Analyzing Time to Student Course Withdrawal Patterns for Predictive Modeling

Gillian M. Nicholls[1]

Abstract – The University of Alabama in Huntsville has a generous policy for course withdrawal. Many students attempt a course for several weeks before withdrawing if they have trouble. University student throughput is hampered and course resources are underutilized. In Fall 2011, 16.2% of the undergraduate engineering economy students withdrew. A study was done to collect data about the students' gender, academic backgrounds, transfer student status, major, homework grades, exam grades, and whether students were taking the class for the first time. Homework assignment submission and exam-taking were used to establish a date at which students stopped actively participating in the class. The time to event data was analyzed to identify significant factors predicting withdrawal. Gender and transfer student status were statistically significant predictors at the start of classes. Once course milestone Exam 1 and 2 scores become available, gender and transfer status become less valuable predictors.

Keywords: Survival Analysis, STEM education, persistence

INTRODUCTION

The number of students graduating with a degree in Science, Technology, Engineering, or Mathematics (STEM) is an important subject for the economic well-being of the United States and other nations. As such, the entry into a STEM program of study and persistence to degree acquisition is of great societal interest. Persistence can be modeled on a "macro" scale as students complete successive courses over a period of years to earn their degrees. Persistence can also be examined on a "micro" scale as students submit assignments and take exams in the process of completing a single course. This study analyzes persistence throughout a single course and identifies factors that can be used to identify students who might have difficulty successfully completing it. The ability to successfully predict which students are at greatest risk of withdrawing from a course will better enable instructors, administrators, and policy makers to offer support to these students.

The University of Alabama in Huntsville (UAHuntsville) maintains a very generous course withdrawal policy with students able to withdraw at will at any point up until a deadline approximately one month before the end of a semester. This policy allows students to register for a class, attempt it for several weeks, and then withdraw if they feel they will not be successful. A withdrawal appears on the student's transcript as a "W" but it does not directly affect the Grade Point Average (GPA). Students cannot obtain a refund of tuition expenses for the class if they withdraw, but there is no penalty for an aborted attempt at the class. However, there are some undesirable results from a course withdrawal for both the student and the university. The student has lost the value of the tuition dollars for the class thereby increasing the cost of the education and any student loan balance; slowed his/her progression through the program; may experience difficulties with taking courses in a non-standard sequence; and placed a less attractive entry on his/her academic transcript. The university has assigned classroom resources (seats and instructional capacity) that were not fully utilized, and the slower progression of students through the program hampers the university's student throughput. Another potential "cost" of high withdrawal rates are that the more difficulty STEM students have in completing required coursework, the more opportunities there are for students to become discouraged and leave the STEM path.

[1] UAHuntsville, Technology Hall N143, Huntsville, AL 35899, gillian.nicholls@uah.edu

The undergraduate Engineering Economy course at UAHuntsville is a good candidate for micro scale engineering education research since it is a required course for most engineering majors, it is taught year round with sections that fill up very quickly, and the withdrawal rate is high. The course is required for engineering students majoring in the aerospace, civil, electrical, industrial, mechanical, and optical disciplines. A course cannot be used as a pre-requisite for subsequent coursework unless the student earns a C or better. While this course is not a required pre-requisite for the chemical or computer engineering students, many opt to take the course anyway to fill out their credit load per semester. This course has no pre-requisites other than sophomore standing, and is often taken in the first or second semester by students transferring into UAHuntsville after completing their first two years of required courses at a community college.

Literature Review

Extensive academic research has been conducted to advance understanding of how students are drawn towards or away from STEM; identify factors that suggest which students will have difficulties in their studies; and determine the factors that affect students' persistence in STEM to ultimately achieve a degree. Sax [17] identified factors that predicted students earning a bachelors degree in STEM choosing to continue their careers in a STEM graduate degree program and found gender was a significant factor. Females were more likely to pursue a STEM career if they had a philosophical commitment to scientific contribution or if their mother had a research career. Leslie, McClure, and Oaxaca [11] developed predictive models of STEM degree acquisition and found that lower levels of self-concept and self-efficacy hampered the rate of participation by females and racial/ethnic minorities in STEM. Students with a parent involved in STEM as a role model had a greater likelihood of being drawn towards STEM. Besterfield-Sacre, Atman, and Shuman [3] assessed the impact of personal self-confidence, attitudes, and expectations of freshmen engineering students on their willingness to remain in STEM. They found that students with a lower interest in engineering were much more likely to leave STEM despite good academic performance. Adelman [2] examined longitudinal data following students from high school through college and concluded students choosing engineering were those that had demonstrated greater academic skill and performance. This was true for students of both genders. Nicholls, Wolfe, Besterfield-Sacre, Shuman, and Larpkiattaworn [15] found a number of significant differences between students choosing STEM vs. Non-STEM majors. These included high school GPA, standardized test scores, student's self-rating of academic skills, and their level of personal interest in STEM. Seymour and Hewitt [18] found females were more likely than males to have been influenced towards STEM by a person close to them and more highly weighted personal satisfaction than career goals. They also found that students who remained in STEM tended to have a greater interest in STEM and developed better coping mechanisms.

Persistence in STEM has been examined in multiple ways. Smyth and McArdle [19] found persistence was greater in cases where students more closely matched the mean student performance in high school and standardized test scores. Zhang, Anderson, Ohland & Thorndyke [21] identified the same factors and also gender, race/ethnicity, and citizenship as significant predictors of STEM persistence to graduation. Mendez, Buskirk, Lohr, and Haag [12] confirmed the predictive value of high school and freshman year GPA as well as the number of STEM courses taken during the freshman year in examining STEM persistence. Prior research in predicting STEM persistence was reported in Nicholls, Wolfe, Besterfield-Sacre, and Shuman [14]. Variables that measured students' skills in science, mathematics, and reading and level of academic commitment were significant predictors of STEM persistence. Family encouragement and support for academic success was another significant predictor.

Survival analysis techniques have been successfully utilized in studying STEM persistence. Willett and Singer [20] applied survival analysis to analyze high school student dropout and teacher attrition. Ronco [16] used survival analysis to study the phenomena of college students dropping out. Civian [1] reported the process of students earning a doctorate showed significant differences by race and citizenship with white and American students taking longer to complete a degree than nonwhite or foreign students. Chimka, Reed-Rhoads, and Barker [4] used proportional hazards models to examine engineering students' persistence to graduation and found standardized test scores in math and science combined with gender were significant factors after controlling for major, hometown population size, and state of residence. Min, Zhang, Long, Anderson, and Ohland [13] found significant differences by race/ethnicity and gender in when students leave engineering. Scores on SAT math exams were stronger predictors of the risk of departing STEM than scores on SAT verbal exams. Females and white students were at greater risk of departing the STEM path earlier than other students. Overall, the prior research indicates that survival analysis is a strong tool for identifying significant factors in student persistence to degree and determining the points at which support for continuing is most needed.

DATA COLLECTION

The data for this study was collected from two sections of Engineering Economy taught in Fall 2011 by two different professors. Both professors were full time faculty members of the Industrial & Systems Engineering & Engineering Management (ISEEM) Department. One professor was tenured and the other was tenure-track. A total of 179 students were enrolled across the two sections past the two-week add/drop period. Students who dropped the course were excluded from the study. The course design elements included in person lectures, regular homework assignments, and four non-cumulative exams. Homework problems were assigned in advance and then submitted by logging into the Course Management System (CMS) and filling out an automated survey that asked questions about the answers to the problems. Students were given the option of repeating each homework survey up to three times per assignment with the average percentage score of the attempts recorded as the final score for that assignment. The exams were conducted in class and featured conceptual questions and quantitative problems in a multiple choice format.

The ANGEL™ CMS owned by Blackboard, Inc. was the source of data about homework completion and grades. Exams were manually graded and the scores were logged into the CMS. The CMS gradebook was downloaded in Microsoft Excel™ to obtain the students' grades and begin the process of organizing the database. Exam and homework scores were recorded on a percentage basis and treated as continuous variables (0 - 100). Students were also asked to fill out an initial background survey posted in ANGEL at the start of the semester. This survey collected data about students' major(s), how many semesters they had been at UAHuntsville, credit load during the semester, transfer student status, and dual enrollment at a school besides UAHuntsville during the semester. The survey data was also downloaded initially into Excel. Missing data from students that did not complete the background survey was collected from UAHuntsville's student information database. Transcript records from the student database were also used to confirm whether or not each student was attempting engineering economy for the first time or whether his/her Fall 2011 enrollment was a second or third attempt. Nine students had already attempted the course at least one time before. Demographic data was not collected with the exception of gender and transfer student status. Table 1summarizes the student outcome by the two major demographic groups, gender and transfer student status. In Fall 2011, 29 of 179 students (16.20%) withdrew from the Engineering Economy classes. A total of 49 out of 179 students (27.37%) either withdrew or earned a grade (D or F) that will require retaking the course. This outcome means a lost opportunity for many students and a great opportunity to improve.

Table 1 Course Outcome by Student Group

Transfer Status	Student Gender	Course Grade Outcome			Total
		Withdrawn	D or F	A, B, or C	
No	Female	3	1	11	15
	Male	2	5	51	58
Yes	Female	10	4	14	28
	Male	14	10	54	78
Total		29	20	130	179

All personally identifiable data about the students was removed from the dataset prior to analysis. Unique but anonymous student identification code numbers were used to distinguish between the individual students. Additional variables were constructed to reflect the students' course outcomes, date of last active course participation, type of last active course participation, time (in days) to last active participation, and whether the time to withdrawal was observed or "censored" [Klein, 10]. The additional variables constructed reflected ordinal categorical, nominal categorical, binary, and continuous data. Some of the binary variables were created to reflect a dichotomous categorization of another variable. For example, a binary variable of "First Semester at UAHuntsville" was created to reflect whether Fall 2011 was (value = 1) or was not (value = 0) the student's first semester after enrollment at UAHuntsville. A binary variable was constructed to indicate whether each student had completed or withdrawn from the course. Another binary variable was constructed to indicate whether each student had stopped actively participating or not during the semester.

The major event of interest in this analysis was the time at which a student stopped actively participating in the class. Each student record was scrutinized to determine whether all exams and homeworks were completed, and if not, when the student's last activity took place. The majority of the students continued participating up through the last exam and completed the course with a C or better or with an unsuccessful outcome (D or F). In this case, the event of interest (withdrawal) did not take place and the time to last active participation was recorded as 100 days, the length of the course. These observations were recorded as having been "censored" indicating that the event of interest was not observed during the study. If a student took part in the exams and homework assignments before the withdrawal deadline and then withdrew, his/her time to last participation was set to 73 days, the university's deadline for withdrawal. Several students that withdrew had stopped completing assignments and/or taking exams prior to the withdrawal deadline. In these cases, the time to last active participation was set to the later of the last exam or homework assignment completed. One of the 29 students that withdrew from the class completed Exam 4 and then retroactively withdrew past the original deadline. This student's time to last active participation was recorded as the date of Exam 4, a total of 98 days into the semester. Three students remained in the class past the withdrawal deadline but stopped actively participating after either the second or third exam. These students were treated as having functionally withdrawn from the course for this analysis. Students may functionally withdraw while remaining in the course because they missed the deadline and lack reasons for a late withdrawal or because they must keep a certain credit load to be eligible for financial aid. The earliest withdrawal time was 28 days, the time at which Exam 1 was taken. It should be noted that the homework deadlines were the same across both sections, and the course withdrawal deadline was constant across the university. The exam dates were closely synchronized but varied by one day across sections. This difference was judged to be minor and so the exam dates for the larger section 1 were used in calculating the time to last active exam participation across both sections.

ANALYSIS AND FINDINGS

The data was analyzed with IBM® SPSS PASW Statistics 18™. The examination of the data with survival analysis techniques began by constructing a Kaplan-Meier life table [Kaplan, 6] of the records and the hazard function. The hazard function [Klein, 8] of continuous survival data plots the probability that the event of interest occurs in the next time interval given that it has not occurred up to that point in time. In this analysis, it depicts the probability that a student having remained active in the class to a certain point is at risk of functionally withdrawing in the short time interval immediately after that point in time. Figure 1 shows the overall hazard function for the sample of 179 students. There are points where the hazard function increases sharply indicating a higher risk of students ending their active participation in the class. These points are the date of Exam 1 (28 days), Exam 2 (56 days), Homework #7 due date (65 days), the Withdrawal deadline (73 days), and Exam 3 (77 days). Figure 2 shows the same data when broken down by the factor of student gender. Male and female students start off at comparable risk, but the hazard function for female students rises quickly after Homework #5 (47 days). This is the first homework assignment that requires students to have synthesized the core concepts of the time value of money, cash flow diagram interpretation, single vs. multiple repeating cash flows, and present worth analysis that underlie the remainder of the course. Figure 3 shows the hazard function for transfer students vs. non-transfer students. The results in this case are even more striking. Transfer students are at a much higher risk of withdrawing from the course as early as Exam 1. This finding remains when genders is added as a strata of the transfer student status. Transfer students are at a higher risk of withdrawal for both genders. Females students are a higher risk of withdrawal regardless of transfer status.

Table 2 shows the mean survival time [Klein, 9] in days for the students broken down by gender and transfer student status. Female transfer students have the shortest mean time to last active participation with 84.467 days and a standard error of 3.869 days as compared to female non-transfer students with 92.933 days and a standard error of 3.736 days. The standard errors for the male students and the overall categories were less than 2.0 days. This is a result of the sample having a smaller proportion of females than males (43 of 179 or 24.02%) as is common in engineering classes and much more variability in withdrawal times. Note that although females were only 24.02% of the section enrollment, they made up 13 of the 29 students (44.83%) that withdrew and 1 of 3 students that functional withdrew.

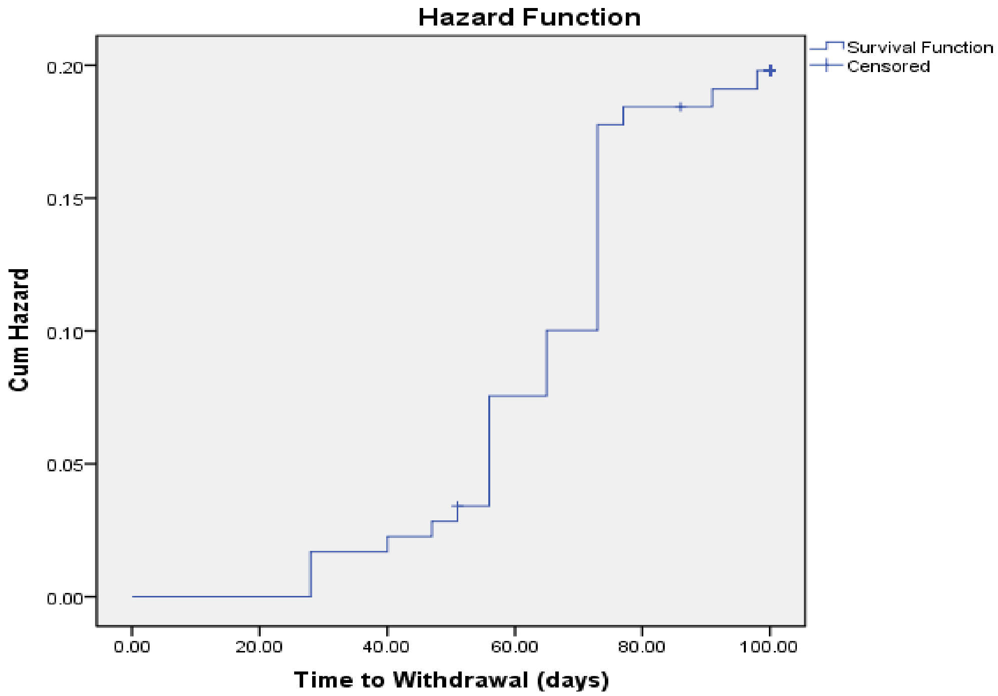

Figure 1 The Hazard Function for the Entire Sample Size

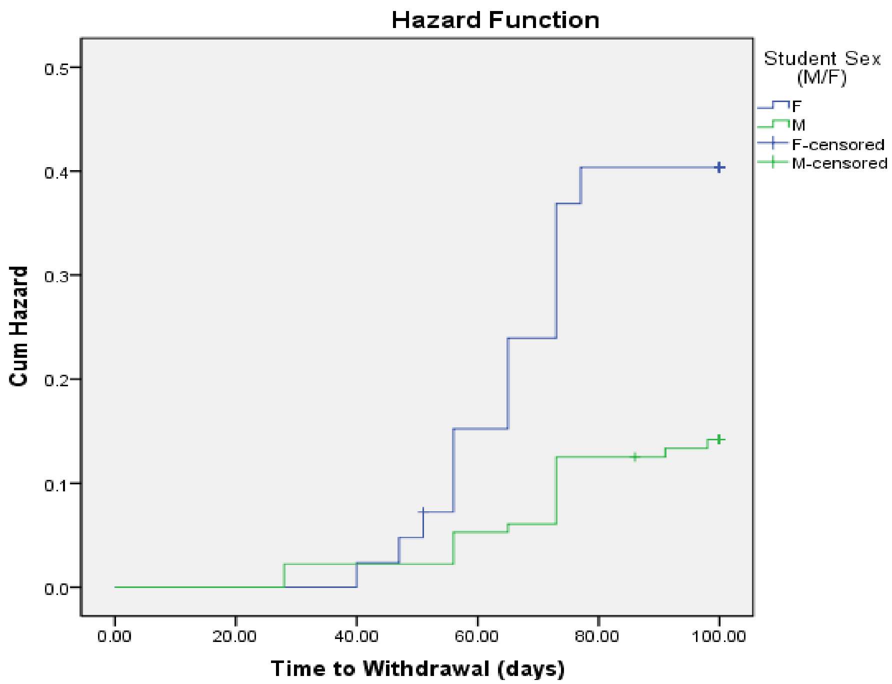

Figure 2 The Hazard Function Factored by Student Gender

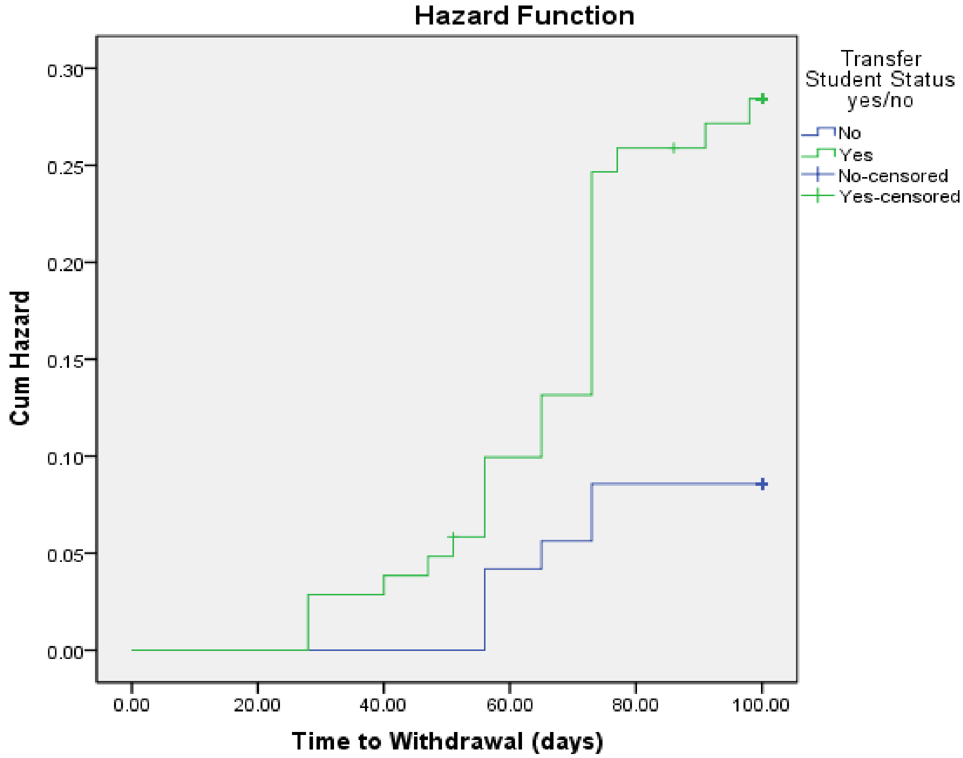

Figure 3 The Hazard Function Factored by Transfer Student Status

Table 2 Mean Survival Time in Days by Student Category

Gender	Transfer Student	Non-Transfer Student	Overall
Female	84.467	92.933	87.455
Male	93.088	98.017	95.190
Overall	90.858	96.973	93.359

Cox Regression Analysis

Models were fitted using the Cox Proportional Hazards method [Cox, 5] to identify the most significant covariates for predicting time to course withdrawal. A forward selection method based on the likelihood ratio statistic was utilized with the *p*-value limits for a variable to enter or leave the model set to 0.05 for entering and 0.10 for leaving. The overall model strength was evaluated with -2 Log Likelihood and Chi-square statistics [Klein, 7]. Individual variables in the model were evaluated with Wald statistics.

The first model fitted examined variables available at the start of the course including student gender, transfer student status, repetitive enrollment, binary variables for engineering majors, dual major status, and first semester at UAHuntsville. The final model was found to be significant with only two variables selected: student gender and transfer student status. Both were extremely strong predictors with the model indicating that female and transfer students were at much greater risk of withdrawing.

The second model fitted examined variable that were available partway through the course including scores for Homework assignments 1-5 and Exam 1 in addition to all the initial variables. This model demonstrated that actual in-course performance outweighed the other variables in predictive utility. The final model selected Exam 1 and Homework 4 scores for inclusion. Both variables were extremely strong predictors and the overall model strength was excellent. The results confirmed that not surprisingly, higher scores on Exam 1 and Homework 4 were negatively associated with the risk of course withdrawal.

The third model was fitted with the consideration of still more variables that became available later in the course: scores for Homework assignments 6 and 7 as well as Exam 2. This model found Exams 1 and 2 as well as Homework 7 to be the most significant predictors. This makes logical sense as the materials on these Exams encapsulated the core concepts from the class and students having difficulties on them would be likely to struggle if they remained in the class. While the exams were not cumulative in the chapter coverage, the core concepts carry through from early chapters throughout the later chapters. Each successive model improved in statistical strength as more and more performance data became available.

The findings indicate that is possible to identify students early in the semesters who are at greatest risk of withdrawing and that as early performance data becomes available; the students can be more precisely identified to offer support. The time points of greatest risk of withdrawal match key dates in the semester including the timing of exams, critical homework assignments, and obviously the university deadline for course withdrawal.

FUTURE WORK

There are several areas in which future research is envisioned. There are additional academic performance and demographic variables that can be tested as covariates in predictive modeling of student withdrawal vs. completion and success vs. failure. These include prior GPA scores, standardized test scores, student age, race/ethnicity, traditional vs. non-traditional students, employment status, and motivational level. Subsequent surveys of students in Engineering Economy have attempted to gather information about students self-rating of their facility with word problems requiring interpretation rather than solely quantitative analysis. Survey questions have been added to ascertain what grade outcome would be satisfactory and acceptable to the students as well as whether withdrawing is even an option. Controlling for prior academic performance has the potential to improve the predictive utility of the modeling as students are just entering a course. Lecture attendance is another variable that could be tracked for modeling purposes and to more closely identify when a student ceased actively participating.

Transfer students and/or part-time students are an increasing fact of academia as financial pressures and other commitments encourage students to seek ways to reduce the cost of financing a bachelor’s degree. Transfer students in general have been the subject of research, but the subset of transfer students seeking STEM degrees has not received as much attention. With more students transferring in from less expensive two year colleges or simply transferring due to relocation, career changes, better cultural fit etc. it’s important to ensure this category of students entering the STEM pipeline are supported to become more successful.

Another direction of future work is expanding the size of the dataset for analysis. This is ongoing as additional sections are taught each semester at UAHuntsville. Over time, the phenomena of students repeating the course multiple times may produce enough records for a meaningful examination of the challenges these students face and a means of identifying them before this pattern emerges. The findings may spark discussion of the withdrawal policy and whether it could be adjusted to promote better student outcomes.

Ultimately, the goal of this micro scale STEM persistence research is to develop greater accuracy in predicting the macro scale persistence of students through core classes towards achieving a STEM degree. STEM degree acquisition is a process in which universities apply constrained resources to develop incoming students into graduates. Micro and macro scale persistence directly affect a university’s student throughput. Better persistence modeling should lead to a greater ability to model student throughput and the STEM pipeline.

REFERENCES

[1] Adelman, Cliff, “*Women and men of the engineering path: A model for analyses of undergraduate careers*,” PLLI 98-8055, U.S. Department of Education, Office of Educational Research and Improvement, Washington, DC: Government Printing Office, 1998.

[2] Besterfield-Sacre, Mary, Cynthia J. Atman, and Larry J. Shuman, “Characteristics of freshmen engineering students: Models for determining student attrition in engineering,” *Journal of Engineering Education* 86 (2), 1997, pgs. 477-489.

[3] Chimka, Justin R., Teri Reed-Rhoads, and Kash Barker, “Proportional hazards models of graduation,” *Journal of College Student Retention: Research Theory and Practice*, 9, (2), pgs. 221-232.

[4] Civian, Janet Trabucco, “*Examining duration of doctoral study using proportional hazards models*,” an unpublished dissertation, Harvard University Graduate School of Education, 1990.

[5] Cox, D.R., "Regression models and life-tables," *Journal of the Royal Statistical Society, Series B (Methodological)*, 34, (2), 1972, pgs 182-220.

[6] Kaplan, E.L. and P. Meier, "Nonparametric estimation from incomplete observations," *Journal of the American Statistical Association*, 53, 1958, pgs 457-481.

[7] Klein, John P. and Melvin L. Moeschberger, *Survival Analyis Techniques for Censored and Truncated Data* 2nd edition, New York City, NY, Springer-Verlag, 2003, pgs 63-74.

[8] Klein, John P. and Melvin L. Moeschberger, *Survival Analyis Techniques for Censored and Truncated Data* 2nd edition, New York City, NY, Springer-Verlag, 2003, pgs 27-32.

[9] Klein, John P. and Melvin L. Moeschberger, *Survival Analyis Techniques for Censored and Truncated Data* 2nd edition, New York City, NY, Springer-Verlag, 2003, pgs 32-36.

[10] Klein, John P. and Melvin L. Moeschberger, *Survival Analyis Techniques for Censored and Truncated Data* 2nd edition, New York City, NY, Springer-Verlag, 2003, pgs 243-283.

[11] Leslie, Larry L., Gregory T. McClure, and Ronald L. Oaxaca, "Women and minorities in science and engineering: A life sequence analysis," *The Journal of Higher Education*, 69 (3), 1998, pgs. 239-276.

[12] Mendez, Guillermo, Trent D. Buskirk, Sharon Lohr, and Susan Haag, "Factors associated with persistence in science and engineering majors: An exploratory study using classification trees and random forests," *Journal of Engineering Education* 97, (1), 2008, pgs 57-70.

[13] Min, Youngkyoung, Guili Zhang, Russell A. Long, Timothy J. Anderson, and Matthew W. Ohland, "Nonparametric survival analysis of the loss rate of undergraduate engineering students," *Journal of Engineering Education* 100, (2), 2011, pgs 349-373.

[14] Nicholls, Gillian M., Harvey Wolfe, Mary Besterfield-Sacre, and Larry J. Shuman, "Predicting STEM degree outcomes based on eighth grade data and standard test scores," *Journal of Engineering Education* 99, (3), 2010, pgs 209-223.

[15] Nicholls, Gillian M., Harvey Wolfe, Mary Besterfield-Sacre, Larry J. Shuman, and Siripen Larpkiattaworn, "A method for identifying variables for predicting STEM enrollment," *Journal of Engineering Education* 96, (1), 2007, pgs 33-44.

[16] Ronco, S.L., "Meandering ways: Studying student dropout with survival analysis," *Proceedings of the Association for Institutional Research 34th Annual Forum*, New Orleans, LA, 1994.

[17] Sax, Linda J., "The impact of college on post-college commitment to science careers: Gender differences in a nine-year follow-up of college freshmen," *Proceedings of the Annual Meeting of the Association for the Study of Higher Education*, Memphis, TN, 1996.

[18] Seymour, Elaine and Nancy M. Hewitt, *Talking about Leaving: Why Undergraduates Leave the Sciences*, Boulder, CO: Westview Press, 1997.

[19] Smyth, Frederick L. and John J. McArdle, "Ethnic and gender differences in science graduation at selective colleges with implications for admission policy and college choice," *Research in Higher Education*, Vol.l 45, No. 4, June 2004, pgs. 353-381.

[20] Willett, John B. and Judith D. Singer, "From whether to when: New methods for studying student dropout and teacher attrition," *Review of Educational Research*, 61 (4), 1991, pgs. 407-450.

[21] Zhang, Guili, Timothy J. Anderson, Matthew W. Ohland, and Brian R. Thorndyke, "Identifying factors influencing engineering student graduation: A longitudinal and cross-institutional study," *Journal of Engineering Education* 93, (4), 2004, pgs 313-320.

Gillian M. Nicholls

Dr. Nicholls is an Assistant Professor of Industrial & Systems Engineering & Engineering Management and a 2009-2010 Gray Faculty Fellow at the University of Alabama in Huntsville. Her research interests are in applying statistical analysis and optimization to supply chain management, transportation management, and engineering education. She holds the B.S. in Industrial Engineering (Lehigh University), Masters in Business Administration (Penn State University), M.S. in Industrial Engineering (University of Pittsburgh.), and Ph.D. in Industrial Engineering (University of Pittsburgh).