating increases for both groups to its highest probability and the gap closes to just below 29 percentage points. The gap stays the same for year six and then starts to widen again through year eight to 35 percentage points.

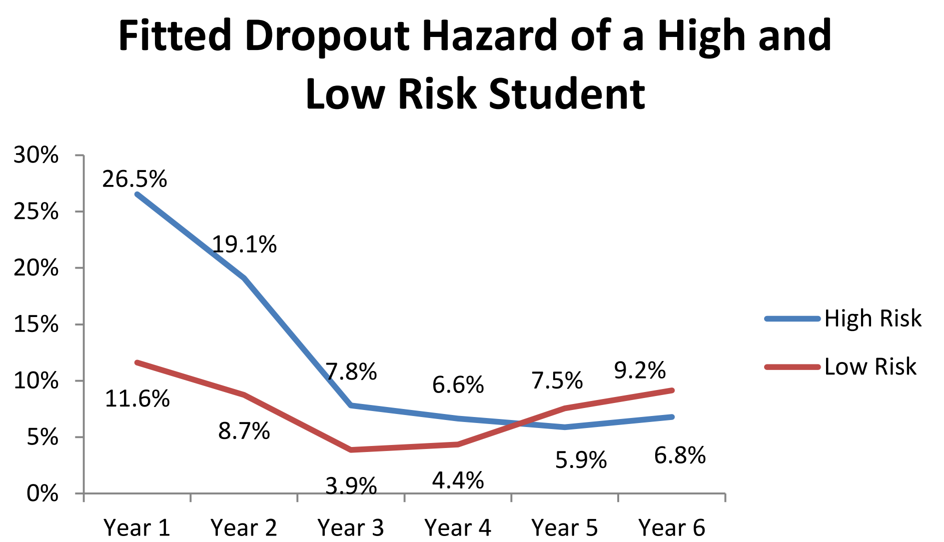

Figure 4.29

Figure 4.29 displays the fitted dropout hazards for the same two low and high risk prototypes. The likelihood of dropout after the first year is considerably higher for the high risk (26.5%) prototype compared to the low risk (11.6%) prototype. From year one to year two, the conditional probability of dropout decreases for both prototypes from year one to year two by a substantial amount (25% for low risk and 28% for high risk). The hazards decrease significantly (55% and 59%) again from year two to year three. After year three, the hazard then begins to increase for the low risk students and continues to decrease (slowly) for the high risk students so that the rates flip and the low risk prototype actually has a higher hazard in year five and six.

The purpose of this chapter is to develop a model that describes the longitudinal nature of dropout and graduation of new freshmen cohorts who started in the fall of 2002 through the fall of 2007. First, exploratory analysis was done to identify which predictors possibly interacted with each other and time. Next, competing models were statistically evaluated to inform the time specification and inclusion of possible interactions. Finally, a descriptive model was discussed in terms of the coefficients' odds ratios and the model's predicted probabilities. The final chapter of this paper will discuss implications of this descriptive model.

CHAPTER 5: IMPLICATIONS AND CONCLUSIONS

This chapter reviews some implications of the model that describes the longitudinal nature of dropout and enrollment and concludes with suggestions for future work.

Implications

The implications of the exploratory analysis, modeling building process, and descriptive model are discussed in this section. First, conclusions from the model fitting process are discussed in terms of how they relate to institutional graduation rates. Second, from a methodological perspective the analytical critic of main effects bias in empirical work is discussed and evaluated in light of the three step modeling building processed utilized in this paper.

Importance of Exploratory Analysis of Observed Hazards

The purpose of the exploratory analysis in this study was mainly to inform the descriptive modeling process, since the research question guiding this study was to describe the longitudinal nature of enrollment and dropout of students at CSU using a competing risk event history model. However, graphically displaying the observed graduation and dropout hazards along with graduation proportions can be useful without regard to modeling. For instance, institutional retention and graduation reports tend to report graduation and retention rates across various demographics by cohort. The figures displayed in the first section of chapter four display the observed variation in graduation and dropout for multiple cohorts of students. These types of graphs might be a positive addition to the current retention and graduation tables currently used to show rates for individual cohorts across demographic variables.

Interpretations of Model Coefficients

The results from table 4.15 are in accordance with what would be expected based on institutional retention studies and previous empirical work regarding the relationships between student characteristics and student dropout or graduation. First generation status, Pell Grant recipient status, and minority status are positively associated with dropout and negatively associated with graduation. Results from this study indicate that the first generation status and Pell Grant recipient status is complicated because the association between first generation status and student outcomes is dependent on whether or not the student is a Pell grant recipient (and vice versa). Although, including this interaction is not common in empirical work, there are reports that indicate the importance of focusing policy efforts on students who are both low-income and first generation to identify populations that are at the greater risk of non-completion (Engle & Tinto, 2008; Thayer, 2000; Wei & Horn, 2009).

Results also show that the associations between residency and student outcomes along with gender and student outcomes are dependent on the academic year of study. Institutional bivariate reports of graduation rates indicate that non-residents have a lower overall graduation compared to residents. A benefit of event history analysis is that this technique allows exploration of the timing of that association. As the discussion of results in chapter four showed, the association between residency status and student success varies based on the students' academic year. Perhaps strategies such as front loading institutional financial aid might be helpful for non-resident students, since these students are much more likely to drop out in years one and two compared to their later years. Similarly, institutional retention studies also show that females have higher first year retention rates, higher four year graduation rates, and similar six year graduation rates compared to male students. Again, the benefit of event history analysis is that association between gender and student outcomes can be discussed at each year of undergraduate education so that the variation in associations is easily identified. Based on results from chapter four, it might be more difficult for male students to complete their degrees in five years or less; therefore, male students might be in need of more support to accomplish early completion of their degrees compared to their female peers.

Implications of Fitted Graphs

There are important implications regarding graduation and dropout trajectories that can be made from the fitted graphs presented in chapter four. Figure 4.28 shows the extremes of graduation trajectories for new freshmen at CSU. First, all students become less likely to graduate the longer they persist past year five, if they do not graduate by year five since the fitted graduation hazard peaks at year five.

Second, the rate that the conditional hazard decreases after year five is faster for students with high risk characteristics. For instance, a high risk student's conditional probability of graduation decreases 47% from year five to year eight, while a low risk student's conditional probability of graduation decreases 19% from year five to year eight. The implications of these results indicate that it is important to encourage as many students as possible to graduate by year five. Each year a student remains enrolled after year five decreases their probability of graduating at all. This is even more important for students who have demographic characteristics such as minority status, first generation status, or a Pell Grant recipient. Students with these characteristics are even less likely to graduate if they remained enrolled past year five compared to students who do not come from these socioeconomic backgrounds. Results from the exploratory analysis also indicated that students with these characteristics had later peaks in their observed graduation hazards; therefore, it will be even more challenging to get the higher risk students to graduate within five years, since they tend to be graduating later already. However, these students are the ones who will benefit the most from graduating within five years, making the effort even more necessary.

Figure 4.29 displays the fitted dropout hazard. Interestingly, this graph shows that low risk students actually have higher hazards of dropout in later years. However, the increased risk of dropout in later years for low risk students is less detrimental to institutional graduation rates, since the majority of these students have graduated by year five; therefore, the risk set this conditional probability is being applied to is a relatively small risk set. What is detrimental to graduation rates is the greatly increased hazard of dropout predicted for the high risk prototype. For instance, the predicted probability for a high risk student of dropping out in year one is 129% higher than the predicted probability of dropout for a low risk student. Unlike the fitted graduation hazards, after year five, the fitted dropout trajectories show that high risk and low risk prototypes decrease their dropout hazard at a similar rate. However, due to starting at a much higher hazard, a high risk student has a substantially higher dropout hazard until year three. An important conclusion from the dropout hazard graph is that a larger proportion of high risk students are taken out of the risk set prior to experiencing the higher graduation hazard in year five, which makes the total number of high risk graduates considerably less. Institutional policy should focus on decreasing dropout for high risk students throughout years one through three because the conditional hazard of dropout remains higher for at-risk students throughout year three.

Exploring Interactions

A common critique of empirical work utilizing regression models in higher education research is the lack of attention to interactions (Dowd, 2008). Models often do not include interactions and there is almost never discussion surrounding the analytical decision to only include main effects in the models. The three step modeling process used in this study offers an excellent opportunity to evaluate how the coefficients' interpretation would be altered if interactions were excluded from the model without sufficient evaluations of their inclusion. This section reviews the differences in interpretations between the baseline model compared to the final descriptive model.

Pell and First Generation

If first generation and Pell Grant recipient interaction had not been included in the model, the conclusions would have been different as a result of main effects bias (Singer & Willet, 2003). For instance, when reviewing the coefficients produced by step two, the interpretation of the odds ratios for graduation would indicate that first generation students have similar odds of graduating compared to non-first generation students and that Pell recipients have about 18% lower odds of graduating compared to non-Pell recipients. Similarly, the interpretation of the odds ratios regarding dropout produced in step two would conclude that Pell recipients have similar odds of dropout compared to the odds of non-Pell recipients and that first generation students have 40% higher odds of dropout compared to non-first generation students. These interpretations would be incorrect because the change in odds associated with first generation status is dependent on whether or not the student is also a Pell Grant recipient.

Residency by Time

Similar to the Pell recipient and first generation status interaction, conclusions regarding the association between residency and the hazards of dropout and graduation differ substantially if the interaction is not included. In step two when residency is entered into the model without the time interaction, results indicated that non-residents have 14% lower odds of graduating compared to residents and 28% higher odds of dropping out. This conclusion is biased since the model with the interaction indicated that non-residents actually have a much larger magnitude in their reduced odds for graduation in years five, six and seven. Likewise, the conclusion of the higher odds of dropout is incorrect because when the effect of residency is allowed to vary by year, the direction of the association flips from positive to negative. This analysis indicates that if non-resident students are retained through their second year, they have lower odds of dropping out in later years compared to resident students after controlling for other variables related to demographics and academic preparation.

Gender by Time

The difference between the associations between gender and the graduation and dropout hazards in table 4.15 and 4.12 again suggest that the model developed in step two would be subject to main effects bias if possible interactions had not been explored and added to the final model. In this case, gender results from table 4.12 suggest that females have 86% higher odds of graduating compared to males when table 4.15 suggests that increased odds for females is only applicable in the earlier years. Similarly, results from table 4.12 suggest that the odds of dropout are very similar for male and female students; however, females have higher odds of dropping out in their first year and also in year five. This complexity in the graduation and dropout hazards by gender is an important consideration for institutional policy regarding student success initiatives. The female students who make it to year two are much more likely to graduate; however, if a female student does not graduate by year five, she is again at elevated risk of dropout.

There is not one correct model; models are just more and less useful (Berk, 2004). The purpose of this discussion is to emphasize that interpretations will differ with inclusion or exclusion of interactions and additional predictors. The inclusion of variables into a model must be informed by substantive knowledge and exploration of the observed data. The final descriptive model developed in this paper is most certainly not the one and only true model, but hopefully the model is useful for describing the trajectories of student enrollment and graduation based on students' characteristics and academic preparation.

Future Work and Conclusion

The theoretical premise of this paper is that students come to an institution with a set of characteristics, experiences, and preparation that influences the way they navigate the college environment. As an institution, it is our responsibility to craft an environment that influences students to behave in ways that will most positively impact their likelihood for success (Offenstein, Moore, & Shulock, 2010). The purpose of the descriptive model is to provide a model that adequately describes various enrollment trajectories so that the impact of different student behaviors or institutional policies can be evaluated with the additional inclusion of predictors that are related to the college experience. Of course, causality cannot be implied; however, using a longitudinal model should illuminate the timing, direction, and magnitude of associations between behaviors and outcomes such as graduation.

Recent policy discussions and analysis requests for Institutional Research at CSU have focused on the credit completion behavior of new freshmen in their first year and how that behavior is associated with eventual graduation. Cross-sectional logistic regression analysis indicates that there is a strong positive association between completing 30 or more credits in the first two semesters and six-year graduation compared to the association between completing a slightly lower credit load of 24 to 29 and six-year graduation. In light of the conclusions drawn from the fitted graduation hazards, these cross-sectional results make intuitive sense. If a student completes 30 credits their first year, they are on track to graduate in four years; therefore, they probably will not remain enrolled past year five and experience the decreased conditional probability of graduation after year five. However, this cross sectional analysis could be improved upon with an event history analysis. Does the timing of completing 15 credits a semester impact a student's likelihood for graduation? It is possible that completing 15 credits a semester is most important during a student's first two years and then lower credit loads (when they are completing more difficult upper level courses) have little or no impact on their hazard of graduation. The addition of longitudinal techniques for institutional policy analysis will increase the capacity of Institutional Research to answer questions about the timing of student behaviors and the association of this behavior and student success.

This study used a competing risk event history analysis to describe the longitudinal nature of student dropout and graduation at CSU. The premise of the model is to describe how the hazards vary based on student demographics so that institutional policy can be more informed regarding how to influence students to behave in ways that improve their likelihood for success. The model presented in this paper is intended to promote the use of longitudinal analytical methods to inform student success at CSU. Hopefully, this descriptive model will start conversations about future research questions that seek to understand when behaviors influence student success. As this model shows, different characteristics will influence the graduation and dropout trajectories and must be accounted for when using models to describe the associations between student behavior and student success.

REFERENCES

Advisory Committee on Student Financial Aid. (2010). The rising price of inequality: How inadequate grant aid limits college access and persistence *Report to congress and the secretary of education*. Washington DC.

Allison, P. (1984). *Event history analysis: Regression for longitudinal event data*. Beverly Hills, CA: Sage.

Allison, P. (2010). *Survival analysis using SAS: A practical guide*. Cary, NC: SAS Institute Inc.

Astin, A. (1993). *What matters in college? Four critical years revisited*. San Francisco: Jossey-Bass.

Aud, S., Hussar, w., Snyder, M., Bianco, T., Fox, K., Frohlich, L., Kemp, J., Drake, L. (2010). *The Condition of Education 2010*. Washington DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Bahr, P. (2009). Educational attainment as process: Using hierarchical discrete-time event history analysis to model rate of progress. *Research in Higher Education, 50*(7), 691-714. doi: 10.1007/s11162-009-9135-x

Baum, S., Ma, J., & Payea, K. (2010). Education pays 2010: The benefits of higher education for individuals and society. *Trends in Higher Education Series*. New York, New York: College Board Advocacy & Policy Center.

Bean, J. (1988). (untitled). *The Journal of Higher Education, 59*(6), 708-711.

Bean, J. (1982). Conceptual models of student attrition: How theory can help the institutional researcher. *New Directions for Institutional Research, 36*, 17-33.

Berk, R. A. (2004). *Regression analysis a constructive critique*. Thousand Oaks, Calif.: Sage Publications.

Bettinger, E. (2004). *How financial aid affects persistence.* National Bureau of Economic Research, Working Paper Series No. 10242.

Bound, J., & Turner, S. (2002). Going to war and going to college: Did World War II and the G.I. Bill increase educational attainment for returning veterans? *Journal of Labor Economics*, *20*(4), 784-815.

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

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

Carini, R., Kuh, G., & Klein, S. (2006). Student Engagement and Student Learning: Testing the Linkages. *Research in Higher Education, 47*(1), 1-32. doi: 10.1007/s11162-005-8150-9

Chen, R. (2008). Financial aid and student dropout in higher education: A heterogeneous research approach. *Higher Education: Handbook of Theory and Research,* 23, 209-239.

Chen, R., & DesJardins, S. (2008). Exploring the effects of financial aid on the gap in student dropout risks by income level. *Research in Higher Education, 49*(1), 1-18. doi: 10.1007/s11162-007-9060-9

Chen, X., & Carroll, D. (2005). First-generation students in postsecondary education: A look at their college transcripts. *Postsecondary Education Descriptive Analysis Report.* Washington DC: National Center for Education Statistics.

Chickering, A. W. (1993). *Education and identity*. San Francisco: Jossey-Bass Publishers.

Chun-Mei, Z., Kuh, G., & Carini, R. (2005). A comparison of international student and American student engagement in effective educational practices. *Journal of Higher Education*, *76*(2), 209-231.

Clark, M., & Cundiff, N. (2011). Assessing the effectiveness of a college freshman seminar using propensity score adjustments. *Research in Higher Education*, *52*(6), 616-639. doi: 10.1007/s11162-010-9208-x

Cook, B., & King, J. (2007). Status report on the Pell Grant program. Washington DC: American Council on Education Center for Policy Analysis.

Cragg, K. (2009). Influencing the probability for graduation at four-year institutions: A multi-model analysis. *Research in Higher Education, 50*(4), 394-413.

Dai, L. (2008). Degree attainment of undergraduate student borrowers in four-year institutions: A multilevel analysis. *Journal of Student Aid*, *37*(3).

DesJardins, S. (2003). Event history methods: Conceptual issues and an application to student departure from college. *Higher Education: Handbook of theory and research, 18*.

DesJardins, S., Ahlburg, D., & McCall, B. (1999). An event history model of student departure. *Economics of Education Review, 18*(3), 375-390.

DesJardins, S., Ahlburg, D., & McCall, B. (2002). A temporal investigation of factors related to timely degree completion. *The Journal of Higher Education, 73*(5), 555-581.

DesJardins, S., & Bell, A. (2006). Using economic concepts to inform enrollment management. *New Directions for Institutional Research, 2006,* 59-74. doi: 10.1002/ir.196

Desjardins, S., Kim, D., & Rzonca, C. (2003). A nested analysis of factors affecting bachelor's degree completion. *Journal of College Student Retention: Research, Theory and Practice, 4*(4), 407-435.

DesJardins, S., & McCall, B. (2010). Simulating the effects of financial aid packages on college student stopout, reenrollment spells, and graduation chances. *The Review of High Education, 33*(4), 213-541.

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

DesJardins, S., & Toutkoushian, R. (2005). Are students really rational? The development of rational thought and its application to student choice. *Higher Education: Handbook of theory and research, 20.*

Dowd, A. (2008). Dynamic interactions and intersubjectivity: Challenges to causal modeling in studies of college student debt. *Review of Educational Research, 78(2),* 232-259.

Doyle, W. (2006). Adoption of merit-based student grant programs: An event history analysis. *Educational Evaluation and Policy Analysis, 28*(3), 259-285.

Dynarski, S. (1999) *Does aid matter? Measuring the effect of student aid on college attendance and completion*. National Bureau of Economic Research, Working Paper Series No. 7422.

Dynarski, S. (2005). *Building the stock of college-educated labor*. National Bureau of Economic Research, Working Paper Series No. 11604.

Dynarski, S., & Scott-Clayton, J. (2008). *Complexity and targeting in federal student aid: A quantitative analysis.* National Bureau of Economic Research, Working Paper Series No. 13801.

Ehrenberg, R. (2004). Econometric studies of higher education. *Journal of Econometrics, 121,* 19-37.

Engle, J., & Tinto, V. (2008). Moving beyond access: College success for low-income, first-generation students. Washington DC: Pell Institute for the Study of Opportunity in Higher Education.

Frank, T. (September 15, 2011). *Fall address to the university*. Colorado State University. Fort Collins, Colorado. Retrieved from http://www.president.colostate.edu/pdf/2011-fall-address.pdf

Gordon, J., Ludlum, J., & Hoey, J. (2008). Validating NSSE against student outcomes: Are they related? *Research in Higher Education, 49*(1), 19-39. doi: 10.1007/s11162-007-9061-8

Hansen, W. L. (1983). Impact of student financial aid on access. *Proceedings of the Academy of Political Science, 35*(2), 84-96.

Hossler, D. (1984). Enrollment management. New York: College Entrance Examination Board.

Hu, S., & St. John, E. (2001). Student persistence in a public higher education system: Understanding racial and ethnic differences. *The Journal of Higher Education, 72*(3), 265-286.

Institutional Research. (2010). Freshman retention study 2010-11. Fort Collins, Colorado: Colorado State University.

Ishitani, T. (2003). A longitudinal approach to assessing attrition behavior among first-generation students: Time-varying effects of pre-college characteristics. *Research in Higher Education, 44*(4), 433-449. doi: 10.1023/a:1024284932709

Ishitani, T. (2006). Studying attrition and degree completion behavior among first-generation college students in the United States. *The Journal of Higher Education, 77*(5), 861-885.

Ishitani, T. (2008). How do transfers survive after “transfer shock”? A longitudinal study of transfer student departure at a four-year institution. *Research in Higher Education, 49*(5), 403-419. doi: 10.1007/s11162-008-9091

Ishitani, T., & Desjardins, S. (2002). A longitudinal investigation of dropout from college in the United States. *Journal of College Student Retention: Research, Theory and Practice, 4*(2), 173-201.

Lott, J., Gardner, S., & Powers, D. (2009). Doctoral student attrition in the STEM fields: An exploratory event history analysis. *Journal of College Student Retention: Research, Theory and Practice 11*(2), 247-266.

Johnson, I. (2006). Analysis of stopout behavior at a public research university: The multi-spell discrete-time approach. *Research in Higher Education, 47*(8), 905-934. doi: 10.1007/s11162-006-9020-9

Kane, T. (1994). College entry by blacks since 1970: The role of college costs, family background, and the returns to education. *The Journal of Political Economy, 102*(5), 878-911.

Kane, T. (2003). *A quasi-experiment of the impact of financial aid on college-going*. NBER Working Paper Series. National Bureau of Economic Research. Cambridge.

Kim, D. (2004). The effect of financial aid on students' college choice: Differences by racial groups. *Research in Higher Education, 45*(1), 43-70.

Kuh, G. (2009). The national survey of student engagement: Conceptual and empirical foundations. *New Directions for Institutional Research, 2009*(141), 5-20. doi: 10.1002/ir.283

Kuh, G., Cruce, T., Shoup, R., Kinzie, J., & Gonyea, R. M. (2007). *Unmasking the effects of student engagement on college grades and persistence*. Paper presented at the American Educational Research Association. Chicago, Illinois.

Kuh, G., Kinzie, J., Buckley, J.A., Bridges, B. K., & Hayek, J. C. (2006, June). What matters to student success: A review of the literature: National Postsecondary Education Cooperative (NPEC) Commissioned Paper.

Kuh, G., Kinzie, J., Cruce, T., Shoup, R., & Gonyea, R. M. (2007). Connecting the dots: Multi-faceted analyses of the relationships between student engagement results from the NSSE and the institutional practices and conditions that foster student success *Revised Final Report prepared for Lumina Foundation for Education*.

Lacy, M. G. (2006). An explained variation measure for ordinal response models with comparisons to other ordinal R^2 measures. *Sociological Methods & Research, 34*(4), 469-520. doi: 10.1177/0049124106286329

Lacy, M., & Long, M. (2007). *Binary response versus event history models of student graduation and attrition in higher education*. Paper presented at the Midwest Sociological Society, Chicago, Illinois.

Leslie, L. L., & Brinkman, P. T. (1988). *The economic value of higher education*. New York: American Council on Education.

Long, J. S. (2006). *Regression models for categorical dependent variables using Stata*. College Station, Texas: StataCorp LP.

Lott, J., Gardner, S., & Powers, D. (2009). Doctoral student attrition in the STEM fields: An exploratory event history analysis. *Journal of College Student Retention: Research, Theory and Practice 11*(2), 247-266.

Monks, J. (2009). The impact of merit-based financial aid on college enrollment: A field experiment. *Economics of Education Review, 28*(1), 99-106. doi: 10.1016/j.econedurev.2008.03.002

Moore, C., & Shulock, N. (2009). Student progress toward degree completion: Lessons from the research literature. Sacramento, California: The Institute for Higher Education Leadership & Policy.

Most, D. (2008). Patterns of doctoral student degree completion: A longitudinal analysis. *Journal of College Student Retention 10*(2).

Murnane, R., & Willett, J. (2011). *Methods matter: Improving causal inference in educational and social science research*. Oxford: Oxford.

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

Offenstein, J., Moore, C., & Shulock, N. (2010). Advancing by degrees: A framework for increasing college completion. Washington DC: The Education Trust.

Pascarella, E., & Terenzini, P. (2005). *How college affects students: A third decade of research*. San Francisco: Jossey-Bass.

Paulsen, M. & St. John E. (1997). The financial nexus between college choice and persistence. *New Directions for Institutional Research, 95*, 65-82.

Pike, G., Hansen, M., & Lin, C.-H. (2011). Using instrumental variables to account for selection effects in research on first-year programs. *Research in Higher Education*, *52*(2), 194-214. doi: 10.1007/s11162-010-9188-x

Porter, S., Rumann, C., & Pontius, J. (2011). The validity of student engagement survey questions: Can we accurately measure academic challenge? *New Directions for Institutional Research, 2011*(150), 87-98. doi: 10.1002/ir.391

Radcliffe, P., Huesman, R., & Kellogg, J. (2006). *Modeling the incidence and timing of student attrition: A survival analysis approach to retention analysis*. Paper presented at the Association for Institutional Research in the Upper Midwest, Bloomington, MN.

Ronco, S. (1996). How enrollment ends: Analyzing the correlates of student graduation, transfer and dropout with a competing risks model *AIR Professional File* (Vol. 61). Tallahassee, Florida.

Scott, M., & Kennedy, B. (2005). Pitfalls in pathways: Some perspectives on competing risks event history analysis in education research. *Journal of Educational and Behavioral Statistics*, *30*(4), 413-442. doi: 10.3102/10769986030004413

Seftor, N., & Turner, S. (2002). Back to school: Federal student aid policy and adult college enrollment. *The Journal of Human Resources*, *37*(2), 336-352.

Singer, J., & Willett, J. (1991). Modeling the days of our lives: Using survival analysis when designing and analyzing longitudinal studies of duration and the timing of events. *Psychological Bulletin, 110*(2), 268-290. doi: 10.1037/0033-2909.110.2.268

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

Singer, J., & Willett, J (2003). *Applied longitudinal data analysis modeling change and event occurrence*. Oxford: Oxford University Press.

St. John, E. (1992). Workable models for institutional research on the impact of student financial aid. *Journal of Student Financial Aid, 22*(3), 13-25.

St. John E. (2000). The impact of student aid on recruitment and retention: What the research indicates. *New Directions for Student Services*, *2*(89), 61-75.

St. John, E., Andrieu, S., Oescher, J., & Starkey, J. (1994). The influence of student aid on within-year persistence by traditional college-age students in four-year colleges. *Research in Higher Education, 35*(4), 455-480.

St. John, E., Hu, S., & Tuttle, T. (2000). Persistence by undergraduates in an urban public university: Understanding the effects of financial aid. *The Journal of Student Financial Aid, 30*(2), 23-37.

St. John, E., Hu, S., & Weber, J. (2000). Keeping public colleges affordable: A study of persistence in Indiana’s public colleges and universities. *The Journal of Student Financial Aid, 30*(1), 21-32.

St. John, E., Hu, S., & Weber, J. (2001). State policy and the affordability of public higher education: The influence of state grants on persistence in Indiana. *Research in Higher Education, 42*(4), 401-428.

St. John, E., Kirshstein, R., & Noell, J. (1991). The effects of student financial aid on persistence: A sequential analysis. *Review of Higher Education, 14*(3), 383-406.

St. John, E., Musoba, G., & Simmons, A. (2003). Keeping the promise: The impact of Indiana’s twenty-first century scholars’ program. *Review of Higher Education, 27*(1), 103-123.

StataCorp. (2010) Stata statistical software: Release 11. College Station, TX: StataCorp LP.

Stratton, L., O'Toole, D., & Wetzel, J. (2008). A multinomial logit model of college stopout and dropout behavior. *Economics of Education Review, 27*(3), 319-331. doi: 0.1016/j.econedurev.2007.04.003

Swail, W. (2002). Higher Education and the New Demographics: Questions for Policy. *Change, 34*(4), 14-23.

Thayer, P. (2000). Retention of students from first generation and low income backgrounds. Washington DC: Council for Opportunity in Education.

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

Tinto, V. (1993). *Leaving college rethinking the causes and cures of student attrition*. Chicago University of Chicago Press.

Tinto, V. (2004). Student retention and graduation: Facing the truth, living with the consequences *Occasional Paper*. Washington DC: Pell Institute for the Study of Opportunity in Higher Education.

Tinto, V. (2006). Research and practice of student retention: What's next? *Journal of College Student Retention: Research, Theory and Practice, 8*(1), 1-19. doi: 10.2190/4YNU-4TMB-22DJ-AN4W

USA.gov. (2011). The White House. Retrieved November 7, 2011, from http://www.whitehouse.gov/issues/education/higher-education

Wei, C., & Horn, L. (2009). A profile of successful Pell Grant recipients: Time to bachelor's degree and early graduate school enrollment *Statistical Analysis Report*. Washington DC: National Center for Education Statistics.

Wetzel, J., O'Toole, D., & Peterson, S. (1999). Factors affecting student retention probabilities: A case study. *Journal of Economics and Finance, 23*(1), 45-55.

Zhao, C., & Kuh, G. (2004). Adding Value: Learning communities and student engagement. *Research in Higher Education, 45*(2), 115-138.