Student Retention as a Model of Change

Kenneth L. Thompson, Ph.D.

March 30, 2017

MILLSAPS COLLEGE

Today's Topics

- Student Persistence
- Logistic Regression
- Path Analysis
- Latent Growth Modeling
- Event Hazard Analysis
- Next Steps

Student Persistence

Retention literature initially focused Tinto's model of student integration (1975, 1983)

- Family background
- Individual attributes
- Goal commitment
- Student engagement

As student populations have become more diverse, studies have expanded to include more factors

Student Persistence at Millsaps

First-Year to Second-Year Retention ~80%

First-Year to Fourth-Year Retention ~71%

4-Year Graduation Rate ~62%

Transfer to Another Institution ~33%

Questions

Millsaps Enrollment Management Committee –

Three questions about retention data:

1. Does is tell us whether students are leaving?
2. Does it tell us when students are leaving?
3. Does it tell us why students are leaving?

Luckily, the answer is yes!

Predictors of Persistence

To answer those questions, we traditionally include factors such as:

- Gender
- Ethnicity
- Major
- High School GPA
- ACT/SAT scores
- Student surveys
- Pell Grant recipient
- Current GPA

We use those factors to predict whether a student will be enrolled/not enrolled at a certain point, such as beginning of second year or beginning of fourth year.

Logistic Regression

Because enrolled/not enrolled at a certain point represents a dichotomous state, logistic regression is a popular tool for modeling retention.

Logistic regression attempts to predict the probability that a student belongs to the mutually exclusive enrolled and not enrolled groups.

Logistic regression makes no assumptions about predictors being normally distributed, linearly related, or equal variances within each group.

Logistic Regression

Popular approach for modeling retention.

Cons:

Cannot model predictors temporally

Cannot model residual variance

Path Analysis

Allows for considering causal order among predictors.

Three criteria for a causal relationship:

Temporal precedence

X precedes Y

Covariation of the cause and effect

If X then Y

In not X then not Y

No plausible alternative explanations

Both X and Y explained by C

Predictors of Persistence

- Gender typically not significant when other indicators present
- Ethnicity often found to be a significant predictor
- High school GPA/college GPA consistently a significant predictor
- SAT/ACT not typically significant when predicting one year retention, but SAT has been shown to be a significant predictor of long-term persistence
- Campus residency has been shown to be a significant predictor
- Financial aid has been shown to be a significant predictor

Persistence typically defined as enrolled at beginning of semester or on census day.

Path Analysis

Gender

Ethnicity

Major

HS GPA

Freshman GPA

ACT/SAT

Socio-Economic Status

Financial Aid

?

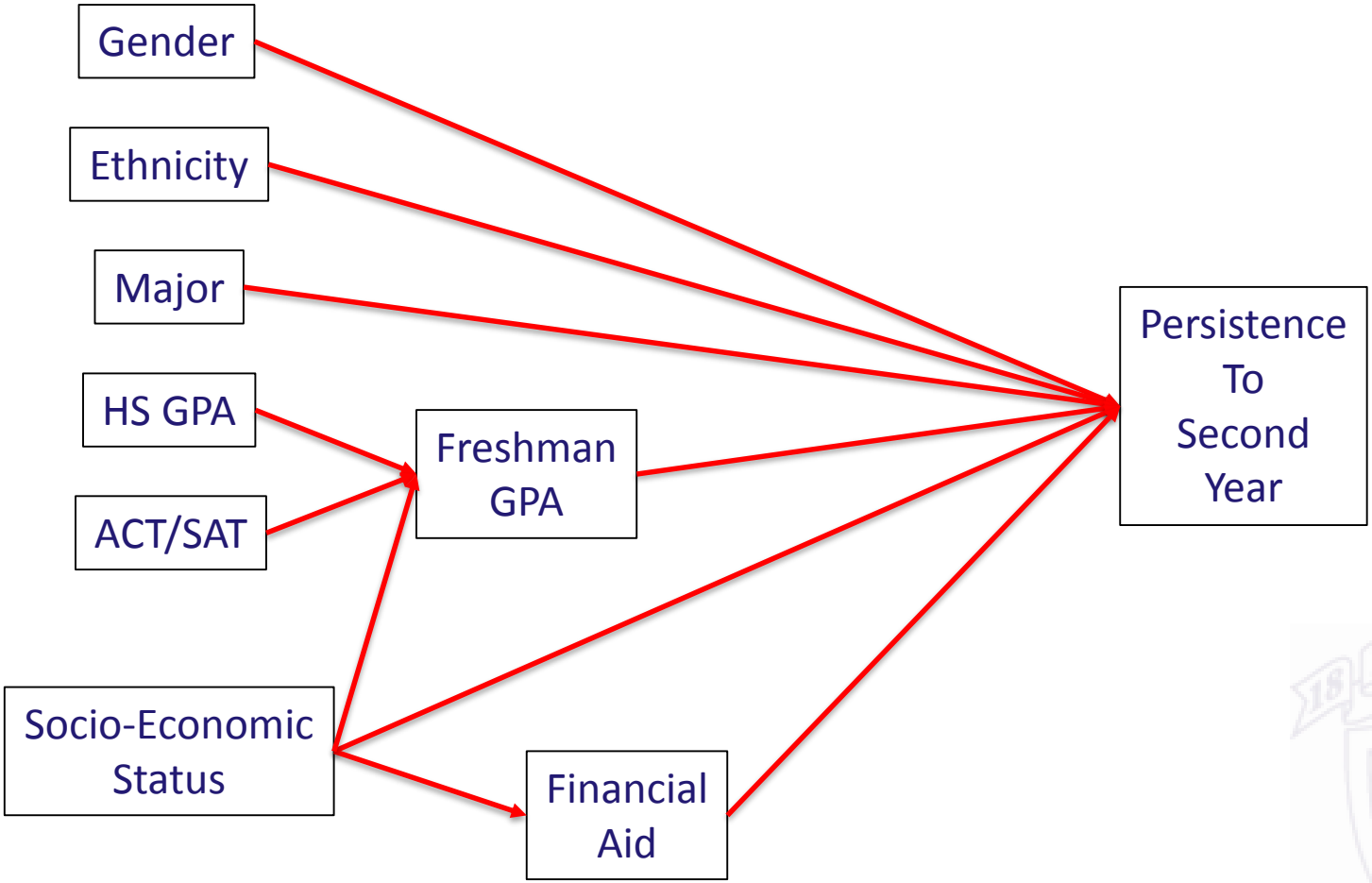
Persistence To Second Year

Path Analysis

Allows for testing temporal order to address causality

Allows measurement of direct and indirect effects of one variable on another

Path Analysis



Path Analysis

Allows for testing temporal order to address causality

Allows measurement of direct and indirect effects of one variable on another

Cons:

Recursive (unidirectional)
Cross-sectional data does not allow studying change

Why is Change Important?

Student persistence models assume that a student will withdraw from college when he perceives that the cost/benefit ratio has shifted in favor of something other than college (Lancaster, 1990, pp. 97-98).

When we study students' movement from being enrolled to not being enrolled, we are observing their movement from one state to another, but we're making inferences about their choices (Lancaster, 1990, p. 5).

That is, using observed phenomena to draw conclusions about students' internally optimized decisions (DesJardins & Moye, 2000).

Analyzing Longitudinal Data & Change

Latent Growth Modeling and Event History Analysis allow for modeling longitudinal data and change.

Singer & Willet (2003) –

Three model features required for modeling change over time:

1. Multiple waves of data
2. Meaningful metric of time
3. Outcome that changes systematically

Two types of questions addressed:

1. How does each person change over time?
2. What predicts difference among people in their change?

Analyzing Longitudinal Data & Change

Latent Growth Modeling and Event History Analysis allow for modeling longitudinal data and change.

Singer & Willet (2003) –

Three model features required for modeling change over time:

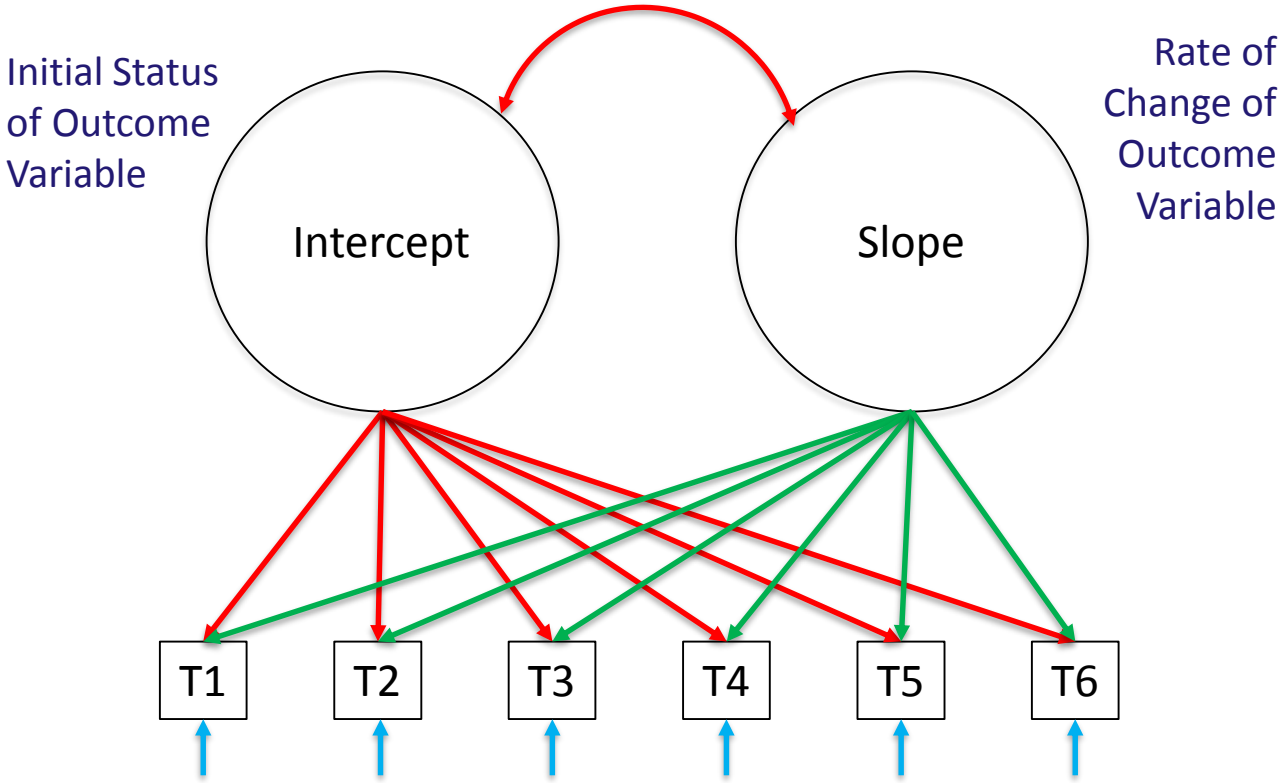
1. Multiple waves of data
2. Meaningful metric of time
3. Outcome that changes systematically

Two types of questions addressed:

1. How does each person change over time?
2. What predicts difference among people in their change?

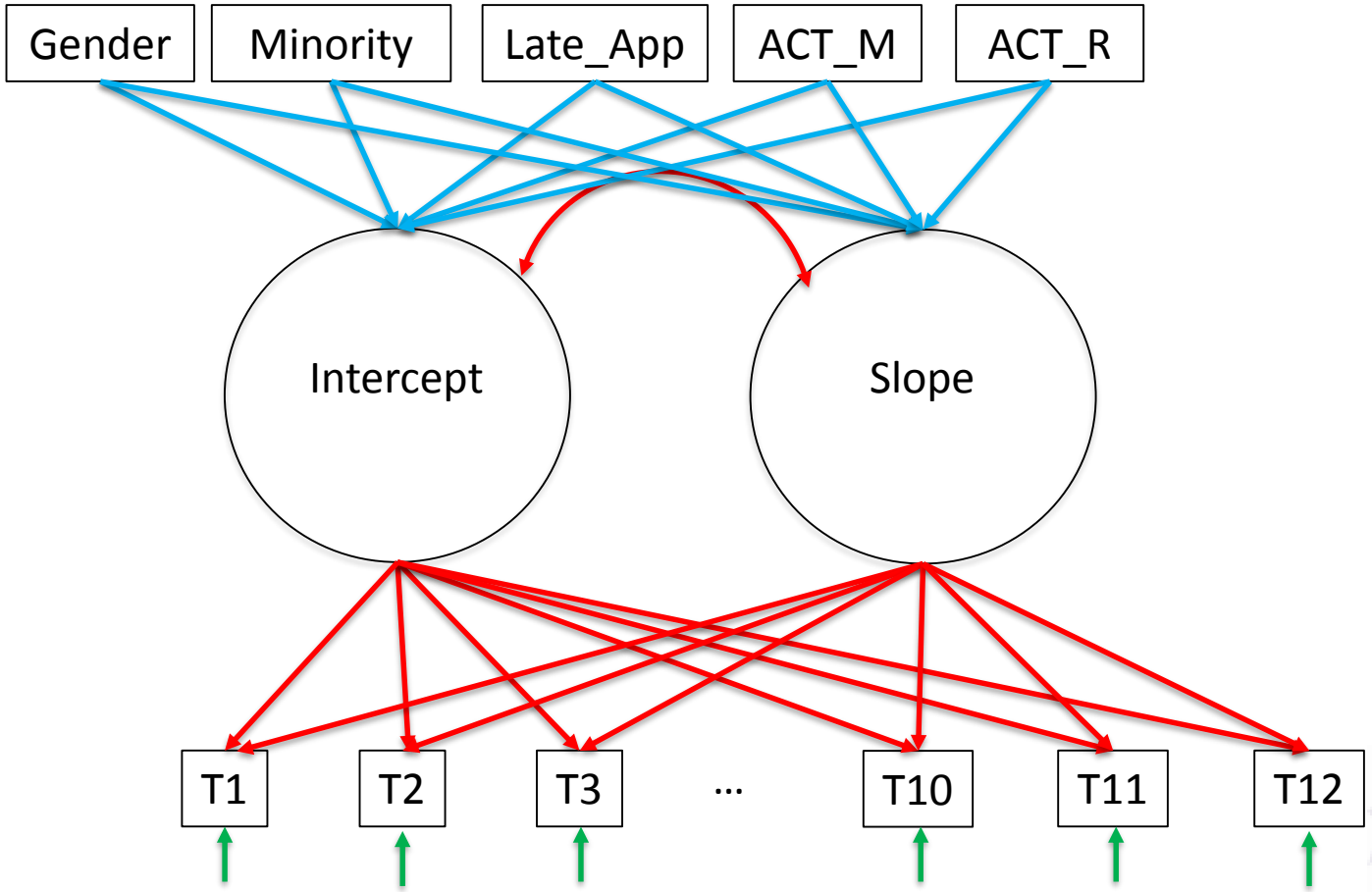
Latent Growth Modeling

Describes individual linear development trajectories
Capture individual differences in trajectories over time
Can describe change at the group level



Compares sample and predicted covariance matrices.

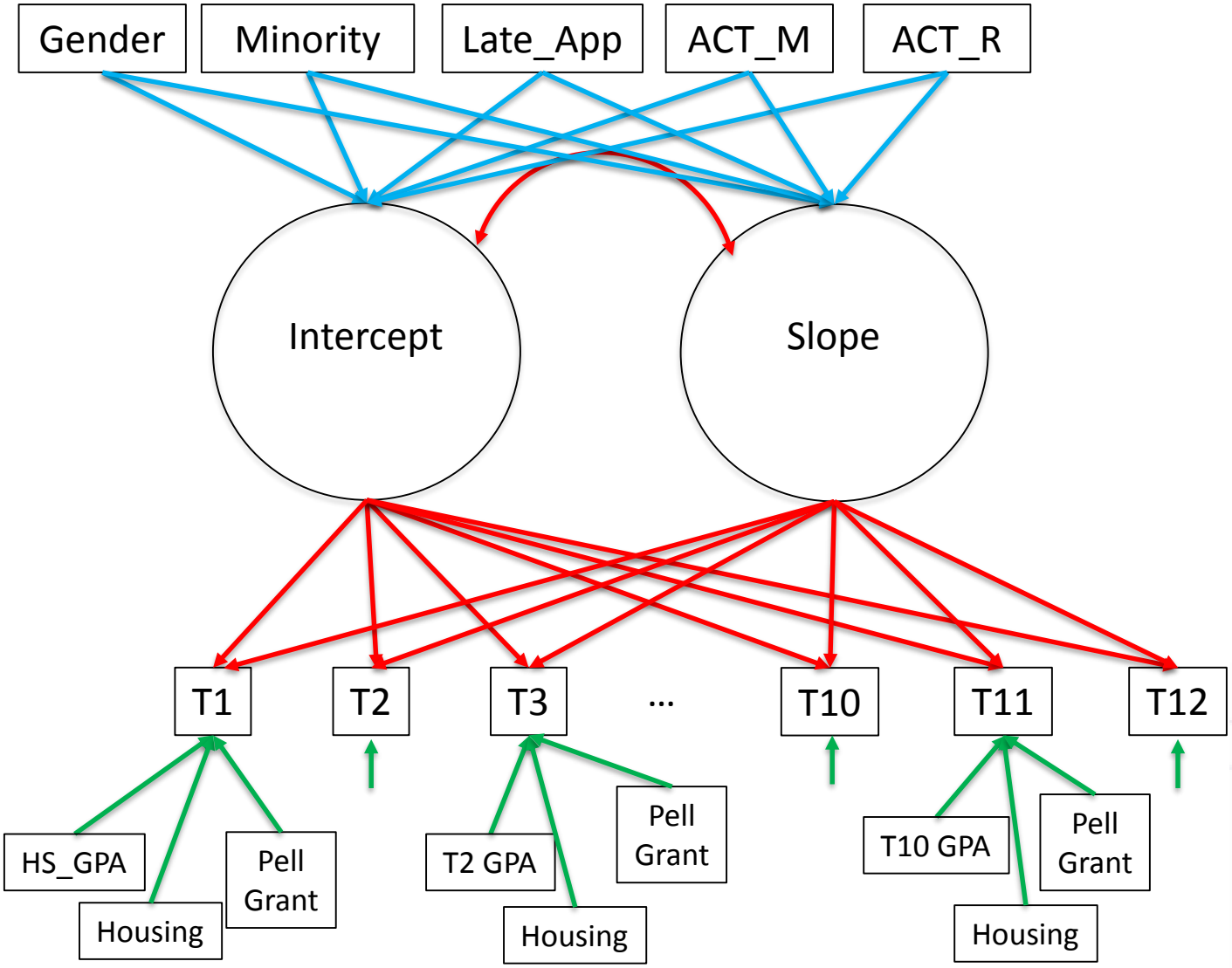
Latent Growth Modeling



What affects each time point?

Allows us to look at change over time.

Latent Growth Modeling



Latent Growth Model

Allows us to look at changes over time. Incorporates time-invariant predictors

- gender
- minority
- application date
- ACT scores

with time-varying predictors

- GPA
- housing assignment
- Pell grant recipient
- major declared

Relying on covariance matrices. Millsaps problem: too little variance.

Event History Analysis

Models transition from one state to another
Addresses factors that led to the transition
Addresses time to transition

Examples –
Recidivism: In Jail → [Out of Jail → In Jail]

Economics: Start a Business → Out of Business

Politics: Party in control → Party not in control

Health: HIV positive → AIDS symptoms

Millsaps: Enrolled → Not Enrolled

Event History Analysis

Three requirements for event history analysis:

1. Target event to study
2. Defined "beginning of time"
3. Meaningful time metric for measuring when (if) the event occurs

Event occurrence is defined as moving from state to another state.

Required redefining enrollment based on census day.

Initial enrollment based on enrollment on census day.

Subsequent enrollment based on credit hours earned.

Event History Analysis

Allows us to compute a "hazard function" – conditional probability that a student will experience the event (not persisting) in a given semester, given that the student didn't experience the event (not persisting) in a prior semester.

Each individual has a unique hazard function for each period that describes his/her risk for not enrolling.

Result in a model that will distinguish each student based on his/her hazard function and associated predictors.

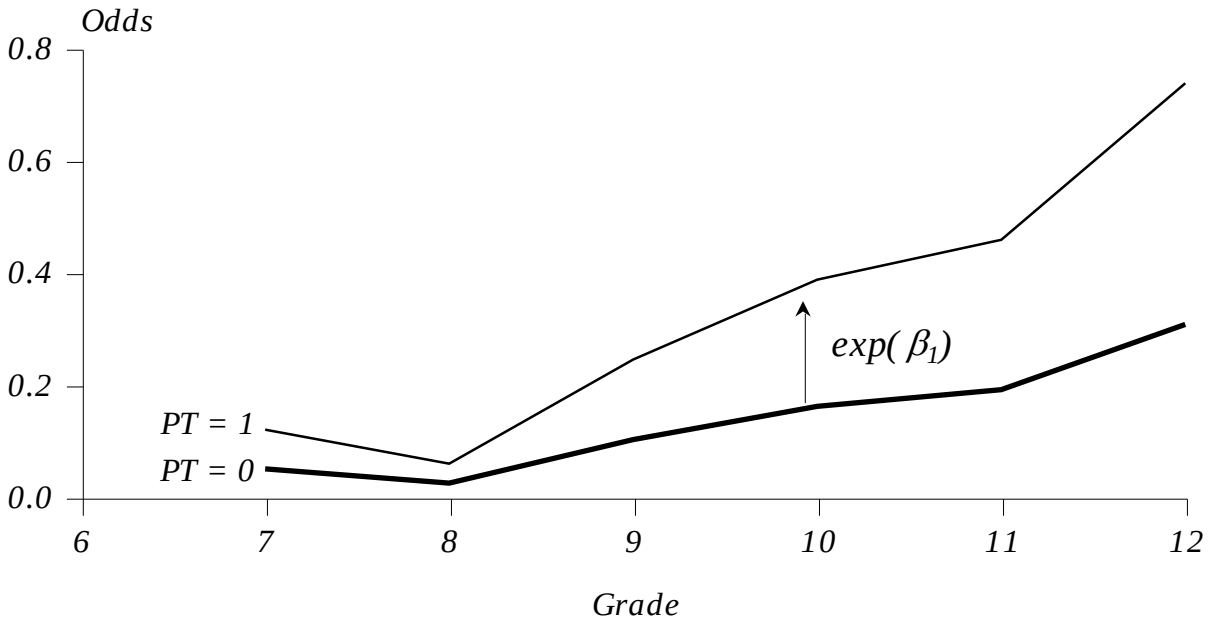
Event History Analysis

Allows us to compute a "hazard function" – conditional probability that a student will experience the event (not persisting) in a given semester, given that the student didn't experience the event (not persisting) in a prior semester.

Each individual has a unique hazard function for each period that describes his/her risk for not enrolling.

Result in a model that will distinguish each student based on his/her hazard function and associated predictors.

Hazard Functions



Study Phases

Phase I: Create Hazard Model using time-invariant and time varying predictors to model odds of not returning.

Phase II: Evaluate Hazard Model to determine whether predictors need to be modified (including predictors such as athlete first semester but not subsequent semesters).

Phase III: Expand Hazard Model to to include polytomous outcomes:

- Student graduated
- Student dropped out without Enrolling another institution
- Student transferred to another institution
- Student was a stop-out

Thank You!