

STOPOUTS AND READMITS:  
USING STUDENT RECORD AND NSC DATA TO PREDICT RE-ENROLLMENT

Jerret K. LeMay  
Research Analyst, Office of Institutional Research & Planning  
Binghamton University

Abstract

Previous cohort analyses demonstrate the importance of understanding stopout behavior, transfer behavior, and the multi-institutional portfolio that has become undergraduate education for many students. This study investigates an aggregate body of students who stopped out from Binghamton University. A person-period dataset incorporating student record data and National Student Clearinghouse (NSC) data is created and a logistic regression model is devised to predict the number of re-enrolling students.

The variables with the largest parameter estimates pertain to the number of semesters a student missed, and whether the student transferred to a two-year institution in New York. The introduction of each of the nine variables in the final model improved the fit and quality of the model and its aggregate predictive accuracy.

Introduction

Retention, persistence, and degree attainment are core components in the research of higher education. They are often considered the gilded measures of an institution or system's "success." Most studies in this arena, whether institutional, statewide, or national in perspective, are cohort analyses. Specifically, they tend to look at a group of first-time, full-time, degree-seeking freshmen and measure their status at the end of a pre-determined time period – be it retention to the second year, graduation by year four or six, or degree attainment by a particular age. What happens between the standard measurement periods often goes unaccounted.

For practical purposes, cohort analyses are not necessarily applicable or useful in enrollment management. Not everyone enters a four-year institution as a first-time, full-time, degree-seeking freshman. The student body at an institution at any given point in time may contain several cohorts of this type, along with cohorts of students who transferred-in, and students who may stop in and stop out at their leisure.

This study investigates an aggregate group of undergraduates who stopped out from Binghamton University, a highly selective public four-year institution. The goal is to predict, at the aggregate level, the number of undergraduates who will re-enter the institution in a given semester using student record data and data from the National Student Clearinghouse (NSC).

## Literature Review

Whether the research is couched in the language of retention/persistence or its opposite, student departure, most of the literature in this area focuses on student characteristics as determinants of success (DesJardins, Ahlburg & McCall, 2002, p.556). Common findings highlight the role of good academic preparation, GPA, socioeconomic status, financial aid, and even gender in the retention/departure dichotomy (for examples, see DesJardins et al., 2002; Bean & Metzner, 1985; Ronco, 1996; Berkner, He, & Cataldi, 2002). A laundry list of “risk factors” has even developed, which generally includes part-time enrollment, delaying entry into postsecondary education after high school, not having a regular high school diploma, having children, being a single parent, being financially independent of parents, and working full time while enrolled (Berkner et al., 2002).

From an institutional perspective, students who leave before completing a degree are considered dropouts (Berkner et al., 2002). From a system or national perspective, however, we can see the forks in the road – students may transfer to another institution(s), take a brief time off, or dropout of higher education altogether.

“For a variety of reasons, many institutions of higher education have never systematically tracked students who leave institutions before completing a degree” (DesJardins & Pontiff, 1999, p.1). DesJardins and Pontiff (1999) point out that tracking students beyond the walls of the matriculating institution has gained importance in the wake of the Student Right to Know and Campus Security Act (SRK) and the imposition of the IPEDS Graduation Rate Survey (GRS). While some institutions may prepare internal reports on tracking students who depart, there are few public studies documenting when leavers return to higher education and where these students decide to attend (DesJardins & Pontiff, 1999, p.1).

Well documented in the retention and attrition literature is that most students leave college early in their college careers (DesJardins & Pontiff, 1999, p.10). In studying 1995-96 beginning postsecondary students, Bradburn (2002) found that attrition is greatest in the first year of college, except at four-year institutions, where there was no difference in departure rates in the first two years.

Working with a slightly different cohort definition than Bradburn (2002), Berkner, He, and Cataldi (2002) found that among students beginning at four-year institutions with a bachelor’s degree goal, the cumulative percentage who had left their first institution *and transferred to a different one* was 10% by the end of the first year, 18% by the end of the second year, 22% by the end of the third year, and then about 23% through the end of the sixth year.

Among the 1995-96 beginning postsecondary student cohort, Berkner et al. (2002) found that 40% of students enrolled in at least one postsecondary institution other than the institution at which they started. About one-third (32%) of the beginners transferred from

their first institution to a different one, and 11% were sometimes co-enrolled, taking courses at more than one institution at the same time. About one-fourth of the students who began at four-year institutions and 42% of those who began at public two-year institutions transferred during the six years of the study.

Given that over a quarter of students who begin their post-secondary education at a four-year institution transfer to another (McCormick & Carroll, 1997, as cited in Porter, 2000), transfer-out behavior cannot be ignored (Porter, 2000). This is especially true when the data reveals that stopouts come back – not only to higher education, but often to the same institution that they left.

In *Stopouts or Stayouts? Undergraduates Who Leave College in Their First Year*, Horn (1998) found that 16% of 1989-90 beginning postsecondary students who enrolled in the four-year sector left before the beginning of their second year, but a majority (64%) of them returned to higher education within five years. Among these stopouts, 42% returned to the *same institution*, while 58% transferred elsewhere. Stopouts from four-year institutions who returned to the same institution returned to education earlier than their transferring counterparts: more than half (57%) returned in their second year, compared with 40% of transfers (Horn, 1998). Of students who left public four-year institutions and later returned, the vast majority (94%) did so in years two and three (Horn, 1998, p.14).

The fact that students return to the same institution they left runs counter to what we have come to “know” about stopout and transfer behavior, that is, “...students transfer in order to find a better fit with the institutional environment, whether that environment is defined in terms of academic program offerings, course availability academic standards, finances, or institutional culture” (Ronco, 1996, p.2).

Transferring, returning, or dropping out are not the terminal options for students who leave an institution. Students may stop out, transfer, and/or return to an institution multiple times, potentially spending periods of time co-enrolled at multiple institutions. Little research has been conducted on these multiple-transfer students, students who transfer more than once during their collegiate careers (Kearney, Townsend, & Kearney, 2002), perhaps because until recently, we've never known where the stopouts who returned to their first institution of attendance have been during the stopout period (Adelman, 1999).

In *Answers in the Tool Box*, Adelman (1999) investigated what factors contribute most to long-term bachelor's degree completion of students who attend four-year colleges (even if they also attend other types of institutions). This longitudinal study followed students who were high school sophomores in 1980 through their academic careers until they reached the age of thirty. In doing so, it provides a rare glimpse of long-term enrollment patterns of students – investigating and tracking enrollment beyond their first, second, or even third institutions.

The proportion of undergraduate students attending more than one institution swelled from 40% to 54% (and among bachelor's degree recipients, from 49% to 58%) during the 1970s and 1980s, with even more dramatic increases in the proportion of students attending more than two institutions (Adelman, 1999, pp.42-45). Adelman asserts that the increasing tendency for students to attend two, three, or more colleges (sometimes in alternating patterns, sometimes simultaneously) in the course of their undergraduate careers is overlooked in both policy and research.

The following are other transfer-related highlights from Adelman's study:

- 61% of those who attended two schools returned to the first institution of attendance at some point, as did 48% of those who attended three or more institutions (p.44).
- Students beginning in highly selective four-year colleges and those starting out in open door institutions have the highest rates of multi-institutional attendance (p.45).
- Sixteen percent of postsecondary students (and 18% of bachelor's degree completers) engaged in alternating or simultaneous enrollment patterns. Some 70% of this group attended three or more institutions (pp.45-46).

Given these and other findings, Adelman built a case that transfer behavior, as it has traditionally been viewed, has been replaced by "portfolio building" in an "open market," noting that the classic form of transfer is an extremely effective route to bachelor's degree completion. His penultimate linear regression model included two variables pertaining directly to transfer behavior, with one of the key factors involving multi-institutional attendance being whether the student *returned* to the referent first institution (Adelman, 1999, p.46). The number of institutions attended by students was found to have no effect on degree completion (Adelman, 1999, p.68). Regardless of transfer behavior, continuous enrollment was found to be very important, as the completion rate for continuously enrolled students is two times that for non-continuously enrolled students.

The literature thus reveals that transfer behavior is prevalent (e.g. Berkner et al.), and more common than in the past (e.g. Adelman). It also illustrates that stopouts often return to the same institution they left (e.g. Horn, Adelman), and relatively quickly (e.g. Horn). Students who leave an institution should not be viewed so much as dropouts, but rather as potential returners. It is in this context that this paper addresses what is a practical concern for enrollment management efforts – predicting the number of students who will return.

## Methods

Undergraduate student record data from Fall 1992 through Spring 2003 were compiled in what Ronco (1996) referred to as a person-period dataset. Existing student record information from semesters in which the students were enrolled was supplemented by inserting blank records for subsequent semesters in which they were not enrolled. A stopout indicator was created to flag instances where a student did not receive a degree and did not enroll at BU in the next major semester. The dataset was then limited only to those students who had stopped out at some point.

Information was submitted to the National Student Clearinghouse's (NSC's) Enrollment Search service for each stopout instance in each student's history. In all, data on 12,654 instances of degree-seeking and non-degree undergraduate stopouts were submitted. NSC allows submission of an extra field to be used at the discretion of the submitter. That field was populated with a student identifier code and a semester variable (year/semester) so that returned NSC data could be merged with the student record data.

Data returned from NSC included 17,254 records. Dates of initial enrollment were converted to a semester format used by BU. Business rules were also established and implemented for instances that appeared to represent co-enrollment.

The student record data and NSC data were merged and variables were created and/or recoded to produce independent variables. The student record data supplied demographic variables, as well as GPA data, major, student level, financial aid data, matriculation status, etc. The NSC data provided variables indicating instances of transfer and the type and location of transfer institution(s).

Last known student record attributes were retained or "dragged forward" through semesters in which students were not enrolled. For example, the gender flag "female" was populated in blank records. As the NSC data pertained to only the initial enrollment at an institution (rather than term-by-term data), the same "dragging forward" procedure was applied to variables derived from the NSC data. Additional flags were created, including "semesters missed" count, a count of the number of stopout instances for each individual, and an institution count for each instance of transfer. The dependent variable (ugread) was also created to signify instances where a student was readmitted as an undergraduate student in the next semester.

The model being devised would predict re-enrollment / readmittance in a Fall semester. Therefore, the dataset was pared down to data available for non-enrolled students as of spring semesters. In order to allow an adequate number of cases for time-based variables (such as missing ten or more semesters), only the previous five springs (1999-2003) were included.

As an initial step, frequencies were run comparing returning students (ugread=1) with non-returning students. These were expressed as a percentage, and the difference in the percentages calculated. For instance, 49% of returning students had indicated that BU was their institution of first choice, compared to 27% of non-returning students, for a difference of 22 points.

Following the initial investigation of frequencies, correlations were run with the dependent variable (ugread) and all of the potential independent variables. These were sorted by groupings of their *p* values (<.1, .1-.199, .2-.299, etc.) and their strength (absolute value of the coefficient).

Logistic regression was utilized for the multivariate model because the dependent variable was binary (0/1). Since the goal of the project was to predict, on the aggregate, the number of students who would be readmitted, three criteria were established for keeping variables in the model: the variable needed to be significant, it had to improve the fit and quality of the model (decreasing -2 log likelihood, increasing c), and it had to improve the aggregate predictive accuracy such that the absolute difference between the predicted and the actual number of readmitted students decreased.

## Results

When the frequencies were run, the variables with the largest differences between returning students and non-returning students related to the number of semesters a student had missed, financial aid information, the type of institution to which the student transferred (if they did transfer), and whether BU was the institution of first choice. The correlations reflected what was observed in the frequencies, and also provided a sorted list for introducing variables into the multivariate model.

Table 1 displays the results of an intercept-only model where no independent variables were included in the model statement. This provided a baseline against which various models could be compared. The -2 log likelihood was 6167.684, and the aggregate predictive error was 77 (Table 2).

Table 1

### *Logistic Regression Output for Intercept-Only Model*

(response variable ugrad: 1=readmitted as undergraduate 0=not)  
(n=39,858)

-2 Log L = 6167.684

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.1929	0.0414	10270.1	<.0001

Table 2

### *Aggregate Predictive Error of Intercept-Only Model*

Spring	Readmits next semester	Predicted	Difference	Absolute difference
1999	85	88.03	3.03	3.03
2000	122	103.65	-18.35	18.35
2001	140	119.87	-20.13	20.13
2002	131	132.95	1.95	1.95
2003	115	148.50	33.50	33.50

76.96

In accordance with the first criterion, all variables in the final model were significant at the .1 level (see Table 3). The second criterion was also met, as the  $-2$  log likelihood was reduced considerably, and a  $c$  of .865 was attained. The introduction of each variable produced an improvement in these measures as the model was built.

Table 3

*Logistic Regression Output for Final Model*

(response variable ugrad: 1=readmitted as undergraduate 0=not)  
(n=39,858)

-2 Log L Intercept Only Model	6167.684
-2 Log L Final Model	4881.798
$c$	0.865
Chi-Squared	1289.886
Significance	<.0001

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.2566	0.0834	2607.1542	<.0001
Missed 1 semester	1	1.6903	0.0902	351.4925	<.0001
Medium financial need	1	0.2232	0.1073	4.3250	0.0376
Trans to NY 2yr	1	0.9225	0.0935	97.2525	<.0001
Missed 10+	1	-2.5633	0.2478	106.9605	<.0001
2 years at BU	1	0.6296	0.1080	33.9980	<.0001
Missed 7-10	1	-1.8597	0.2073	80.5039	<.0001
2nd semester senior	1	-0.6909	0.1444	22.8928	<.0001
PLUS Loan amount	1	0.000046	0.000028	2.7773	0.0956
BU first choice	1	0.1971	0.1025	3.6973	0.0545

The variables with the largest parameter estimates are all measures of the number of semesters a student missed (missed one semester, missed 7-10, missed 10+). As time passes, stopouts are less likely to return. The fourth largest effect comes from a variable derived from the NSC data indicating that the student transferred to a two-year institution in New York. Students who achieve second-semester senior status are less likely to return, perhaps because they have found employment in their field. It is interesting to note that the only grant or loan variable to stay in the model represents a parent loan.

The aggregate predictive error for the final model was 55 (Table 4), compared to the baseline intercept-only model's 77 (Table 2). The introduction of each variable produced a decrease in this number as the model was built.

Table 4  
*Aggregate Predictive Error of the Final Model*

Spring	Readmits next semester	Predicted	Difference	Absolute difference
1999	85	100.75	15.75	15.75
2000	122	113.00	-9.00	9.00
2001	140	127.34	-12.66	12.66
2002	131	125.00	-6.00	6.00
2003	115	126.92	11.91	11.91
				55.32

Several variables were significant and improved the quality and fit of the model, but did not aid in improving the aggregate predictive ability. As the model was being built, it was run on the entire five-spring dataset, as well as being run on each spring separately. Indicators of significance, standard errors, and the parameter estimates were collected in a dataset and reviewed. Those variables that attained significance and produced a better model but did not aid in improving the aggregate predictive ability of the model displayed variation from year to year in terms of their significance, standard errors, and/or the strength and direction of the parameter estimates.

Parameter estimates from the final model were applied to data from the spring of 2004, and 138 returning undergraduates were predicted for the Fall 2004 semester. However, an average of eighteen students return in Fall semesters who stopped out prior to 1992, and thus are not accounted for in the original dataset nor the regression above. Adding this average to the predicted 138 would yield an overall prediction of 156. Actual readmitted undergraduates for Fall 2004 numbered 169. Table 5 displays the results of the logistic regression method and different methods that have been used in the past to estimate the number of re-enrolling/readmitted undergraduates. The new methods performed better than any of the previous attempts.

Table 5  
*Results and Comparison With Previous Methods*

	<b>Prediction</b>	<b>Actual</b>	<b>Difference</b>
<b>New method</b>	156	169	-13
<b>Old methods:</b>			
Previous year	134	169	-35
3 year average	147	169	-22
5 year average	140	169	-29

## Discussion / Conclusions

In a break from previous studies, this analysis opted not to explore theoretical underpinnings of student retention and attrition of freshman cohorts. Instead, it utilized a person-period data structure that included student record and NSC variables for all students who stopped out between Fall 1992 and Spring 2003, regardless of whether or not they entered Binghamton University as full-time, degree-seeking freshmen. The focus was of an applied nature, with practical application in Enrollment Management efforts.

Some data elements commonly found to be determinants of success or risk factors were simply not available for inclusion in this study. For instance, there were no indicators for employment or family situations. Employment situations *following departure* might be especially instructive.

The NSC's Enrollment Search service now provides a key element that was missing in previous research – the ability to track students after they leave an institution. The coverage is not perfect, however. The NSC now boasts 91% “coverage,” meaning that the vast majority of students enrolled in postsecondary education in the U.S. are enrolled in institutions that participate in the NSC program. Coverage in prior years was not as extensive, so there may be data concerns the further back one delves into history. In this study, considerable back-year data was required, and still there was a need to supplement the model prediction with estimates for pre-1992 stopouts.

The NSC data only supplies the date of a student's initial enrollment. As DesJardins and Pontiff iterated in their 1999 study, term-by-term information might be more instructive. This may be especially true in light of Adelman's finding that continuous enrollment is a significant predictor of degree attainment. Degree attainment data might also aid in building a better model. Still, the NSC data currently available is a considerable improvement over what was available in the past, and produced one of the largest parameter estimates in the model.

Focusing on practical application, this study eschewed the common freshman cohort analysis. The end result was an improvement in the prediction of re-enrolling/readmitted students. For Binghamton University, the number of readmitted undergraduate students is a very small portion (1.5%) of the undergraduate student body. The method of integrating student record and NSC data in a person-period dataset presented here could prove useful for enrollment management efforts at other institutions, especially those that serve a more transient undergraduate populace. It could also be used for other purposes such as testing theoretical precepts in the current literature, or in marketing efforts that rely on identifying individuals most likely to return to an institution.

## References

- Adelman, C. (1999). *Answers in the Tool Box: Academic Intensity, Attendance Patterns, and Bachelor's Degree Attainment*. Washington, DC: U.S. Department of Education.
- Bean, J.P. and Metzner, B.S. (1985). A Conceptual Model of Nontraditional Undergraduate Student Attrition. *Review of Educational Research* 55 (4): 485-540.
- Berkner, L, He, S., and Cataldi, E.F. (2002). *Descriptive Summary of 1995 –96 Beginning Postsecondary Students: Six Years Later*. Washington DC: U.S. Department of Education. National Center for Education Statistics.
- Bradburn, E.M. (2002). *Short-Term Enrollment in Postsecondary Education: Student Background and Institutional Differences in Reasons for Early Departure, 1996 –98*. Washington DC: U.S. Department of Education. National Center for Education Statistics.
- DesJardins, S.L., and Pontiff, H. (1999). Tracking Institutional Leavers: An Application. *AIR Professional File* 71.
- DesJardins, S.L., Ahlburg, D.A., and McCall, B.P. (2002). A Temporal Investigation of Factors Related to Timely Degree Completion. *The Journal of Higher Education* 73(5): 555-581.
- Horn, L. (1998). *Stopouts or Stayouts? Undergraduates Who Leave College in Their First Year*. Washington DC: U.S. Department of Education. National Center for Education Statistics.
- Kearney, G.W., Townsend, B.K., and Kearney, T.J. (2002). Multiple-Transfer Students in a Public Urban University: Background Characteristics and Interinstitutional Movements. *Research in Higher Education* 36(3): 323-44.
- Porter, S.R. (2000). Including Transfer-Out Behavior in Retention Models: Using the NSLC Enrollment Search Data. *Presented at the 40<sup>th</sup> Annual Association of Institutional Research Forum, Cincinnati, Ohio, May 21-24, 2000*.
- Ronco, S.L. (1996). How Enrollment Ends: Analyzing the Correlates of Student Graduation, Transfer and Dropout with a Competing Risks Model. *AIR Professional File* 61.