

RESEARCH REPORT

Institutional Research & Enterprise Data Management



August 2022

FIRST SPRING AND SECOND FALL RETENTION OF THE FIRST-TIME IN COLLEGE COHORT

Mark A. Davenport, PhD, & Karen M. Blackwell

***About the Research Brief/Report Series:** The Research Brief/Report Series provides a review of the activities of the Data Analytics and Research arm of the Office of Institutional Research and Enterprise Data Management. Each publication in the Research Report Series contains two documents: A single, executive summary called the Research Brief, and a follow-on document called the Research Report that contains the full report. You can request a copy of these documents from the Office of Institutional Research and Enterprise Data Management.*

OVERVIEW

This Research Report provides information regarding retention rates among UNCG's first-time in college cohorts from Fall 2017 through Fall 2021. Retention rates for both the first Spring term and second Fall term were examined. A comprehensive set of academic ability, demographic, and socioeconomic status variables was employed in this analysis. A logistic regression was performed to explore the impacts of these variables on first Spring and second Fall retention.

RESEARCH QUESTIONS

This Research Report compares the pre-academic and academic profiles of first-time in college (FTIC) students who retain into their first Spring and second Fall with those who do not. There are three main research questions (RQ) addressed in this report.

- RQ1. Has the composition of the first time in college cohort changed across the past five years?*
- RQ2. How does the relationship between pre-college predictors and retention change when we account for college academic performance?*
- RQ3. Have these relationships identified in RQ2 changed over the last five cohort years?*

RQ1: Has the composition of the first time in college cohort changed across the past five years?

- Although the FTIC cohorts have remained between 66% and 68% female across all five fall cohorts, they have become decidedly more diverse, ethnically. White students made up 41% of the cohort in 2017, but only 32% in the 2021 cohort (UNC System Office).

- The cohorts have also become less rural (33% in 2017 to 30% in 2021), less financially able (unmet need average from \$2,446 in 2017 to \$3,689 in 2021), and less academically prepared (high school GPA of 3.8 in 2017 and 3.6 in 2021).

RQ2 & RQ3: How does the relationship between pre-college predictors and retention change when we account for college academic performance and has this changed over the last five years?

- The inclusion of UNCG GPA tends to reduce or eliminate the influence of pre-college predictors, particularly high school GPA. However, the relationship between pre-college independent variables (hereafter, IVs) and spring retention is complex. By the second fall, pre-college predictors appear to lose their influence even before we account for college GPA.
- High school GPA was lower for students who did not retain. The difference was most pronounced by second fall; high school GPA was 3.52 for those who did not return compared to 3.78 for those retained.
- Whereas most ethnic groups' spring retention rates have been relatively stable, Black retention rates have trended up slightly in the past few years. By the second fall, rates had peaked in fall 2019 and dropped for minorities again by 2020.
- There has been a slight drop in the proportion of rural students in the later cohorts (33.1% in 2017 to 30.5% in 2021), and a corresponding drop in first-year retention rates for that group; from 76.2% for the 2017 cohort to 73.0% for the 2021 cohort.
- The proportion of first-generation students increased from 37.5% in 2017 to 49.0% in 2021. Although the percent of first-generation students appears to be increasing, that increase may be tied to the fact that we are gathering first-generation data on more students than we did in the past, as the proportion of 'NULL' responses to these items dropped from 591 in 2017 to 120 in 2021.
- The proportion of Pell-eligible students retaining in spring has remained stable, however in fall, rates were less stable with an increase from 76.2% in 2017 to a high of 80.4% in 2019 but by 2021 retention dropped to 73.0%.
- Unmet need has increased substantially since 2017 from about \$2,446, on average, to \$3,689 in 2021. Students in the 2017 cohort who retained in spring had, on average \$1,149 lower unmet need than those who did not retain. That difference decreased slightly to about \$856 by 2021. For fall retention, the difference decreased to about one-half of the spring amount; on average, about \$500.
- For the fall 2018 cohort the average primary expected family contribution (efc) for those not retained in spring or fall exceeded the primary efc for those who were retained. For all other cohorts, those who were retained had a higher average primary efc (on average, over \$2,000 more) than did those who were not retained. We do not show Primary efc in a table in this report, but it is used as an IV in the prediction studies.
- For those not-retained in spring, their first fall attempted course hours average was roughly half a credit-hour less than those who were retained. This was true for all cohorts. By fall, this relationship disappeared, and the average attempted hours was roughly the same for retained and not-retained students across all cohort years. We do not show attempted hours in a table in this report, but it is used as an IV in the prediction studies.
- The percent of spring non-satisfactory SAP warnings was relatively stable at roughly 22.6% across cohorts. We do not show non-satisfactory SAP warnings in a table in this report, but it is used as an IV in the prediction studies.

DATA

The population for this research was UNCG's first-time in college cohorts from Fall 2017 through Fall 2021. First-time in college cohorts consist of degree-seeking undergraduate students who are entering college for the first time. Population counts as reported in Figure 1 can be confirmed using other public data sources. However, we identified

several cases in these cohorts that had to be removed because their inclusion was not conducive to the logistic regression procedure that is the core of this project. All subsequent figures are based on this analysis dataset, not the population set.

Due to test optional admission test scores in more recent years because of the Covid pandemic, SAT and ACT scores were not included in the analyses. Data were pulled from the Student Data Mart as well as directly from Banner. Imputations were performed for the unmet need and primary expected family contribution variables where necessary, and the techniques applied are explained in Appendix 1.

FINDINGS

Overall retention rates were calculated for the first spring and second fall semesters for each cohort year (**Figure 1**). On average 92.1% of students retain into the first spring. This rate has been consistent over the past five cohorts.

FIGURE 1. FIRST SPRING AND SECOND FALL RETENTION RATES FOR NEW FTIC COHORTS

Population					
Cohort Starting Term	New Freshman	Retained to First Spring		Retained to Second Fall	
	N	n	%	n	%
Fall 2017	2,791	2,574	92.2%	2,127	76.2%
Fall 2018	2,988	2,760	92.4%	2,250	75.3%
Fall 2019	2,743	2,552	93.0%	2,208	80.5%
Fall 2020	2,432	2,224	91.4%	1,888	77.6%
Fall 2021	2,543	2,330	91.6%	1,854	72.9%
5-Year Avg	2,699	2,488	92.1%	2,065	76.5%
Analysis Sample					
Cohort Starting Term	New Freshman	Retained to First Spring		Retained to Second Fall	
	N	n	%	n	%
Fall 2017	2,759	2,546	92.3%	2,103	76.2%
Fall 2018	2,958	2,734	92.4%	2,231	75.4%
Fall 2019	2,697	2,507	93.0%	2,168	80.4%
Fall 2020	2,400	2,194	91.4%	1,863	77.6%
Fall 2021	2,504	2,297	91.7%	1,827	73.0%
5-Year Avg	2,664	2,456	92.2%	2,038	76.5%

Retention into the second Fall drops to 76.5% on average, indicating a loss of close 25% of the initial cohort class. Fall retention rates have fluctuated within 5% of the average over the cohort years.

Figure 2 presents the high school academic ability profile of new students. The academic potential of incoming FTIC students was examined to identify which new students may be at risk of dropping out sometime during their first year. These metrics were compared for students who did and did not retain into

their first spring as well for students who did and did not retain into their second fall. Students who did not retain into their first spring on average had a 3.54 high school GPA compared to 3.73 for those who continued. For second fall the difference was more pronounced; high school GPA was 3.52 for those who dropped out compared to 3.78 for those retained. The average difference in high school GPA for those who retained into the spring was .19 and for fall retention the difference climbed to close to a quarter of a grade point average at .26. Overall, the average high school GPA has fallen from 3.84 in Fall 2017 to 3.59 for those in the most recent Fall 2021 cohort indicating a less academically prepared cohort of students. High School Percentile Score was calculated by dividing high school rank by high school

class size. Cohort students who retained into their first spring were at the 35% top percentile of their class compared to about 42% for those that did not return. A similar pattern was seen with fall retention where students were in the top 34% of high school class compared to the top 42% for those who left. Although relatively stable between Fall 2017 and Fall 2020 at about 33%, the Fall 2021 cohort is on average at about 37%, a drop in high school percentage score of 4% again indicating a less academically prepared class.

UNCG’s first-time cohorts are typically two-thirds female and one-third male. Retention rates on average differ by gender. Although retention rates into first spring are slightly higher for women (92.2% to 93.2%) compared to men

FIGURE 2. HIGH SCHOOL GPA AND HIGH SCHOOL PERCENTILE RANK

High School Data	Cohort Starting Term	Overall Freshman Cohort	Retained to First Spring		Retained to Second Fall	
			No	Yes	No	Yes
High School GPA	Fall 2017	3.84	3.61	3.86	3.64	3.90
	Fall 2018	3.84	3.65	3.86	3.61	3.92
	Fall 2019	3.66	3.59	3.66	3.52	3.69
	Fall 2020	3.67	3.50	3.68	3.49	3.72
	Fall 2021	3.59	3.37	3.61	3.35	3.68
	5-Year Avg	3.72	3.54	3.73	3.52	3.78
High School Percentile (UNCG Calculated)	Fall 2017	34.6	41.3	34.0	40.7	32.6
	Fall 2018	35.1	40.9	34.6	41.3	33.1
	Fall 2019	34.6	38.1	34.4	41.1	33.1
	Fall 2020	34.7	41.0	34.0	41.7	32.5
	Fall 2021	39.5	46.0	39.0	46.7	36.8
	5-Year Avg	35.7	41.5	35.2	42.3	33.6

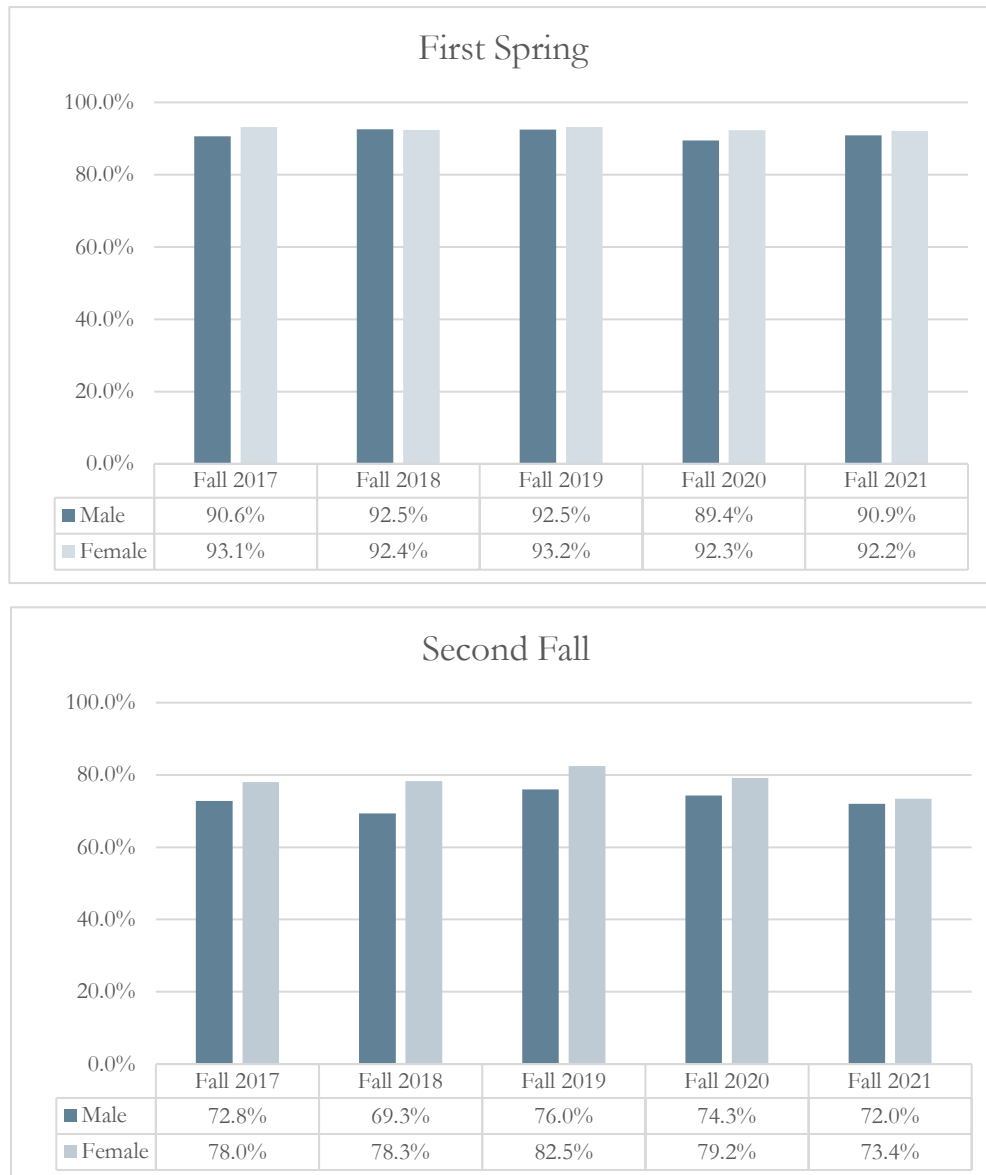
(89.4% to 92.5%), the difference is more pronounced by the end of the first year as (73.4% to 82.5%) of women came back for their second year while only (69.3% to 76.0%) of men persist between 2017 and 2020 (**Figure 3A**). Most concerning is the severe dip that one-year retention took in 2021, having dropped by 2% from 2020 to 2021 for males and by almost 6% between 2020 and 2021 for female students.

Figure 3B illustrates that African American, Asian, and Hispanic students all retain at a higher percentage than Whites into the first spring across all cohort years. The pattern remains the same for second fall retention with White students returning at lower rates than African American, Asian, and Hispanic students until the Fall 2021 cohort where Black and Hispanic retention actually drops below that of Whites. On average, African American students retain at a 3% higher rate into the first spring and 4% higher by the second fall. Of particular note is the precipitous drop in the African American rate between 2020 and 2021 (more than 10% in the one-year rate between fall of 2020 and 2021). Asians retain at a 5% greater rate in the spring and 6% by the second fall. And Hispanics return at a 2% higher rate than Whites in both spring and second fall semesters. The precipitous drop in one-year retention for African American, Asian, and

Hispanic students between 2020 and 2021 is particularly concerning in comparison to the steady rate for White students in the same time period.

Recognizing that students' socioeconomic status can impact performance and retention, **Figures 4a – Figures 4c**, provide retention data on low-income, rural, and first-generation students, respectively. Students were considered low-

**FIGURE 3A.
RETENTION RATES BY GENDER**



income if they were Pell eligible. The data in **Figure 4a1** show that retention rates for Pell eligible students are about the same on average as for non-Pell eligible students in the first spring term. The biggest difference occurs in the most recent cohort year where in spring Pell eligible students returned at 91% compared to non-Pell eligible students at 92.8% and the difference by second fall widened to 77.3% of non-Pell eligible students returning to just 69.8% of Pell students, which is over 7% lower. Again, the greatest differences between Pell- and non-Pell-eligible students' retention rates are for the 2020 and 2021 cohorts.

Over 93% of UNCG’s new first-time population are in-state students and typically about 20% come from Guilford County, where UNCG is located (UNC System Office). Many of North Carolina’s counties are rural and economically

FIGURE 3B.
RETENTION RATES BY RACE/ETHNICITY

Retention	Cohort Starting Term	American Indian or Alaska Native	Asian	Black or African American	Hispanic or Latino	Native Hawaiian or Other Pacific Islander	Nonresident Alien	Two or More Races	Unknown	White
Retained First Spring	Fall 2017	76.9%	96.5%	93.1%	92.4%	100.0%	93.8%	94.9%	100.0%	90.7%
	Fall 2018	91.7%	94.5%	93.3%	94.9%		100.0%	92.7%	85.7%	90.4%
	Fall 2019	87.5%	95.9%	95.2%	93.5%	100.0%	93.3%	93.4%	98.0%	90.2%
	Fall 2020	71.4%	95.3%	93.5%	92.2%	100.0%	83.3%	88.5%	90.3%	89.5%
	Fall 2021	75.0%	95.6%	91.8%	91.9%	100.0%	87.5%	90.3%	92.6%	91.3%
Retained Second Fall	Fall 2017	46.2%	77.3%	79.1%	75.0%	0.0%	87.5%	79.0%	85.7%	73.7%
	Fall 2018	83.3%	77.9%	76.0%	78.4%	0.0%	72.7%	77.1%	71.4%	73.2%
	Fall 2019	62.5%	83.7%	85.2%	80.6%	100.0%	86.7%	86.8%	74.5%	75.0%
	Fall 2020	71.4%	82.9%	81.3%	77.0%	66.7%	83.3%	68.9%	77.4%	75.4%
	Fall 2021	50.0%	79.7%	70.8%	71.8%	0.0%	65.6%	72.9%	70.4%	75.4%

distressed. **Figure 4a2** shows an increase in average unmet need from \$2,446 in 2017 to \$3,689 in 2021. Consequently, an examination of students from rural counties provided additional insight into their socioeconomic status. Students were identified as ‘rural’ if they came from a rural North Carolina county as designated by the NC Department of Commerce, which considers each county’s average unemployment rate, median household income, percentage growth in population, and adjusted property tax base per capita. Rural students (**Figure 4b**) across the cohort terms in both first spring and second fall retain at lower rates than those from non-rural counties by 1% to 2%, but, again, as was seen with Pell eligibility, the most recent cohort year sees the gap widening even more with non-rural students retained at a 2.8% higher rate in first spring and close to 3% by second fall for the 2019 and 2020 cohorts. By fall 2021, the difference shrinks again to about 1%.

Where data were available, the first-generation college criterion was considered (**Figure 4c**). This is a self-reported metric, based on students’ disclosure of their parents’ or guardians’ highest attained level of education. A first-generation student is a student whose parents or guardians have not earned at least a bachelor’s degree. First-generation college students are especially at risk during their first year of enrollment because they face additional challenges that can impact their performance and retention; they typically have less financial and social support and are less prepared for college than their inter-generational peers. Within the new first-time cohorts, first-generation students had lower retention rates than students whose parents had obtained at least a bachelor’s degree. First spring retention rates for first-generation students was typically on average 2.3% lower than non-first-generation students, but in the most recent cohort year the

difference widened to 4.2%. For the 2017 cohort, the difference in fall retention was 6.6%, dropping to around 2% for 2018 and 2019, then climbing again to 5.1% in 2020 and more than 9% in the most recent cohort year.

FIGURE 4A1. SOCIOECONOMIC PROFILE – PELL ELIGIBLE

PELL ELIGIBLE						PELL ELIGIBLE					
Term	First Spring Not Retained		First Spring Retained		Total	Term	Second Fall Not Retained		Second Fall Retained		Total
	n	%	n	%			n	%	n	%	
Fall 2017	213	7.7%	2546	92.3%	2,759	Fall 2017	656	23.8%	2,103	76.2%	2,759
No	88	7.3%	1,113	92.7%	1,201	No	274	22.8%	927	77.2%	1,201
Yes	125	8.0%	1,433	92.0%	1,558	Yes	382	24.5%	1,176	75.5%	1,558
Fall 2018	224	7.6%	2,734	92.4%	2,958	Fall 2018	727	24.6%	2,231	75.4%	2,958
No	91	7.4%	1,139	92.6%	1,230	No	292	23.7%	938	76.3%	1,230
Yes	133	7.7%	1,595	92.3%	1,728	Yes	435	25.2%	1,293	74.8%	1,728
Fall 2019	190	7.0%	2,507	93.0%	2,697	Fall 2019	529	19.6%	2,168	80.4%	2,697
No	90	7.5%	1,104	92.5%	1,194	No	247	20.7%	947	79.3%	1,194
Yes	100	6.7%	1,403	93.3%	1,503	Yes	282	18.8%	1,221	81.2%	1,503
Fall 2020	206	8.6%	2,194	91.4%	2,400	Fall 2020	537	22.4%	1,863	77.6%	2,400
No	83	8.0%	952	92.0%	1,035	No	213	20.6%	822	79.4%	1,035
Yes	123	9.0%	1,242	91.0%	1,365	Yes	324	23.7%	1,041	76.3%	1,365
Fall 2021	207	8.3%	2,297	91.7%	2,504	Fall 2021	677	27.0%	1,827	73.0%	2,504
No	76	7.2%	976	92.8%	1,052	No	239	22.7%	813	77.3%	1,052
Yes	131	9.0%	1,321	91.0%	1,452	Yes	438	30.2%	1,014	69.8%	1,452
Grand Total	1,040	7.8%	12,278	92.2%	13,318	Grand Total	3,126	23.5%	10,192	76.5%	13,318

FIGURE 4A2. SOCIOECONOMIC PROFILE – UNMET NEED

UNMET NEED						UNMET NEED					
Term	First Spring Not Retained		First Spring Retained		Total Avg	Term	Second Fall Not Retained		Second Fall Retained		Total Avg
	n	Avg	n	Avg			n	Avg	n	Avg	
Fall 2017	197	\$3,506	2,360	\$2,357	\$2,446	Fall 2017	599	\$2,824	1,958	\$2,330	\$2,446
Fall 2018	211	\$4,044	2,572	\$2,811	\$2,905	Fall 2018	673	\$3,286	2,110	\$2,783	\$2,905
Fall 2019	179	\$3,605	2,330	\$2,578	\$2,652	Fall 2019	488	\$2,963	2,021	\$2,577	\$2,652
Fall 2020	183	\$3,880	2,069	\$2,996	\$3,068	Fall 2020	489	\$3,523	1,763	\$2,942	\$3,068
Fall 2021	183	\$4,477	2,115	\$3,621	\$3,689	Fall 2021	635	\$4,300	1,663	\$3,455	\$3,689
Grand Total	953	\$3,902	11,446	\$2,853	\$2,934	Grand Total	2,884	\$3,399	9,515	\$2,793	\$2,934

FIGURE 4B. SOCIOECONOMIC PROFILE – RURAL

RURAL					
Term	First Spring Not Retained		First Spring Retained		Total
	n	%	n	%	
Fall 2017	213	7.7%	2,546	92.3%	2,759
No/NA	135	7.3%	1,712	92.7%	1,847
Yes	78	8.6%	834	91.4%	912
Fall 2018	224	7.6%	2,734	92.4%	2,958
No/NA	137	6.9%	1,854	93.1%	1,991
Yes	87	9.0%	880	91.0%	967
Fall 2019	190	7.0%	2,507	93.0%	2,697
No/NA	118	6.4%	1,732	93.6%	1,850
Yes	72	8.5%	775	91.5%	847
Fall 2020	206	8.6%	2,194	91.4%	2,400
No/NA	140	8.4%	1,523	91.6%	1,663
Yes	66	9.0%	671	91.0%	737
Fall 2021	207	8.3%	2,297	91.7%	2,504
No/NA	129	7.4%	1,611	92.6%	1,740
Yes	78	10.2%	686	89.8%	764
Grand Total	1,040	7.8%	12,278	92.2%	13,318

RURAL					
Term	Second Fall Not Retained		Second Fall Retained		Total
	n	%	n	%	
Fall 2017	656	23.8%	2,103	76.2%	2,759
No/NA	434	23.5%	1,413	76.5%	1,847
Yes	222	24.3%	690	75.7%	912
Fall 2018	727	24.6%	2,231	75.4%	2,958
No/NA	484	24.3%	1,507	75.7%	1,991
Yes	243	25.1%	724	74.9%	967
Fall 2019	529	19.6%	2,168	80.4%	2,697
No/NA	348	18.8%	1,502	81.2%	1,850
Yes	181	21.4%	666	78.6%	847
Fall 2020	537	22.4%	1,863	77.6%	2,400
No/NA	356	21.4%	1,307	78.6%	1,663
Yes	181	24.6%	556	75.4%	737
Fall 2021	677	27.0%	1,827	73.0%	2,504
No/NA	464	26.7%	1,276	73.3%	1,740
Yes	213	27.9%	551	72.1%	764
Grand Total	3,126	23.5%	10,192	76.5%	13,318

FIGURE 4C. SOCIOECONOMIC PROFILE – FIRST GENERATION

FIRST GENERATION					
Term	First Spring Not Retained		First Spring Retained		Total
	n	%	n	%	
Fall 2017	213	7.7%	2,546	92.3%	2,759
No	70	6.2%	1,064	93.8%	1,134
N/A	44	7.4%	547	92.6%	591
Yes	99	9.6%	935	90.4%	1,034
Fall 2018	224	7.6%	2,734	92.4%	2,958
No	90	8.0%	1,035	92.0%	1,125
N/A	57	7.5%	699	92.5%	756
Yes	77	7.1%	1,000	92.9%	1,077
Fall 2019	190	7.0%	2,507	93.0%	2,697
No	65	5.5%	1,113	94.5%	1,178
N/A	38	7.8%	450	92.2%	488
Yes	87	8.4%	944	91.6%	1,031
Fall 2020	206	8.6%	2,194	91.4%	2,400
No	88	7.5%	1,091	92.5%	1,179
N/A	17	15.5%	93	84.5%	110
Yes	101	9.1%	1,010	90.9%	1,111
Fall 2021	207	8.3%	2,297	91.7%	2,504
No	67	5.8%	1,089	94.2%	1,156
N/A	17	14.2%	103	85.8%	120
Yes	123	10.0%	1,105	90.0%	1,228
Grand Total	1,040	7.8%	12,278	92.2%	13,318

FIRST GENERATION					
Term	Second Fall Not Retained		Second Fall Retained		Total
	n	%	n	%	
Fall 2017	656	23.8%	2,103	76.2%	2,759
No	240	21.2%	894	78.8%	1,134
N/A	129	21.8%	462	78.2%	591
Yes	287	27.8%	747	72.2%	1,034
Fall 2018	727	24.6%	2,231	75.4%	2,958
No	267	23.7%	858	76.3%	1,125
N/A	179	23.7%	577	76.3%	756
Yes	281	26.1%	796	73.9%	1,077
Fall 2019	529	19.6%	2,168	80.4%	2,697
No	219	18.6%	959	81.4%	1,178
N/A	98	20.1%	390	79.9%	488
Yes	212	20.6%	819	79.4%	1,031
Fall 2020	537	22.4%	1,863	77.6%	2,400
No	227	19.3%	952	80.7%	1,179
N/A	39	35.5%	71	64.5%	110
Yes	271	24.4%	840	75.6%	1,111
Fall 2021	677	27.0%	1,827	73.0%	2,504
No	253	21.9%	903	78.1%	1,156
N/A	41	34.2%	79	65.8%	120
Yes	383	31.2%	845	68.8%	1,228
Grand Total	3,126	23.5%	10,192	76.5%	13,318

Data in **Figure 5** depict academic outcomes of new first-time students after their first year. Not surprisingly, GPA was lower for students who did not retain to the first spring term and the second fall term compared to students who did retain. On average, the first fall term GPA for students who retained to their first spring was 2.92, compared to 1.44 for students who did not return in spring. The first term GPA of students who did not retain into the first spring has steadily decreased over the five cohort terms. For those who did retain into the first spring, the term GPA ranged from 2.86 to 3.08 in the earlier years but dropped significantly with the Fall 2021 cohort to 2.76. The first spring term

FIGURE 5. TERM AND CUMULATIVE GPA

Average	Cohort Starting Term	TOTAL		Retained to First Spring				Retained to Second Fall			
		n	mean	No		Yes		No		Yes	
				n	mean	n	mean	n	mean	n	mean
First Fall Term GPA	Fall 2017	2,719	2.78	184	1.61	2,535	2.86	627	1.78	2,103	3.03
	Fall 2018	2,911	2.81	198	1.58	2,713	2.90	700	1.78	2,230	3.03
	Fall 2019	2,665	2.92	167	1.57	2,498	3.01	504	2.11	2,167	3.18
	Fall 2020	2,370	2.95	182	1.27	2,188	3.08	518	1.76	1,857	3.19
	Fall 2021	2,468	2.64	180	1.18	2,288	2.76	653	1.49	1,824	3.03
First Spring Term GPA	Fall 2017	2,520	2.80	3	3.12	2,517	2.80	437	1.78	2,083	3.02
	Fall 2018	2,699	2.74	0		2,699	2.74	502	1.71	2,197	2.97
	Fall 2019	2,396	3.21	1	3.00	2,395	3.21	302	2.52	2,094	3.31
	Fall 2020	2,169	2.95	1	0.00	2,168	2.95	339	1.86	1,830	3.15
	Fall 2021	2,269	2.72	5	0.78	2,264	2.72	471	1.52	1,798	3.03
First Year Cumulative GPA	Fall 2017	2,730	2.74	184	1.61	2,546	2.82	627	1.78	2,103	3.03
	Fall 2018	2,930	2.73	199	1.59	2,731	2.81	700	1.78	2,230	3.03
	Fall 2019	2,671	2.98	167	1.57	2,504	3.07	504	2.11	2,167	3.18
	Fall 2020	2,375	2.88	182	1.27	2,193	3.01	518	1.76	1,857	3.19
	Fall 2021	2,477	2.62	180	1.19	2,297	2.73	653	1.49	1,824	3.03

GPA for students who retained to their second fall was 3.10 on average, while the first spring term GPA for students who did not return the next fall was 1.88. The average first year cumulative GPA for students who returned for their second fall was 3.09, while for students who completed their first year but did not return for the second fall had a 1.78 cumulative GPA. A few students show a spring GPA despite the fact that Banner shows they were not retained into spring.

ANALYSIS

Questions of retention/attrition are commonly addressed using logistic regression. We follow that convention here as well. The technical aspects of logistic regression are described elsewhere (Hosmer, Lemeshow & Sturdivant (2013), and Agresti (2013)). We will briefly describe the models but will concentrate our attention on the interpretation of results.

The steps we took to conduct these analyses were based on a set of assumptions: 1) Retention at UNCG has been well-researched in the past and we have ample evidence to believe that a particular set of predictors are related to retention; 2) The entry of these predictors into any regression framework follows a reasonable order of entry based on temporal contiguity (demographic predictors first, high school predictors second, and last, financial and other predictors measured at the time of admission).

To address our three research questions, we tested two sets of identical models: one to predict first spring retention (Spring Study), and a second to predict second fall retention (Fall Study). For each set of predictors, we ran an initial model without any UNCG GPA entered in the model (Model 1), and a subsequent model with UNCG GPA entered (Model 2); such that Spring Model 1 is the spring retention model WITHOUT the UNCG GPA as a covariate, Fall Model 2 is the fall retention model WITH the UNCG GPA covariate, etc. The intent of Model 2 for the spring and fall retention studies was to see how academic performance changes our ability to model retention using only pre-admission/admission data. These four models were tested across multiple cohort years, beginning with the 2017 first-time, first-year new student cohort. In total, we tested twenty models. Green cells appearing in Figures 7A to 8E represent statistically significant ($p < .05$ or less) coefficients. The sign of the coefficient estimate, i.e., whether it is positive or negative, describes the relationship between the independent variable (hereafter, IV) and the odds of the dependent variable (hereafter, DV). A positive coefficient estimate indicates a positive relationship between the IVs and the DV outcome being modeled, while a negative coefficient estimate indicates an inverse relationship. For example, a positive coefficient estimate for the IV ‘male’ would indicate that being male is associated with greater odds of retention, while a negative coefficient estimate would indicate that being male is associated with lower odds of retention. To understand the effect of each variable on the odds of retention in practical terms, each coefficient is exponentiated to calculate the odds ratio. Odds ratios quantify the strength of the association between two variables, or the odds that an outcome will occur given a particular condition (being a first-generation student) compared to the odds of the outcome occurring in the absence of that condition (not being a first-generation student).

First Spring Retention Study

Figure 6 provides a list of predictors used in this study presented in the order they were entered into the model. In this study, we were not interested in finding the ‘best’ of competing models. Rather, we were primarily interested in testing known predictors against changes in cohort and term. Figures 7A to 7E provide the robust statistics for spring study models 1 and 2 and Figures 8A to 8E provide the statistics for the fall study models. Two models were developed to estimate first Spring retention. The first spring retention study included all the IVs that are shown in Figure 6, entered in that order. The first model was intended to provide insight into student characteristics without the statistical weight of UNCG academic performance. Each model was tested using SAS’ PROC LOGISTIC procedure with stepwise selection of variables. Significant IVs in each model are shown in Figures 7A to 7E in green. Non-significant variables were also tested and included in the table in red. The values shown in these tables are the robust statistics provided by the imputation process described in Appendix 1. Each stepwise model resulted in a well-fitting model. However, we were less concerned with this aspect of the analysis. We concentrated on interpreting the pattern of coefficients and the manner in which the pattern of significance changed across cohort years.

Model 2 was assembled and tested in the same way as Model 1 and is presented on the right side of each Figure. The second models included the additional variable 1st Term Fall GPA to control for students’ academic outcome at the end of their first semester. The addition of the first Fall term GPA variable changed the statistical significance of some IVs in the model. However, a students’ GPA at the end of their first Fall semester will naturally be an important factor in whether or not they continue into their first Spring, and its inclusion in this model allowed for the identification of student characteristics that still significantly impacted Spring retention after accounting for academic performance during the first Fall semester. Results from Spring Models 1 and 2 are discussed in detail below.

**FIGURE 6. VARIABLES USED IN LOGISTIC REGRESSION MODELS
IN ORDER OF ENTRY**

Variable	Description	Type	Use
DEPENDENT VARIABLES			
Enrolled in 1st Spring	1 if retained in first Spring, 0 if otherwise	Discrete	Dependent
Enrolled in 2nd Fall	1 if retained in second Fall, 0 if otherwise	Discrete	Dependent
IVs: COVARIATES			
1 st Term Fall GPA	First fall term GPA (Model 2 only)	Continuous	Independent/Covariate
1 st Year Cum GPA	Cumulative first year GPA (Model 2 only)	Continuous	Independent/Covariate
IVs: ETHNICITY			
Black	1 if Black, 0 if White	Discrete	Independent
Asian	1 if Asian, 0 if White	Discrete	Independent
Hispanic	1 if Hispanic, 0 if White	Discrete	Independent
Other Race	1 if other race/ethnicity, 0 if White	Discrete	Independent
IVs: DEMOGRAPHICS			
Male	1 if male, 0 if female	Discrete	Independent
Rural vs N/A	1 if from rural county, 0 if county is N/A	Discrete	Independent
Rural vs Not Rural	1 if from rural county, 0 if county is Not Rural	Discrete	Independent
IVs: STATE AT ADMISSION			
High School GPA at Admission	High School GPA used for admission	Continuous	Independent
1 st Generation vs N/A	1 if 1 st generation, 0 if N/A	Discrete	Independent
1 st Generation vs Not 1 st Generation	1 if 1 st generation, 0 if Not 1 st Generation	Discrete	Independent
Pell Eligible	1 if Pell Eligible in first fall, 0 if Not eligible	Discrete	Independent
Unmet Need in \$1,000	Amount of unmet need in first year, in \$1,000	Continuous	Independent
Primary etc	Expected family contribution in first year, in \$1,000	Continuous	Independent
IVs: UNCG ACADEMIC RECORD			
Hours attempted in 1 st fall	Credit hours attempted in first fall term	Continuous	Independent
Total hours attempted in 1 st year	Total credit hours attempted in 1 st fall and spring terms	Continuous	Independent
SAP Spring Term	1 is satisfactory progress for spring term, 0 is otherwise	Discrete	Independent

Figure 7A summarizes results for Spring Models 1 and 2 for the 2017 cohort. Consistent with what we already knew about minority retention rates, Model 1 shows the odds of retaining for Black students were 1.91 times higher than for their White counterparts; 3.40 times higher for Asian students over their White counterparts; 1.71 times higher for Hispanic students over their White counterparts. Model 2 shows that, while accounting for the fall term UNCG GPA, these odds diminished slightly but remained significant. The odds of Black student retention were 1.92 times higher than for White students; the odds of Asian student retention were 3.16 times that for White students. Hispanic students no longer differed in their odds of retention over White students. Students in the ‘Other Races’ category, in Model 2, were 2.26 times more likely to retain than White students.

For the 2017 cohort, none of the demographic variables were significant.

As expected, high school GPA was a significant and positive predictor of spring retention; a 1.0 increase in GPA raised the odds of spring retention by a factor of 2.68. Stated as a percent, every additional one-point increase in high school GPA increases the odds of returning for the spring semester by 167.81%; for every tenth of a GPA point, the odds of returning for spring semester increase by 16.78%. In Model 2, first-term fall UNCG GPA negates the influence of high school GPA. This fall term GPA is a significant predictor, increasing the odds of retention by almost 19% for every .1 increase in fall GPA.

**FIGURE 7A. SPRING RETENTION LOGISTIC REGRESSION RESULTS*:
2017 COHORT**

Variable	Model 1				Model 2			
	Coefficient	Odds Ratio	Standard Error	p-value	Coefficient	Odds Ratio	Standard Error	p-value
Constant	-3.441	0.0320	0.9235	.0002	-1.3996	0.2467	1.1354	.2177
IVs: COVARIATES								
1 st Term Fall GPA					1.0632	2.8956	1.1354	<.0001
IVs: ETHNICITY								
Black vs White	0.6484	1.9125	0.1875	.0005	0.6522	1.9198	0.2104	.0019
Asian vs White	1.2245	3.4025	0.4798	.0107	1.1492	3.1557	0.5043	.0227
Hispanic vs White	0.5379	1.7124	0.5379	.0288	0.4576	1.5803	0.2690	.0889
Other Race vs White	0.5423	1.7200	0.2959	.0668	0.8146	2.2583	0.3456	.0184
IVs: DEMOGRAPHICS								
Male vs Female	-0.2435	0.7839	0.1561	.1187	-0.1125	0.8936	0.1753	.5210
Rural vs N/A	0.0604	1.0623	0.3666	.8691	-0.5842	0.5576	0.3935	.1377
Rural vs Not Rural	-0.2205	0.8021	0.1636	.1779	0.0866	1.0905	0.1851	.6398
IVs: STATE AT ADMISSION								
High School GPA at Admission	0.9851	2.6781	0.1612	.0001	-0.1236	0.8837	0.2047	.5462
1 st Generation vs N/A	-0.2051	0.8146	0.2124	.3341	-0.1510	0.8598	0.2387	.5269
1 st Generation vs Not 1 st Generation	-0.4696	0.6253	0.1825	.0101	-0.3736	0.6883	0.2023	.0649
Pell Eligible vs Not Pell Eligible	0.2141	1.2387	0.2193	.3289	0.4331	1.5420	0.2400	.0712
Unmet Need in \$1,000	-0.0603	0.9415	0.0212	.0045	-0.0597	0.9420	0.0240	.0130
Primary efc in \$1,000	0.0096	1.0096	0.0089	.2810	0.0083	1.0084	0.0094	.3749
IVs: UNCG ACADEMIC RECORD								
Hours attempted in 1 st fall	0.1574	1.1705	.0463	.0007	0.1197	1.1272	0.0582	.0399

* Results shown are logistic model results from 25 imputed datasets; results of the 25 model runs were combined using SAS's PROC MIANALYZE process. The description of the imputation process is provided in Appendix 1.

First-generation students were less likely to retain in spring before accounting for fall GPA; the odds of returning in spring were 1.60 times higher for non-first-generation students than for first-generation students. Interestingly, unmet need appeared to be unaffected by the accounting of fall GPA and remained relatively unchanged between Model 1 and Model 2. In both models, every increase in unmet need by \$1,000 was associated with a significant decrease in the odds of spring retention, a 5.8% decrease in odds for spring retention for both models. As well, fall GPA had little effect on the influence of course hours attempted on spring retention. In both models, a one-hour increase in course credit hours attempted was associated with an increased likelihood of spring retention; a 17% increase for Model 1, and a 12% increase for Model 2.

**FIGURE 7B. SPRING RETENTION LOGISTIC REGRESSION RESULTS*:
2018 COHORT**

Variable	Model 1				Model 2			
	Coefficient	Odds Ratio	Standard Error	p-value	Coefficient	Odds Ratio	Standard Error	p-value
Constant	-3.3615	0.0347	0.9060	.0002	-.8772	.4159	1.1309	.4380
IVs: COVARIATES								
1 st Term Fall GPA					1.2622	3.5330	0.0834	<.0001
IVs: ETHNICITY								
Black vs White	0.5514	1.7357	0.1759	.0017	0.6491	1.9139	0.2052	.0016
Asian vs White	0.5464	1.7270	0.3876	.1587	0.4288	1.5354	0.4269	.3151
Hispanic vs White	0.8001	2.2258	0.2538	.0016	0.8617	2.3671	0.2881	.0028
Other Race vs White	0.4106	1.5077	0.2965	.1661	0.4636	1.5898	0.3461	.1804
IVs: DEMOGRAPHICS								
Male vs Female	0.2263	1.2540	0.1590	.1547	0.5426	1.7205	0.1871	.0037
Rural vs N/A	0.5554	1.7426	0.3486	.8734	-0.6438	0.5253	0.4037	.1108
Rural vs Not Rural	-0.3857	0.6800	0.1559	.0134	-0.3030	0.7386	0.1820	.0959
IVs: STATE AT ADMISSION								
High School GPA at Admission	0.8400	2.3164	0.1503	<.0001	-0.4948	0.6097	0.1984	.0126
1 st Generation vs N/A	0.0461	1.0472	0.1885	.8067	0.1896	1.2088	0.2183	.3850
1 st Generation vs Not 1 st Generation	0.0843	1.0880	0.1797	.6389	0.3455	1.4127	0.2084	.0972
Pell Eligible vs Not Pell Eligible	-0.2652	0.7671	0.2075	.2013	-0.1070	0.8985	0.2434	.6603
Unmet Need in \$1,000	-0.0632	0.9388	0.0176	.0003	-0.0812	0.9220	0.0209	.0001
Primary efc in \$1,000	-0.0114	0.9887	0.0063	.0688	-0.0158	0.9843	0.0071	.0270
IVs: UCG ACADEMIC RECORD								
Hours attempted in 1 st fall	0.1953	1.2157	0.0467	<.0001	0.1593	1.1726	0.0572	.0054

* Results shown are logistic model results from 25 imputed datasets; results of the 25 model runs were combined using SAS's PROC MIANALYZE process. The description of the imputation process is provided in Appendix 1.

Figure 7B summarizes results for Models 1 and 2 for the 2018 cohort. As with the 2017 cohort, Black students were about 1.9% more likely to retain than white students, this is true for both models 1 and 2. Unlike for the 2017 cohort, Asian students were not significantly more likely than White students to retain in spring given either model. Hispanic students were more than 2 times more likely to retain in spring than White students in both models. As we saw in 2017, the inclusion of the fall term GPA in Model 2 does not have much influence on the odds for retention of either Black students or Hispanic students.

Interestingly, only for the 2018 cohort does gender ever appear to be a significant predictor of spring retention, but only in Model 2. Although the Figure 7B shows males 1.72 times more likely to retain in spring, we know from experience that our female retention rate tends to be higher than it is for males. Non-rural students were 1.47 more likely to retain in spring before accounting for fall GPA. This relationship disappeared in Model 2.

In 2018, high school GPA was a significant predictor of spring retention (13.16 times greater likelihood of retention for every .1 increase in GPA). However, the inclusion of fall GPA greatly reduced the predictive power of high school GPA but it was still a significant predictor of spring retention in Model 2 although the relationship reversed. This likely indicates the collinear nature (See Appendix 2) of high GPA and UCG GPA. Therefore, we should not interpret the negative coefficient as it is likely anomalous.

For Model 1 unmet need was significant; in Model 2 both unmet need and primary efc were significant. Again, as in 2017, course hours were predictive of spring retention.

**FIGURE 7C. SPRING RETENTION LOGISTIC REGRESSION RESULTS*:
2019 COHORT**

Variable	Model 1				Model 2			
	Coefficient	Odds Ratio	Standard Error	p-value	Coefficient	Odds Ratio	Standard Error	p-value
Constant	-2.3841	0.0922	0.9766	.0145	0.1196	1.1270	1.2200	.9219
IVs: COVARIATES								
1 st Term Fall GPA					1.2495	3.4886	0.0881	<.0001
IVs: ETHNICITY								
Black vs White	0.8280	2.2887	0.2090	<.0001	0.8157	2.2608	0.2471	.0010
Asian vs White	1.2686	3.5559	0.4772	.0079	1.4755	4.3732	0.5758	.0104
Hispanic vs White	0.6307	1.8789	0.2482	.0110	0.5382	1.7129	0.2830	.0572
Other Race vs White	0.6461	1.9081	0.2979	.0301	0.5756	1.7782	0.3445	.0947
IVs: DEMOGRAPHICS								
Male vs Female	-0.0475	0.9536	0.1688	.7784	0.2876	1.3332	0.2025	.1555
Rural vs N/A	-0.2418	0.7852	0.3570	.4982	-0.4271	0.6524	0.4209	.3102
Rural vs Not Rural	-0.3828	0.6819	0.1714	.0255	-0.0773	0.9256	0.2005	.5998
IVs: STATE AT ADMISSION								
High School GPA at Admission	0.4391	1.5513	0.1848	.0175	-0.8605	0.4230	0.2523	.0006
1 st Generation vs N/A	-0.5081	0.6016	0.2336	.0295	-0.2696	0.7637	0.2746	.3263
1 st Generation vs Not 1 st Generation	-0.6372	0.5288	0.1939	.0010	-0.5240	0.5921	0.2256	.0202
Pell Eligible vs Not Pell Eligible	0.1995	1.2208	0.2148	.3531	0.2534	1.2884	0.2514	.3134
Unmet Need in \$1,000	-0.0691	0.9332	0.0206	.0008	-0.0900	0.9139	0.0247	.0003
Primary efc in \$1,000	-0.0096	0.9904	0.0063	.1319	-0.0157	0.9844	0.0073	.0305
IVs: UNCG ACADEMIC RECORD								
Hours attempted in 1 st fall	0.2545	1.2898	0.0473	<.0001	0.1972	1.2180	0.0570	.0006

* Results shown are logistic model results from 25 imputed datasets; results of the 25 model runs were combined using SAS's PROC MIANALYZE process. The description of the imputation process is provided in Appendix 1.

Figure 7C summarizes results for Models 1 and 2 for the 2019 cohort. Consistent with what we already knew about minority retention rates, Model 1 shows the odds of retaining for Black students were 2.29 times higher than for their white counterparts; 3.56 times greater for Asian students over their white counterparts; 1.88 times greater for Hispanic students over their white counterparts; and 1.91 time greater for other races over their white counterparts. In Model 2, the relative significance of Black and Asian ethnicity remains significant after accounting for fall GPA.

Rural students were less likely to retain in spring before accounting for fall GPA.

High school GPA was a significant and positive predictor of spring retention; a 1.0 increase in GPA increasing the odds of spring retention by a factor of 1.55. As with the 2018 cohort we should not interpret the odds ratio for Model 2 due to multicollinearity. First-generation students were less likely to retain in spring in both models. In both models, every increase in unmet need by \$1,000 was associated with significant decrease in the odds of spring retention. Primary efc also became significant in Model 2. As in earlier cohorts, the number of course hours attempted was a significant predictor of spring retention for both models.

**FIGURE 7D. SPRING RETENTION LOGISTIC REGRESSION RESULTS*:
2020 COHORT**

Variable	Model 1				Model 2			
	Coefficient	Odds Ratio	Standard Error	p-value	Coefficient	Odds Ratio	Standard Error	p-value
Constant	-1.8653	0.1548	0.9307	.0500	0.5356	1.7085	1.1419	.6390
IVs: COVARIATES								
1 st Term Fall GPA					1.2613	3.5300	0.0763	<.0001
IVs: ETHNICITY								
Black vs White	0.8007	2.2271	0.1997	<.0001	0.8052	2.2371	0.2447	.0010
Asian vs White	0.9742	2.6490	0.0440	.0269	0.5638	1.7573	0.4797	.2399
Hispanic vs White	0.6217	1.8621	0.2257	.0059	0.7132	2.0405	0.2778	.0102
Other Race vs White	0.1196	1.1270	0.2520	.6350	0.4085	1.5046	0.3202	.2020
IVs: DEMOGRAPHICS								
Male vs Female	-0.1990	0.8195	0.1589	.2103	0.0195	1.0197	0.1956	.9107
Rural vs N/A	-0.2777	0.7575	0.3070	.3658	-0.4054	0.6667	0.3838	.2908
Rural vs Not Rural	-0.1001	0.9047	0.1670	.5490	0.1564	1.1693	0.2026	.4402
IVs: STATE AT ADMISSION								
High School GPA at Admission	0.8716	2.3907	0.1697	<.0001	-0.4604	0.6310	0.2258	.0415
1 st Generation vs N/A	-0.7742	0.4611	0.3081	.0120	-0.5238	0.5923	0.3733	.1608
1 st Generation vs Not 1 st Generation	-0.2973	0.7428	0.1762	.0915	-0.1780	0.8369	0.2133	.4042
Pell Eligible vs Not Pell Eligible	-0.2085	0.8118	0.2215	.3466	0.0579	1.0596	0.2668	.8286
Unmet Need in \$1,000	-0.0167	0.9834	0.0187	.3714	-0.0070	0.9930	0.0215	.7444
Primary efc in \$1,000	-0.0030	0.9970	0.0066	.6544	-0.0002	0.9998	0.0082	.9815
IVs: UCG ACADEMIC RECORD								
Hours attempted in 1 st fall	0.0846	1.0883	0.0456	.0637	0.0257	1.0260	0.0547	.6388

* Results shown are logistic model results from 25 imputed datasets; results of the 25 model runs were combined using SAS's PROC MIANALYZE process. The description of the imputation process is provided in Appendix 1.

Figure 7D summarizes results for Models 1 and 2 for the 2020 cohort. Representing the first of the 'COVID years', the 2020 cohort looks remarkably different than the prior three cohorts. The ethnicity factors were consistent with what we found in the prior cohorts: Model 1 showing the odds of retaining Black students is 2.23 times higher than for their white counterparts; 2.65 times greater for Asian students over their white counterparts; and 1.86 times greater for Hispanic students over their white counterparts.

However, only first generation versus 'first-generation marker not-given' was significant. As we don't really know what 'not-given' means, this factor is hard to interpret.

Also, in this year only, course hours attempted was not a significant predictor of spring retention. Generally speaking, what were appearing to be familiar predictors of spring retention for the earlier, pre-COVID cohorts, (high school GPA, unmet need, and course hours attempted) were no longer significant for this cohort.

**FIGURE 7E. SPRING RETENTION LOGISTIC REGRESSION RESULTS*:
2021 COHORT**

Variable	Model 1				Model 2			
	Coefficient	Odds Ratio	Standard Error	p-value	Coefficient	Odds Ratio	Standard Error	p-value
Constant	-4.0024	0.0183	0.8926	<.0001	-0.8433	0.4303	1.0595	0.426
IVs: COVARIATES								
1 st Term Fall GPA					1.1063	3.0231	0.0815	<.0001
IVs: ETHNICITY								
Black vs White	0.5430	1.7212	0.2002	.0067	0.7696	2.1589	0.2338	0.0010
Asian vs White	0.8495	2.3385	0.4169	.0416	0.8630	2.3703	0.4455	0.0527
Hispanic vs White	0.5103	1.6658	0.2376	.0317	0.6408	1.8980	0.1100	0.0180
Other Race vs White	0.2658	1.3045	0.2819	.3457	0.1877	1.2065	0.3185	0.5557
IVs: DEMOGRAPHICS								
Male vs Female	-0.0457	0.9553	0.1608	.7764	-0.1560	0.8556	0.1803	0.3869
Rural vs N/A	-0.4784	0.6198	0.3274	.1440	-0.7859	0.4557	0.3832	0.0403
Rural vs Not Rural	-0.3053	0.7369	0.1652	.0645	-0.2124	0.8086	0.1886	0.2602
IVs: STATE AT ADMISSION								
High School GPA at Admission	0.9676	2.6316	0.1601	<.0001	-0.4277	0.6520	0.2055	0.0374
1 st Generation vs N/A	-0.7632	0.4662	0.3133	.0149	-0.6936	0.4998	0.3514	0.0484
1 st Generation vs Not 1 st Generation	-0.4170	0.6590	0.1799	.0204	-0.3409	0.7111	0.2040	0.0948
Pell Eligible vs Not Pell Eligible	0.1991	1.2203	0.2363	.3997	0.3618	1.4359	0.2709	0.1817
Unmet Need in \$1,000	-0.0043	0.9957	0.0198	.8283	0.0051	1.0051	0.0220	0.8166
Primary efc in \$1,000	0.0108	1.0109	0.0092	.2423	0.0111	1.0112	0.0108	0.3034
IVs: UNCG ACADEMIC RECORD								
Hours attempted in 1 st fall	0.1989	1.2201	0.0394	<.0001	0.1544	1.1670	0.0483	0.0014

* Results shown are logistic model results from 25 imputed datasets; results of the 25 model runs were combined using SAS's PROC MIANALYZE process. The description of the imputation process is provided in Appendix 1.

Figure 7E summarizes results for Models 1 and 2 for the 2021 cohort. Again, as with earlier cohorts and minority retention rates, Model 1 shows the odds of retaining for Black students were 1.72 times higher than for their White counterparts; 2.34 times greater for Asian students over their White counterparts; and 1.67 times greater for Hispanic students over their White counterparts.

High school GPA was a significant and positive predictor of spring retention as was first-generation. First-generation students were less likely to retain in spring Model 1. In Model 2 the significance of high school GPA and first-generation versus 'NA' should not be interpreted for reasons described earlier. As with prior cohorts, the number of course hours attempted remained a significant predictor of spring retention.

Second Fall Retention Models

The fall retention models are similar to the spring retention models with two exceptions: 1) an additional IV was available from the set of UNCG academic-related variables in the form of a rating for ‘satisfactory academic progress’, or SAP, for the spring term. This SAP factor was given a value of ‘1’ if the student had made satisfactory progress by the end of spring term or a ‘0’ if, for any reason, they were rated as NOT having made ‘satisfactory’ progress. Reasons for unsatisfactory progress might include being at risk of losing financial aid due to grades, being placed on probation or suspension, not earning sufficient course hours, etc. Additionally, the fall-term GPA covariate used in the spring retention models was replaced with the first year cumulative GPA.

Other than the two new IVs described above, the processes used to assemble and test the fall retention models were the same as for the spring retention models.

FIGURE 8A. FALL RETENTION LOGISTIC REGRESSION RESULTS*:
2017 COHORT

Variable	Model 1				Model 2			
	Coefficient	Odds Ratio	Standard Error	p-value	Coefficient	Odds Ratio	Standard Error	p-value
Constant	-4.1476	0.0158	0.6873	<.0001	-3.2593	0.0384	0.7848	<.0001
IVs: COVARIATES								
1 st Year Cum GPA					1.5025	4.4929	0.0925	<.0001
IVs: ETHNICITY								
Black vs White	0.6848	1.9834	0.1352	<.0001	0.7589	2.1359	0.1473	<.0001
Asian vs White	0.1507	1.1626	0.2459	0.5399	0.3008	1.3509	0.2610	0.2489
Hispanic vs White	0.3071	1.3595	0.1731	0.0760	0.3161	1.3718	0.1869	0.0909
Other Race vs White	0.3476	1.4157	0.1979	0.0790	0.5954	1.8138	0.2179	0.0063
IVs: DEMOGRAPHICS								
Male vs Female	-0.0337	0.9669	0.1111	0.7614	0.1951	1.2154	0.1236	0.1146
Rural vs N/A	0.3952	1.4847	0.2884	0.1706	-0.0331	0.9674	0.3078	0.9145
Rural vs Not Rural	0.0143	1.0144	0.1165	0.9023	0.1846	1.2027	0.1284	0.1504
IVs: STATE AT ADMISSION								
High School GPA at Admission	0.5240	1.6888	0.1158	<.0001	-0.2953	0.7443	0.1382	0.0325
1 st Generation vs N/A	-0.0622	0.9397	0.1459	0.6699	0.0530	1.0544	0.1609	0.7417
1 st Generation vs Not 1 st Generation	-0.2477	0.7806	0.1264	0.0500	-0.1809	0.8345	0.1384	0.1912
Pell Eligible vs Not Pell Eligible	0.1520	1.1642	0.1517	0.3161	0.0420	1.0429	0.1656	0.7995
Unmet Need in \$1,000	-0.0188	0.9814	0.0166	0.2582	-0.0091	0.9909	0.0176	0.6066
Primary efc in \$1,000	-0.0038	0.9962	0.0049	0.4314	-0.0072	0.9928	0.0063	0.1746
IVs: UNCG ACADEMIC RECORD								
Hours attempted in 1 st fall	0.1010	1.1063	0.0351	0.0040	0.0678	1.0702	0.0406	0.0947
SAP Spring Term	2.3261	10.2379	0.1200	<.0001	0.6718	1.9578	0.1543	<.0001

* Results shown are logistic model results from 25 imputed datasets; results of the 25 model runs were combined using SAS’s PROC MIANALYZE process. The description of the imputation process is provided in Appendix 1.

Figure 8A summarizes results for Models 1 and 2 for the 2017 cohort. Unlike the spring 2017 results, only the Black students’ fall retention differs from their White counterparts, and this effect remains in Model 2. Students of other races were also more likely to retain than their White counterparts in Model 2. High school GPA, hours attempted, and SAP were significant in Model 1, in Model 2 high school GPA and SAP remained significant. SAP standing is a very strong predictor of fall retention. In Model 1 those in good standing were ten times more likely to retain. Even after accounting for first year GPA in Model 2, those in good SAP standing were still twice as likely to retain.

FIGURE 8B. FALL RETENTION LOGISTIC REGRESSION RESULTS*:
2018 COHORT

Variable	Model 1				Model 2			
	Coefficient	Odds Ratio	Standard Error	p-value	Coefficient	Odds Ratio	Standard Error	p-value
Constant	-4.8384	0.0079	0.6815	<.0001	-3.3185	0.0362	0.7661	<.0001
IVs: COVARIATES								
1 st Year Cum GPA					1.3860	3.9988	0.0865	<.0001
IVs: ETHNICITY								
Black vs White	0.6325	1.8823	0.1293	<.0001	0.6056	1.8324	0.1388	<.0001
Asian vs White	0.2018	1.2236	0.2510	0.4213	-0.0014	0.9986	0.2680	0.9980
Hispanic vs White	0.5095	1.6645	0.1658	0.0021	0.6400	1.8965	0.1804	0.0004
Other Race vs White	0.5927	1.8089	0.2126	0.0053	0.4719	1.6030	0.2310	0.0411
IVs: DEMOGRAPHICS								
Male vs Female	-0.1232	0.8841	0.1078	0.2531	0.0043	1.0043	0.1174	0.9706
Rural vs N/A	-0.1436	0.8662	0.2500	0.5656	-0.7218	0.4859	0.2719	0.0079
Rural vs Not Rural	-0.1558	0.8557	0.1123	0.1653	-0.0871	0.9166	0.1220	0.4753
IVs: STATE AT ADMISSION								
High School GPA at Admission	0.8053	2.2374	0.1103	<.0001	-0.0047	0.9953	0.1318	0.9715
1 st Generation vs N/A	0.0860	1.0898	0.1358	0.5266	0.0843	1.0880	0.1473	0.5669
1 st Generation vs Not 1 st Generation	-0.1238	0.8836	0.1258	0.3252	-0.0510	0.9503	0.1363	0.7084
Pell Eligible vs Not Pell Eligible	0.2492	1.2830	0.1462	0.0883	0.3258	1.3851	0.1594	0.0410
Unmet Need in \$1,000	-0.0140	0.9861	0.0139	0.3123	-0.0126	0.9875	0.0150	0.4020
Primary efc in \$1,000	-0.0001	0.9999	0.0048	0.9792	0.0011	1.0011	0.0054	0.8318
IVs: UNGC ACADEMIC RECORD								
Hours attempted in 1 st fall	0.0485	1.0497	0.0345	0.1605	-0.0088	0.9912	0.0388	0.8218
SAP Spring Term	2.4908	12.0709	0.1227	<.0001	0.9272	2.5274	0.1544	<.0001

* Results shown are logistic model results from 25 imputed datasets; results of the 25 model runs were combined using SAS's PROC MIANALYZE process. The description of the imputation process is provided in Appendix 1.

Figure 8B summarizes results for Models 1 and 2 for the 2018 cohort. Unlike the fall 2017 results the ethnicity IVs become relevant again, with Black, Hispanic, and other race students being more likely to retain than their White counterparts.

As in 2017, SAP status was very influential in Model 1 and remains influential in Model 2 even after including cumulative GPA. As a side note, the only time Pell was significant in any model for any cohort was in 2018.

**FIGURE 8C. FALL RETENTION LOGISTIC REGRESSION RESULTS*:
2019 COHORT**

Variable	Model 1				Model 2			
	Coefficient	Odds Ratio	Standard Error	p-value	Coefficient	Odds Ratio	Standard Error	p-value
Constant	-4.0463	0.0175	0.7293	<.0001	-3.7831	0.0228	0.7953	<.0001
IVs: COVARIATES								
1 st Year Cum GPA					1.1594	3.1880	0.0896	<.0001
IVs: ETHNICITY								
Black vs White	0.8265	2.2853	0.1416	<.0001	0.8132	2.2551	0.1514	<.0001
Asian vs White	0.4888	1.6304	0.2582	0.0583	0.5921	1.8078	0.2329	0.0363
Hispanic vs White	0.3608	1.4345	0.1681	0.0319	0.4054	1.4999	0.1794	0.0238
Other Race vs White	0.5515	1.7359	0.2016	0.0062	0.5449	1.7244	0.2168	0.0029
IVs: DEMOGRAPHICS								
Male vs Female	-0.2116	0.8093	0.1132	0.0616	-0.0259	0.9744	0.1232	0.9333
Rural vs N/A	-0.1670	0.8462	0.2533	0.5097	-0.1833	0.8325	0.2750	0.5050
Rural vs Not Rural	-0.1856	0.8306	0.1201	0.1222	0.0006	1.0006	0.1298	0.9951
IVs: STATE AT ADMISSION								
High School GPA at Admission	0.5178	1.6783	0.1323	<.0001	0.0227	1.0230	0.1512	0.8807
1 st Generation vs N/A	-0.0939	0.9104	0.1590	0.5549	0.0768	1.0798	0.1717	0.6547
1 st Generation vs Not 1 st Generation	-0.1881	0.8285	0.1306	0.1497	-0.0069	0.9931	0.1408	0.9607
Pell Eligible vs Not Pell Eligible	0.2607	1.2978	0.1519	0.0861	0.1616	1.1754	0.1627	0.3207
Unmet Need in \$1,000	-0.0168	0.9833	0.0152	0.2678	-0.0099	0.9901	0.0162	0.5402
Primary efc in \$1,000	-0.0028	0.9972	0.0046	0.5434	-0.0037	0.9963	0.0049	0.4517
IVs: UCG ACADEMIC RECORD								
Hours attempted in 1 st fall	0.1336	1.1429	0.0366	0.0003	0.9974	2.7112	0.0402	0.0131
SAP Spring Term	1.8688	6.4805	0.1267	<.0001	0.1277	1.1362	0.1329	0.4852

* Results shown are logistic model results from 25 imputed datasets; results of the 25 model runs were combined using SAS's PROC MIANALYZE process. The description of the imputation process is provided in Appendix 1.

Figure 8C summarizes results for Models 1 and 2 for the 2019 cohort. For fall 2019, the significant ethnicity IVs for Model 1 were the same as in 2018. In Model 2 all of the ethnicity IVs were significant.

As in 2017, academic-related IVs were significant for Model 1 and remained significant in Model 2, except for SAP spring term status which dropped out of significance after the covariate cumulative GPA was added in Model 2.

**FIGURE 8D. FALL RETENTION LOGISTIC REGRESSION RESULTS*:
2020 COHORT**

Variable	Model 1				Model 2			
	Coefficient	Odds Ratio	Standard Error	p-value	Coefficient	Odds Ratio	Standard Error	p-value
Constant	-4.1072	0.0165	0.7240	<.0001	-3.4820	0.0307	0.8013	<.0001
IVs: COVARIATES								
1 st Year Cum GPA					1.3152	3.7255	0.0847	<.0001
IVs: ETHNICITY								
Black vs White	0.8827	2.4174	0.1502	<.0001	0.8616	2.3669	0.1654	<.0001
Asian vs White	0.4604	1.5847	0.2747	0.0937	0.4019	1.4947	0.2956	0.1739
Hispanic vs White	0.5492	1.7319	0.1690	0.0012	0.5222	1.6857	0.1869	0.0052
Other Race vs White	0.1896	1.2088	0.2029	0.3500	0.2692	1.3089	0.2325	0.2470
IVs: DEMOGRAPHICS								
Male vs Female	-0.0870	0.9167	0.1208	0.4714	0.0637	1.0658	0.1349	0.6367
Rural vs N/A	-0.2003	0.8185	0.2423	0.4084	-0.2641	0.7679	0.2730	0.3323
Rural vs Not Rural	-0.0647	0.9373	0.1249	0.6044	0.0331	1.0337	0.1389	0.8118
IVs: STATE AT ADMISSION								
High School GPA at Admission	0.6935	2.0007	0.1327	<.0001	-0.0463	0.9548	0.1537	0.7634
1 st Generation vs N/A	-0.7183	0.4876	0.2503	0.0041	-0.5088	0.6012	0.2877	0.0770
1 st Generation vs Not 1 st Generation	-0.3225	0.7243	0.1308	0.0136	-0.1432	0.8666	0.1451	0.3235
Pell Eligible vs Not Pell Eligible	0.2722	1.3128	0.1673	0.1038	0.1590	1.1723	0.1836	0.3864
Unmet Need in \$1,000	0.0065	1.0065	0.0142	0.6473	0.0202	1.0204	0.0156	0.1927
Primary etc in \$1,000	0.0065	1.0065	0.0054	0.2306	0.0064	1.0064	0.0061	0.2932
IVs: UNCG ACADEMIC RECORD								
Hours attempted in 1 st fall	0.0545	1.0560	0.0349	0.1178	0.0454	1.0464	0.0396	0.2513
SAP Spring Term	2.2210	9.2165	0.1304	<.0001	0.3544	1.4253	0.1795	0.0484

* Results shown are logistic model results from 25 imputed datasets; results of the 25 model runs were combined using SAS's PROC MIANALYZE process. The description of the imputation process is provided in Appendix 1.

Figure 8D summarizes results for Models 1 and 2 for the 2020 cohort. In 2020, Black and Hispanic student groups were more likely to retain in fall than their White counterparts.

Although high school GPA and first-generation were significant predictors for fall retention Model 1, that significance faded when cumulative GPA was added in Model 2. In Model 1, higher high school GPA and NOT being a first-generation student were associated with a higher likelihood of retention. As with all of the earlier fall models, satisfactory academic progress was a strong fall retention predictor, especially for Model 1.

**FIGURE 8E. FALL RETENTION LOGISTIC REGRESSION RESULTS*:
2021 COHORT**

Variable	Model 1				Model 2			
	Coefficient	Odds Ratio	Standard Error	p-value	Coefficient	Odds Ratio	Standard Error	p-value
Constant	-4.7635	0.0085	0.7138	<.0001	-3.1957	0.0409	0.8065	<.0001
IVs: COVARIATES								
1 st Year Cum GPA					1.3962	4.0398	0.0856	<.0001
IVs: ETHNICITY								
Black vs White	0.6251	1.8684	0.1486	<.0001	0.6137	1.8473	0.1633	0.0002
Asian vs White	0.4296	1.5366	0.2550	0.0920	0.4591	1.5826	0.2752	0.0883
Hispanic vs White	0.3063	1.3584	0.1716	0.0742	0.4206	1.5229	0.1910	0.0276
Other Race vs White	0.2177	1.2432	0.2132	0.3072	0.1091	1.1153	0.2323	0.6387
IVs: DEMOGRAPHICS								
Male vs Female	0.1297	1.1385	0.1159	0.2632	0.1569	1.1699	0.1278	0.2194
Rural vs N/A	-0.6566	0.5186	0.2312	0.0045	-0.9332	0.3933	0.2580	0.0003
Rural vs Not Rural	-0.0032	0.9968	0.1125	0.9794	0.0892	1.0933	0.1352	0.5095
IVs: STATE AT ADMISSION								
High School GPA at Admission	0.6167	1.8528	0.1205	<.0001	-0.3105	0.7331	0.1445	0.0315
1 st Generation vs N/A	-0.3372	0.7138	0.2593	0.1918	0.0448	1.0458	0.2928	0.8785
1 st Generation vs Not 1 st Generation	-0.2568	0.7735	0.1258	0.0397	-0.1260	0.8816	0.1381	0.3615
Pell Eligible vs Not Pell Eligible	0.1975	1.2184	0.1630	0.2256	0.0528	1.0542	0.1798	0.7693
Unmet Need in \$1,000	-0.0004	0.9996	0.0138	0.9757	-0.0078	0.9922	0.0154	0.6120
Primary efc in \$1,000	0.0012	1.0012	0.0052	0.8141	-0.0029	0.9971	0.0059	0.6257
IVs: UNCG ACADEMIC RECORD								
Hours attempted in 1 st fall	0.1092	1.1154	0.0342	0.0014	0.0821	1.0856	0.0399	0.0398
SAP Spring Term	2.4820	11.9652	0.1258	<.0001	0.8137	2.2562	0.1609	<.0001

* Results shown are logistic model results from 25 imputed datasets; results of the 25 model runs were combined using SAS's PROC MIANALYZE process. The description of the imputation process is provided in Appendix 1.

Figure 8E summarizes results for Models 1 and 2 for the 2021 cohort. Black and Hispanic students for the most part were more likely to return for their second year than White students. The significance seen with the spring Models of Asian students retaining at higher rates than White students is no longer true for fall retention. We are aware that the one-year Black retention rate for the fall 2021 cohort dropped significantly compared to prior years as has Hispanic retention. When all other factors are accounted for in our models the coefficient for these two groups are positive suggesting an increasing rate rather than decreasing rate. This result is being investigated further.

Also similar to the 2020 cohort, NOT being first generation was associated with more likely retention in fall Model 1, but not for fall Model 2. However, unlike the 2020 cohort, Rural students were less likely to retain than students for whom we have no rural indicator.

Consistent with earlier cohorts, academic factors appear to have the largest influence on both spring and fall retention. Unlike fall retention for earlier cohorts, fall course hours remains significant even after accounting for cumulative gpa.

SUMMARY AND CONCLUSIONS

To help summarize the many models we ran and to better describe the changes we observed in time, we created a stoplight heatmap (Figure 9). This figure reports spring and fall models side-by-side showing all years running from left to right. IVs are shown along the y-axis in the same order as the earlier figures. Again, this order follows the order of entry into the original regression models. If an IV was reported as significant at less than .05 in Figures 7 or 8, that is recorded in Figure 9 as a green cell; non-significant IVs are shown as red cells. If the direction of a significant coefficient was positive, a '+' sign is shown within the cell. If negative, the cell contains a '-' sign.

As expected, the covariates used in Model 2 for both spring and fall retention studies are the most consistent predictors of retention for all cohorts. For minorities, the pattern is complex. When comparing ethnic groups, the odds of Black student retention are higher than those of White students for most cohort years. For Hispanic students, the odds of retention are higher for most cohorts from 2018 onward. The precipitous drop in minority retention for Blacks and Hispanics shown in Figure 3B, although not obvious from the multivariate model is of real concern and needs to be investigated further. Asian students tend to have higher retention in the spring, but the significance drops when we looked at fall retention. The 'other races' group tends to be small, and the pattern of significance is hard to explain.

In terms of demographics, no consistent trends appear and those few significant results that appear for the spring study tended to disappear for the fall study.

For the state of admission IVs, high school GPA is a consistent predictor of retention for both spring models and fall Model 1. It is interesting that, while the inclusion of cumulative GPA does not eliminate the influence of high school GPA for the spring study, it does for the fall study for the 2018 to 2020 cohorts. Whereas financial need IVs appear to aid the prediction of spring retention, they don't provide much help for predicting fall retention. The lone exception, and the only time Pell eligibility was significant was for the 2018 fall Model 2 cohort where 1st year cumulative GPA was included.

Satisfactory Academic Progress (SAP) for the spring term was significant for all cohorts and models in the fall study with the exception of the spring 2019 cohort for fall Model 2.

FIGURE 9. SIGNIFICANT VARIABLES FOR ALL COHORTS FOR MODELS 1 AND 2

	Spring 1 Retention Model 1					Spring 1 Retention Model 2					Fall 2 retention Model 1					Fall 2 Retention Model 2				
	2017	2018	2019	2020	2021	2017	2018	2019	2020	2021	2017	2018	2019	2020	2021	2017	2018	2019	2020	2021
IVs: COVARIATES																				
1 st Term Fall GPA						+	+	+	+	+						+	+	+	+	+
IVs: ETHNICITY																				
Is Black vs White	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Is Asian vs White	+	-	+	+	+	+	-	+	+	+	+	-	+	+	+	+	-	+	+	+
Is Hispanic vs White	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
All Others vs White			+			+							+			+	+	+		
IVs: DEMOGRAPHICS																				
Male vs Female							+													
Rural vs N/A										-					-		-			-
Rural vs Not Rural		-	-																	
IVs: STATE AT ADMISSION																				
High School GPA at Admission	+	+	+	+	+		-	-	-	-	+	+	+	+	+	-				-
1 st Generation vs N/A																				
1 st Generation vs Not 1 st Generation																				
Pell Eligible vs Not Pell Eligible																	+			
Unmet Need in \$1,000																				
Primary etc in \$1,000																				
IVs: UNGC ACADEMIC RECORD																				
Hours attempted in 1 st fall	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
SAP Spring Term											+	+	+	+	+	+	+	+	+	+

COVID

An important aspect of the timing of this research is the unprecedented influence of COVID on the 2020 and 2021 academic years. COVID has created an ‘asterisk’ in the timeline for institutions across the globe. UNCG went online exclusively for more than a year and families found themselves without work, or at work without daycare for the same time period. In an attempt to temper the economic impact of the pandemic, the federal government infused both the institution and its students and employees with financial stimulus payments meant to soften the impact of pandemic-induced financial contraction. The end results of this chaos have yet to be fully assessed and it may be several years before we fully understand these effects. But the first step in understanding how the pandemic affected UNCG, and how UNCG should respond now and in the future is assessing how COVID impacted our academic community in 2020 and 2021.

Examining Figure 9, the most apparent difference between the COVID years of 2020 and 2021 is the sudden absence of a significant effect for the financial IVs, unmet need and primary etc. The simple explanation is that those with the most need either did not return at all, or that stimulus funds were used to reduce the financial strain that would normally be evident in that student population. Another interesting trend is the sudden non-significance of 1st fall hours attempted for the 2020 cohort spring models. Further examination of how the pandemic affected our academic community will help UNCG better prepare for the future and may provide insights for improving institutional processes going forward that will help UNCG better navigate times of stress.

LIMITATIONS

Given the multitude of factors that may impact retention, it is emphasized that the models presented here are a starting point for studying these questions and are by no means exhaustive or definitive. This study was limited by both time and data constraints. Additional variables that may be considered for future research include how many hours the student spent working outside of school, the distance a student must commute to UNCG, or the number of courses a freshman enrolls in that have historically high Drop-Fail-Withdrawal (DFW) rates.

We feel that the inclusion of Starfish data will be an important predictor in future studies. Those data were not included here because prior to 2019 those data are limited.

Although the results of the logistic regressions provide valuable insight into some significant variables related to retention, it would also be beneficial to explore retention rates among various combinations of demographic and socioeconomic variables, for example retention rates among female Asian students from urban areas compared to female Asian students from rural areas.

Despite these limitations, this Research Report contributes a strong foundation for exploring the relationships between a variety of student variables and first year retention. It is important to continue providing time trend analyses so the institution may better prepare to act rather than just ‘react’.

APPENDIX 1: THE IMPUTATION PROCEDURE

Multiple imputation is a method for dealing with missing data that is superior to most other common and convenient methods of dealing with missing data—that is, replacing the empty cell with a mean or a neighboring value, dropping all the cases with missing data, or dropping the variable altogether. Multiple imputation as implemented by SAS, uses a statistical method to produce a reasonable replacement value based on the other variables in the dataset. The process for doing this is described below.

The multiple imputation procedure we used began with the creation of a regression model, wherein the DV is the target variable with the missing data. SAS includes the following recommendation regarding the imputation model: ‘...generally, you should include as many variables as you can when doing multiple imputation.’ This means including variables that may be correlated with the variables to be imputed that may not be included in the analysis model. In our case, we excluded the class variables (we assumed that student gender would be unrelated to SES); race might be, but its effect would likely be accounted for by our inclusion of other economic variables in the imputation process.

SAS also includes a recommendation by Van Buuren, Boshuizen, and Knook (1999, p. 687) that three sets of imputation covariates be considered: variables that are included in the actual analysis model, variables not included in the model but that correlate highly with the variables to be imputed, and variables that correlate to the missingness of the imputed variables. We have tried to honor this suggestion as best we could.

We imputed unmet need and primary efc (both having about 6% missingness). The primary mechanism of missingness in this case was related to the variables themselves. Both unmet need and primary efc are both indicators of family economic status and both are derived from FAFSA information. If students are missing these data, it is almost certainly likely this is due to not having completed a FAFSA. The primary reason a student would not complete a FAFSA is because they do not need financial aid. If the data were not missing for these students, they would most certainly have very low or no unmet need and high expected family contribution.

We imputed values for unmet need and primary efc separately. The only difference is we included unmet need to impute primary efc, and primary efc to impute unmet need. Following the advice of Van Buuren, et al (1999), we included variables from the actual model that might correlate with each variable (Race, First Generation, Rural and Pell Eligible) and we included variables from outside the model that likely correlate to unmet need and primary efc, but also relate to the reason for missingness (wealthier students are more likely to be missing data). These exterior variables are Median Household Income and the estimated poverty rate of the zip code where the student attended high school (if they attended high school) or, in the absence of a high school zip code, the zip code for the students’ permanent address at the time of admission. These data elements were derived from US Census and NCES data.

We used SAS PROC MI’s FCS Predicted Means Matching method (REGPMM) for the imputation process. REGPMM is recommended for imputing continuous variables and can be used where the missingness pattern is either monotonic or arbitrary. Although unmet need and primary efc are likely to be missing for the same student due to those variables being reliant on FAFSA completion, most of the other variables in both the imputation and the logistic models are complete data. PROC MI’s internal test of monotonic missingness indicated no pattern to the missing data. So, we used the arbitrary process.

The first step we ran used all the variables to impute unmet need. In the second step, we used the same variables with the addition of the now complete (imputed) unmet need data to impute primary efc. This dataset was then used in all subsequent analyses.

APPENDIX 2: A NOTE ABOUT MULTICOLLINEARITY

We examined the intercorrelations between the IVs, including the DV, and eliminated from model consideration several data points that we had available, such as high school percentile rank and low income as they showed high correlation with other IVs that we had already committed to the model, such as high school GPA and unmet need.

As a final check of the complete model, we chose to run both complete models using PROC REG to get variance inflation factor (VIF) and tolerance (TOL) values for each IV. Neither of these important multicollinearity measures are provided in PROC LOGISTIC. Fortunately, these measures are computed between the IVs only. Allison (2012) suggests that linear regression can be applied to the model without regard to the DV and the VIF and TOL measures interpreted as they appear. Allison (2012) suggests that the tolerance should be less than .40; the highest values for the IVs included in our complete model were for Pell eligibility (.51) and primary efc (.50). The VIF for both IVs is almost 2.0, implying that the intercorrelation of these variables (likely with each other) may result in standard errors for these variables being inflated by roughly 1.5. On examination, (Figures 7 and 8) none of the standard errors appeared out-of-line, although the GPA measures did reflect a high level of intercorrelation. These two variables were very important to the model, so they were retained. Given the VIF and TOL statistics, we concluded that multicollinearity is not problematic enough to remove any of the variables that entered the models. However, we point out that the reversal of sign in the high school GPA variable and the UNCG GPA variables when entered together in Model 2 is likely an artifact of the pre-existing inter-relatedness of the two GPA variables. We made note in the text where we saw this issue and advised the reader to interpret those relationships with caution.

REFERENCES

- Agresti, A. (2013) *Categorical Data Analysis*. 3rd Edition, John Wiley & Sons Inc., Hoboken, NJ.
- Allison, P. D. (2012) *Logistic Regression Using SAS: Theory and Applications*, 2nd Edition. SAS Institute Inc., Cary, NC.
- Hosmer Jr., D.W., Lemeshow, S. and Sturdivant, R.X. (2013) *Applied Logistic Regression*, 3rd Edition. John Wiley & Sons, Inc., Hoboken, NJ. <https://doi.org/10.1002/9781118548387>
- Van Buuren, S, Boshuizen, H. C., & Knook, D. L. (1999) Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine*, *18*, 681-694.
- PROC MI and PROC MIANALYZE documentation can be found here:
- SAS Institute Inc. 2015. *SAS/STAT® 14.1 User's Guide*. Cary, NC: SAS Institute Inc.
- UNC System Office
https://myinsight.northcarolina.edu/t/Public/views/db_enroll/Demographics?iid=1&%3AisGuestRedirectFromVizportal=y&%3Aembed=y