

# USING PREDICTIVE MODELING TO INCREASE SIX-YEAR GRADUATION

Jorge Martinez

Director of Research & Reporting, Enrollment Services

Caroline Neary

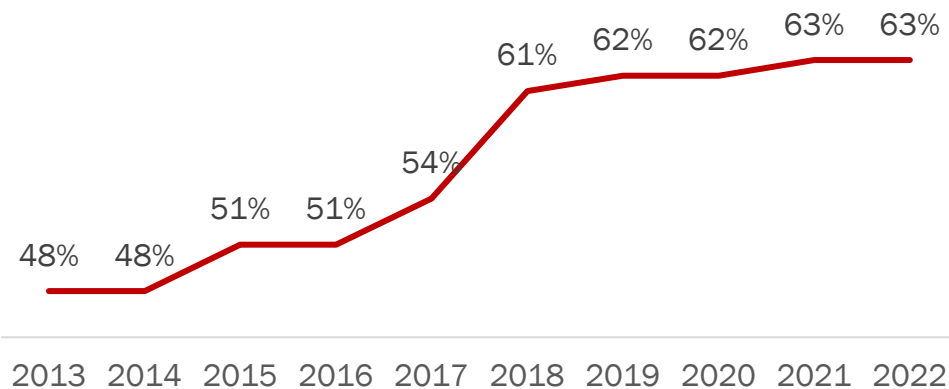
Senior Research Analyst, Undergraduate Student Success

# UH: INSTITUTIONAL OVERVIEW

- Large, 4-year, public, urban university
  - 37k undergraduate students
  - 72% attend full-time
  - 45% first-generation
  - 33% Hispanic
  - 40% receive Pell grant

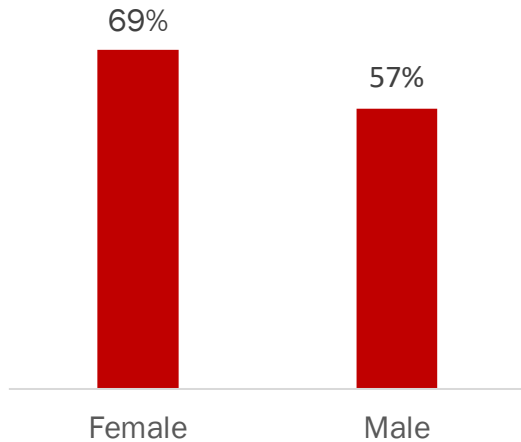
# BACKGROUND

## 6-Year Graduation Rate

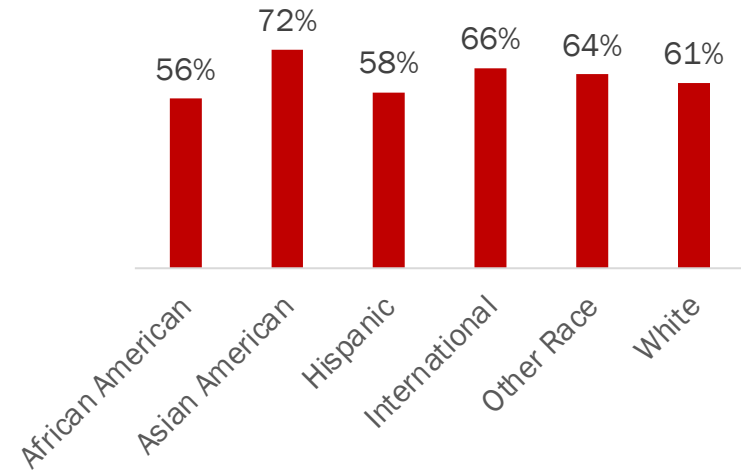


# 6-YEAR GRADUATION RATES: DEMOGRAPHIC

Gender

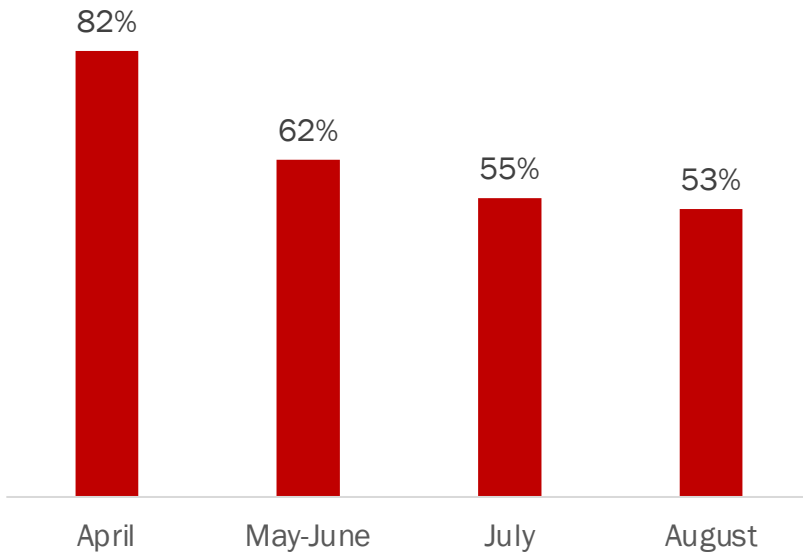


Race/Ethnicity

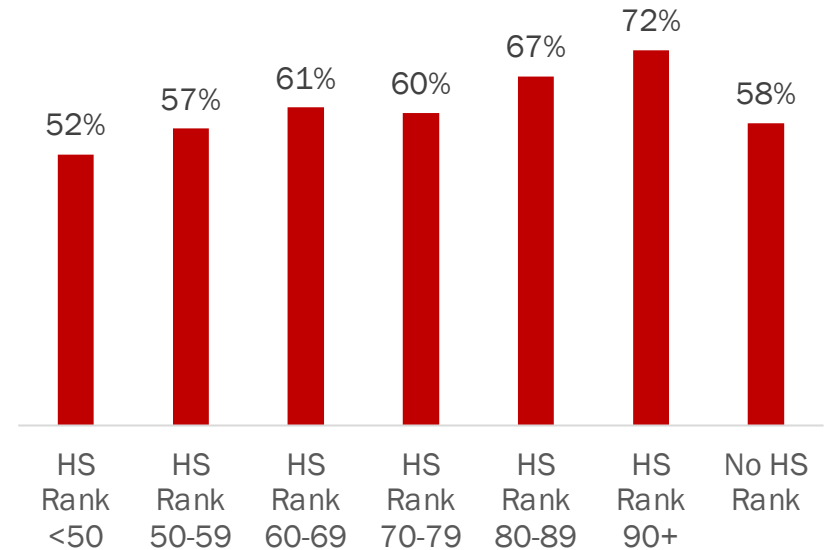


# 6-YEAR GRADUATION RATES: PRE-COLLEGE

Orientation Month

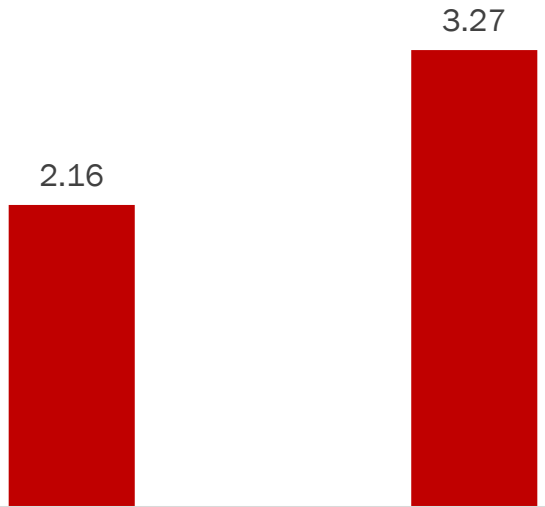


High School Rank



# 6-YEAR GRADUATION RATES: ACADEMIC

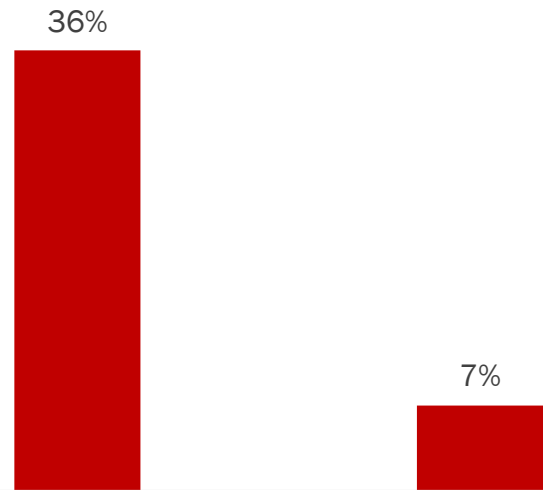
Average Cum. GPA



Did not Graduate

Graduated

Average DWF Ratio

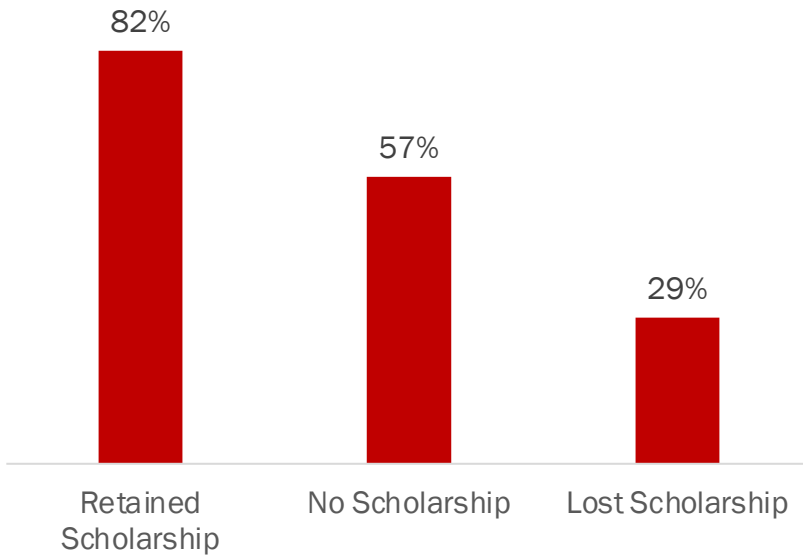


Did not Graduate

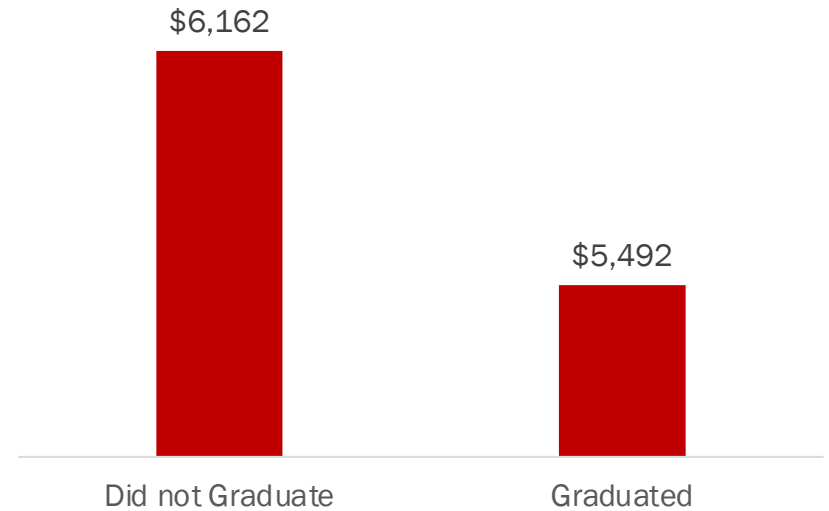
Graduated

# 6-YEAR GRADUATION RATES: FINANCIAL

## Scholarship



## Average Unmet Need



# PHASE 1: LOGISTIC REGRESSION MODEL

## Objective:

Utilize logistic regression analysis to identify relationships between student characteristics and six-year graduation.

## Population:

Fall 2012, Fall 2013, and Fall 2014 FTIC Cohorts  
(N=10,579)



# PREDICTORS

	Academic	Financial	Admissions	Demographics
Strong Predictors	DWF Grade Ratio Total Credits Passed	Lost Scholarship		
Moderate Predictors	Test/Transfer Credits Percent Full-Time Cumulative GPA Change of College	No Scholarship		Race/Ethnicity Residence County
Non-Significant Variables		Pell Eligibility Total Loans Unmet Financial Need	HS Class Rank Orientation Month SAT Score	Gender First Generation

# ACTIONABLE CONCLUSIONS

Compared to students from Harris County and its adjacent counties, students from other Texas counties were less likely to graduate in six years.

- ACTION: Support and outreach for these students (about 14% of FA20 cohort)

Students who lost or never had a merit scholarship were less likely to graduate in six years.

- ACTION: Expand first year academic scholarship opportunities, e.g., retention scholarship

Students enrolled full-time for a higher percentage of terms were more likely to graduate in six years.

- ACTION: Continue to encourage full-time enrollment, e.g., UHin4

Students with a higher ratio of D, W, and F grades to all grades were less likely to graduate in six years.

- ACTION: Expand support for students/instructors in high DWF rate courses, e.g., Gateways to Completion, LAUNCH

# PHASE II: SURVIVAL ANALYSIS

- Helps us answer questions like
  - How long can we expect patients to survive with certain medical conditions?

# PHASE II: SURVIVAL ANALYSIS

- Helps us answer questions like
  - How long can we expect **students to graduate** with **different characteristics** (gender, college, first generation status)?
  - What proportion of students are expected to graduate **by a specific academic year**?
  - What variables/factors/interventions are likely to **increase or decrease** time to graduation?

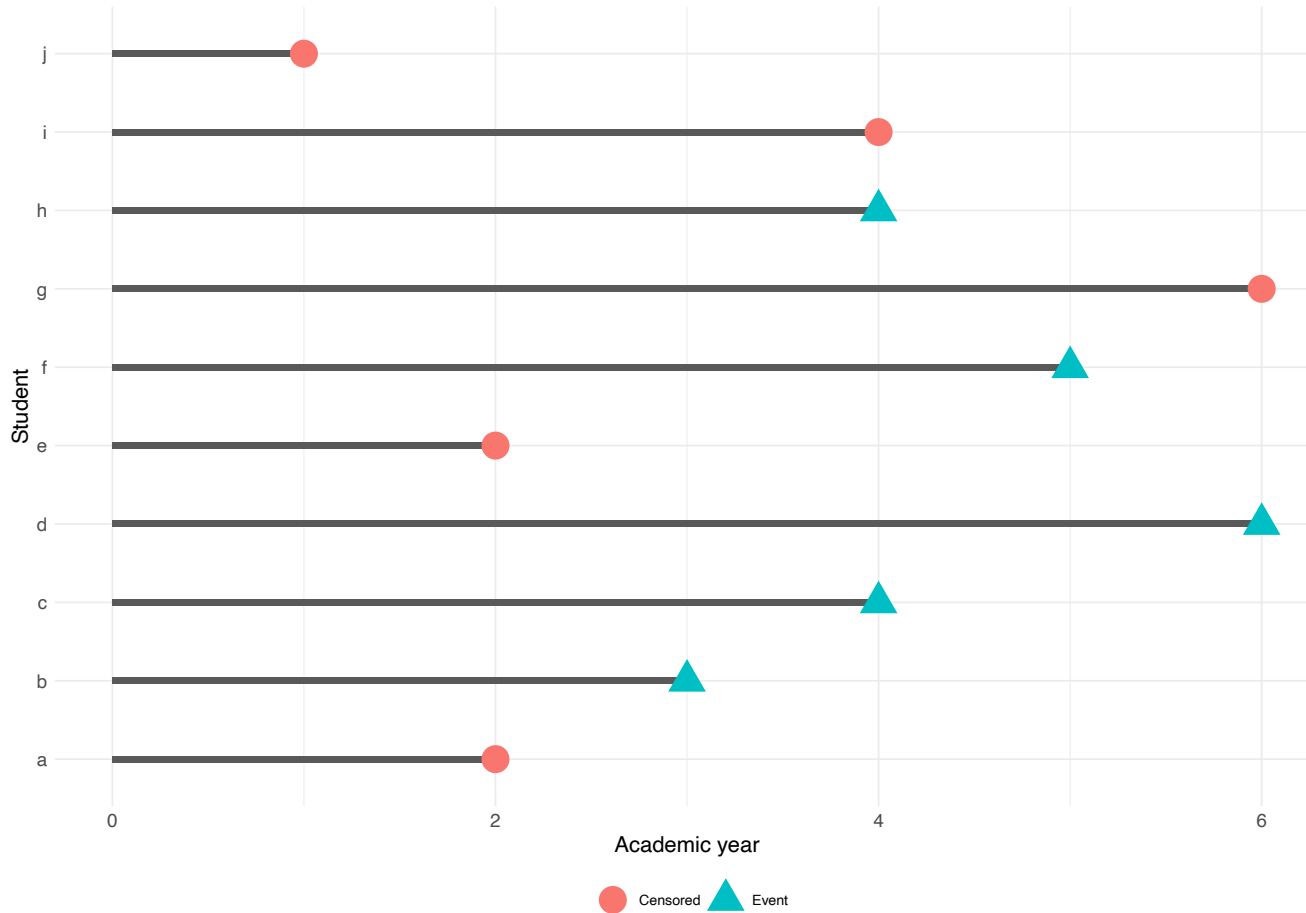
# PHASE II: SURVIVAL ANALYSIS

- From classification to **degree velocity**
  - Logistic regression (graduated Y/N)
  - Survival analysis (time-to-degree)
- Model time until an event occurs
  - Compare between groups
  - How event correlates with quantitative variables
- Also known as **Event History Analysis**

# CENSORING

- Censoring is a type of **missing data** problem
  - The **event never occurs** during the study window
  - Student **drops out** of the study for various reasons
  - You only know if the individual **survived up to the loss of follow-up**

# TIME-TO-EVENT



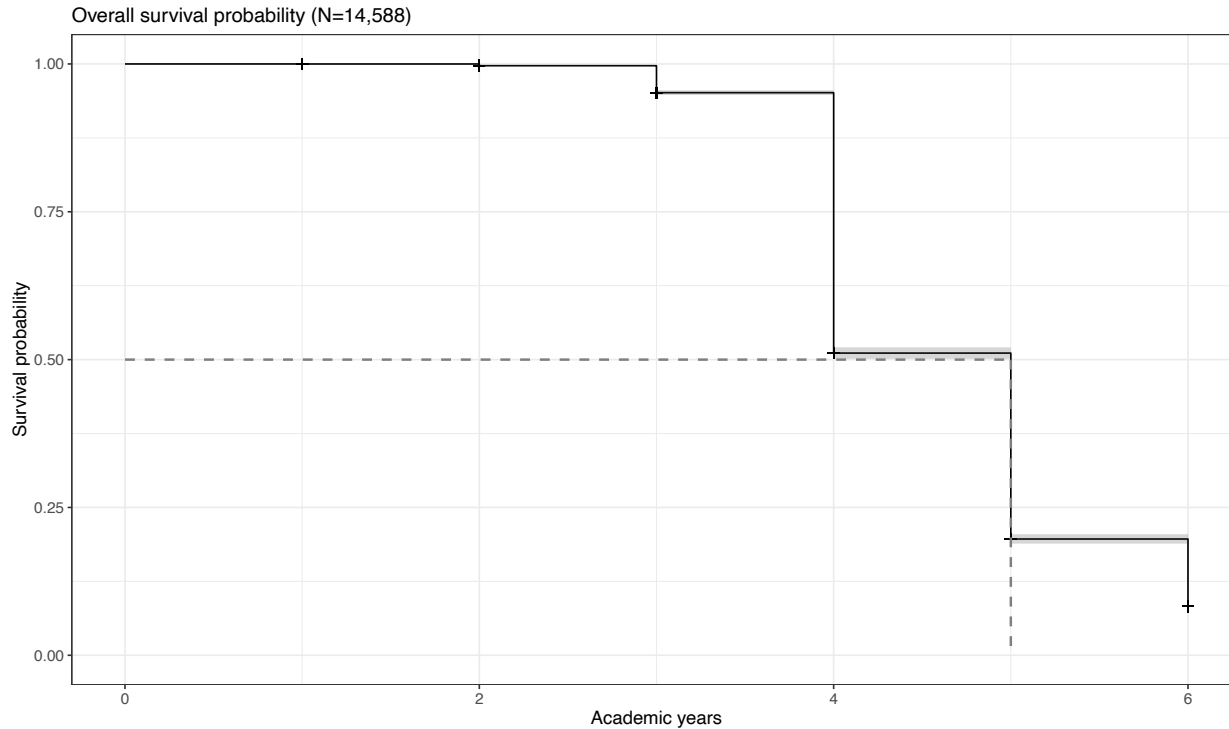
# SURVIVAL FUNCTION

- Survival function is the probability an individual survives up to and including time  $t$ .

Academic Years	# risk	# event	# censored	Survival probability	Std. err.	Upper	Lower
1	14,588	1	1797	0.9999	0.0001	1.0000	0.9998
2	12,790	37	1607	0.9970	0.0005	0.9980	0.9961
3	11,146	510	817	0.9514	0.0021	0.9554	0.9475
4	9,819	4545	457	0.5110	0.0096	0.5207	0.5015
5	4,817	2964	323	0.1966	0.0206	0.2047	0.1888
6	1,530	887	643	0.0826	0.0364	0.0887	0.0769



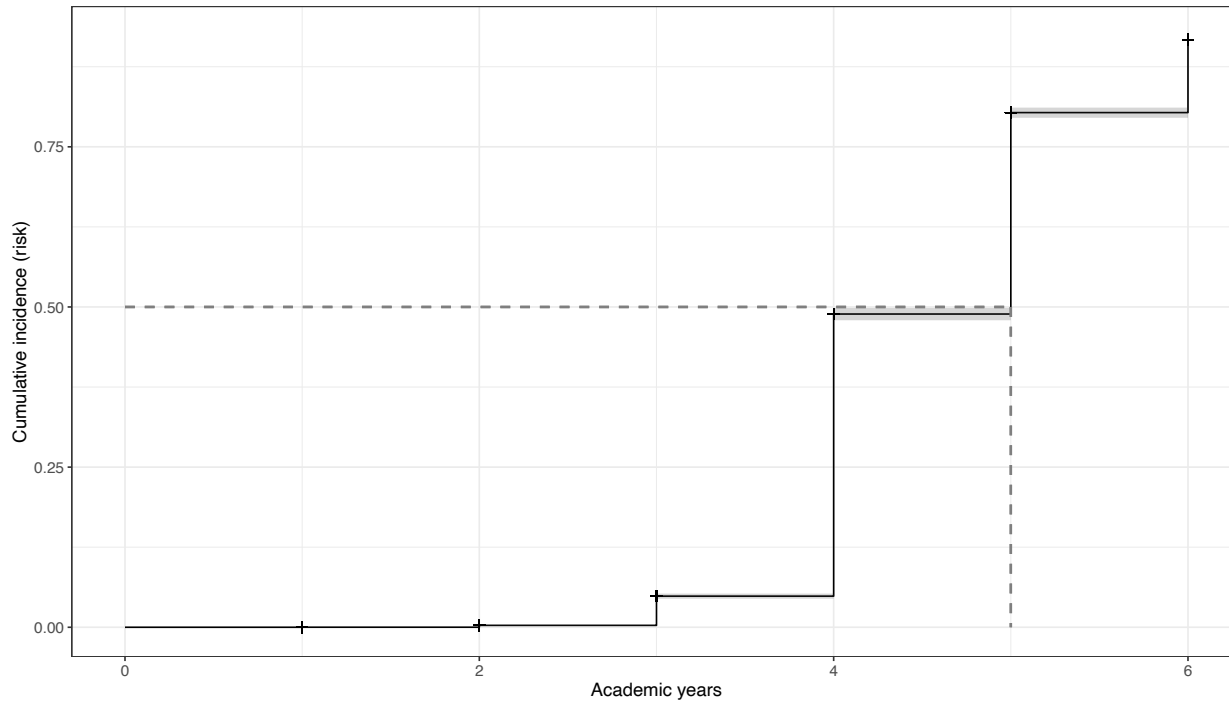
# KAPLAN-MEIER SURVIVAL CURVE



At Risk	14588	12790	9819	1530
Censored	0	3404	4678	5644
Events	0	38	5093	8944

# CUMULATIVE INCIDENCE

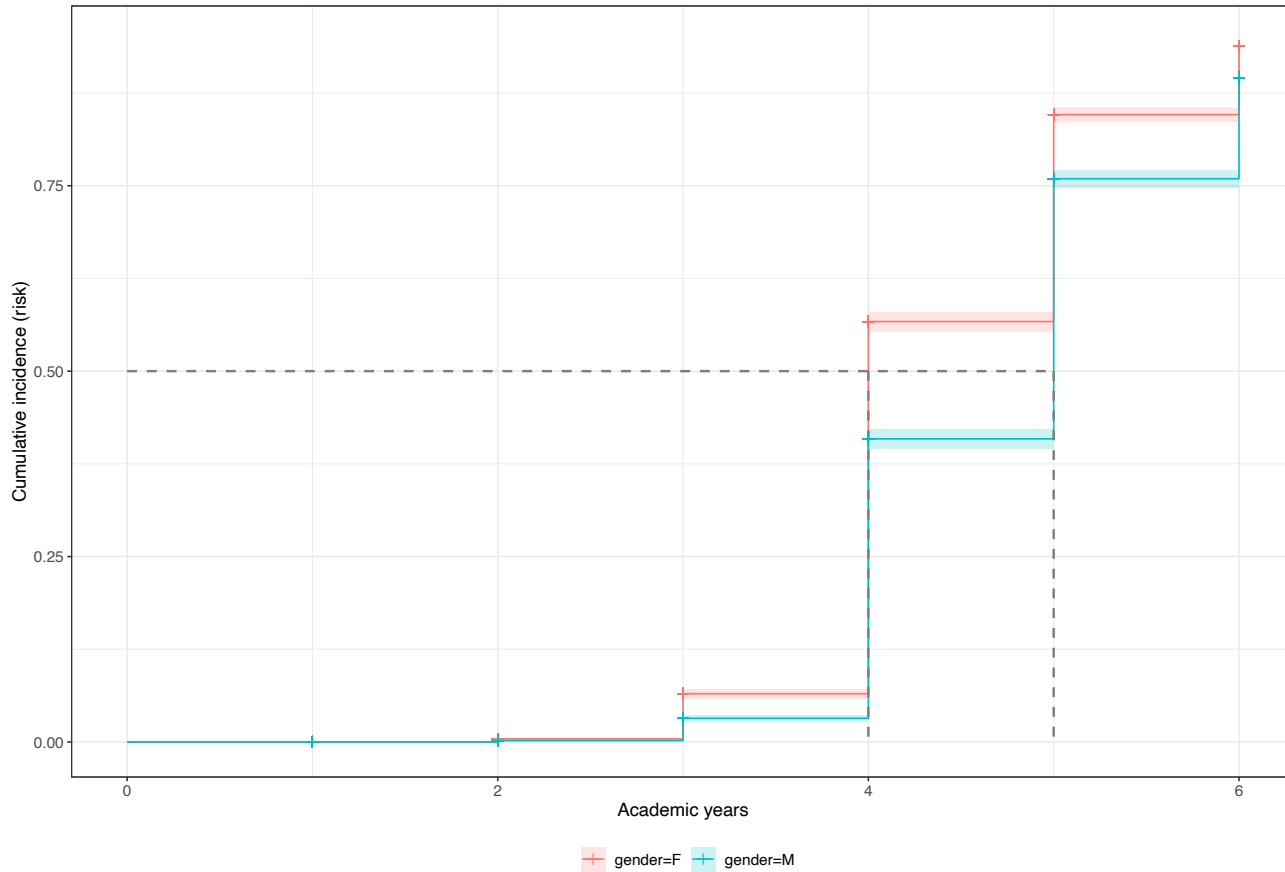
Cumulative incidence for graduation (N=14,588)



At Risk	14588	12790	9819	1530
Censored	0	3404	4678	5644
Events	0	38	5093	8944

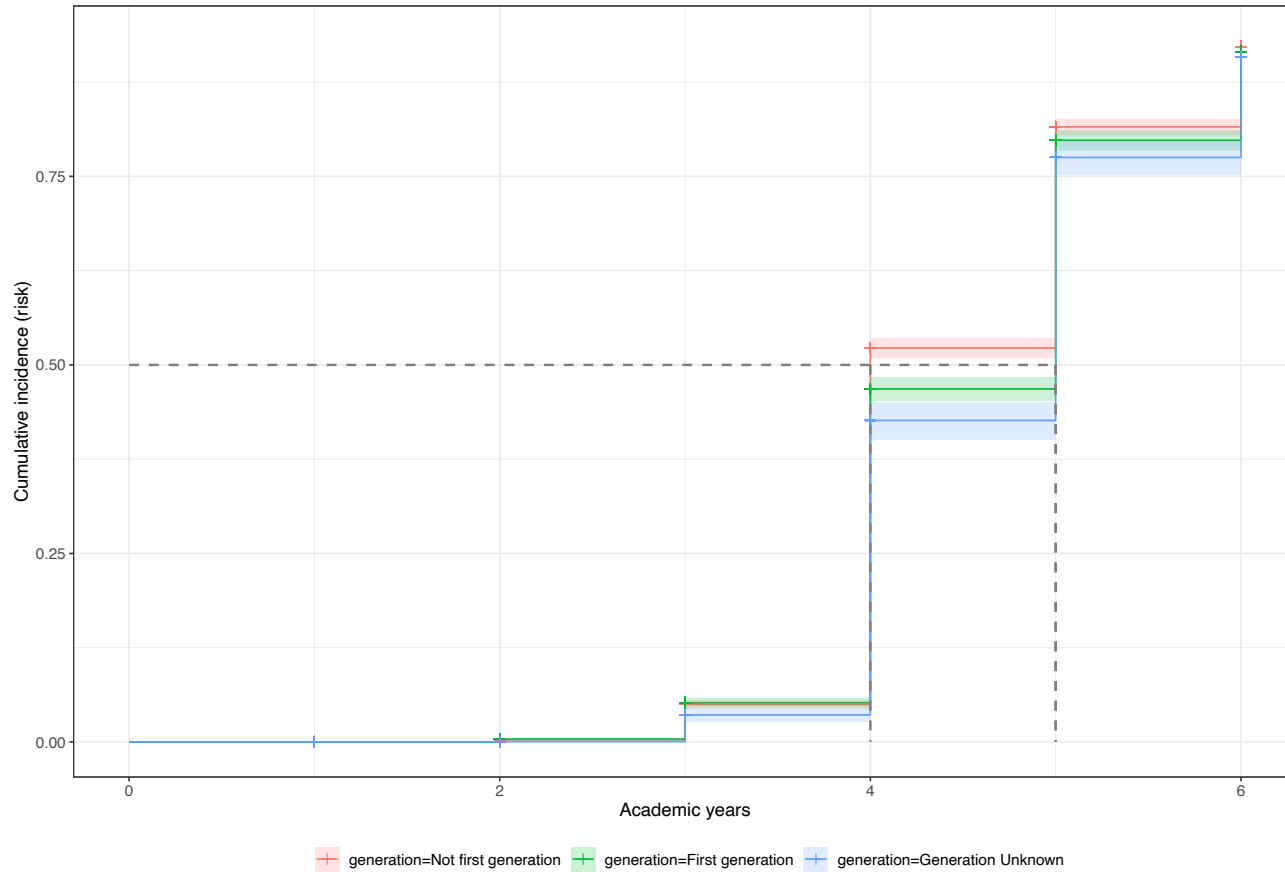
# GENDER

Cumulative incidence for graduation (N=14,588)



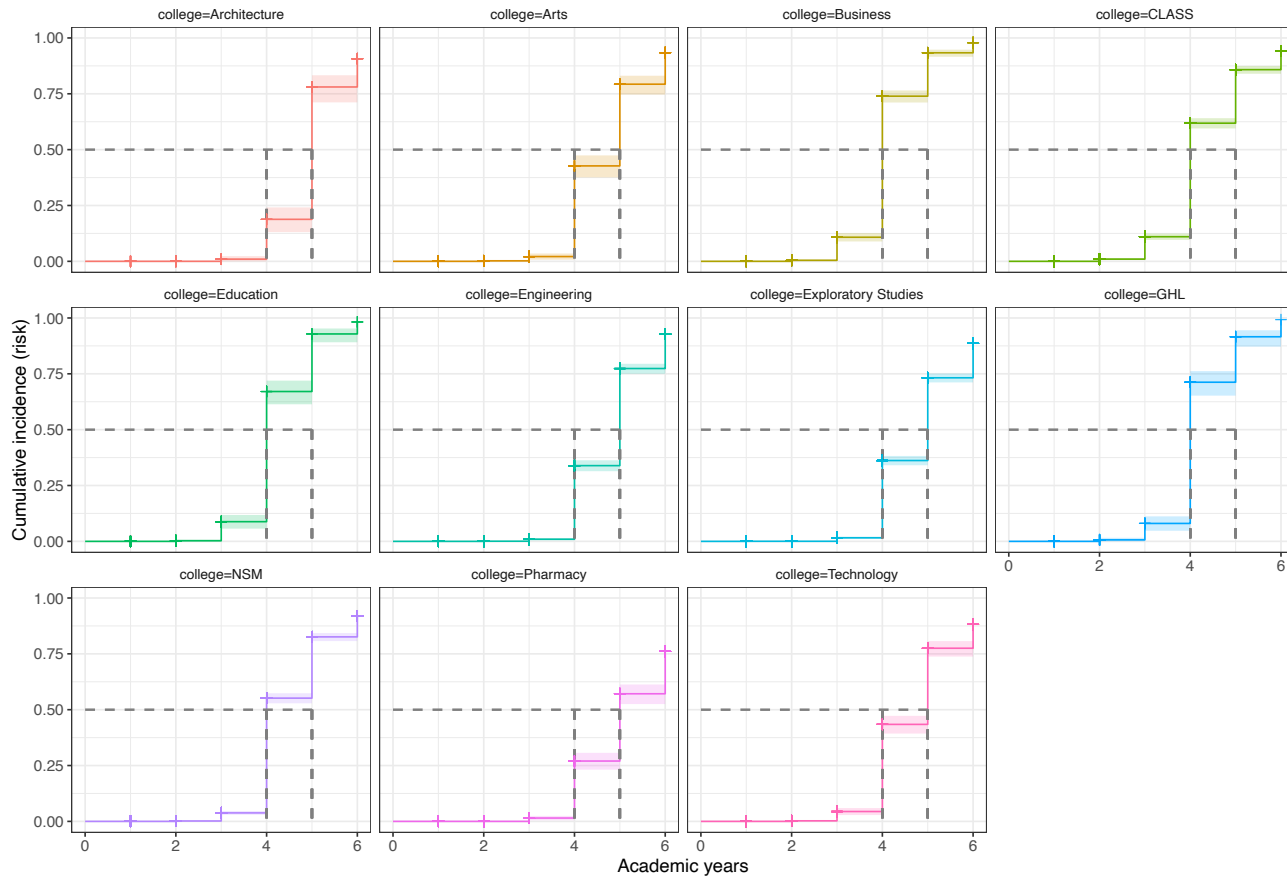
# GENERATION

Cumulative incidence for graduation (N=14,588)



# COLLEGE

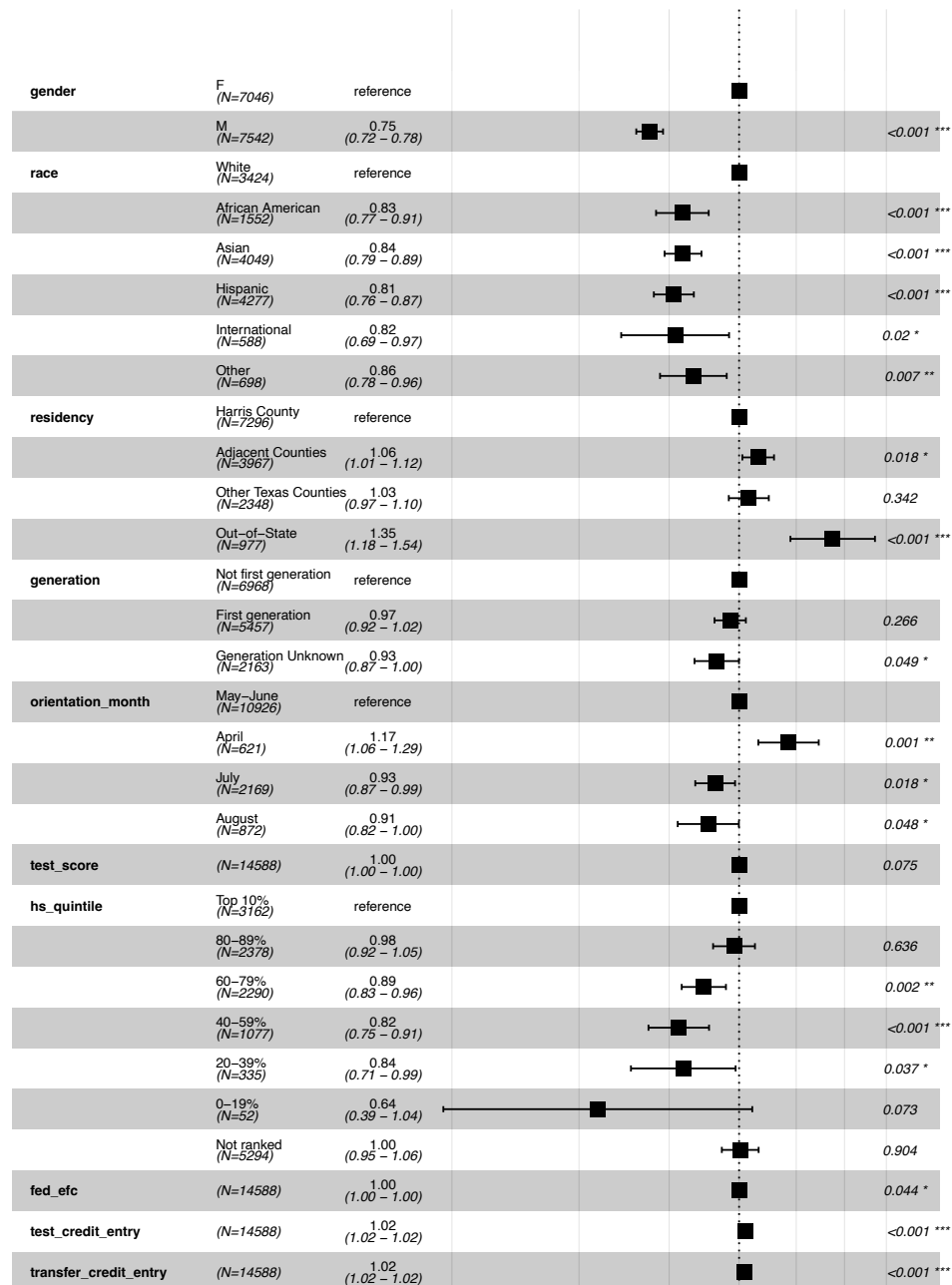
Cumulative incidence for graduation (N=14,588)



# COX PROPORTIONAL HAZARD MODELS

- Survival function helps us **compare rates between categorical** values
- We need the **hazard function** to estimate models with covariates and covariates that are numeric
- The **hazard** is the instantaneous event rate at a particular time point  $t$ .
- **Hazard ratio** is the ratio of two rates between two levels of a predictor (or unit increase in continuous predictor)

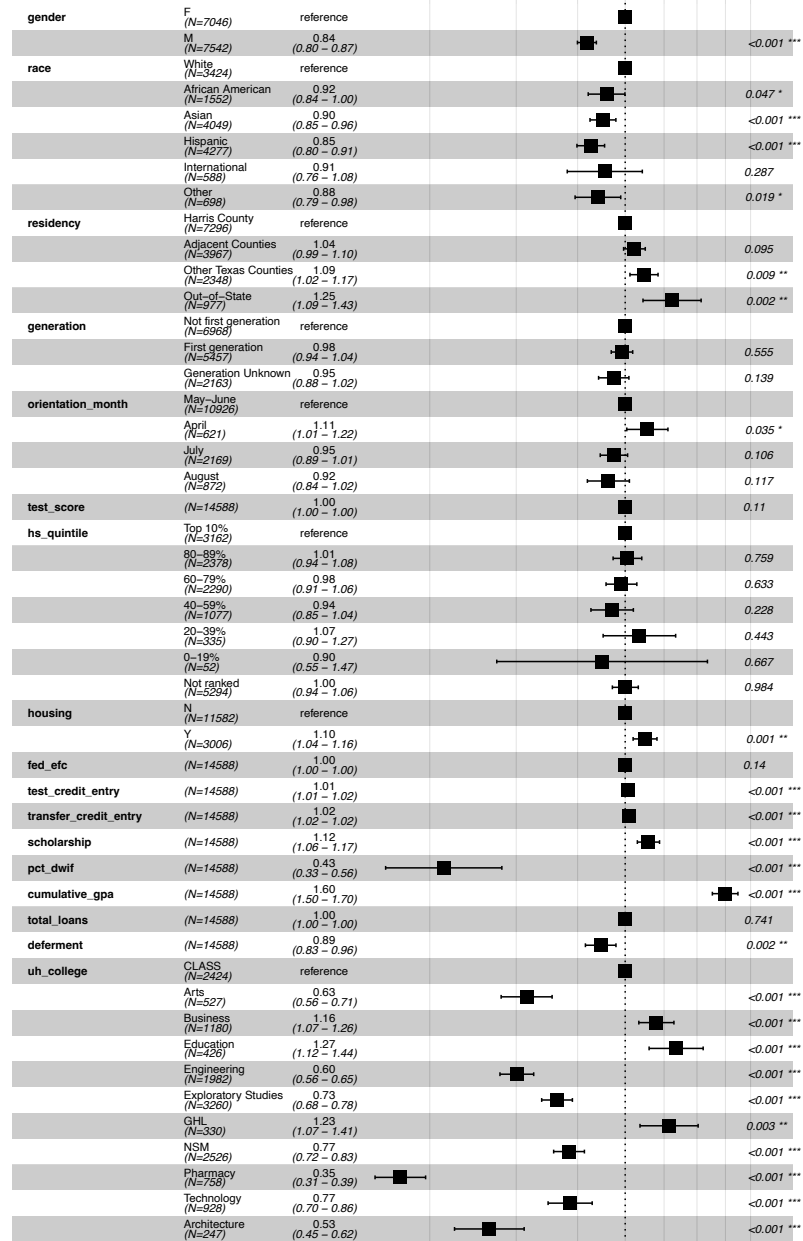
# Incoming Characteristics Model Hazard Ratios



# Events: 8944; Global p-value (Log-Rank): 5.0894e-205  
 AIC: 150460.19; Concordance Index: 0.67

0.4 0.6 0.8 1 1.2 1.4 1.6

# UH Model Hazard Ratios



# Events: 8943; Global p-value (Log-Rank): 0  
AIC: 148837.26; Concordance Index: 0.76

0.4 0.6 0.8 1 1.2 1.4 1.6 1.8



# MODEL COMPARISONS

Incoming vs. First Year survival models:

- Pre-college characteristics no longer significant once more college characteristics were incorporated into the first-year model

# MODEL COMPARISONS

## Logistic Regression vs. Survival Analysis:

- Gender, race/ethnicity = African American, and race/ethnicity = Hispanic became significant in the survival analysis
- Being from further away from UH became significant in the survival analysis with a positive relationship to graduation

# NEXT STEPS

- Decide on the most parsimonious model
- Expand analysis term-by-term
- Time-varying covariates
- Incorporate course data
- Use to identify students for outreach/intervention at specific times

# LIMITATIONS

The variables in the model are limited to the data accessible on UH students. The model does not capture variables like student engagement or sense of belonging; it cannot capture individual student experiences and struggles. It also does not capture the daily efforts of undergraduate student success staff, such as advising, outreach, and tutoring.

# CONTACT INFORMATION

Jorge Martinez

[jxm@uh.edu](mailto:jxm@uh.edu)

Caroline Neary

[csneary@uh.edu](mailto:csneary@uh.edu)