Essays on the Economics of Education and Health

by Joshua Averett Price

This thesis/dissertation document has been electronically approved by the following individuals:

Ehrenberg,Ronald Gordon (Chairperson)

Jackson,Clement Kirabo (Minor Member)

Cawley,John H. (Minor Member)

ESSAYS ON THE ECONOMICS OF EDUCATION AND HEALTH

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Joshua Averett Price

August 2010

ESSAYS ON THE ECONOMICS OF EDUCATION AND HEALTH

Joshua Averett Price, Ph. D.

Cornell University 2010

This dissertation contains three distinct chapters. Each chapter utilizes a different type of data set and implements a different method to identify causal relationships regarding current issues in the economics of education and health.

The first chapter analyzes persistence of minority and female students in science, technology, engineering, and math (STEM) majors in college. I utilize student-course level data for all students attending public four-year universities to determine whether having an instructor of the same race or same gender affects persistence of black and female students within STEM fields. I implement an instrumental variable strategy to control for self selection into courses. Results indicate that black students who have an introductory STEM course taught by a black instructor are significantly more likely to persist in a STEM field after the first year. However, female students are less likely to persist when their introductory STEM courses are taught by female instructors.

The second chapter (co-authored with Dr.John Cawley at Cornell University) is an evaluation of a program which offers financial incentives for weight loss. We analyze data from a company which provides a year-long health promotion program that offered financial rewards for weight loss. The types of incentive program varies by employer, with some offering steady payments for weight loss and others requiring participants to post a bond that is refundable based on achievement of weight loss goals. Comparing outcomes across groups, we find modest weight loss for participants

after one year.

The third chapter analyzes NCAA's Proposition 16 which changed the admission requirements for freshmen student-athletes at Division I colleges. Using institutional level data, I examine how requiring higher SAT scores and high school GPA for eligibility standards affects enrollment and graduation rates for Division I colleges. I implement a difference-in-differences approach using Division II schools and non-student-athletes as the comparison groups. The results indicate that after Proposition 16, Division I schools recruited fewer black freshmen student-athletes when compared with Division II schools. Additionally, I find that higher admission standards did not increase graduation rates at Division I schools.

BIOGRAPHICAL SKETCH

Joshua Price has made many stops en route to his doctoral degree, which has formed a foundation for much of his research on issue concerning higher education. After graduating high school, he enrolled at Utah State University with the hopes of making the tennis team. After his freshmen year, he spent two years in Santiago, Chile as a representative for the Church of Jesus Christ of Latter Day Saints. Living internationally and learning Spanish was an invaluable experience in his character development.

Upon returning, he enrolled in a community college with the hopes of raising his grades and transferring into a four-year university. Some teachers took time to work with him to excel in his courses, and after two semesters he transferred to Brigham Young University. While at BYU, Joshua declared Economics as his major but was determined to go on to law school upon graduation. After taking a probability and statistics course, Josh was hooked, and decided that pursuing a Ph.D. was what he wanted to do. He began working as a Research Assistant and began taking incredible amounts of math courses (even during the summer break). In 2005 he graduated with a B.A. in Economics and a minor in Math and Spanish. And to make his experience at BYU even better, he met and married the love of his life, Kelli.

Two months after graduating, Josh and Kelli packed the car up and moved to Ithaca, NY to pursue Joshua's graduate degree in Economics at Cornell University. After the first year of graduate school, Joshua transferred to the Ph.D. program in Policy Analysis and Management. This was one of the best things that happened to him, as the PAM department was extremely supportive of his research. He received is M.A. in Economics in 2007 and was able to get a picture in his robes next to his older brother, Joey, who finished his Ph.D. in Economics. In 2009, Joshua received his M.A. in Policy Analysis and Management and he spent his summer as a visiting lecturer at BYU. He is scheduled to graduate with a Ph.D. in Policy Analysis and Management in 2010.

The addition of his two sons McKay and Matthew (with the third son, Spencer, due a week before graduation), his 11 intramural championships in football, basketball, and softball with the Stormin' Mormons, and most importantly, being married to Kelli has made life outside of academia so wonderful.

Josh will begin his academic career at the University of Texas at Arlington, as an assistant professor, where the next chapter of his life will be written

To my wife Kelli,

without her support and love, I never would have been able to accomplish this

ACKNOWLEDGMENTS

I would like to thank my advisor, Ron Ehrenberg for providing me with an amazing opportunity to pursue my education and research at Cornell University. From him I have learned how to grow from failure, how to enjoy success, and how to really help others achieve their goals. I consider it a great blessing and privilege to have been his student and aspire to provide similar opportunities to other students in the future.

I would also like to thank the other advisors and mentors who have shared their time, insight, and knowledge with me. I have learned so much about how to conduct empirically sound research while working with John Cawley and appreciate his willingness to work on a project with me. His dedication to the PAM program is appreciated by myself and many other PAM graduate students. The time I spent with him at the Obesity Conference in Baton Rouge opened my eyes to understand why he is so well respected among other academics. Furthermore, I thank Kirabo Jackson and George Jakubson for the insight they provided in their classes and for letting me stop by and ask questions.

Also deserving thanks are all those who make up the PAM department. I feel that they have been completely supportive of helping me accomplish my academic goals. I would like to acknowledge all those who have participated in the Cornell Higher Education Institute; the weekly meetings have been instrumental in increasing my understanding of research. And thank you to Darrie, who was always there to help or just to talk.

Since birth I have been a step or two behind my brother Joey, and grad school has been no different. While I may suffer from younger brother syndrome, I prefer to think of Joey as my guinea pig; always willing to go first just to make sure it is safe. He has always taken time to make sure I was being productive; from letting me tag along at conferences, to the way I dress, to working together on research projects, to random phone calls and emails asking STATA questions, he has been a good big brother.

I would also like to acknowledge my parents, who at a young age instilled in me the appreciation of an education and a good work ethic. They continue to show their love and support toward me. I would also like to thank my other siblings, for they put up with my noise and antics for many years and continue to be kind and supportive; thank you Dusty, Annie, and Katie.

Finally, I would like to acknowledge my wife and boys. Kelli took the biggest risk in marrying me and moving so far from home to Ithaca, NY. She has served many roles while I have been in grad school (Bread winner, Editor, Counselor, Cheerleader, Mother and Wife) and has excelled at every one of them. And to my boys McKay and Matthew (and soon Spencer), you make my day when I come home and hear you scream. Your energy and enthusiasm amazes me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1: The Effect of Instructor Race and Gender on Student Persistence in STEM Fields

CHAPTER 1

THE EFFECT OF INSTRUCTOR RACE AND GENDER ON STUDENT PERSISTENCE IN STEM FIELDS

Abstract

The objective of this study is to determine if minority and female students are more likely to persist in a science, technology, engineering, and math (STEM) major when they enroll in classes taught by instructors of their own race or gender. Using data from public four-year universities in the state of Ohio, I analyze first semester STEM courses to see if the race or gender of the instructor effects persistence of initial STEM majors in a STEM field after the first semester and first year. Results indicate that black students are more likely to persist in a STEM major if they have a STEM course taught by a black instructor. Similar to previous findings, female students are less likely to persist when more of their STEM courses are taught by female instructors.

Introduction

Historically, women and racial minorities have been underrepresented in science and engineering occupations. In an effort to increase the number of women and minorities in these occupations, many recent policies have focused on increasing the number of female and minority students who enter college in science, technology, engineering, or math (STEM) fields. However, the NSF's Science and Engineering Indicators report (2008) finds that students who begin college as STEM majors have a lower probability of receiving a degree in their initial field than students in other majors. Of even greater concern is that women and historically disadvantaged racial minorities who initially intend to major in a STEM field are the least likely to persist toward a degree in one of these fields.

It is hypothesized that students experience better educational outcomes when they are able to interact and associate with faculty who are of their own race or gender. Provided that this hypothesis was true, one approach to increase the persistence of women and minority students is to increase the number of female and minority faculty in STEM departments. Previous studies indicate that female instructors in STEM courses do not increase the likelihood that female students take more STEM courses in subsequent semesters (Bettinger & Long, 2005; Carrel et al, 2010). These studies even suggest that female students who take STEM courses taught by female instructors are less likely to major in STEM fields. There is abundant research on instructor gender matching, but a lack of research on instructor race matching at the college level. There is research that links positive academic outcomes to having an own race teacher in elementary and secondary school (Klopfenstein, 2005; Dee, 2004), yet there is little research to date that identifies a causal effect of having an own-race instructor on academic outcomes in college. The primary objective of this study is to estimate the effect of having own-race instructors on persistence in STEM fields using within institution variation of the number of black faculty assigned to teach introductory courses. A secondary objective is to examine the effects of own-gender instructors on persistence of initial STEM majors.

The outcome of interest in the current study is persistence in a STEM field major, with persistence being defined as entering college with the intent of majoring in a STEM field and remaining in a STEM field major in subsequent semesters. The focus is placed on the intermediate measurement of intended major in order to identify when students begin to transition from STEM to non-STEM fields while in college. Many students change their major during their first years of college (40 percent after the first year and 74 percent after the second year). As findings from previous research indicate that faculty have the strongest influence on students within the first years of their college experience (Canes & Rosen 1995; Solnick, 1995), this study focuses on the student-faculty interaction occurring during the first semester of the freshmen year.

I use a linear probability model to estimate the effect of STEM instructor's race on student persistence in STEM fields. The key explanatory variable is the number of STEM courses in which a student enrolls in the first semester that are taught by black instructors. I use data from the Ohio Board of Regents, which includes course enrollment data for first time freshmen who enrolled between 1998 and 2002 in all public 4-year institutions in the state of Ohio. One of the empirical challenges of identifying casual effects of instructors is that students may differentially select into courses based on the race or gender of the instructor. For example, the data used in this study indicate that courses taught by a black instructor have a 2.4 to 10.4 percent higher fraction of black students enrolled in the course.

To address this selection issue, I instrument for whether a STEM course is taught by a black instructor with the fraction of STEM courses taught by black instructors at a given institution during a given semester. The fraction of STEM courses taught by black instructors within an institution varies due to several factors such as recent hires, course assignments, sabbaticals, and faculty leaving the institution. I also include a measure for the total number of black faculty in STEM fields within each institution to capture factors that may change within an institution over time. I use a similar instrumental variable to estimate the causal effects of female instructors on persistence.

My IV estimates indicate that having a black instructor increases the likelihood that black students persist in a STEM field. However, In contrast to past studies using elementary and secondary school student samples (Dee, 2004), I find that for college students, black instructors have no effect on persistence of white students in STEM fields. These results illustrate the positive effect that own-race instructors can have on academic outcomes early on in college for underrepresented minorities. In addition, I find that female instructors do not have a positive effect on the likelihood that female students persist in a STEM field.

Background

Several theories have been presented to explain why having an instructor with similar racial or gender characteristics might increase a student's academic performance. One of the explanations most thoroughly studied in past literature is the idea that faculty members serve as a mentor or role model for students, and having a mentor or role model of the same gender or race increases academic performance (Jacobi, 1991; Crisp & Cruz, 2009). A consistent finding over time is that students tend to choose role models who have the same characteristics as themselves (Erku & Mokros, 1984; Jacobi, 1991; and Karunanayake & Nautu, 2004).

Multiple studies have examined whether having a teacher of the same gender increases academic performance of students in elementary and secondary school.

Using data from the National Longitudinal Survey of Youth of 1979, Nixon and Robinson (1999) show that girls who have a female teacher in high school have higher levels of educational attainment. Using data from the National Education Longitudinal Study of 1988, Dee (2007) finds that having an own-gender teacher in 8^{th} grade increases test scores and student engagement with academic subjects. Additionally, in a sample of high school aged students in Sweden, Holmlund and Sund (2006) show a positive association between grades of female students and taking classes from female teachers.

Several studies have examined whether female role models have positive effects on the academic performance of female college students. In the past it has been difficult to identify specific role models and mentors due to data constraints, so studies have utilized the gender composition of university departments as a proxy for role models or possible mentors. Rothstein (1995) utilizes data from the National Longitudinal Study of 1972 and find a positive association between the percentage of female faculty at the school and the probability that a female student attains an advanced degree. Canes and Rosen (1995) look within three institutions and find no evidence that the share of women on a department's faculty leads to an increase in the share of female majors within that department. Robst, Keil, and Russo (1998) use data from the State University of New York at Binghamton and examine the fraction of classes that are taught by female instructors. Their results indicate a positive correlation between the percentage of science and math courses taught by female instructors and retention of female students in those majors.

Although these studies describe a positive association between female instructors and academic outcomes for female students, they may not identify a causal relationship. More recent studies have focused on estimating the causal effect of an instructor's gender on academic outcomes of female college students, while accounting for possible selection issues. Results of these more recent studies remain inconclusive as to whether there are short term benefits for female students from having a female instructor in initial STEM courses.

Hoffman and Oreopoulos (2009) examine the effect of female instructors in first semester courses on academic performance and the number of additional same-subject courses taken. They assert that introductory courses are chosen independent of gender of instructors and provide evidence that sections taught by female instructors (within a specific course) do not have a significantly higher share of female students than sections taught by male faculty. They show that female students' average grade performance is not significantly higher when their introductory courses are taught by female instructors, but male student performance decreases with female instructors. Pooling men and women and estimating the effect of same-sex instructor, they show that having an own-gender instructor in a math or science course actually decreases female grade performance and the number of same-subject courses taken in later years.

Carrel et al (2009) rely on random assignment of both students and faculty that is unique to the Air Force Academy. They find that high ability female students, defined by SAT scores, who have their introductory STEM courses taught by a female instructor perform better in these and additional courses and are more likely to receive a degree in a STEM field. However, when examining all female students, they find that having a female instructor in a STEM course increases performance in that course but has no significant effect on performance in subsequent courses and no effect on graduating in a STEM field. Due to the unique nature of the university where the data for this particular study comes from, the results may not generalize to other academic institutions.

Bettinger and Long (2007) use a more representative sample, including all public 4-year universities in the state of Ohio, to examine the effect that female instructors have on female students in STEM fields. They look at whether having a female instructor for a course in a particular field increases the probability of taking additional courses in that field and receiving a degree in that field. To address differential selection into courses, they instrument for having a course taught by a female instructor with a measure of the fraction of courses within a department taught by female instructors. They find mixed results of the effect of female STEM faculty on female students. For example; female faculty have a positive effect on female students taking additional courses in mathematics and geology fields but a negative effects in the fields of biology and physics.

Examining the relationship between having an instructor of the same race and student academic outcomes, Rask and Bailey (2002) find that minority students who take more courses in a field from a professor of the same race are more likely to major in that field. Additional studies have examined the effect that minority teachers have on the academic outcomes of minority students in elementary and secondary school. Ehrenberg & Brewer (1995) use data from the mid 1960's and find evidence that black high school students who take classes from black are receive higher test gains on average. Using more recent data, Ehrenberg, Goldhaber, and Brewer (1995) show that a teacher's racial characteristics have no effect on how much students learn between the 8th and 10th grades. Using data from Texas high schools, Klopfenstein (2005) finds that increasing the percentage of black math teachers has a significant effect on the probability that black students in a geometry class will enroll in more advanced math classes. Using data from Tennessee's Project STAR, Dee (2004) utilizes the project's random assignment of students and teachers to classes and finds that having an own-race teacher significantly increases the math and reading test scores of black students. In addition, teachers are more likely to give higher ratings to students who are of the same race as themselves (Ehrenberg et al. 1995, Dee 2005). While there is documented evidence for a positive effect of teacher-student racial matching in elementary and secondary school, the relationship that exists between professor and student may be quite different at the college level. There is a need for research on the effect of having a same-race teacher for college students.

This study contributes to the existing literature examining the effects of own-race and own-gender instructors on academic outcomes of STEM majors in two ways. First, it is the first quasi-experimental study that identifies the own race instructor effect on persistence in a STEM field for college students. Second, it re-examines the effect that female instructors have on the decision of college students to persist toward a degree in a STEM field in a different manner than previous studies. It changes the level of analysis from course-student level to student-semester level. This allows me to simultaneously control for other courses that the student enrolls in within a given semester. Additionally, this study uses variation in the number of black and female instructors assigned to teach STEM courses within an institution over time to identify the causal effect of an own-race or own-gender instructor on persistence.

Data

The data for this study comes from the Ohio Board of Regents, which collects data from all public universities within the State of Ohio. One strength of such a data set is that it contains a large number of observations of underrepresented groups of students in STEM fields detailed enough to match students to the instructor of each course in which they enrolled. The data consist of first-time freshmen who enrolled in one of the 13 public 4-year universities in the state of Ohio between 1998 and 2002. Three sources of student-level data are included in the present analysis: (1) information the school receives when the student first enrolls, including gender, race, age, standardized test score (ACT or SAT) and state of residence; (2) information the school records each term, such as term grade point average and intended major field of study, and (3) the courses in which each student enrolled for each term up to six years after initial enrollment. In addition course records identify the instructor of each course. I then merge in administrative faculty files containing information on each faculty member, including race, gender, tenure status, rank, and highest degree earned. This allows me to match each student with the instructor of each course in which they enrolled.

One of the difficulties of examining the effect of minority instructors on academic outcomes is that many data sets have a small number of observations of either the number of minority students or minority faculty. The Ohio data used in this study includes information on 14,448 black students and 1,613 black faculty, a sample size that makes it possible to estimate the effects of having a black instructor on academic outcomes for black college students. Another advantage of using data from Ohio is that the demographic characteristics of students who attend public 4-year universities in the state are similar to nationally representative samples.[1]

The five cohorts of first-time freshmen included in the data utilized in this analysis include over 155,000 students, of whom 22.1 percent initially intended to major in a STEM field. Throughout this paper, I aggregate subfields into a general STEM or non-STEM classification. Table 1.1 examines initial major choice and shows that female students initially constitute a lower percentage of STEM majors than non-STEM majors. Additionally, ACT scores are 2.5 points higher (≈90 SAT points) among STEM majors.[2] Significant differences arise when examining the fraction of
B

[1] For a more detailed argument of external validity see Bettinger (2007)
[2] The ACT is a college entrance exam similar to the SAT. A 21.7 on the ACT is approximately equivalent to a 1000 and a 24.2 is equivalent to a 1090 on the SAT.

Table 1.1 Summary Statistics

			STEM		Non-STEM		
Variable	Mean		Mean		Mean		p-value
female	0.538		0.344		0.593		0.000
white	0.828		0.824		0.829		0.035
black	0.093		0.083		0.096		0.000
asian	0.022		0.038		0.018		0.000
other	0.057		0.055		0.057		0.066
ACT score	22.3		24.2		21.7		0.000
	[4.284]		[4.252]		[4.137]		
Engineering	0.088						
Life/Physical Science	0.083						
Technology/Math	0.050						
Business	0.149						
Communication	0.050						
Education	0.098						
Humanities	0.195						
Social Science	0.096						
Vocational	0.086						
Unknown	0.105						
N	156,056		34,687		121,369		

students from a particular subgroup who initially declare a STEM major. Among men in the sample, 31.8 percent initially declare a STEM major compared to only 14.3 percent of female students. In terms of initial racial differences; 22.3 percent of white students initially declare a STEM field major compared to 20 percent of black students.

Faculty characteristics also differ between STEM fields and non-STEM fields. As shown in Table 1.2, about a third of STEM field faculty is female compared to half in non-STEM fields. While the fraction of white faculty is similar between STEM and non-STEM fields, within STEM fields there is a lower fraction of black faculty compared with non-STEM fields. STEM fields have a higher proportion of faculty who have earned doctoral degrees, are of higher academic rank, are tenured and are employed full-time. Within STEM fields, there are significant gender and racial

Table 1.2 Characteristics of the Faculty

		Full Sample		STEM	Non-STEM	
Variable		Mean		Mean	Mean	
Female		0.447		0.299	0.501	***
White		0.871		0.874	0.869	
Black		0.055		0.031	0.064	***
Asian		0.037		0.064	0.027	***
Other race		0.037		0.031	0.040	***
Highest Degree						
	Ph.D.	0.344		0.432	0.311	***
	Masters	0.326		0.251	0.355	***
	Other degree	0.314		0.300	0.317	***
Rank						
	Professor	0.119		0.168	0.101	***
	Associate	0.117		0.142	0.108	***
	Assistant	0.129		0.127	0.130	
	Other rank	0.619		0.546	0.644	***
Apointment						
	Full-time	0.206		0.228	0.198	***
	Part-time	0.288		0.268	0.296	***
	Grad assistant	0.261		0.223	0.273	***
Observations		32,555		8,737	23,818	

Note: Asterisks represent significant difference in means, *** 1%, ** 5%, and *1

differences in faculty characteristics. As shown in Table 1.3, the same proportion of black faculty in STEM fields have a Ph.D. as white faculty, but they are less likely to be full professor. Female faculty are less likely to have a Ph.D., be a full or associate professor, and be full-time employed.

If the ultimate policy goal is to increase the number of female and minority students who major in a STEM field, then outcomes of interest should include indicators that are correlated with receiving a degree in a STEM field. Previous studies have examined grade performance and probability of enrolling in additional courses in a particular field as indicators for earning a degree (Hoffman & Oreopoulos, 2009; Carrell et al., 2009; Bettinger & Long, 2005). These outcomes may not provide the

Table 1.3 Racial and Gender Differences of Faculty within STEM Fields

Variable		White Mean	Black Mean			Male Mean	Female Mean	
Highest Degree								
	Ph.D.	0.340	0.340			0.441	0.241	***
	Masters	0.333	0.407	***		0.295	0.382	***
	Other degree	0.327	0.253	***		0.264	0.376	***
Rank								
	Professor	0.123	0.067	***		0.185	0.044	***
	Associate	0.115	0.149	***		0.148	0.086	***
	Assistant	0.125	0.159	***		0.131	0.131	
	Other rank	0.637	0.626			0.536	0.739	***
Apointment								
	Full-time	0.205	0.271	***		0.240	0.175	***
	Part-time	0.297	0.291			0.266	0.321	***
	Grad assistant	0.272	0.203	***		0.218	0.324	***
Observations		28,358	1,784			17,632	14,227	

Note: Asterisks represent significant difference in means, *** 1%, ** 5%, and *10

best measures of intent to earn a degree in a STEM field since enrolling in additional STEM courses may be the result of a general education requirement needing to be fulfilled and not necessarily due to interest in that field or intent to graduate in it. Whether the individual intends to major in a STEM field major is a better indicator that can be used to show progress toward the goal of receiving a degree in a STEM field. Therefore, the outcome which is of most interest in this study is whether a student who initially intends to major in a STEM field continues in a STEM field as his or her intended major in subsequent terms in which he or she enrolls.

Patterns of Persistence

In this study, persistence is defined as continuing on in the field of the initial major during subsequent semesters that the student is enrolled in classes. The data for this measure are constructed from administrative records that contain the student's intended major for each term the student is enrolled. The focus of this analysis is on

STEM fields in general; thus changing majors within STEM fields is counted as persisting in a STEM field (i.e. a student who initially declares a major in chemistry and then changes to a biology major is considered as persisting in a STEM field). The same is true for students who transfer within non-STEM field majors.

Lower persistence rates exist among those students who initially enter STEM fields compared with those who initially enter non-STEM fields. Table 1.4 shows the cumulative distribution of persisting in initial major, changing majors, and dropping out of school. Among initial STEM majors, 91.6 percent remain STEM majors by the second semester of their freshman year. However, only 71.8 percent of initial STEM majors remain in a STEM field by the beginning semester of their sophomore year. Persistence rates for non-STEM majors are significantly higher, with 95.8 percent persisting in a non-STEM field after the first semester and 83.4 percent after the first year[3]. Also, a larger fraction of students in non-STEM majors drop out of college compared with students in STEM majors. Among initial STEM majors, those individuals who either change majors or drop out of school, 14 percent do so after the first semester, 47 percent do so within in the first year of school, and 75 percent within the first two years.

In addition to differences in persistence rates across fields of study, there are significant differences in persistence rates between gender and racial groups within STEM field majors. Table 1.5 shows the unconditional means of outcomes after each semester. The results in the top panel of Figure 1.1 indicate that, even after controlling for institution and cohort, white students are more likely to persist in STEM fields than

[3] The persistence rates between STEM and non-STEM majors is significantly different at the 1% level

Table 1.4 Cumulative Distribution of Persistence

Initial Non-STEM Majors (N=121,369)			
	Remain in Non-STEM	Change to STEM	Dropout
1st Semester	0.958	0.011	0.032
2nd Semester	0.834	0.028	0.142
3rd Semester	0.802	0.032	0.166
4th Semester	0.738	0.040	0.222
5th Semester	0.727	0.040	0.233
6th Semester	0.688	0.040	0.272
Initial STEM Majors (N=34,687)			
	Remain in STEM	Change to Non-STEM	Dropout
1st Semester	0.916	0.065	0.018
2nd Semester	0.718	0.174	0.108
3rd Semester	0.654	0.217	0.129
4th Semester	0.551	0.272	0.177
5th Semester	0.530	0.287	0.183
6th Semester	0.484	0.298	0.218

black students. However, controlling for a measure of prior achievement by including ACT test scores, the white-black persistence gaps decreases by almost one-half. This provides suggestive evidence that prior preparation is an important factor in explaining the racial differences of persistence in STEM fields. The bottom panel shows that males are more likely to persist in STEM fields than female students. However, ACT test scores do not explain the difference in persistence rates between males and females as the male-female persistence gap is virtually unchanged when controlling for test scores.

While there are racial and gender differences in persistence in STEM fields, there also exist such differences in non-STEM fields. As shown in Figure 1.2, there is a racial gap in 3-year

Table 1.5 Cumulative Distribution of Persistence by Race and Gender
Conditional on being an initial Non-STEM major

White			
	Remain in STEM	Change to Non-STEM	Dropout
1st Semester	0.911	0.064	0.016
2nd Semester	0.723	0.172	0.095
3rd Semester	0.661	0.214	0.116
4th Semester	0.562	0.270	0.159
5th Semester	0.542	0.283	0.166
6th Semester	0.497	0.296	0.198
Black			
	Remain in STEM	Change to Non-STEM	Dropout
1st Semester	0.875	0.089	0.036
2nd Semester	0.574	0.211	0.214
3rd Semester	0.505	0.266	0.229
4th Semester	0.380	0.307	0.314
5th Semester	0.359	0.328	0.313
6th Semester	0.302	0.326	0.372
Male			
	Remain in STEM	Change to Non-STEM	Dropout
1st Semester	0.931	0.051	0.018
2nd Semester	0.751	0.139	0.110
3rd Semester	0.694	0.173	0.133
4th Semester	0.597	0.220	0.182
5th Semester	0.578	0.235	0.187
6th Semester	0.530	0.247	0.223
Female			
	Remain in STEM	Change to Non-STEM	Dropout
1st Semester	0.888	0.093	0.018
2nd Semester	0.654	0.243	0.104
3rd Semester	0.578	0.302	0.121
4th Semester	0.462	0.372	0.166
5th Semester	0.439	0.387	0.174
6th Semester	0.396	0.395	0.209

persistence rates in both STEM and non-STEM fields. This measure of persistence counts individuals who change to non-STEM majors and dropouts as not persisting, yet individuals who dropout and individuals who change majors may be very different. To examine the decision of persisting in STEM majors versus changing majors, I condition on not dropping out of school. Once conditioning on not dropping out, the persistence gaps remains among STEM fields but black are marginally more likely to persist in non-STEM fields. This indicates that the

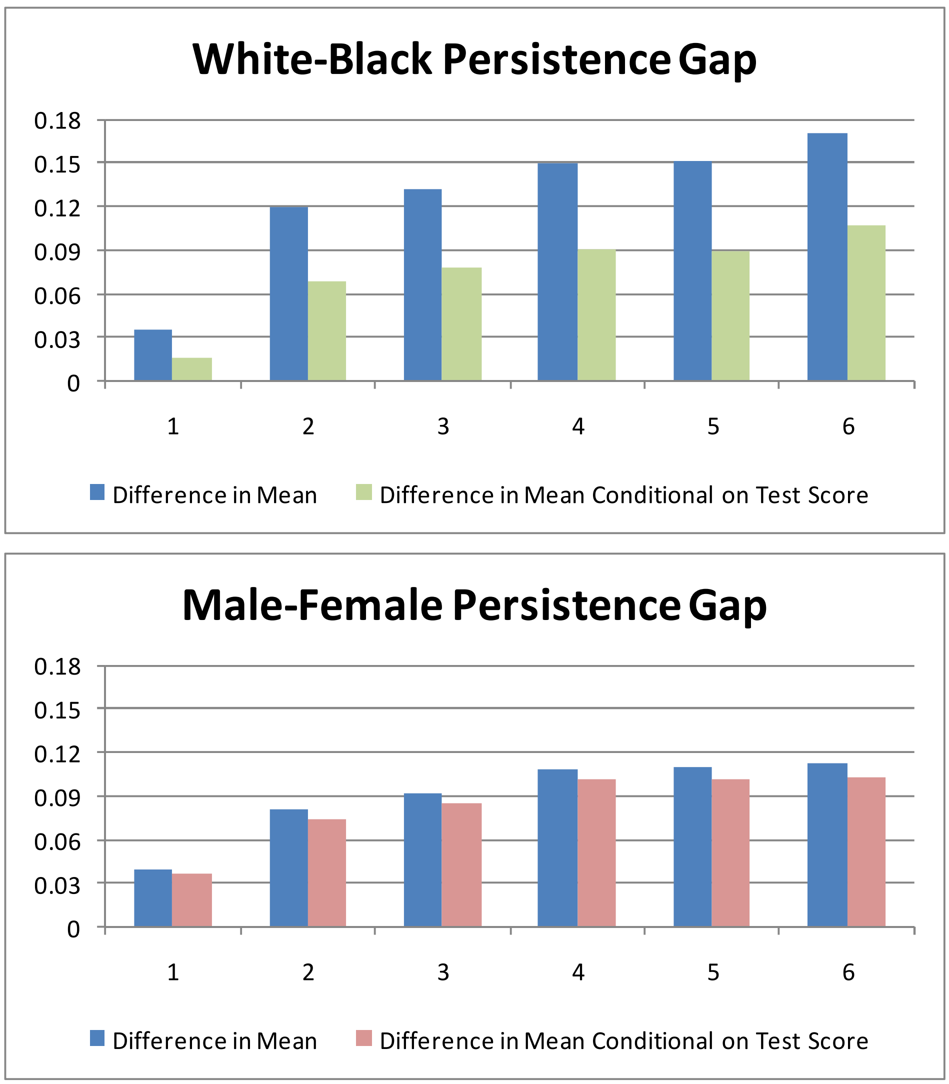


Figure 1.1 Racial and Gender Persistence Gap of Initial STEM Majors by Semester

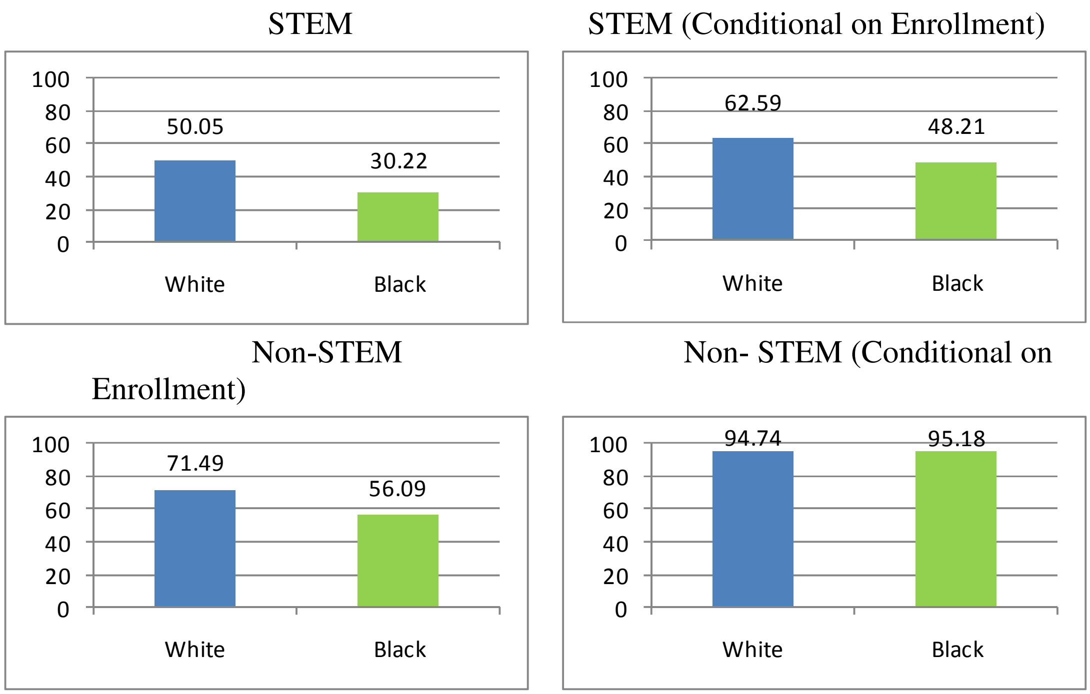


Figure 1.2 Three Year Persistence Rates by Race

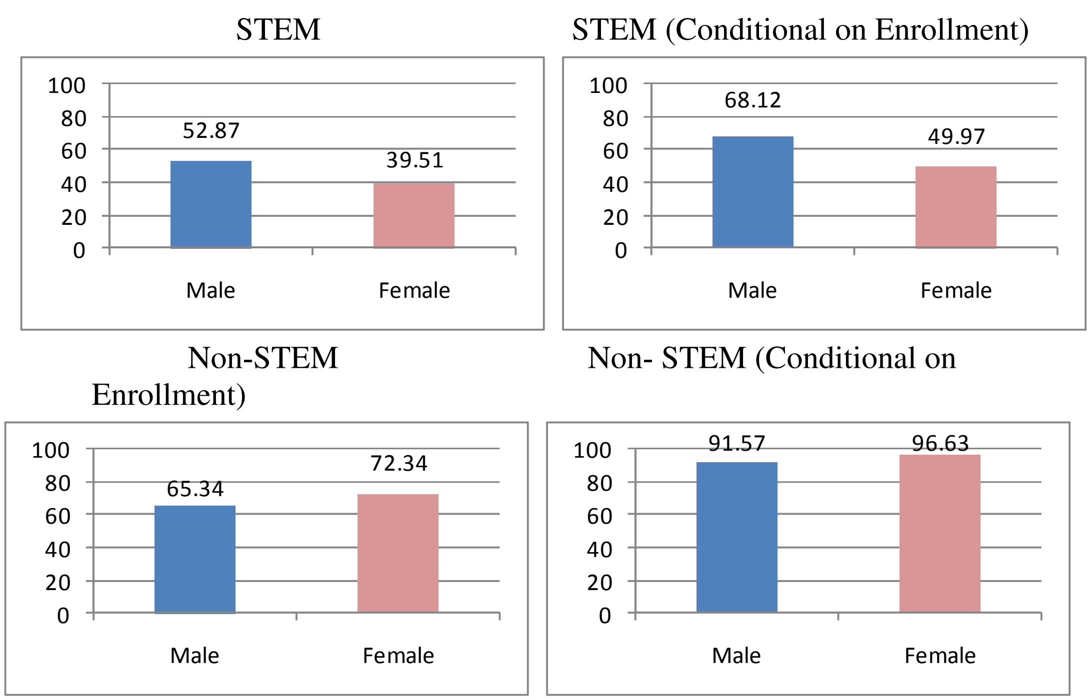


Figure 1.3 Three Year Persistence Rates by Gender

unconditioned persistence gap among non-STEM majors is being driven by students who drop out. Figure 1.3 shows that men are more likely to persist in STEM fields, but less likely to persist in non-STEM fields.

Methods

The objective of this study is to test whether students who have their STEM courses taught by an instructor with similar racial or gender characteristics are more likely to persist in a STEM field major. To test this hypothesis, I focus on the first semester courses of students who initially declare a STEM major. The basic econometric model is represented with the following equation:

$$\begin{aligned} Persistence_{ijk} = {} & \beta_1(Black)_{ijk} + \beta_2(Number\ of\ black\ STEM\ instructors)_{ijk} + \\ & \beta_3(Number\ of\ black\ STEM\ instructors * Black)_{ijk} + \\ & \beta_4(Number\ of\ STEM\ courses)_{ijk} + \beta_5(Number\ of\ Non-STEM\ courses)_{ijk} + \\ & X_{ijk} + \lambda_{ijk} + \theta_i + \delta_j + \delta_k + u_{ijt} \end{aligned}$$

where $Persistence_{ijk}$ is a binary outcome equal to one if student *i* at school *j* in cohort *k* is a STEM major in the second semester given that student i's initial major was in a STEM field. The key variable of interest is the number of black STEM instructors, which gives the effect of black instructors on white students, and the interaction of number of black STEM instructors and black, which yields the effect of black STEM instructors on black students. The vector X_i controls for student characteristics such as race, gender, ACT test score[4], and state of residence. Also included in the equation are controls for observable characteristics of instructors such as rank, tenure status, full-time, and graduate assistant (λ). To account for structural differences between majors within STEM fields, θ is a set of dummy variables for the initial major of student i. There may also be specific programs implemented by individual universities that may affect a student's decision to remain in a STEM field

[4] A dummy variable is included to account for the 16% of the sample who have missing ACT scores.

major; thus I also include institutional fixed effects (δ_j), and cohort fixed effects (δ_k) to account for differences over time. I also use this same model to estimate the own gender effect of instructors on persistence.

I start by assuming that first semester courses are chosen independent of the characteristics of the instructors of the course. Based on this assumption, I use a linear probability model to estimate the correlation between the number of STEM courses that are taught by black instructors and the outcome of persisting in a STEM field. However, there are some possible reasons why the assumption that students randomly sort into classes in their first semester may not be a valid assumption. For example, although students sign up for classes before coming to campus, they can access information about potential instructors online or there may be opportunities to switch classes during the first week of school.

The selection into courses may occur at two different levels. First, students may decide to take a course within a field of study, and then enroll in a course which is taught by an instructor of similar characteristics, or what I refer to as between-course selection. The second type of selection occurs when students decide on a course, and then enroll in a section of the course based on faculty race or gender (within-course selection). Following the method of Hoffman and Oreopoulos (2009), I test both type of selection by examining the relationship between the race of the instructor and the racial composition of the students in the course and show results in Table 1.6. Within field selection shows that once controlling for faculty characteristics, institution and cohort fixed effects, and field fixed effects, STEM courses taught by a black instructor have a 5.7 percentage point increase in the fraction of black students

Table 1.6 Selection into Courses by Race
Outcome is the Fraction of Students Who are Black

	Non-STEM Course				STEM Course			
Black Instructor	0.190***	0.105***	0.104***	0.024***	0.214***	0.059***	0.057***	0.028***
	[0.003]	[0.002]	[0.002]	[0.002]	[0.005]	[0.005]	[0.004]	[0.004]
Constant	0.105***	0.089***	0.078***	0.253***	0.098***	0.078***	0.029***	0.094**
	[0.001]	[0.002]	[0.003]	[0.083]	[0.001]	[0.002]	[0.003]	[0.038]
Observations	63370	63370	63370	63370	25332	25332	25332	25332
R-squared	0.08	0.43	0.44	0.76	0.06	0.42	0.44	0.73
faculty characteristics		x	x	x		x	x	x
Institution FE		x	x	x		x	x	x
Cohort FE		x	x	x		x	x	x
Field FE			x				x	
Course FE				x				x

Standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

enrolled in the course. With an average class size of 32.86, this increase represents, on average, 1.87 more black students when a course is taught by a black instructor. Within-course selection shows even less selection, once including course fixed effects, a black instructor is correlated with a 2.8 percentage point increase in the fraction of to about .92 of a student on average. Though modest in magnitude, this does provide suggestive evidence that black students do select into courses taught by black instructors.

Similar to the analysis examining selection into courses based on race of instructor, I examine the gender of the instructor influences the gender composition of the class (see Table 1.7). Between-course selection shows that having a female instructor in a STEM course increases the fraction of female students by 3.7 percentage points (1.2 more female students). Within-course selection shows that having a female instructor is not correlated with the fraction of female students who enroll in the course, indicating that selection may be less of an issue when examining the effect of own-gender instructor on persistence.

To address possible selection bias issues, I use the fraction of STEM courses taught by black instructors at an institution to instrument for the number of STEM courses taught by black instructors. Since institutional and cohort fixed effects are included in the model, the variation of the instrument comes from within institution changes over time in the number of courses taught by black faculty and the total number of courses offered. This variation can be driven by recent hires, course assignments, sabbaticals, job loss, or other within institution factors. I also control for the total number of black STEM instructors at each institution in the first stage equation to proxy for time-varying institutional factors that might be correlated with the type of instructors assigned to introductory courses and a student's decision to persist in a STEM field.

Table 1.7 Selection into Courses by Gender
Outcome is the Fraction of Students Who are Female

	Non-STEM Course				STEM Course			
Female Instructor	0.064***	0.068***	0.066***	0.012***	0.060***	0.053***	0.037***	0.003
	[0.002]	[0.002]	[0.002]	[0.001]	[0.003]	[0.003]	[0.003]	[0.002]
Constant	0.530***	0.514***	0.531***	0.658***	0.435***	0.359***	0.097***	0.549***
	[0.001]	[0.004]	[0.005]	[0.133]	[0.002]	[0.007]	[0.007]	[0.056]
Observations	57608	57608	57608	57608	20259	20259	20259	20259
R-squared	0.03	0.06	0.08	0.64	0.02	0.07	0.31	0.78
faculty characteristics		x	x	x		x	x	x
Institution FE		x	x	x		x	x	x
Cohort FE		x	x	x		x	x	x
Field FE			x				x	
Course FE				x				x

Standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

This instrument is similar to that used by Bettinger and Long (2005), but can be seen as an improvement because it aggregates fields to classify them as STEM versus non-STEM. Bettinger and Long conduct their analysis on more refined measures of field of study (i.e. physics, chemistry, biology, etc.) and use proportion of courses taught by female faculty to instrument for having a female instructor. While this controls for selection within a field, there may be selection across closely related fields of study based on faculty characteristics. For example, the choices of students may not just be between sections of the same chemistry course, but between different courses within STEM fields. Thus, aggregating to a higher level better accounts for the type of selection that occurs.

Results

Effect of Racial Matching

The baseline model examines the relationship between the number of black instructors in STEM courses and persistence of students in a STEM field after the first semester and after the first year. OLS makes the assumption that factors related to a student enrolling in a class taught by a black instructor are not correlated with persistence. Results in Table 1.8 indicate that under OLS assumptions, black students are equally as likely to persist as non-black students. Additionally, the number of black STEM instructors has a positive influence on persistence of non-black students, as the coefficient on Black STEM Instructor indicates a 2.5 percentage point increase in persistence. The interaction term between Black STEM Instructors and Black yields the effect of the number of STEM courses taught by black faculty on the persistence of black students. After the first semester, there is not a statistical relationship between the number of black instructors and persistence on black STEM majors. Looking at persistence after the first year of school, OLS results indicate that black students are

4.4 percentage points less likely to persist after one year. However, each additional black instructor that a black student has in a STEM course is correlated with a 5.5 percentage point increase in persistence, which closes the white-black persistence gap. Other results from the baseline model indicate that higher ACT scores are correlated with increased rates of persistence. Also, holding constant the number of non-STEM courses, each additional STEM course in which a student enrolls is associated with a 1.9 and 5.5 percentage point increase in persistence after the first semester and first year respectively.

OLS estimations. To account for this possible bias, I instrument for the number of black STEM instructors with the fraction of STEM courses that are taught by black faculty at the institution. The lower panel of Table 1.8 shows the first stage estimation results for the number of black STEM instructors. Since the baseline equation includes both the number of black STEM instructors and the number of black instructors interacted with black, I use the instrument and the instrument interacted with black (student race) to create the first stage estimations. In both cases the instruments are highly correlated with the number of black instructors in STEM courses and have large F statistics. The second stage results show that there continues to be no statistical effect on persistence after the first semester. However, examining transitions that occur by the end of the first year; it appears black students are significantly more likely to persist when more of their initial STEM courses are taught by black instructors.

While Table 1.8 looks at the number of black STEM instructors, Table 1.9 redefines the explanatory variable of interest as having at least one black instructor in an introductory STEM course. 89 percent of black students who have at least 1 black

Table 1.8 Regression Results: Outcome is Persist in STEM Field Major (Number of Black STEM Instructors)

	[0.006]	[0.008]		[0.010]	[0.013]
Black STEM Instructor	0.025***	-0.014		0.027**	-0.114
	[0.008]	[0.074]		[0.012]	[0.118]
Black STEM Instructor * Black	-0.024	0.101		0.055**	0.552***
	[0.017]	[0.073]		[0.028]	[0.118]
ACT	0.004***	0.004***		0.011***	0.011***
	[0.000]	[0.000]		[0.001]	[0.001]
Female	-0.036***	-0.036***		-0.077***	-0.077***
	[0.003]	[0.003]		[0.005]	[0.005]
Number of Non-STEM Courses	-0.001	-0.002		-0.014***	-0.015***
	[0.003]	[0.003]		[0.005]	[0.005]
Number of STEM Courses	0.019***	0.018***		0.055***	0.053***
	[0.003]	[0.003]		[0.004]	[0.004]
Constant	0.847***	0.850***		0.473***	0.487***
	[0.014]	[0.014]		[0.022]	[0.023]
R-squared	0.03	0.03		0.06	0.05
First Stage Estimation					
Frac Courses Taught by Black Faculty		0.863			0.863
		[0.143]			[0.143]
F-Stat		12.08			12.08
Frac Courses Taught by Black Faculty*Black		0.99			0.99
		[0.024]			[0.024]
F-Stat		81.48			81.48
Observations	34,687	34,687		34,687	34,687
Standard errors in brackets					
* significant at 10%; ** significant at 5%; *** significant at 1%					

Table 1.9 Regression Results: Outomce is Persistin STEM Field Major
(At Least One Black Faculty)

	After First Semester		After First Year	
	OLS	IV	OLS	IV
Black	-0.002	-0.012	-0.045***	-0.086***
	[0.006]	[0.008]	[0.010]	[0.014]
Black STEM Instructor	0.028***	-0.011	0.030**	-0.11
	[0.008]	[0.088]	[0.013]	[0.142]
Black STEM Instructor * Black	-0.021	0.117	0.081**	0.637***
	[0.020]	[0.084]	[0.031]	[0.135]
ACT	0.004***	0.004***	0.011***	0.011***
	[0.000]	[0.000]	[0.001]	[0.001]
Female	-0.036***	-0.036***	-0.077***	-0.077***
	[0.003]	[0.003]	[0.005]	[0.005]
Number of Non-STEM Courses	-0.001	-0.002	-0.014***	-0.015***
	[0.003]	[0.003]	[0.005]	[0.005]
Number of STEM Courses	0.019***	0.018***	0.055***	0.053***
	[0.003]	[0.003]	[0.004]	[0.004]
Constant	0.846***	0.850***	0.473***	0.488***
	[0.014]	[0.014]	[0.022]	[0.023]
R-squared	0.03	0.03	0.06	0.05
First Stage Estimation				
Frac Courses Taught by Black Faculty		0.0638		0.0638
		[0.130]		[0.130]
F-Stat		9.78		9.78
FracCourses Taught by Black Faculty*Black		0.886		0.886
		[0.021]		[0.021]
F-Stat		84.64		84.64
Observations	34,687	34,687	34,687	34,687
Standard errors in brackets				
* significant at 10%; ** significant at 5%; *** significant at 1%				

STEM instructor have only one black instructor teach a STEM course. Results outlined in Table 1.9 follow the same pattern as Table 1.8, including a statistically insignificant effect of having a black instructor on persistence of black students after the first semester. Having at least one black instructor in a STEM course has a positive and significant effect on persistence of black students after the first year, with the point estimate becoming larger when using instrumental variables.

Another specification redefines persistence as being equal to one if the individual persists in a STEM field and zero if the individual changes majors, or in other words, persistence conditional upon not dropping out. Table 1.10 examines this outcome with the key explanatory variable being defined as the number of black STEM instructors. For black students it appears that there is no effect of additional STEM courses taught by black instructors on persistence after the first semester but a positive and significant effect on persistence after the first year. Table 1.11 examines the binary measure of having at least one black faculty, instead of the count variable, and the results are very similar. In all these specifications the point estimates on the interaction of Black STEM Instructor and Black student are smaller in magnitude than when they are not conditioning upon enrolling in school.

The general finding is that black instructors do not have a significant impact on black students' persistence in a STEM field after the first semester. It does seem to be the case that having a black instructor has a significant impact on persistence of black students after the first year. The magnitude of the effect is hard to pin down. OLS results would suggest the effect is between five to eight percentage points. IV results would suggest that the real effect, once accounting for selection bias, is much larger. Maybe equally important is that black instructors do not have a negative impact on non-black students. The effect was ranged from two to three percentage points under OLS, but was insignificant with instrumental variables. There is evidence to suggest

Table 1.10 Regression Results: Outcome is Persist in STEM Field Major Conditional on Enrollment (Number of Black Stem Instructors)

	After First Semester		After First Year	
	OLS	IV	OLS	IV
Black	0.007	0	0	-0.025*
	[0.005]	[0.007]	[0.009]	[0.014]
Black STEM Instructor	0.020***	-0.061	0.015	-0.071
	[0.007]	[0.069]	[0.012]	[0.106]
Black STEM Instructor * Black	-0.022	0.075	0.054*	0.395***
	[0.016]	[0.069]	[0.028]	[0.137]
ACT	0.003***	0.003***	0.009***	0.008***
	[0.000]	[0.000]	[0.001]	[0.001]
Female	-0.036***	-0.036***	-0.097***	-0.097***
	[0.003]	[0.003]	[0.005]	[0.005]
Number of Non-STEM Courses	-0.008***	-0.008***	-0.023***	-0.023***
	[0.003]	[0.003]	[0.005]	[0.005]
Number of STEM Courses	0.017***	0.016***	0.043***	0.042***
	[0.002]	[0.002]	[0.004]	[0.004]
Constant	0.885***	0.888***	0.632***	0.639***
	[0.013]	[0.013]	[0.021]	[0.022]
R-squared	0.03	0.02	0.05	0.05
First Stage Estimation				
FracCourses Taught by Black Faculty		0.842		0.842
		[0.144]		[0.144]
F-Stat		11.72		11.72
FracCourses Taught by Black Faculty*Black		0.97		0.97
		[0.025]		[0.025]
F-Stat		77.42		77.42
Observations	34,057	34,057	30,946	30,946
Standard errors in brackets				
* significant at 10%; ** significant at 5%; *** significant at 1%				

Table 1.11 Regression Results: Outcome is Persist in STEM Field Major Conditional on Enrollment (At Least One Black Faculty)

	After First Semester			After First Year	
	OLS	IV		OLS	IV
Black	0.007	0		0	-0.026*
	[0.005]	[0.008]		[0.009]	[0.015]
Black STEM Instructor	0.024***	-0.07		0.019	-0.075
	[0.007]	[0.082]		[0.012]	[0.126]
Black STEM Instructor * Black	-0.022	0.084		0.067**	0.439***
	[0.018]	[0.079]		[0.032]	[0.156]
ACT	0.003***	0.003***		0.009***	0.008***
	[0.000]	[0.000]		[0.001]	[0.001]
Female	-0.036***	-0.036***		-0.097***	-0.097***
	[0.003]	[0.003]		[0.005]	[0.005]
Number of Non-STEM Courses	-0.008***	-0.008***		-0.023***	-0.023***
	[0.003]	[0.003]		[0.005]	[0.005]
Number of STEM Courses	0.017***	0.016***		0.043***	0.042***
	[0.002]	[0.003]		[0.004]	[0.004]
Constant	0.884***	0.889***		0.632***	0.639***
	[0.013]	[0.013]		[0.021]	[0.022]
R-squared	0.03	0.02		0.05	0.05
First Stage Estimation					
FracCourses Taught by Black Faculty		0.651			0.651
		[0.132]			[0.132]
F-Stat		9.9			9.9
FracCourses Taught by Black Faculty*Black		0.879			0.879
		[0.022]			[0.022]
F-Stat		81.32			81.32
Observations	34,057	34,057		30,946	30,946
Standard errors in brackets					
* significant at 10%; ** significant at 5%; *** significant at 1%					

that not having an own –race teacher has negative impacts in the K-12 setting (Dee, 2007), yet it is evident that this is not the case at the college level.

Effect of Gender Matching

The baseline model of the effect of the number of female STEM instructors on persistence of initial STEM majors is shown in Table 1.12. The OLS model shows that female students are 3.9 percentage points less likely to persist in a STEM field than male students after the first semester. Having a female instructor in a STEM course has a small negative effect on the persistence of male students and no effect on the persistence of female students. However, when examining persistence after the first year, female students are 7.3 percentage points less likely to persist and each additional female instructor lowers the persistence of male students by 1.4 percentage point after the first year. There is no statistical relationship between own-gender instructors and persistence of female students after the first year.

As was shown in Table 1.7, it appears that there is between course sorting into courses taught by female instructors but not within course sorting. If the selection process is that students choose a course and then sort into sections based on faculty gender, then IV results should be similar to OLS results. However, in many instances IV results differ greatly from OLS result indicating that the selection into courses occurs across courses rather than within courses. Once controlling for selection into courses, the IV results indicate that female students are significantly less likely to persist when their STEM courses are taught by female instructors. The magnitude is around 26 percentage points, but is imprecisely measured and only significant at the 10% level. This effect is only found after the first semester, as the coefficient in the IV estimation is not statistically significant.

Table 1.12 Regression Results: Outcome is Persist in STEM Field Major (Number of Female STEM Instructors)

	After First Semester		After First Year	
	OLS	IV	OLS	IV
Female	-0.039***	0.118	-0.073***	0.028
	[0.004]	[0.080]	[0.006]	[0.121]
Female STEM Instructor	-0.005*	-0.036	-0.014***	-0.066
	[0.003]	[0.043]	[0.005]	[0.065]
Female STEM Instructor * Femal	0.005	-0.258*	-0.006	-0.169
	[0.004]	[0.135]	[0.007]	[0.203]
ACT Score	0.004***	0.003***	0.012***	0.011***
	[0.000]	[0.001]	[0.001]	[0.001]
black	-0.004	-0.001	-0.039***	-0.037***
	[0.006]	[0.007]	[0.009]	[0.010]
asian	0.027***	0.028***	0.066***	0.066***
	[0.008]	[0.008]	[0.012]	[0.013]
other	-0.002	-0.003	-0.003	-0.003
	[0.007]	[0.007]	[0.011]	[0.011]
Number of Non-STEM Courses	-0.002	-0.001	-0.014**	-0.012**
	[0.003]	[0.004]	[0.005]	[0.006]
Number of STEM Courses	0.017***	0.021***	0.047***	0.050***
	[0.003]	[0.004]	[0.005]	[0.006]
Constant	0.857***	0.812***	0.514***	0.484***
	[0.013]	[0.027]	[0.020]	[0.040]
R-squared	0.03		0.06	0.02
First Stage Estimation				
Frac Courses Taught by Female Faculty		0.619		0.619
		[0.126]		[0.126]
F-Stat		9.84		9.84
Frac Courses Taught by Fem Faculty*Female		0.307		0.307
		[0.082]		[0.082]
F-Stat		7.52		7.52
Observations	34,373	34,373	34,373	34,373
Standard errors in brackets				
* significant at 10%; ** significant at 5%; *** significant at 1%				

These results are fairly robust under other specifications. Table 1.13 redefines the explanatory variable of interest as having at least one female STEM instructor. Tables 1.14 and 1.15 look at the outcome of persistence conditional upon not dropping out of school and use both a count measure and binary measure of female instructors in STEM courses. OLS results indicate that male students have persistence rates 1 to 2 percentage points lower when matched with a female STEM instructor. Once controlling for selection, the point estimates remain negative and get larger in absolute value, but are not statistically significant but are always negative in sign. OLS also indicates that female students are about one percentage point more likely to persist after the first semester when courses in STEM fields are taught by female faculty, and equally as likely to persist after the first year. However, the instrumental variables suggest that female students are significantly less likely to persist in a STEM major after the first semester and first year when courses are taught by female instructors.

In a similar study, Carrel et al (2010) find that on average there are no effects on academic outcomes when female students enroll in STEM courses taught by a female faculty. However, when they look at a subsample of students who are high ability, they find significant and positive effects of high ability females who have female instructors. They use the 75th percentile of the SAT math section (score of 700) as the cutoff for high ability. While the data in this study does not include a detailed breakdown of the ACT exam, I can divide the sample in two ways. First, I set the cutoff at the 75th percentile of the distribution of scores (ACT score of 25). Second, I set the cutoff at 30, which represents high ability students[5]. Table 1.16 defines high ability students as those in the 75th percentile and the results show that having a female instructor has a negative effect on persistence after the first semester and first year.

[5] Using other cutoffs above 30 provides similar results

Table 1.13 Regression Results: Outcome is Persist in STEM Field Major
(At Least One Female Faculty)

	After First Semester		After First Year	
	OLS	IV	OLS	IV
Female	-0.041***	0.06	-0.074***	0.022
	[0.004]	[0.052]	[0.007]	[0.082]
Female STEM Instructor	-0.012***	-0.039	-0.018***	-0.103
	[0.004]	[0.052]	[0.007]	[0.082]
Female STEM Instructor * Female	0.011*	-0.211*	-0.006	-0.211
	[0.006]	[0.115]	[0.010]	[0.181]
ACT Score	0.004***	0.004***	0.012***	0.012***
	[0.000]	[0.001]	[0.001]	[0.001]
black	-0.004	-0.001	-0.039***	-0.036***
	[0.006]	[0.006]	[0.009]	[0.010]
asian	0.027***	0.028***	0.066***	0.066***
	[0.008]	[0.008]	[0.012]	[0.013]
other	-0.002	-0.002	-0.003	-0.004
	[0.007]	[0.007]	[0.011]	[0.011]
Number of Non-STEM Courses	-0.002	-0.003	-0.014***	-0.015***
	[0.003]	[0.003]	[0.005]	[0.005]
Number of STEM Courses	0.018***	0.023***	0.047***	0.055***
	[0.003]	[0.005]	[0.005]	[0.007]
Constant	0.858***	0.836***	0.516***	0.499***
	[0.013]	[0.018]	[0.020]	[0.028]
R-squared	0.03		0.06	0.03
First Stage Estimation				
Frac Courses Taught by Female Faculty		0.286		0.286
		[0.088]		[0.088]
F-Stat		6.54		6.54
Frac Courses Taught by Fem Faculty*Female		0.388		0.388
		[0.056]		[0.056]
F-Stat		13.92		13.92
Observations	34,373	34,373	34,373	34,373
Standard errors in brackets				
* significant at 10%; ** significant at 5%; *** significant at 1%				

Table 1.14 Regression Results: Outcome is Persist in STEM Field Major Conditional on Enrollment (Number of Female STEM Instructors)

	After First Semester		After First Year	
	OLS	IV	OLS	IV
Female	-0.040***	0.126*	-0.098***	0.081
	[0.004]	[0.074]	[0.006]	[0.114]
Female STEM Instructor	-0.008***	-0.023	-0.012***	-0.058
	[0.003]	[0.041]	[0.004]	[0.064]
Female STEM Instructor * Female	0.008**	-0.273**	0.003	-0.294
	[0.004]	[0.125]	[0.006]	[0.192]
ACT Score	0.003***	0.003***	0.009***	0.008***
	[0.000]	[0.001]	[0.001]	[0.001]
black	0.006	0.009	0.007	0.004
	[0.005]	[0.006]	[0.009]	[0.010]
asian	0.025***	0.026***	0.061***	0.062***
	[0.007]	[0.008]	[0.012]	[0.012]
other	0.005	0.004	0.021**	0.016
	[0.006]	[0.007]	[0.010]	[0.011]
Number of Non-STEM Courses	-0.007**	-0.006*	-0.024***	-0.024***
	[0.003]	[0.003]	[0.005]	[0.005]
Number of STEM Courses	0.015***	0.019***	0.040***	0.046***
	[0.003]	[0.004]	[0.005]	[0.006]
Constant	0.890***	0.843***	0.660***	0.611***
	[0.012]	[0.025]	[0.019]	[0.037]
R-squared	0.03		0.05	
First Stage Estimation				
Frac Courses Taught by Female Faculty		0.619		0.619
		[0.126]		[0.126]
F-Stat		9.84		9.84
Frac Courses Taught by Fem Faculty*Female		0.307		0.307
		[0.082]		[0.082]
F-Stat		7.52		7.52
Observations	33,743	33,743	30,652	30,652
Standard errors in brackets				
* significant at 10%; ** significant at 5%; *** significant at 1%				

Table 1.15 Regression Results: Outcome is Persist in STEM Field Major Conditional on Enrollment (At Least One Female Faculty)

	After First Semester		After First Year	
	OLS	IV	OLS	IV
Female	-0.043***	0.061	-0.097***	0.035
	[0.004]	[0.045]	[0.006]	[0.069]
Female STEM Instructor	-0.013***	-0.02	-0.020***	-0.085
	[0.004]	[0.049]	[0.006]	[0.077]
Female STEM Instructor * Female	0.015***	-0.216**	0.003	-0.285*
	[0.006]	[0.099]	[0.010]	[0.153]
ACT Score	0.003***	0.003***	0.009***	0.009***
	[0.000]	[0.000]	[0.001]	[0.001]
black	0.006	0.008	0.007	0.007
	[0.005]	[0.006]	[0.009]	[0.010]
asian	0.025***	0.025***	0.061***	0.061***
	[0.007]	[0.007]	[0.012]	[0.012]
other	0.005	0.004	0.020**	0.018*
	[0.006]	[0.006]	[0.010]	[0.011]
Number of Non-STEM Courses	-0.007**	-0.008**	-0.024***	-0.027***
	[0.003]	[0.003]	[0.005]	[0.005]
Number of STEM Courses	0.016***	0.020***	0.040***	0.049***
	[0.003]	[0.004]	[0.005]	[0.007]
Constant	0.891***	0.867***	0.661***	0.636***
	[0.012]	[0.016]	[0.019]	[0.025]
R-squared	0.03		0.05	
First Stage Estimation				
Frac Courses Taught by Female Faculty		0.286		0.286
		[0.088]		[0.088]
F-Stat		6.54		6.54
Frac Courses Taught by Fem Faculty*Female		0.388		0.388
		[.056]		[.056]
F-Stat		13.92		13.92
Observations	33,743	33,743	30,652	30,652
Standard errors in brackets				
* significant at 10%; ** significant at 5%; *** significant at 1%				

Table 1.16 Regression Results: Outcome is Persist in STEM Field Major (Number of Female STEM Instructors, ACT Score ≥ 25 (75^{th} Percentile))

	After First Semester		After First Year	
	OLS	IV	OLS	IV
Female	-0.033***	0.081	-0.068***	-0.011
	[0.005]	[0.050]	[0.009]	[0.082]
Female STEM Instructor	-0.008**	0.005	-0.014**	-0.057
	[0.004]	[0.047]	[0.007]	[0.078]
Female STEM Instructor * Female	0.002	-0.202**	-0.004	-0.098
	[0.006]	[0.088]	[0.010]	[0.146]
ACT Score	0.005***	0.005***	0.014***	0.013***
	[0.001]	[0.001]	[0.002]	[0.002]
black	-0.001	-0.01	-0.014	-0.018
	[0.015]	[0.016]	[0.025]	[0.026]
asian	0.012	0.009	0.061***	0.060***
	[0.010]	[0.011]	[0.018]	[0.018]
other	0	-0.008	0.001	-0.003
	[0.010]	[0.011]	[0.018]	[0.019]
Number of Non-STEM Courses	-0.001	-0.002	-0.023***	-0.023***
	[0.004]	[0.005]	[0.008]	[0.008]
Number of STEM Courses	0.016***	0.018***	0.056***	0.058***
	[0.004]	[0.005]	[0.007]	[0.008]
Constant	0.825***	0.802***	0.483***	0.471***
	[0.026]	[0.029]	[0.045]	[0.048]
R-squared	0.03		0.04	0.02
First Stage Estimation				
Fract Courses Taught by Female Faculty		0.286		0.286
		[0.088]		[0.088]
F-Stat		6.54		6.54
Frac Courses Taught by Female Faculty*Female		0.388		0.388
		[.056]		[.056]
F-Stat		13.92		13.92
Observations	14,021	14,021	14,021	14,021
Standard errors in brackets				
* significant at 10%; ** significant at 5%; *** significant at 1%				

Table 1.17 Regression Results: Outcome is Persist in STEM Field Major
(Number of Female STEM Instructors, ACT Score ≥ 30)

	After First Semester		After First Year	
	OLS	IV	OLS	IV
Female	-0.033***	0.081	-0.068***	-0.011
	[0.005]	[0.050]	[0.009]	[0.082]
Female STEM Instructor	-0.008**	0.005	-0.014**	-0.057
	[0.004]	[0.047]	[0.007]	[0.078]
Female STEM Instructor * Female	0.002	-0.202**	-0.004	-0.098
	[0.006]	[0.088]	[0.010]	[0.146]
ACT Score	0.005***	0.005***	0.014***	0.013***
	[0.001]	[0.001]	[0.002]	[0.002]
black	-0.001	-0.01	-0.014	-0.018
	[0.015]	[0.016]	[0.025]	[0.026]
asian	0.012	0.009	0.061***	0.060***
	[0.010]	[0.011]	[0.018]	[0.018]
other	0	-0.008	0.001	-0.003
	[0.010]	[0.011]	[0.018]	[0.019]
Number of Non-STEM Courses	-0.001	-0.002	-0.023***	-0.023***
	[0.004]	[0.005]	[0.008]	[0.008]
Number of STEM Courses	0.016***	0.018***	0.056***	0.058***
	[0.004]	[0.005]	[0.007]	[0.008]
Constant	0.825***	0.802***	0.483***	0.471***
	[0.026]	[0.029]	[0.045]	[0.048]
R-squared	0.03		0.04	0.02
First Stage Estimation				
Frac Courses Taught by Female Faculty		0.286		0.286
		[0.088]		[0.088]
F-Stat		6.54		6.54
Frac Courses Taught by Female Faculty*Femal		0.388		0.388
		[.056]		[.056]
F-Stat		13.92		13.92
Observations	14,021	14,021	14,021	14,021
Standard errors in brackets				
* significant at 10%; ** significant at 5%; *** significant at 1%				

Redefining the cutoff of high ability at an ACT score of 30 shows that there are no negative effects on persistence of female student for additional STEM courses which are taught female instructors (See Table 1.17).

The general result that female students are less likely to persist when they have a female instructor in an introductory STEM corresponds with findings in previous studies. Hoffman & Oreopoulos (2009) find that female students who have female instructors in math and science courses get lower grades in those courses. There is evidence to suggest that female students are more sensitive to grades, and are less likely to take additional courses than male students when they receive poorer grades (Rask & Tiefenthaler, 2008). Therefore, if female students receive lower grades in a STEM course when taught by a female instructor, then this could explain why female students are less likely to persist in STEM fields. Subsequently, multiple studies find that female students enroll in significantly fewer classes in STEM fields when initial courses are taught be female faculty (Hoffman & Oreopoulos, 2008; Bettinger & Long, 2005).

Conclusion

The current study seeks to explain why minority and female students are less likely to persist in STEM fields. The black-white persistence gap begins to emerge after the first semester, and continues to grow after each semester. After 3 years, about 30 percent of initial STEM majors who are black are still in a STEM related field, compared to almost half of white students. Prior preparation, as measured by ACT scores, explains about half of the black-white persistence gap for minority STEM majors. This indicates that there is much to do in both preparing underrepresented minorities prior to enrolling in college as well as during the first few years in their college experience to be able to succeed in STEM related majors. There are also distinct patterns of persistence between male and female STEM majors. The gender persistence gap is smaller in magnitude than the black-white persistence gap, but emerges even after the first semester and continues to increase after the sixth semester (40 percent persistence for women versus 53 percent for men). However, controlling for prior preparation does not significantly degrease the gender. This suggests that more could be done within the college setting to improve persistence of women in STEM majors.

The main objective of this study was to test the hypothesis that students who have STEM courses taught by an instructor of their own race or gender are more likely to persist in a STEM major. The empirical evidence provided in this study suggests that black students who enroll in STEM courses taught by black instructors are more likely to persist in a STEM field after the first year. Furthermore, this study also suggests that female students are less likely to persist in a STEM field when courses in these fields are taught by female instructors.

Findings from this study would suggest that increasing the number of black faculty teaching introductory STEM courses would have a positive influence on improving persistence of black students. But the limitation of the study is that it does not identify the mechanism driving the result. If black instructors serve as mentors to black students, then maybe schools could do more to facilitate and foster mentor relationships between students and faculty. If the presence of black instructors in the classroom serve as role models or help improve student's view of self efficacy, then just having black faculty in the department could have the same effect. What is really needed is future research designed to understand the mechanism, such that policy implications can be provided to increase the representation of minority students persisting toward a degree in a STEM field.

REFERENCES

Bettinger, Eric and Bridget Terry Long, 2005. "Do Faculty Serve as Role Modes?" *American Economic Review.* 95(2): 152-157

Bettinger, Eric 2007. "To Be or Not to Be: Major Choice in Budding Scientists" *NBER Working Paper 11593*

Campbell, Toni A. and David E. Campbell, 1997. "Faculty/Student Mentor Program: Effects on Academic Performance and Retention." *Research in Higher Education.* 38(6):727-741

Canes, Brandice J. and Harvey S. Rosen, 1995. "Following in Her Footsteps? Faculty Gender Composition and Women's Choices of College Majors." *Industrial and Labor Relations Review* 48(3): 486-504

Carrel, Scott, Marianne E. Page, and James E. West, 2009. "Sex and Science: How Profesor Gender Perpetuates the Gender Gap." *NBER Working Paper 14959.*

Crisp, Gloria and Irene Cruz, 2009. "Mentoring College Students: A Critical Review of the Literature Between 1990 and 2007." *Research in Higher Education.* 50(6): 525-546

Dee, Thomas, 2004. "Teachers, Race, and Student Achievement in a Randomized Experiment." *The Review of Economics and Statistics 86(1): 195-210.*

Dee, Thomas, 2005. "A Teacher Like Me: Does Race, Ethnicity, or Gender Matter?" *The American Economic Review.* 95(2):158-165

Dee, Thomas, 2007. "Teachers and the Gender Gaps in Student Achievement." *Journal of Human Resources.* 42(3): 528-554

Ehrenberg, Ronald and Dominic Brewer, 1995. "Did Teacher's Verbal Ability and Race Matter in the 1960's?" *Economics of Education Review.* 14(1): 1-21

Ehrenberg, Ronald, Daniel Goldhaber, and Dominic Brewer, 1995. "Do Teacher's Race, Gender, and Ethnicity Matter? Evidence from the National Educational Longitudinal Study of 1988." *Industrial and Labor Relations Review.* 48(3): 547-561

Hoffman, Florian and Philip Oreopoulos, 2007. "A Professor Like Me: The Influence of Instructor Gender on College Achievement." *Journal of Human Resources.* 44(2): 479-494

Holmlund, Helena and Krister Sund, 2008. "Is the Gender Gap in School Performance

Affected by the Sex of the Teacher?" *Labour Economics.* 15: 37-53

Jacobi, Maryann, 1991. "Mentoring and Undergraduate Academic Success: A Literature Review." *Review of Educational Research.* 61(4): 505-532

Karunanayake, Danesh and Margaret M. Nauta, 2004. "The Relationship Between Race and Students' Identified Career Role Models and Perceived Role Model Influence." *The Career Development Quarterly* 52(3):225-234.

Klopfenstein, Kristin, 2005. "Beyond Test Score: The Impact of Black Teacher Role Models on Rigorous Math Taking." *Contemporary Economic Policy.* 23(3): 416-428

National Science Foundation, 2008. *Science and Engineering Indicators.*

National Center of Educational Statistics, 2003. *Science and Engineering Indicators*, National Science Foundation.

Nixon, Lucia A. and Michael D. Robinson, 1999. "The Educational Attainment of Young Women: Role Model Effects of Female High School Faculty." *Demography.* 36(2): 185-194

Rask, Kevin and Elizabeth Bailey, 2002. "Are Faculty Role Models? Evidence from Major Choice in an Undergraduate Institution." *The Journal of Economic Education.* 33(2):99-124

Robst, John, Jack Keil, and Dean Russo, 1999. "The Effect of Gender Composition of Faculty on Student Retention." *Economics of Education Review.* 17(4): 429-439

Rothstein, Donna, 1995. "Do Female Faculty Influence Female Students' Educational and Labor Market Attainments?" *Industrial and Labor Relations Review*48(3):515-530

Solnick, Sara J. 1995. "Changes in Women's Majors from Entrance to Graduation at Women's and Coeducational Outcomes." *Industrial and Labor Relations Review.* 48(3):505-514

CHAPTER 2

OUTCOMES IN A PROGRAM THAT OFFERS FINANCIAL INCENTIVES FOR WEIGHT LOSS[6]

Abstract

A large percentage of Americans are attempting to lose weight at any given time but the vast majority of weight loss attempts fail. Researchers continue to search for safe and effective methods of weight loss, and this paper examines one promising method - offering financial rewards for weight loss.

This paper studies data on 2,407 employees in 17 worksites who participated in a year-long worksite health promotion program that offered financial rewards for weight loss. The intervention varied by employer, in some cases offering steady quarterly rewards for weight loss and in other cases requiring participants to post a bond that would be refunded at year's end conditional on achieving certain weight loss goals. Still others received no financial incentives at all and serve as a control group. We examine the basic patterns of enrollment, attrition, and weight loss in these three groups.

Weight loss is modest. After one year, it averages 1.4 pounds for those paid steady quarterly rewards and 3.6 pounds for those who posted a refundable bond, under the assumption that dropouts experienced no weight loss. Year-end attrition is as high as 76.4%, far higher than that for interventions designed and implemented by researchers.

---

[6] This chapter is co-authored with John Cawley

Introduction

A variety of approaches are being used to treat obesity and encourage weight loss. One promising strategy based on psychology and behavioral economics is to offer financial incentives for weight loss. Obesity is costly to health insurance companies (Finkelstein et al. 2003) and employers (Cawley et al. 2007), so for either or both of those organizations to offer monetary incentives for enrollees or employees to lose weight could be mutually beneficial.

This paper studies data from a firm that coordinates a program of financial incentives for weight loss in various worksites in the U.S. We study attrition and weight loss in three types of incentive programs: one that offers no financial rewards for weight loss, one that offers quarterly payments that rise in value with the amount of weight loss, and a third that takes deposits (bonds) that are only refunded if the employee achieves a specific weight loss goal and also includes a quarterly lottery for those who have lost weight. Relative to previous studies of weight loss in response to financial incentives, strengths of this study include a large sample size (2,407) and a long intervention (one year).

A 2007 Institute of Medicine report on obesity prevention set the immediate next step – which it described as an essential priority action for the near future – as "learning what works and what does not work and broadly sharing that information." (IOM 2007, p. 410). It also notes that "All types of evaluation can make an important contribution to the evidence base upon which to design policies, programs, and interventions." (Ibid, p. 4). This paper makes a contribution to that effort by documenting enrollment, attrition, and weight loss in one interesting and promising intervention. This paper presents basic patterns in the data; a subsequent paper will estimate regression models to test specific hypotheses about attrition and weight loss.

Conceptual Framework and Previous Literature

For obese people, weight loss would likely result in substantial benefits.[7] For example, the health benefits of modest weight loss (defined as 5-10% of starting weight) include decreased blood pressure and cholesterol and a 25% reduction in mortality risk for type 2 diabetics (Vidal, 2002). Weight loss may also improve quality of life (Ford et al. 2001). There may also be financial benefits. Cawley (2004) finds a causal impact of weight on wages, and that obese white females earn roughly 11 percent less than healthy-weight white females. Finkelstein et al. (2003) calculate that, relative to the healthy weight, the obese incur $125 higher annual out-of-pocket health care costs. With two-thirds of Americans overweight or obese (Ogden et al. 2006), and given these potential benefits of weight loss, it may not be surprising that 46 percent of all American women and 33 percent of all American men are trying to lose weight (Bish et al. 2005).

Most people fail in their attempts to lose weight[8], and many of those who are successful in losing weight regain it in a short period of time. [9] For example, in one community-based study of weight gain prevention (Crawford et al., 2000), most

---

[7] There are two ways researchers have sought to measure the benefits of weight loss. The first is to examine changes in outcomes associated with losing weight. The second is to compare the outcomes of individuals of different weight, and assume that the difference in outcomes is due to the difference in weight. Each has its limitations: weight loss studies often lack power, and comparisons across weight levels are confounded by differences in unobserved characteristics. Vidal (2002) assesses the evidence on the benefits of weight loss and concludes that modest weight loss (5-10% of initial body weight) improves cardiovascular risk factors and helps prevent or delay the onset of type 2 diabetes and hypertension.

[8] Some obese individuals are able to lose weight by modifying their behaviors: eating less and exercising more. In the select group enrolled in the Weight Control Registry, all of whom have lost at least 30 pounds and kept it off for at least one year, 44.6% report losing the weight entirely on their own, that is, without the help of a commercial program, physician, or nutritionist (Wing and Phelan, 2005). Clearly such statistics do not generalize to the population; anyone who failed at initial weight loss is ineligible for this registry of people who maintained weight loss for a year.

[9] Conventional wisdom is that virtually no one succeeds at maintaining weight loss. This perception has been traced back to a 1959 study of 100 obese individuals in which only 2 percent maintained loss of 20 pounds or more two years after the treatment (Stunkard and McLaren-Hume 1959; Wing and Phelan 2005). However, the 1959 study was based on a crude diet intervention with negligible support or follow-up so its poor results may not generalize to today's much more intensive interventions.

(53.7%) participants *gained* weight in the first twelve months, three-quarters gained weight over three years, and only 4.6% lost weight and maintained the loss for three years.

Theory and evidence from psychology and behavioral economics provide several explanations for why so many weight loss attempts fail. First, the benefits of weight loss are not salient. For example, foregone quality of life and lost wages are not visible and therefore they are frequently unrecognized as opportunity costs (Bastiat, 1850).

A second possible explanation for repeated failure at weight loss is that the benefits of weight loss may not be immediate. Improvements in health and labor market outcomes may not occur for some time after weight loss, and Ainslie (1975) finds consistent evidence that there is a decline in the effectiveness of rewards as the rewards are delayed from the time of choice.

A third explanation for repeated failure at weight loss is that, contrary to the standard economic model of discounted utility (Samuelson 1937), people may discount hyperbolically, which produces time-inconsistent preferences (Ainsley 1975). In this context, time inconsistent preferences mean that people want to do what is in their long-run interest (lose weight), but they consistently succumb to the temptation to eat and be sedentary. Thaler and Shefrin (1981) describe individual decision-making as a battle between a farsighted planner (who in this context wants to diet) and a myopic doer (who in this context wants to eat and be sedentary).

One intervention, financial rewards for weight loss, may offer a solution to the problems of salience, immediacy, and time-inconsistency. Financial rewards, even though they may be dwarfed in value by the other benefits of weight loss, have the benefit of being salient, with their amount and delivery date known with certainty in exchange for clearly defined objectives. Even small financial incentives can be effective because research has found that people tend not to compare payoffs to their income or wealth but instead “bracket” them - consider them in isolation (Read et al. 1999; Kahneman and Tversky 1979). Lotteries may be particularly cost-effective incentives for healthy behavior. People tend to overweight the probability of unlikely events and underweight the probability of likely events (Kahneman and Tversky 1979), implying that lotteries can be more attractive than certain payments even if the two have equal expected values. Financial rewards can also be paid immediately, before other benefits of weight reduction may be realized.

Financial rewards can also be structured to help people with time-inconsistent preferences stay committed to weight loss. In general, pre-commitment devices may help people with time-inconsistent preferences empower their farsighted planner (Strotz 1956; Laibson 1997). In this context, one could allow people to post a bond that is automatically forfeited if they fail to achieve their weight loss goals. Such a bond allows a person to influence their own future decisions by increasing the punishment for succumbing to short-run temptation. People tend to exhibit loss aversion – they dislike losing their own money more than they like winning an equal amount of someone else’s money (Tversky and Kahneman 1991; Camerer 2005), which suggests that a posted bond may be more effective than a reward of the same size. Using a bond to increase adherence to a weight-loss regimen does not guarantee success. Even individuals who are aware of their time-inconsistent preferences may still be partially naive in that they overestimate their future willpower (O’Donoghue and Rabin 2001) and as a result may either post too small a bond or have too much faith in the bond as a precommitment device.

Motivated by these theories and findings, several businesses now help employers offer financial incentives for employee weight loss. In addition, several businesses help consumers post bonds that are only refunded if one achieves specific weight loss goals. The William Hill betting agency in the U.K. books wagers that the bettor cannot achieve a specified weight loss in a specific period of time and verifies the weight loss with a medical examination (Burger and Lynham 2008).[10] A company named stickK.com[11] that was founded by Yale economists Ian Ayres and Dean Karlan allows people to post bonds that are forfeited if they fail to meet their weight loss goal. However, verification is weak: success in achieving one's goal is determined (and refunds are made) based on either the honor system or through verification by a third party chosen by the bettor, and if the third party doesn't submit a report the self-report of the bettor is accepted.

The contribution of this paper is to examine outcomes in a program that offers various financial rewards (including certain payments, lotteries, and refundable bonds) for weight loss. The outcomes we examine include attrition and weight loss, both in pounds and as a percentage of baseline weight.

A substantial literature confirms that financial incentives influence healthy behaviors. Kane et al. (2004) review 42 studies of the effect of economic incentives on preventive behaviors such as immunization, smoking cessation, and exercise; they find that the economic incentives were effective at changing behavior in 73% of studies. Financial incentives form the basis for an innovative substance abuse treatment program known as contingency management. A meta-analysis found overwhelming evidence that such incentives raise compliance (drug abstinence) by an average of 30 percent (Lussier et al., 2006). Consistent with bracketing, even small financial incentives have proven effective; for example, as little as $2.50 for a single negative test result for cocaine (Higgins et al., 2002).

---

[10] This market is relatively small – the annual number of applications for such bets is roughly 200 (Burger and Lynham, 2008)

[11] The website's Frequently Asked Questions page states that the company's name includes two K's because "K" often symbolizes "contract" in legal writing.

Specific to the current context, there is mixed empirical evidence on the extent to which weight loss is responsive to financial rewards. A recent review and meta-analysis (Paul-Ebhohimhen and Avenell 2007) identified nine published randomized controlled trials (RCTs) that used guaranteed financial incentives (i.e. certain payments, not lotteries) for weight loss, with a follow-up of at least one year. The meta-analysis was unable to reject the null hypothesis of no effect of financial rewards on weight loss; it calculated a mean weight loss of 0.4 kg at 12 months, which was not statistically significant. A broader set of studies (including, e.g., those with non-randomized designs or shorter follow-up) are listed in Appendix Table 2.1.[12]

Relative to past studies, ours has several advantages. This study has a relatively large sample size (2,407); for comparison, the sample size of all published RCTs of financial incentives for weight loss combined totals 424 (treatment N=252, control N=172) (Paul-Ebhohimhen and Avenell, 2007). The intervention studied by this paper also covers a relatively long time period (one year). Moreover, we examine data from a real-world intervention rather than one constructed by and overseen by researchers, which is important because a criticism of studies of weight loss programs is that it is unclear how the results of pilot programs generalize to real-world implementation. A limitation of this study, however, is that it is opportunistic data; individuals were not randomly assigned to different incentive schedules for weight loss.

Description of the Intervention

Our data come from a company (that we will call Company X) that helps employers provide financial incentives for their employees to lose weight; specifically,

[12] There are other studies that offer financial rewards for exercise or for attending weight loss programs, but Appendix Table 2.1 is limited to studies of financial rewards for weight loss.

it monitors employee weight loss and pays the rewards. After an employer contracts with Company X, Company X has a kickoff event in the workplace that explains the program to the employees and encourages them to sign up. Participation is optional. Those who sign up select a physical activity regimen at either the foundation (easiest), intermediate, or advanced level. The program consists of several elements: 1) daily email coaching that includes information about healthy and effective methods of weight loss including decreasing calorie intake and increasing physical activity in a manner consistent with the regimen the enrollee chose at baseline; 2) call center support; 3) weigh-ins at least once a quarter; and 4) financial incentives for achieving specific weight loss targets. Only employees who are overweight (BMI of at least 25) are eligible to receive financial rewards, and no financial rewards will be paid once an employee's BMI falls below 25 (i.e. when the employee falls into the "healthy weight" category).

The weigh-ins take place in HIPAA-compliant[13] kiosks that company X installs in the employer's workplace. Employees enter the privacy-protected kiosk and stand on a scale; their body mass index is recorded and sent over an internet connection to their personal webpage as well as to Company X's database. Participants can weigh themselves as often as they like, and the lowest recorded weight will be counted as that quarter's weight. Financial rewards are paid based on percent of baseline weight lost.[14]

Company X has a standard set of incentives that it proposes, but employers can

[13] The Health Insurance Portability and Accountability Act (HIPAA) regulates the disclosure of health information.

[14] We asked Company X whether people game the system by trying to weigh more at baseline (from which future weight losses are judged). They said that through the cameras installed in their kiosks they do not see people wearing heavier clothes to the baseline weigh-in than to later weigh-ins; in all cases people seem for vanity reasons to remove shoes and sweaters before weighing in. However, Company X acknowledges that they have no way to know if people (e.g.) hid weights in their pockets or shoes before the baseline weigh-in. If people engage in such deception then we would expect to see significant drops in weight at the first weigh-in after baseline but we do not find this pattern in the data.

modify it. In our data, there are three incentive schedules. The first is Company X's standard set of incentives: the employee participants pay no fee (all costs are paid by the employer), and employees receive quarterly payments determined by percent of baseline weight lost to date. Table 2.1 lists the standard set of incentives: payment thresholds occur at each percentage point of weight loss up to 5% (1, 2, 3, 4, 5), then thresholds occur every 5 percentage points (5, 10, 15, 20, 25, 30) up to 30% of weight loss. The payment associated with these thresholds varies; for the first seven (1, 2, 3, 4, 5, 10, 15) the reward is a dollar per percentage point of weight loss. Then the per-percentage-point rewards increase: $25 for losing 20%, $35 for losing 25%, and $50 for losing 30%. These are monthly amounts that are paid quarterly, so someone who loses 5% of his weight and keeps it off for three months receives a $15 check for the quarter ($5 monthly payment x 3 months). Five employers (with a total of thirteen worksites participating) used this standard incentives schedule.

The second incentive schedule, used by one employer (with two worksites participating), is shown in Table 2.2 and includes both a lottery and a deposit contract. The lottery takes place each quarter and the prizes are gift certificates (ten $50 gift cards and ten $50 salon vouchers); only those who had lost some weight since baseline are eligible for the drawing. The deposit contract is that employees must pay $9.95 per month (except the first month, which is free), all of which (11 * $9.95 or $109.45) is refunded at the end of the year if the respondent loses at least 5% of baseline weight by year's end. If the respondent loses 10% or more of their baseline weight, they receive in addition to their refunded fees ($109.45) a $100 bonus, for a total of $229.40. In addition, the "biggest loser" (as a percent of baseline weight) receives a $250 gift certificate at the end of the year.

Table 2.1 Financial Rewards Based on Weight Loss "Standard incentives"

Weight Loss (as % of Baseline Weight)	Dollar Reward Per Month (Paid Quarterly)
1	1
2	2
3	3
4	4
5	5
10	10
15	15
20	25
25	35
30	50

Notes: Only participants with BMI over 25 (that is, those who are overweight or obese) are eligible to receive incentives. Moreover, people can only get incentives for weight loss down to a BMI of 25 – there is no financial incentive for anyone in the healthy weight (18.5 to 25) or underweight (<18.5) BMI categories to lose weight.

Table 2.2 Financial Rewards Based on Weight Loss "Modified incentives"

Weight Loss (as % of Baseline Weight)	Reward (Some Quarterly, Some Annual)
Greater than zero	Entered into quarterly drawing for gift certificates: ten $50 gift cards each quarter and ten $50 salon vouchers each quarter.
5	Complete reimbursement of monthly fees (11 * $9.95 = $109.45), paid at end of year
10	Complete reimbursement of monthly fees (11 * $9.95 = $109.45) plus $100 bonus, paid at end of year
"Biggest loser" (as % of baseline) at worksite	$250 gift certificate, awarded at end of year, plus the appropriate award listed above for the specific amount of weight loss

Notes: Only participants with BMI over 25 (that is, those who are overweight or obese) are eligible to receive incentives. Moreover, people can only get incentives for weight loss down to a BMI of 25 – there is no financial incentive for anyone in the healthy weight (18.5 to 25) or underweight (<18.5) BMI categories to lose weight.

Whether a participant would receive a higher payoff in the standard or modified group depends on both quarter and magnitude of weight loss. In quarters one through three, the standard incentives are more generous than the modified incentives at all levels of weight loss, with the exception that those losing between 0.1% and 0.9% of baseline weight receive no reward in the standard incentives group but are eligible for the lottery for gift cards in the modified incentives group. In quarter four, the standard incentives are more generous for weight loss of between 1% and 4%, but the modified incentives are more generous for weight loss of 5% or more.

The third schedule, used by one employer (with a total of two worksites), offered no incentives for weight loss, but did include one modest incentive to not attrite: participants were promised $20 if they participated for the entire year (i.e. weighed in at least once in each of the four quarters). This group received all of the features of the Company X intervention (daily emails, call center access, weigh-ins at the kiosk) but were offered no incentives for weight loss, making it useful both as a control group for measuring the impact of financial incentives isolated from all the other program elements, and for estimating the impact of the Company X treatment minus the financial incentives.

Figure 2.1 presents a flow diagram of attrition and analysis for all three groups (standard incentives, modified incentives, control) combined.

Hypotheses

Part of our purpose in this paper is exploratory - to measure enrollment, attrition and weight loss in these programs. We focus in particular on attrition and weight loss as outcomes because the NIH Technology Assessment Conference Panel

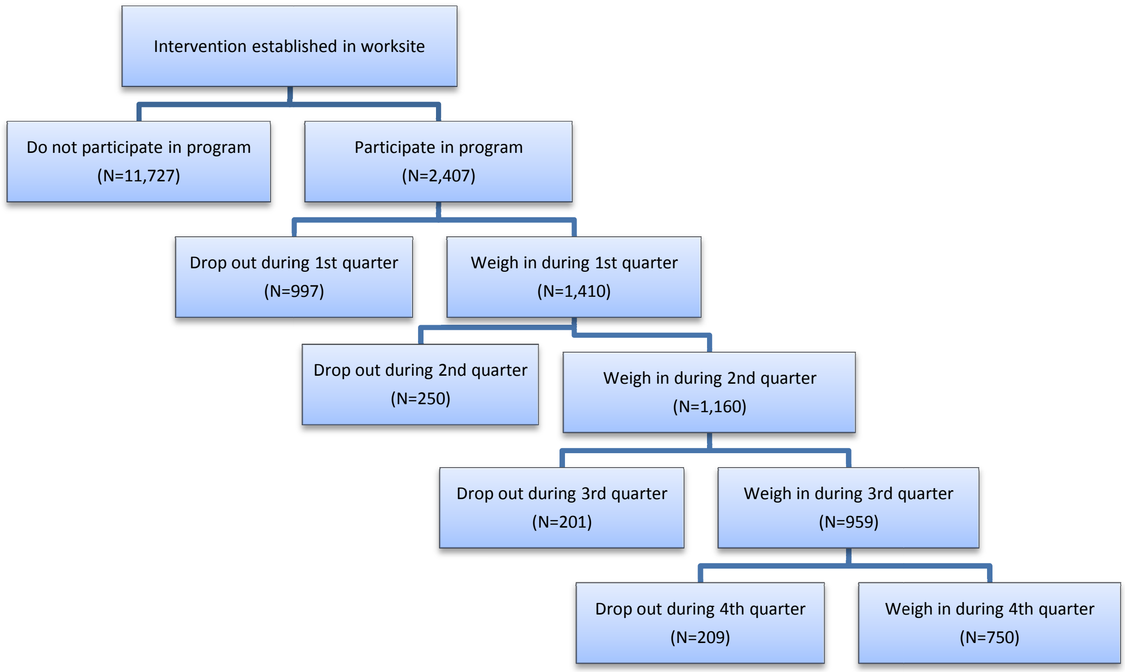


Figure 2.1 Flow Diagram of Attrition and Analysis

(1993) recommends using the percentage of all beginning participants who complete the program, and the percentage of those completing the program who achieve various degrees of weight loss as measures of program success. The NIH considers a loss of 10% of baseline weight in 6 months to one year to be good progress for an obese individual (NHLBI 2000).

Another purpose of this paper is to test the following hypotheses.

Hypotheses Regarding Enrollment

There will be lower enrollment in the program that required people to post forfeitable bonds. The law of demand states that the quantity demanded falls with price. The program that requires people to post a forfeitable bond raises the expected price of participation, assuming that not all possible participants expect a 100% probability of success (and therefore the return of their bond). The published literature confirms that, all else equal, enrollment in weight loss programs is lower if people are asked to post forfeitable bonds (e.g. Jeffery et al. 1978).

Those who are willing to post a forfeitable bond will be better motivated or prepared for weight loss than those not required to post such a bond.

In other words, we expect differential selection – those unwilling to post a forfeitable bond are excluded from the modified incentives group but are not excluded from the control group or standard incentives group. As a result, we expect that the modified incentives group will be better prepared or motivated for weight loss than the other groups.

Hypotheses Regarding Attrition

There will be lower attrition in the program that required employees to post bonds that are refundable based on achievement of weight loss goals. Those willing to post a bond are expected to be more motivated or determined to lose weight. Selection aside, bonds may also increase retention.

Those who attrite will have been relatively unsuccessful at weight loss. Participants enroll with incomplete information about certain costs and benefits of participating. Those that lose relatively little weight may update their prior beliefs and conclude that it is utility maximizing for them to drop out of the intervention. This is especially true of those in the modified incentives group, who are charged a monthly fee for participation that will not be refunded if year-end weight loss is less than 5% of baseline weight.

Hypotheses Regarding Weight Loss

Weight loss will be greater for those offered financial rewards for weight loss. Both the standard incentives group and the modified incentives group were offered financial rewards for weight loss, whereas the control group were not offered any. In other words, we hypothesize that financial rewards are effective in promoting weight loss.

In quarter 4 weight loss will be greater in the modified incentives group than in the standard incentives group.

This prediction is based on the magnitudes of the incentives; the modified incentives group has much greater incentives for 5% and 10% weight loss by the end of quarter 4. Specifically, the modified incentives group receives a refund of their $119.40 bond if at least 5% of weight is lost, with an additional bonus of $100 if 10% of weight is lost. Moreover, those achieving very high weight loss will be in competition for the $250 bonus for being the "biggest loser". In contrast, the standard incentives group is paid $5 per month for losing 5% of starting weight and $10 per month for losing 10% of starting weight (triple those amounts for the entire quarter). Relative to the standard incentives, the modified incentives create greater incentive for weight loss by the end of quarter four.

In addition, there are two reasons that the relative performance of the modified incentives group by the end of quarter 4 might be better than one would expect based on the magnitude of the rewards alone. First, we expect differential selection - those willing to post a bond are likely better prepared or more motivated for weight loss. Second, the research literature on loss aversion indicates that people are more motivated by a risk of losing their own money (as in the modified incentives group) than they are by the prospect of winning someone else's money (as in the standard incentives group).

In quarters 1 through 3, weight loss will be greater in the standard incentives group than in the modified incentives group.

This prediction is also based on the magnitudes of the incentives. In quarters 1 through 3, the standard incentives group is offered $5 per month for 5% weight loss, and $10 per month for 10% weight loss (see Table 2.1 for the full schedule of financial rewards). In contrast, there is no marginal reward for 5% or 10% weight loss in any of the first three quarters for the modified incentives group (those losing any weight at all are eligible for lottery prizes, but there is no additional reward for any weight loss above the trivial amount that makes one eligible for the lottery).

However, there are three reasons that the relative performance of the modified incentives group in quarters one through three might be better than one would expect based on the magnitude of the rewards alone. The first reason is differential selection. The second reason is loss aversion; the fear of losing one's money at year's end may motivate members of the modified incentives group to lose weight in the early quarters, even when there are no quarter-specific rewards for doing so. Third, it may take more than one quarter to achieve 5% or 10% weight loss, so in order to meet their year-end goals members of the modified incentives group may have to lose weight in earlier quarters, even though they have no financial incentives for meaningful weight loss in those quarters.

Methods and Data

A limitation of our data is that they are not the result of a randomized controlled trial. They are opportunistic data, provided to us by Company X. As a result, we face two challenges: 1) assignment to the three treatment groups is nonrandom: the incentive schedules were chosen by the employers; 2) the participation of employees is voluntary; there is selection by employees.

Regarding problem #1 (selection by employers into different incentive schedules), we assume that this is ignorable. In other words, we assume that employer preference for incentive structure is uncorrelated with unobserved employee characteristics that affect attrition and weight loss. Company X told us that the reason that one employer requested the modified incentives schedule (with forfeitable bonds) is because the company didn't want to pay for cash rewards. This would be more problematic if the employer requested the modified schedule because the employer thought it would be more effective for their particular employees.

A related problem is that unobserved employee characteristics may vary systematically across the three groups. Company X designed this intervention for office employees who spend their days in front of computers; it is they, for example, who are most likely to read the daily emails regarding nutrition and physical activity. For the most part, enrollees fit this description. Table 2.3 lists the industries of the employers. The five employers (with a total of 13 worksites) in the standard incentive group include an HMO office, an HMO clinic (in which enrollees are nurses), two bank offices, and an insurance company. The one employer (with a total of two worksites) that instituted the modified incentive schedule is an insurance company, and the one employer (with a total of two worksites) in the control group is a grocery administrative office. Company X tells us that the nurses (who face the standard incentive schedule) have generally been least compliant with the program; they speculate that it may be because they do not work in front of computers all day and thus derive less benefit from the daily emails and the online tracking of measured weight.

Regarding problem #2 (selection by employees into participation), we consider this to be a limitation for generalizing results to the entire population but not a problem in the sense that any similar intervention is also likely to be optional, and so the findings for a set of volunteers is most relevant. All of the studies in Appendix Table 2.1 are all based on volunteers recruited to participate in a weight loss program, and are likewise not a random sample of the general population.

Table 2.3 Description of Employers

Employer	Description	Incentive Schedule
1	HMO clinic - nurses	Standard
2	Banking office	Standard
3	HMO office	Standard
4	Banking office	Standard
5	Insurance office	Standard
6	Insurance company	Modified
7	Grocery administrative office	Control

An additional problem when studying weight loss is that there is attrition from the program. Weight-loss interventions in general (even those without financial rewards) typically have substantial attrition (Ware 2003; Gadbury et al. 2003). There are several strategies for handling the attrition when evaluating interventions. The definitive is the intent-to-treat analysis, which includes all patients in their groups, regardless of whether they received the treatment, deviated from the protocol or withdrew (Ware 2003). However, to implement this one must have follow-up data on all of the dropouts, which is not available in this case. Another option is to conduct a "completers" analysis, which examines data only for those who completed the study. This is likely to be biased toward showing an impact of the treatment, as those most likely to quit are probably those for whom the intervention was least effective (Ware 2003). Another option is last-observation-carried-forward, which assumes that the dropouts remained at their last measured weight. This also likely results in upward bias in estimates of program effectiveness, as weight regain is common (Ware 2003; Serdula et al. 1999). Another option is baseline-carried-forward, which assumes that after attriting the subjects return to their baseline weight. This may cause downward bias in the estimate of efficacy, as weight regain may be incomplete or slow. We present findings for completers analysis, last-observation-carried-forward, and baseline-carried-forward.

The total number of employees in the dataset is 2,407: 1,513 facing the standard incentives, 765 facing the modified incentives, and 129 in the control group with no financial incentives. The data cover 2004-2008. We drop from the sample participants with baseline BMI below 25 because they were not eligible for financial rewards. Thirteen participants in the control group were dropped because they were simultaneously participating in another workplace weight-loss intervention.

We estimate attrition rates by quarter and group. We graph the distribution of weight loss by group and quarter, both for a completers analysis (ignoring dropouts), assuming that dropouts stayed at their last measured weight (last observation carried forward) and assuming that dropouts return to baseline (baseline carried forward). We also calculate the unconditional mean loss in pounds and percent of baseline weight lost by group and quarter, for a completers analysis, last observation carried forward, and baseline carried forward.

Empirical Results

Descriptive statistics:

Table 2.4 presents the summary statistics for participants by group. Our overall sample (N=2,407) consists of 1,513 participants in the standard incentives group, 765 participants in the modified incentives group, and 129 participants in the control group.

In each of these groups, men are a minority: 15.7% of the standard incentives group, 21.2% of the modified incentives group, and 35.7% of the control group. The average age of participants ranges from 43.0 to 46.2 across groups, and average baseline BMI ranges between 31.3 and 32.8 across groups. In each group there is a strikingly high prevalence of morbid obesity (BMI of greater than or equal to 40). In the U.S. as a whole, the morbidly obese constitute 4.8% of the population and 7.3% of all overweight Americans (Ogden et al. 2006). In contrast, the morbidly obese constitute 28.7% of the standard incentives group, 30.5% of the modified incentives group, and 22.5% of the control group.

Enrollment

We hypothesized that: **There will be lower enrollment in the program that required people to post forfeitable bonds.** Table 2.5 lists the percent of the workforce that enrolled in the program, by incentive schedule. Ideally we would know the number of employees with BMI of 25 or higher, because only they are eligible for financial rewards for weight loss. Instead, for the denominator we know only the total number of employees (i.e. those of all BMI). As a result, these are likely to be underestimates of the percentage of those eligible for financial rewards who enrolled in the program. Percent enrollment was 18.6% for the modified incentives (which required a bond), 24.8% for the standard incentives, and 20.3% for the program that offered no financial rewards for weight loss but all of the other program elements (i.e. the control group). The point estimates of enrollment are consistent with our prediction that the requirement of a bond would result in lower enrollment, but the differences are not statistically significant.

We also hypothesized that: **Those who are willing to post a forfeitable bond will be better motivated or prepared for weight loss than those not required to post such a bond.** There are two variables that can give us information about the degree of such differences in selection. The first variable is the level of exercise regimen that the employee chose at the beginning of the program. If those willing to pay the monthly fees in the modified incentives group are more motivated or prepared to lose weight, one should find that they are less likely to choose the easiest exercise regimen. This is confirmed by the data. Table 2.4 indicates that the easiest exercise regimen (called Foundation) was chosen by 60.1% of the standard incentives group but only 55.0% of the modified incentives group, a difference significant at the 1% level. We also expected that the control group, offered \$20 if they participated for the full year, would be less motivated on average and therefore more likely to choose the easiest exercise regimen than those in the modified incentives group, but we do not find this – an even lower percentage of the control group than the modified incentives group (48.8% versus 60.1%) chose the easiest exercise regimen, but the difference is not statistically significant.

The second variable that sheds light on difference in selectivity is the percentage of the program emails that enrollees read. If those willing to pay the monthly fees in the modified incentives group are more motivated or prepared to lose weight, one should find that they read a higher percentage of the program emails. That prediction is confirmed by the data – Table 2.4 indicates that the average percentage of emails read was 51.0% for members of the modified incentives group compared to 45.7% for members of the standard incentives group, a difference significant at the 1% level. (A caveat is that this variable is missing for 51.1% of the standard incentives group – it simply wasn't recorded for certain employers in certain years.)

The control group, being paid to participate, had the lowest email open rate of 28.7%, which is significantly different from both other groups at the 1% level. It is interesting that the control group had the lowest percentage choosing the easiest exercise regimen (which suggests more motivation or better preparation) but the lowest email open rate (which suggests lower commitment).

Overall, the patterns of both exercise regimen and email opening suggest that the group required to post a bond (i.e. the modified incentives group) was selected to be better prepared and more serious about weight loss than the standard incentives group, and therefore should be less likely to attrite and more likely to lose weight.

Table 2.4 Summary Statistics by Group

Variable	Standard Incentives			Modified Incentives			Control Group		
	Obs	Mean	Std. Dev	Obs	Mean	Std. Dev	Obs	Mean	Std. Dev
Initial BMI	1513	32.8	6.24	765	32.8	6.00	129	31.3	5.72
Male	1513	0.157	0.364	765	0.212	0.409	129	0.357	0.481
Age	1513	46.2	10.4	765	43.0	8.8	129	44.4	10.6
Height	1513	65.5	3.41	765	66.1	3.42	129	66.7	4.25
Overweight (30>BMI>=25)	1513	0.412	0.492	765	0.382	0.486	129	0.519	0.502
Obese (40>BMI>=30)	1513	0.301	0.459	765	0.314	0.464	129	0.256	0.438
Morbidly Obese (BMI>=40)	1513	0.287	0.452	765	0.305	0.461	129	0.225	0.419
Foundation exercise regimen	1513	0.601	0.490	765	0.550	0.498	129	0.488	0.502
Intermediate exercise regimen	1513	0.337	0.473	765	0.374	0.484	129	0.426	0.496
Advanced exercise regimen	1513	0.062	0.241	765	0.076	0.265	129	0.085	0.280
Email open rate	740	45.7	36.41	765	51.0	35.09	129	28.7	32.47

Table 2.5 Enrollment Rates

	Control Group (1)	Standard Incentive Group (2)	Modified Incentive Group (3)	p-value (1) equals (2)	p-value (1) equals (3)	p-value (2) equals (3)
Mean	0.203	0.248	0.186	0.613	0.839	0.477
(Std. Dev.)	(0.100)	(0.115)	(0.024)			

Note: Enrollment rates are calculated by the fraction of those who enroll in the program by the total population of the work place. Individuals with BMI<25 may enroll in the program, but receive no payouts.

Attrition

Table 2.6 lists the cumulative percentages dropping out, by quarter, for each group. In the standard incentives group, 51.2% of baseline participants have dropped

Table 2.6 Cumulative Attrition, by Group and Quarter

Quarter	Standard Incentives	Modified Incentives	Control Group
1	51.2%*[t]	24.8%	25.6%
2	62.1%*[t]	33.5%	39.5%
3	72.0%*[t]	39.3%	45.0%
4	76.4%*[t]	57.4%*	48.1%

* represents significant difference with the control group at the 5% level
[t] represents significant difference between standard and modified incentive groups at the 5% level

out by the end of quarter 1, and cumulative attrition rises in the three subsequent quarters to 62.1%, 72.0% and 76.4%. In the modified incentives group, attrition is lower: 24.8% after one quarter, rising in the three subsequent quarters to 33.5%, 39.3%, and 57.4%. Even in the control group, where participants are promised $20 if they weigh in every quarter for a year, attrition is substantial: 25.6% after one quarter, rising in the three subsequent quarters to 39.5%, 45.0%, and 48.1%. When considering the levels of attrition, one should keep in mind that enrollees were already a select sample. Participation was optional, and most employees declined to enroll.

Attrition is typically substantial in weight loss interventions of all kinds (Ware 2003; Gadbury et al. 2003). However, the attrition in these groups is particularly high. For example, a recent review (Paul-Ebhohimhen and Avenuell, 2007) of RCTs involving financial rewards for weight loss found that the maximum attrition in any such study was 57.9% at 13 months, far below what the standard incentives group experienced in 12 months (76.4%) but roughly equal to what the modified incentives group experienced at 12 months (57.4%). This suggests that real-world interventions may experience far higher rates of attrition than those overseen by researchers (who for the purposes of data quality undertake extensive efforts to keep enrollees from attriting), which raises questions about how well the results of pilot studies such as those in Appendix Table 2.1 can be duplicated on a larger scale.

We hypothesized that **There will be lower attrition in the program that required employees to post bonds that are refundable based on achievement of weight loss goals.** The data are consistent with this hypothesis; in every quarter, attrition is significantly lower in the modified than the standard incentives group. For example, after quarter 1 attrition in the modified incentives group is only half that in the standard incentives group (24.8% versus 51.2%). It impossible to tell from our data whether the difference in attrition is due to selection or loss aversion. Selection was evident in the earlier finding that those in the modified incentives group were more likely to choose an advanced physical activity regimen and tend to open more program emails; before entering the program they may have been better prepared and more motivated to lose weight. On the other hand, those in the modified incentives group have “skin in the game” in the form of their deposits, and loss aversion may motivate them to stay in the program.

We also hypothesized that: **Those who attrite will have been relatively unsuccessful at weight loss.** Table 2.7 lists the weight loss by quarter, categorized by whether the participant dropped out in the following quarter or persisted in the program through the following quarter. The table is divided vertically into four panels: full sample, standard incentives group, modified incentives group, control group. Among the full sample, those who drop out in the subsequent quarter have significantly lower average weight loss than those who persist through the next quarter, in quarters 1, 2, and 3. In each case the difference in mean weight loss to date is statistically significant at better than the 1% level. When we divide the sample by incentive schedule, the same pattern exists for those in the modified incentives group: in each of the first three quarters, weight loss to date is significantly lower among those who drop out in the following quarter than those who persist through the following quarter. Note that those in the modified incentives group have the greatest incentive to drop out if they are not making progress, because to persist would require paying monthly fees that one is unlikely to have refunded. The pattern is weaker for the standard incentives group; in quarter 2 future dropouts have significantly lower weight loss than those who persist through the next quarter, but the difference is not statistically significant. In quarter 1 and in quarter 3 the sign is in the opposite direction and the difference is not statistically significant. For the control group, in no quarter do future dropouts have significantly lower weight loss to date than those who will persist in the program.

Weight Loss

The distribution of percent weight loss, by quarter, is shown in Figures 2 (for the standard incentives group), Figures 3 (for the modified incentives group), and Figures 4 (for the control group). The horizontal axis shows the percent of baseline weight lost (rounded down to the nearest percentage point[15]) as of that quarter, and the vertical axis indicates the percentage of that sample. For each group, there is a separate page devoted to the data for each quarter. On each page are three graphs: the top graph is the distribution of weight loss in a completers analysis that ignores dropouts, the middle graph is from a last observation carried forward analysis in which dropouts are assumed to have stayed at their last measured weight, and the bottom

[15] We round down so that everyone indicated as having a specific percent weight loss received exactly the reward associated with that percent weight loss. If we rounded to the nearest percentage point, a participant who lost 4.6% of her starting weight would be rounded to 5% even though she would not have qualified for the financial reward associated with achieving 5% weight loss.

Table 2.7 Weight Loss by Future Attrition Status

Full Sample			
Quarter	Persist in next quarter	Dropout next quarter	ttest p-value
1	4.67 (2.3%)	3.49 (1.8%)	0.004
2	5.73 (2.8%)	3.33 (1.7%)	0.000
3	6.38 (3.1%)	4.23 (2.0%)	0.008
Standard Incentive Group			
Quarter	Persist in next quarter	Dropout next quarter	ttest p-value
1	4.90 (2.5%)	4.07 (2.1%)	0.122
2	6.67 (3.2%)	3.99 (2.0%)	0.003
3	6.93 (3.2%)	8.96 (4.3%)	0.128
Modified Incentive Group			
Quarter	Persist in next quarter	Dropout next quarter	ttest p-value
1	4.66 (2.3%)	1.03 (0.6%)	0.000
2	5.36 (2.6%)	0.76 (0.3%)	0.001
3	6.77 (3.4%)	-1.91 (-1.0%)	0.000
Control Group			
Quarter	Persist in next quarter	Dropout next quarter	ttest p-value
1	3.42 (1.8%)	3.54 (1.5%)	0.929
2	3.46 (1.9%)	1.36 (0.6%)	0.325
3	1.82 (0.9%)	7.99 (4.0%)	0.041

Weight loss in pounds (Percent weight loss in parenthesis) graph is the distribution of weight loss in a baseline-carried-forward analysis that assumes that every dropout returned to their baseline weight. For any given page, a comparison of the top, middle, and bottom graphs confirms that how attrition is handled has a substantial impact on estimated weight loss. In the top graphs (the completers analysis), the distribution of outcomes seems more favorable (although the modal outcome is usually zero weight loss), but in the middle and bottom graphs that include information on dropouts, by far the most common outcome is that respondents lost zero weight (largely driven by the assumption of setting dropouts at baseline weight).

Each of the graphs in Figure 3-4 indicate that more people in the modified incentives group are just over the thresholds of 5% weight loss (at which participants are refunded their year's worth of fees, or $109.46) and 10% weight loss (at which they also receive a $100 bonus), then just under the thresholds. This is less apparent in quarters 1-3 (Figures 3-1, 3-2, and 3-3), when there were no financial rewards tied to those thresholds for the modified incentives group. Moreover, such heaping is not apparent in the distribution associated with the standard incentive schedule, which has more continuous reward thresholds. This suggests that people may be pushing to achieve the substantial payoffs associated with losing 5% or 10% of baseline weight.

We next discuss the evidence regarding our hypotheses regarding weight loss.

Weight loss will be greater for those offered financial rewards for weight loss. We test for differences in unconditional means of weight loss in pounds and percent of baseline weight by quarter and group. We then test for differences in unconditional probability of losing 5% and 10% of baseline body weight. Note that the differences between the treatment groups and the control group can be interpreted as the effect of the financial incentives, distinct from all of the other program elements (e.g. daily emails and call center support) shared by the control group.

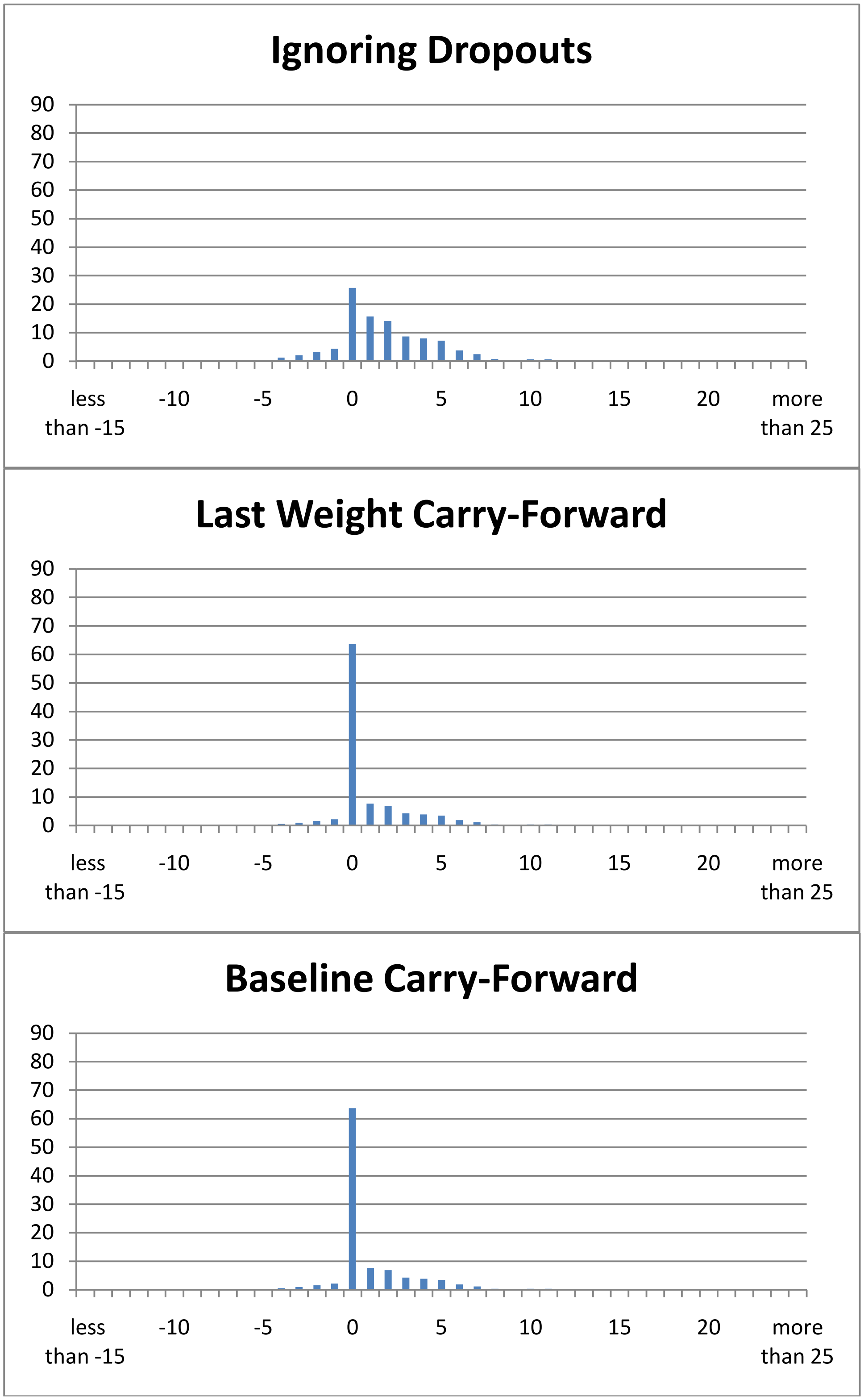


Figure 2.2-1 Distribution of Percent Weight Loss under Standard incentives Quarter 1

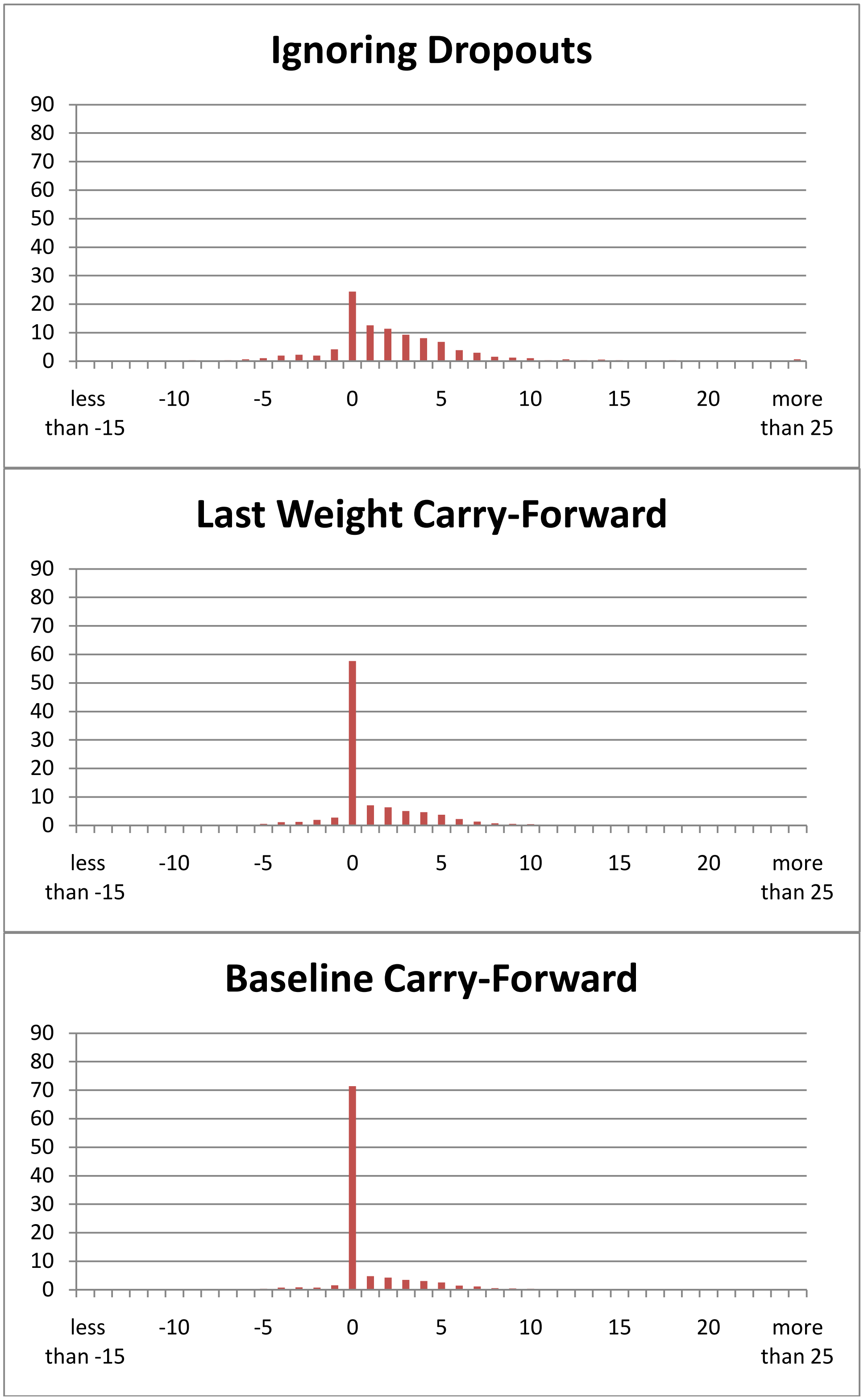


Figure 2.2-2 Distribution of Percent Weight Loss under Standard incentives Quarter 2

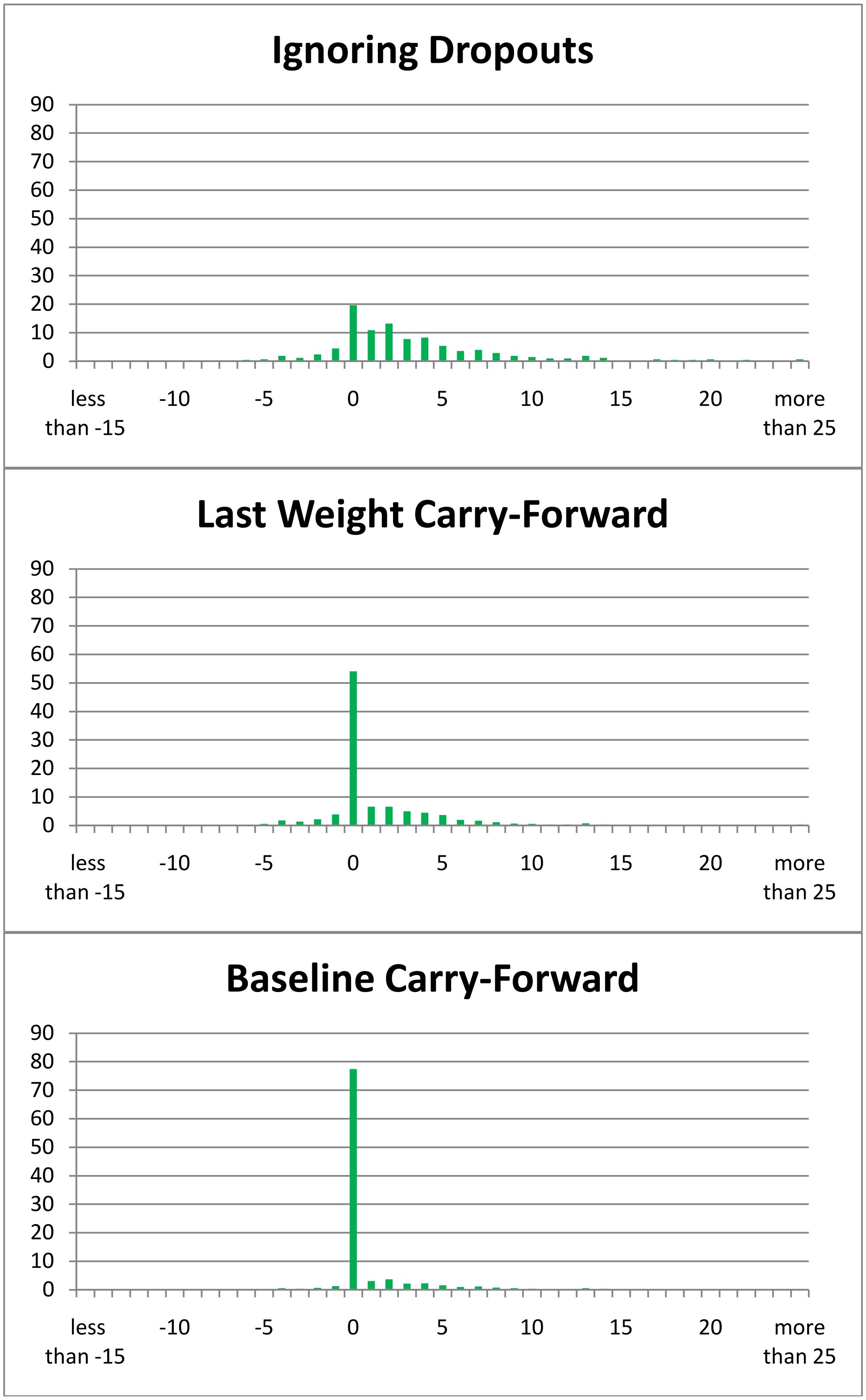


Figure 2.2-3 Distribution of Percent Weight Loss under Standard incentives Quarter 3

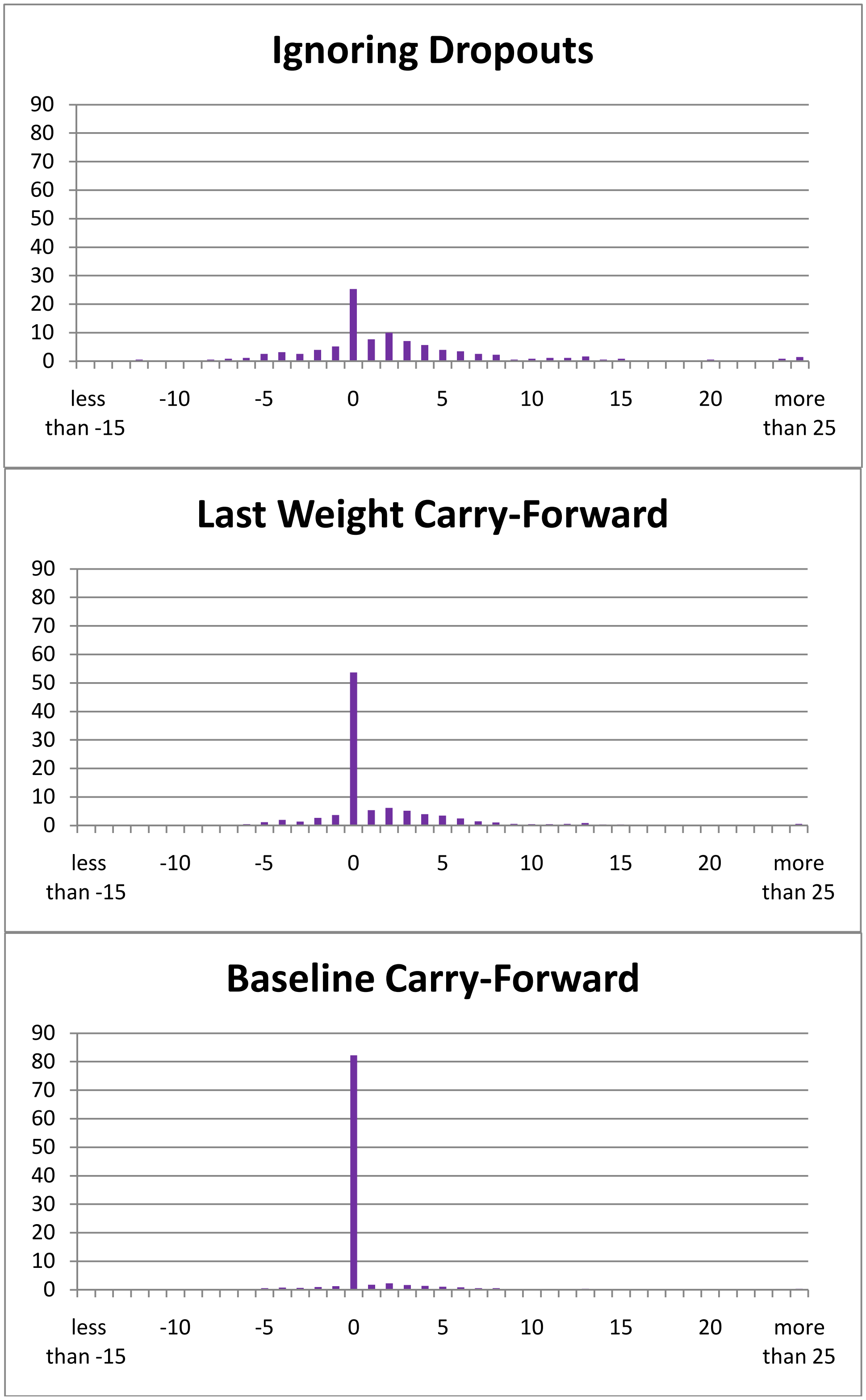


Figure 2.2-4 Distribution of Percent Weight Loss under Standard incentives Quarter 4

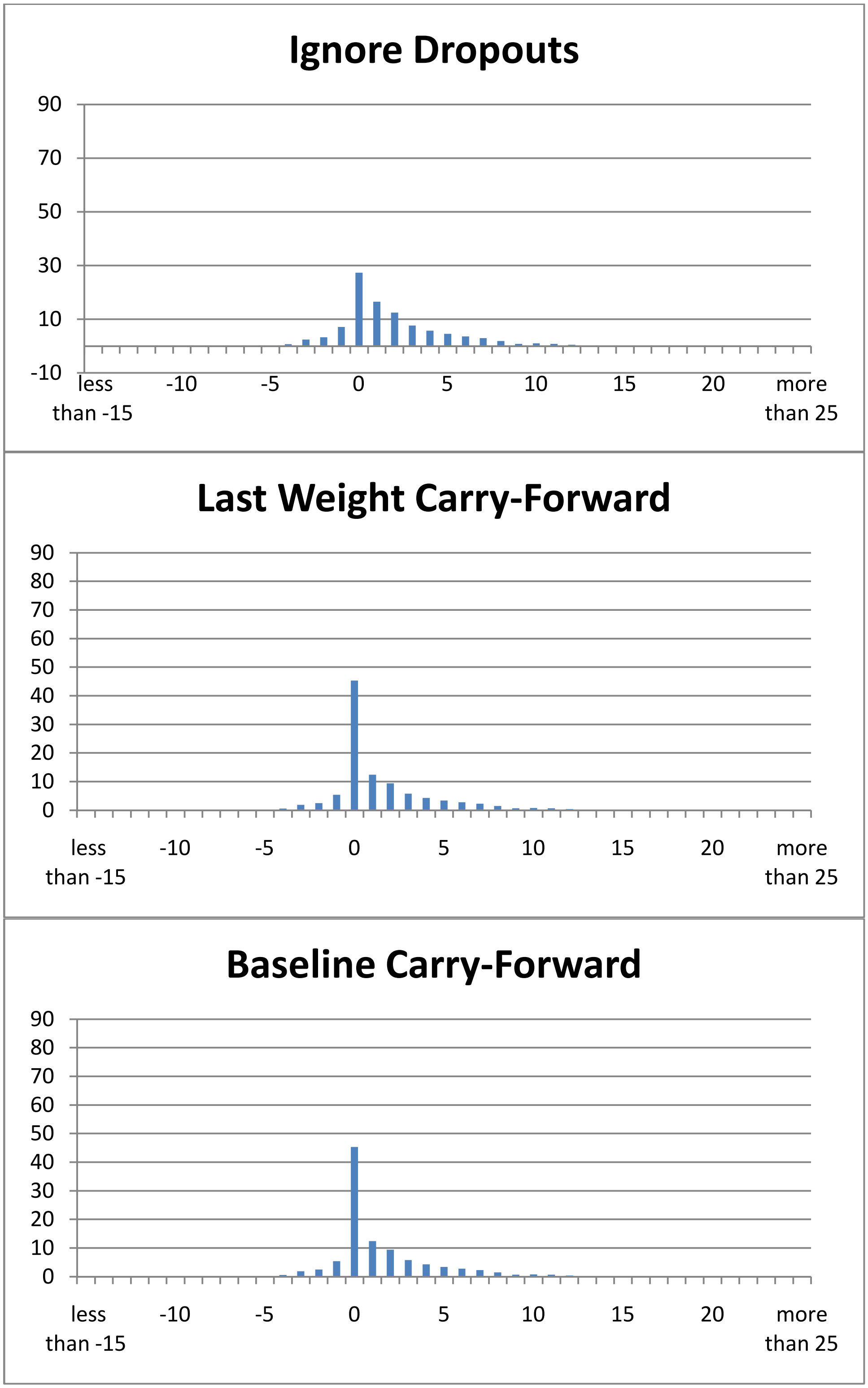


Figure 2.3-1 Distribution of Percent Weight Loss under Modified incentives Quarter 1

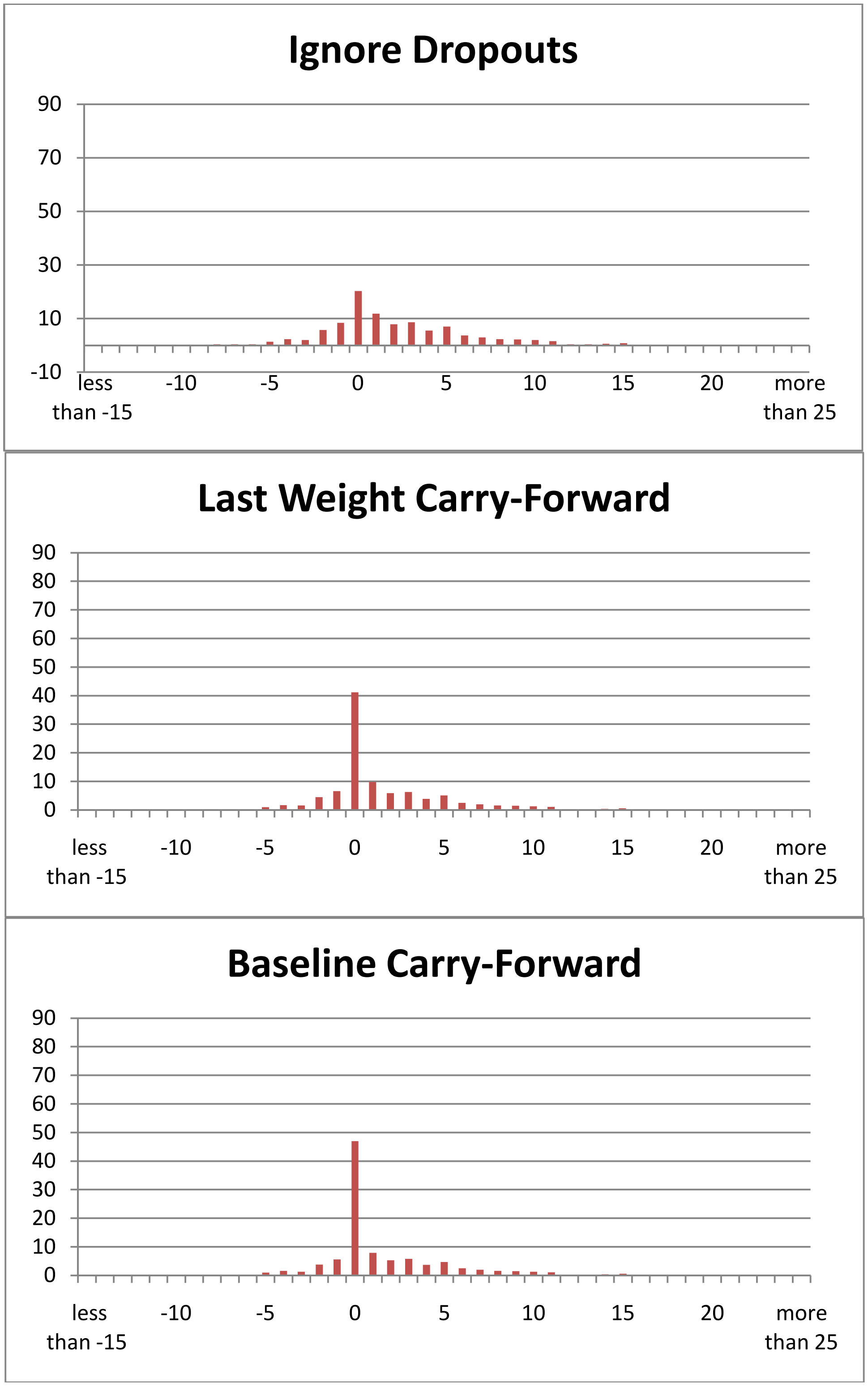


Figure 2.3-2 Distribution of Percent Weight Loss under Modified incentives Quarter 2

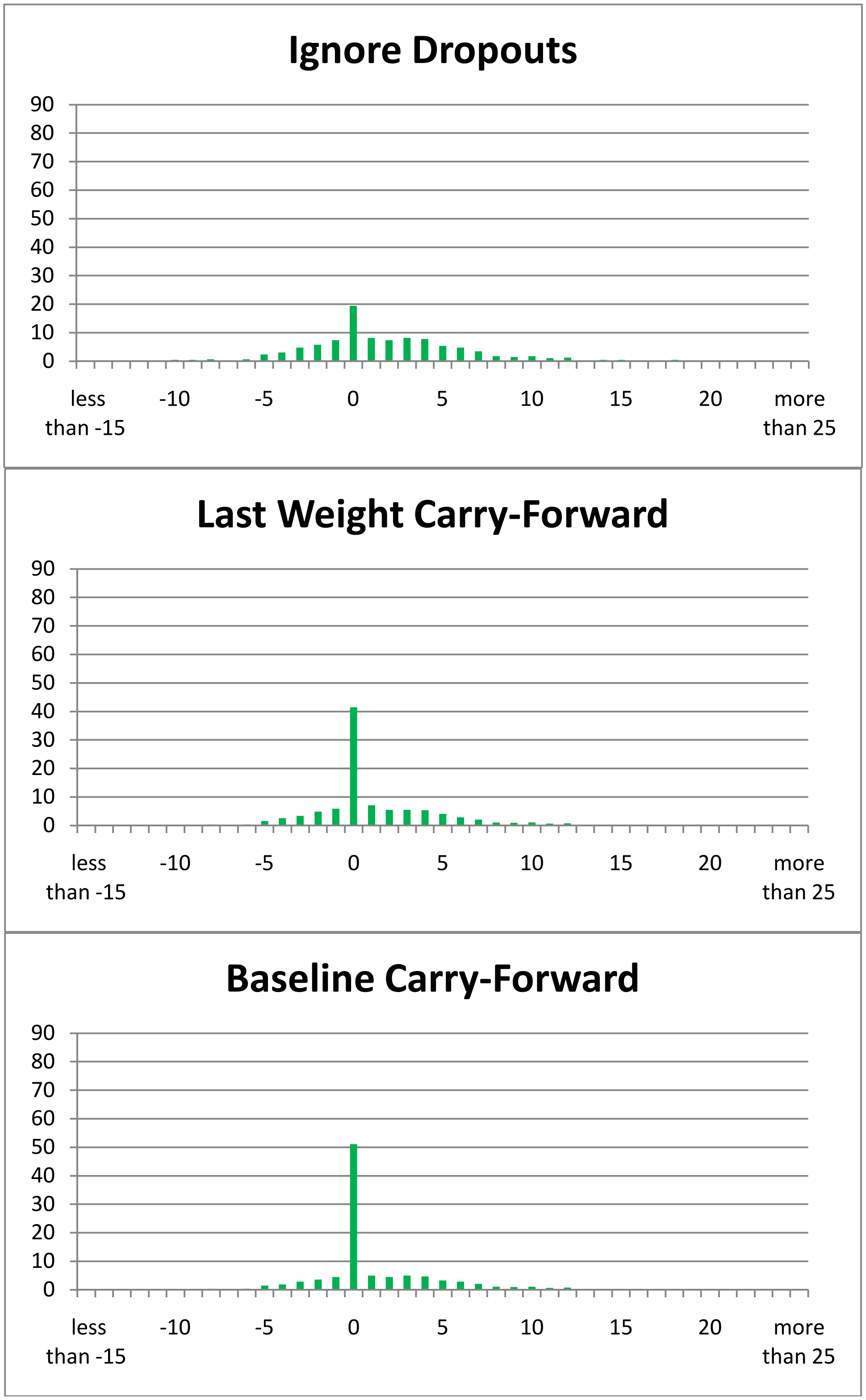


Figure 2.3-3 Distribution of Percent Weight Loss under Modified incentives Quarter 3

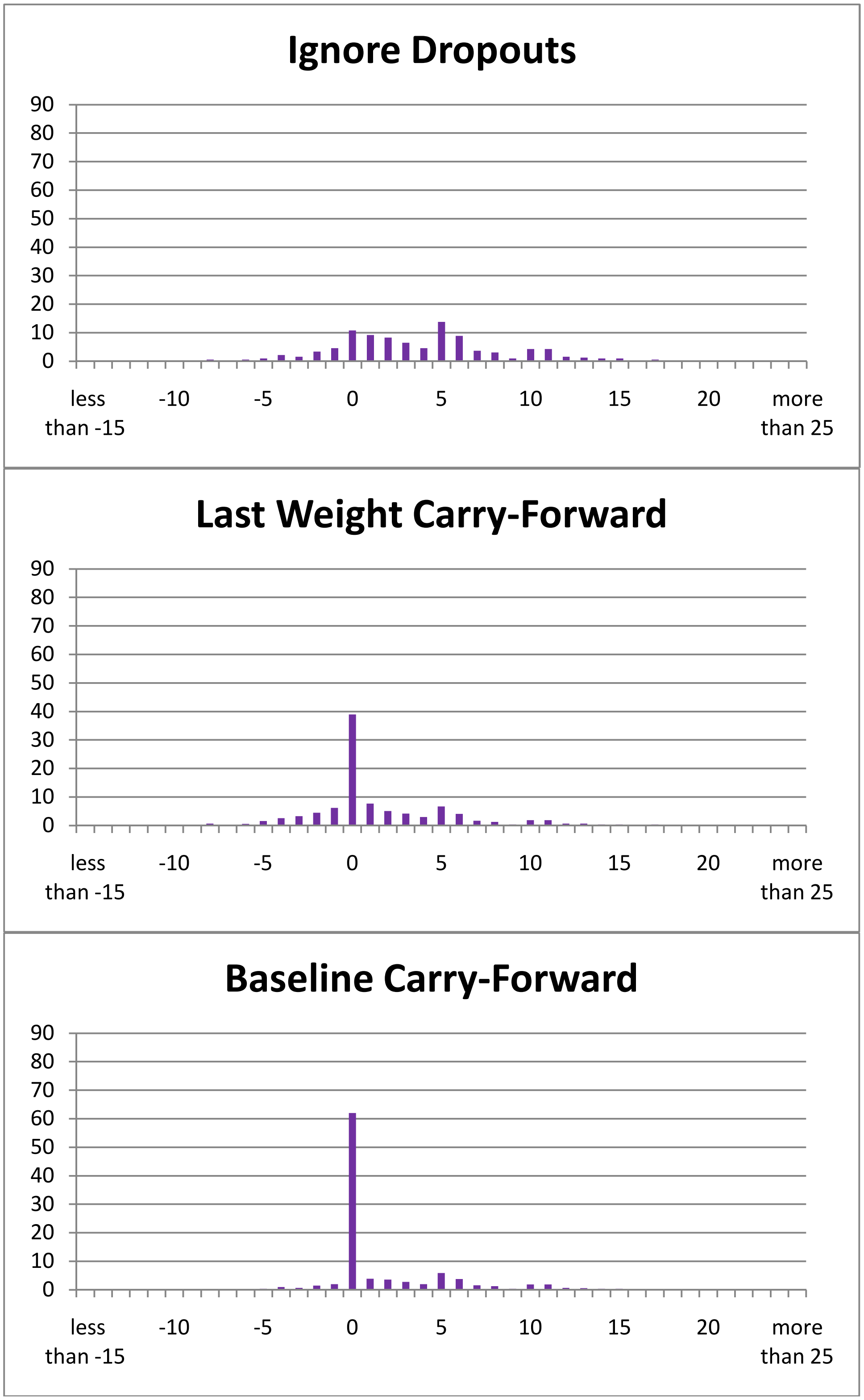


Figure 2.3-4 Distribution of Percent Weight Loss under Modified incentives Quarter 4

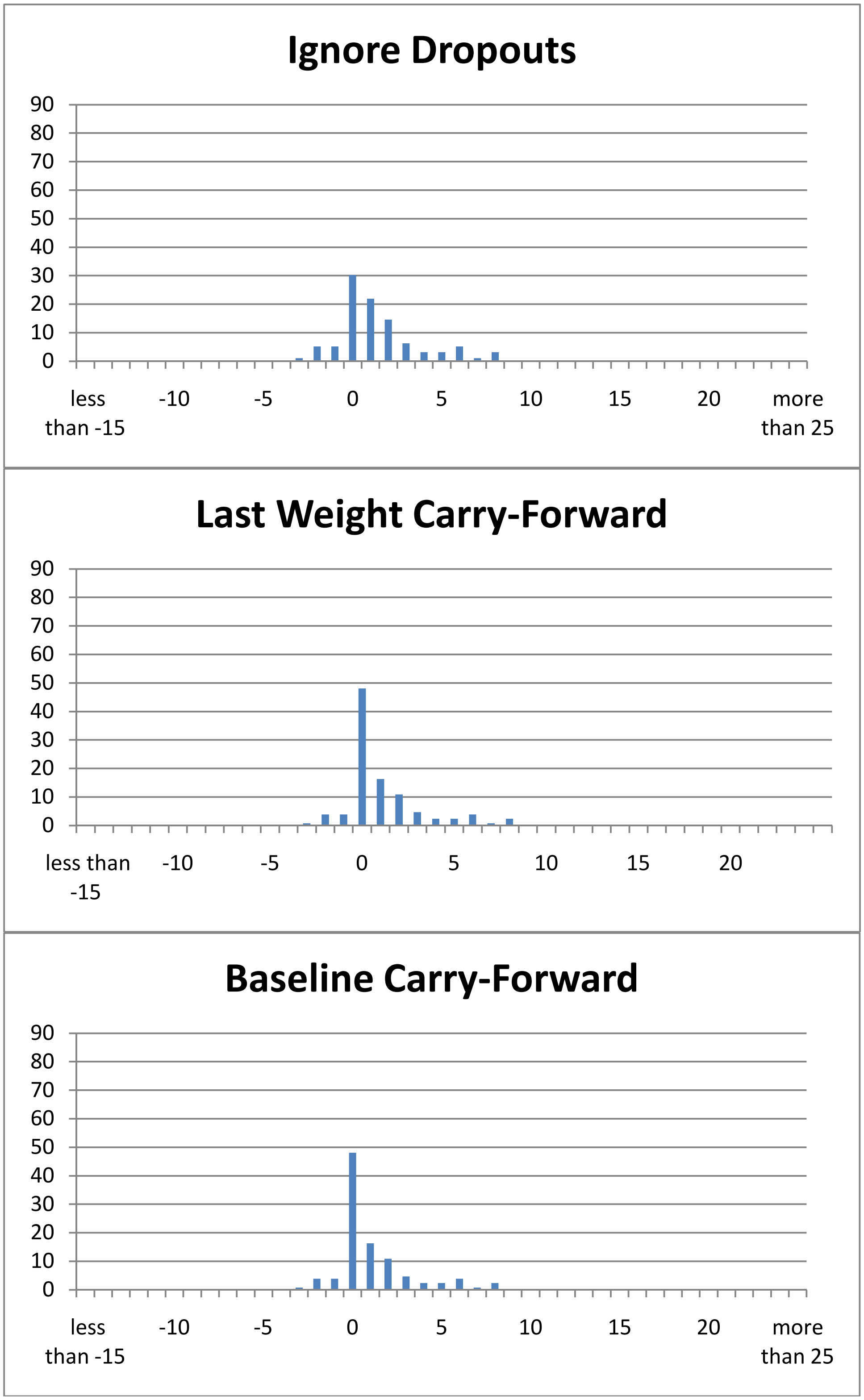


Figure 2.4-1 Distribution of Percent Weight Loss in the Control Group Quarter 1

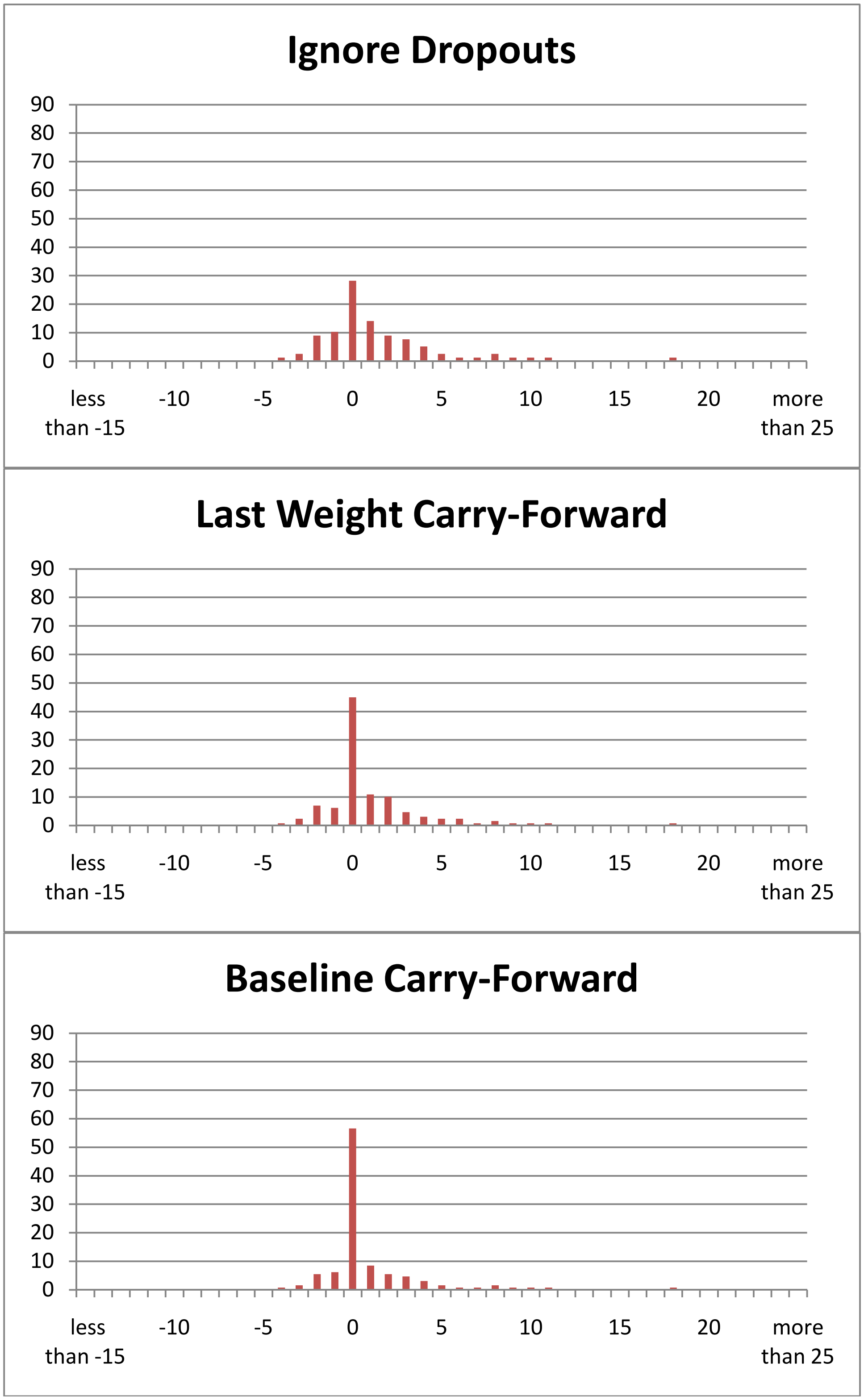


Figure 2.4-2 Distribution of Percent Weight Loss in the Control Group Quarter 2

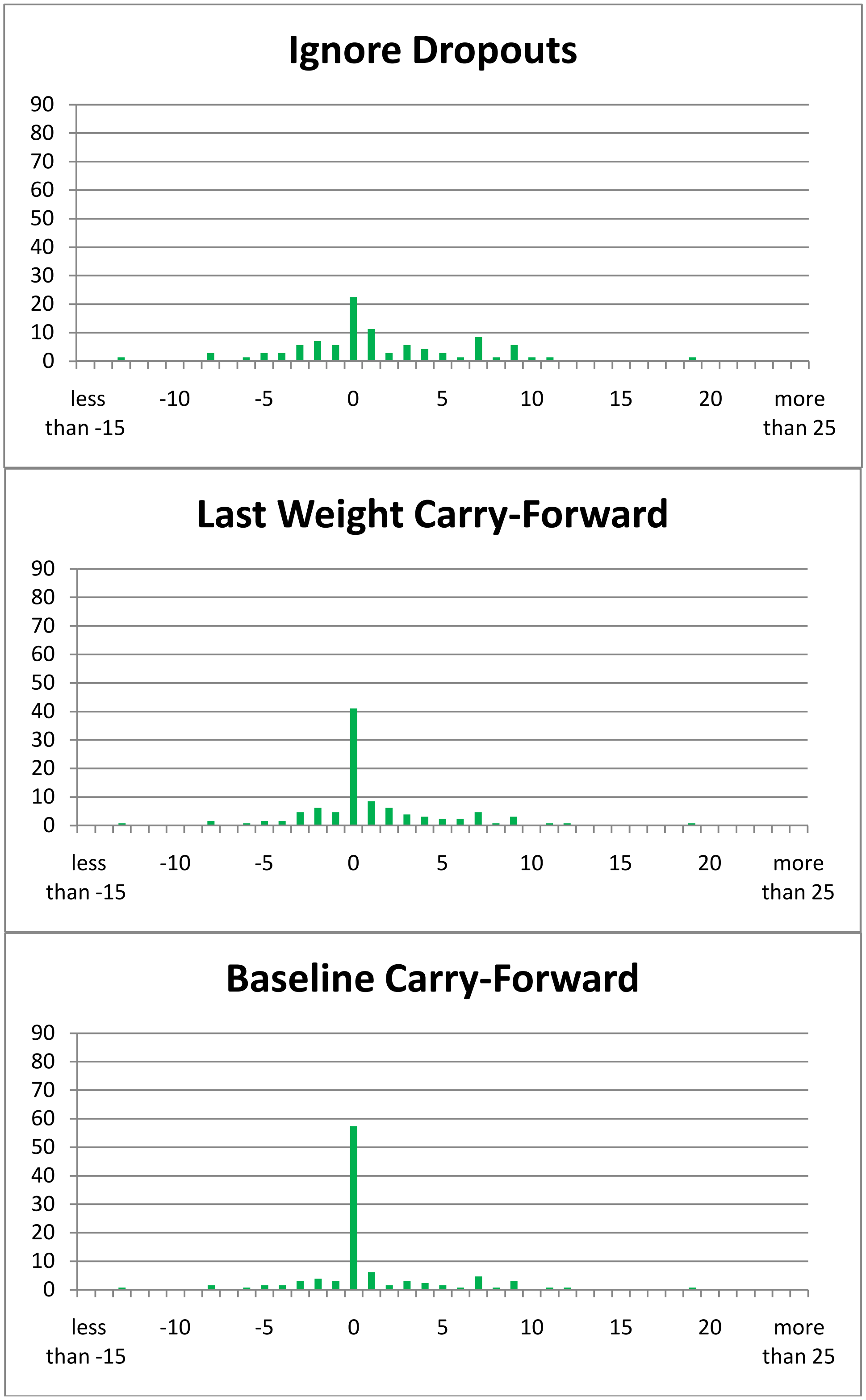


Figure 2.4-3 Distribution of Percent Weight Loss in the Control Group Quarter 3

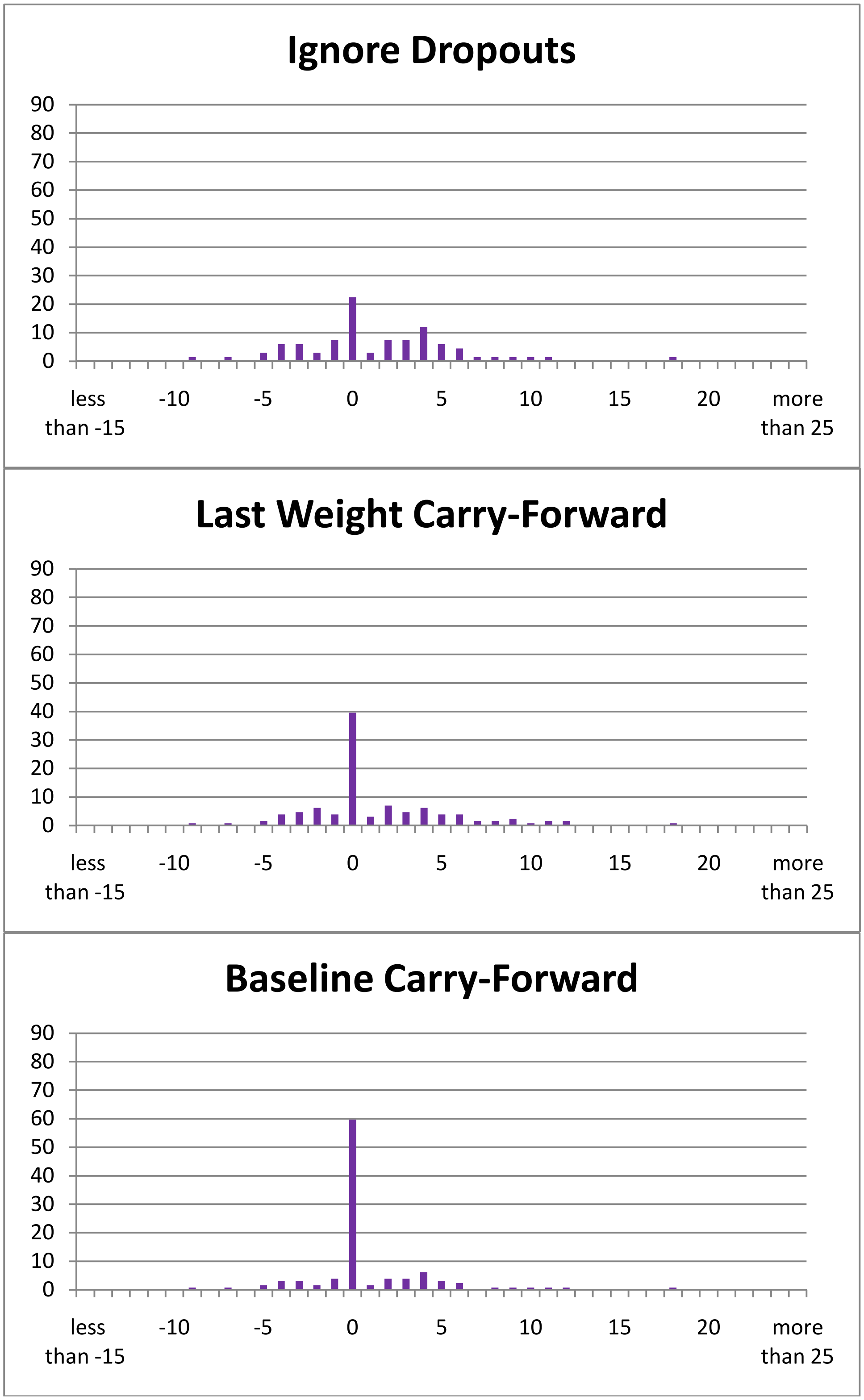


Figure 2.4-4 Distribution of Percent Weight Loss in the Control Group Quarter 4

Table 2.8 lists weight loss in pounds and percent of baseline weight, by group and quarter. The cells also list the minimum and maximum weight loss (a negative minimum weight loss indicates weight gain) for that group in that quarter (the minimum and maximum are not affected by how dropouts are treated, so they are entered in only the leftmost column for each group). Because so many participants drop out, and attrition is correlated with weight loss success, estimates of average weight loss are extremely sensitive to how attrition is handled. We focus here on the baseline carried forward analysis, which assumes that everyone who dropped out went back to their baseline weight.

In the baseline carried forward analysis, average weight loss in the control group is 2.6 pounds (1.29%) in the first quarter, 1.9 pounds (0.98%) in the second quarter, 1.7 pounds (0.82%) in the third quarter, and 1.7 pounds (0.87%) in the fourth quarter. These can be interpreted as the unconditional average effect of the program elements other than financial rewards (e.g. emails, call center access, and weigh-ins), because in a previous randomized experiment, a control group that received no treatment of any kind experienced virtually no change in average weight after 6 or 12 months (Jeffery, Wing, et al., 1993). This suggests that changes in weight observed in the control group measure the effect of all elements of Company X treatment except financial rewards.

In the standard incentives group, average weight loss is 2.2 pounds (1.13%) in the first quarter, 2.1 pounds (1.04%) in the second quarter, 2.2 pounds (1.03%) in the third quarter, and 1.4 pounds (0.64%) in the fourth quarter. We fail to reject the null hypothesis of no difference between the control and standard incentives group; in fact, average weight loss is consistently lower in the standard incentives group than in the control group. Despite the small average weight loss in the standard incentives group, there are some substantial success stories; the maximum weight lost since baseline is 58.8 lbs. in quarter 1, 89.4 lbs. in quarter 2, 109.2 lbs. in quarter 3, and 116.8 lbs. in quarter 4. For any given mean, success stories are balanced by failures; for example, the maximum weight *gain* is 12.6 lbs. in quarter 1, 19.8 lbs. in quarter 2, 25 lbs. in quarter 3, and 25.6 lbs. in quarter 4.

Weight loss in the modified incentives group averages 3.2 pounds (1.55%) in the first quarter, 3.3 pounds (1.58%) in the second quarter, 2.5 pounds (1.21%) in the third quarter, and 3.6 pounds (1.77%) in the fourth quarter. In quarter four (but not earlier quarters) the difference between the modified incentives group and the control group in average weight loss is statistically significant.

We also measure weight loss by success in reaching certain benchmarks. Table 2.9 lists the percent of participants losing 5% of baseline weight, by group and quarter, for completers, last observation carried forward, and baseline carried forward analyses. In the baseline carried forward analysis, the percentage of the control group that lost 5% of their baseline weight, by quarter, was: 9.3%, 7.8%, 13.2%, and 10.1%.

Relative to the control group, it is generally the case that smaller percentages of the standard incentives group achieved 5% weight loss in each quarter (8.3%, 8.2%, 7.9%, and 5.4%); the difference is statistically significant in quarters 3 and 4.

Relative to the control group, higher percentages of the modified incentives group achieved 5% weight loss in each quarter (12.6%, 16.5%, 14.0%, 19.5%); the differences are statistically significant in quarters 2 and 4.

We also examine the probabilities of losing 10% of baseline weight, the outcome that the NIH (1990) recommends for evaluating weight loss programs. Table 2.10 lists the unconditional probabilities of losing 10% of baseline weight by group and quarter, for completers, last observation carried forward, and baseline carried forward analyses. Assuming that dropouts returned to their baseline weight, the percentage of the control group that lost 10% of baseline weight, by quarter, was 0.0%, 2.3%, 2.3%, and 3.1%. These are comparable to the corresponding percentages for the standard incentives group (1.2%, 2.0%, 2.9%, and 2.4%); the differences are not statistically significant. Relative to the control group, higher percentages of the modified incentives group achieved 10% weight loss in each quarter (2.1%, 4.3%, 3.8%, and 6.5%) but the differences are not statistically significant.

We hypothesized that: **In quarter 4 weight loss will be greater in the modified incentives group than in the standard incentives group.** This is true for the unconditional means in Table 2.8. Assuming dropouts return to their baseline weight (baseline carried forward), average year-end weight loss is 3.6 pounds (1.77%) in the modified incentives group compared to 1.4 pounds (0.64%) in the standard incentives group, a difference significant at the 1% level. Table 2.9 indicates that at the end of quarter 4, 19.5% of the modified incentives group had lost 5% or more of their baseline weight, compared to only 5.4% of the standard incentives group, a difference significant at the 1% level. Table 2.10 shows that the percent losing 10% or more of baseline weight was 6.5% in the modified incentives group and only 2.4% in the standard incentives group, a difference significant at the 1% level.

In quarters 1 through 3, weight loss will be greater in the standard incentives group than in the modified incentives group.

Contrary to our prediction, weight loss is greater in the modified incentives group than in the standard incentives group in quarters one through three. Table 2.8 shows that those in the modified incentives group lost an average of 3.2, 3.3, and 2.5 pounds in the first three quarters, compared to the standard incentives group average losses of 2.2, 2.1, and 2.2 pounds. This difference is statistically significant at the 1% level in quarters one and two. Table 2.9 shows that in each case a higher

Table 2.8 Weight Loss in Pounds and Percent of Baseline Weight, by Group and Quarter

Quarter	Standard Incentives			Modified Incentives			Control Group		
	Ignoring Dropouts	**Last Weight Carry-Forward**	**Baseline Carry-Forward**	**Ignoring Dropouts**	**Last Weight Carry-Forward**	**Baseline Carry-Forward**	**Ignoring Dropouts**	**Last Weight Carry-Forward**	**Baseline Carry-Forward**
1	4.6 (2.31%) Min = -12.6 Max = 58.8	2.2[t] (1.13%)	2.2[t] (1.13)	4.2 (2.06%) Min = -11.6 Max = 34.6	3.2 (1.55%)	3.2 (1.55%)	3.4 (1.73%) Min = -6.0 Max = 19.6	2.6 (1.29%)	2.6 (1.29%)
2	5.5*[t] (2.73%) Min = -19.8 Max = 89.4	2.7 (1.34%)	2.1[t] (1.04%)	4.9 (2.38%) Min = -18.4 Max = 52.8	3.3 (1.64%)	3.3 (1.58%)	3.1 (1.62%) Min = -8.8 Max = 30.2	2.4 (1.21%)	1.9 (0.98%)
3	7.77*[t] (3.68%) Min = -25.4 Max = 109.2	3.2 (1.54%)	2.2 (1.03%)	4.1 (2.00%) Min = -29.8 Max = 53.6	2.6 (1.27%)	2.5 (1.21%)	3.0 (1.49%) Min = -22.2 Max = 32	2.2 (1.06%)	1.7 (0.82%)
4	6.1[t] (2.75%) Min = -25.6 Max = 116.8	3.2 (1.52%)	1.4[t] (0.64%)	8.4* (4.15%) Min = -30.6 Max = 61.2	3.3 (1.61%)	3.6* (1.77%)	3.2 (1.68%) Min = -13.8 Max = 30.8	2.9 (1.47%)	1.7 (0.87%)

Note: A positive number indicates weight lost. A negative number (e.g. for the minimum weight loss) indicates weight gain.
* represents significant difference with the control group at the 5% level
[t] represents significant difference between standard and modified incentive groups at the 5% level

Table 2.9 Percent of Respondents Losing 5% of Baseline Weight, by Group and Quarter

Quarter	Standard Incentives			Modified Incentives			Control Group		
	Ignoring Dropouts	Last Weight Carry-Forward	Baseline Carry-Forward	Ignoring Dropouts	Last Weight Carry-Forward	Baseline Carry-Forward	Ignoring Dropouts	Last Weight Carry-Forward	Baseline Carry-Forward
1	17.1%	8.3%[t]	8.3%[t]	16.7%	12.6%	12.6%	12.5%	9.3%	9.3%
2	21.6%	10.8%[t]	8.2%[t]	24.8%*	16.9%	16.5%*	12.8%	10.1%	7.8%
3	28.1%	13.0%	7.9%*[t]	23.1%	15.0%	14.0%	23.9%	15.5%	13.2%
4	22.7%[t]	13.3%[t]	5.4%*[t]	45.7%*	20.9%	19.5%*	19.4%	17.8%	10.1%

* represents significant difference with the control group at the 5% level

[t] represents significant difference between standard and modified incentive groups at the 5% level

Table 2.10 Percent of Respondents Losing 10% of Baseline Weight, by Group and Quarter

Quarter	Standard Incentives			Modified Incentives			Control Group		
	Ignoring Dropouts	Last Weight Carry-Forward	Baseline Carry-Forward	Ignoring Dropouts	Last Weight Carry-Forward	Baseline Carry-Forward	Ignoring Dropouts	Last Weight Carry-Forward	Baseline Carry-Forward
1	2.4%	1.2%	1.2%	2.8%	2.1%	2.1%	0.0%	0.0%	0.0%
2	5.2%	2.2%[t]	2.0%[t]	6.5%	4.3%	4.3%	3.8%	2.3%	2.3%
3	10.4%[t]	3.8%	2.9%	6.3%	4.1%	3.8%	4.2%	2.3%	2.3%
4	10.1%[t]	4.4%[t]	2.4%[t]	15.3%*	6.8%	6.5%	6.0%	4.7%	3.1%

* represents significant difference with the control group at the 5% level

[t] represents significant difference between standard and modified incentive groups at the 5% level

proportion of the modified incentives group than the standard incentives group achieved 5% weight loss: 12.6% versus 8.3% in quarter one, 16.5% versus 8.2% in quarter two, and 14.0% versus 7.9% in quarter three; in each case these differences are statistically significant at the 1% level. Table 2.10 shows that the probability of losing 10% or more of baseline weight is consistently higher in the modified incentives group than the standard incentives group, and the difference is statistically significant in quarter 2.

These results suggest that the effect of greater financial incentives for the standard incentives group is swamped by some combination of more favorable selection into the modified incentives group, loss aversion, and the necessity of starting early to achieve 5% or 10% weight loss by the end of quarter four.

Discussion

A 2007 Institute of Medicine report on preventing obesity set the immediate next step – which it described as an essential priority action for the near future – as "learning what works and what does not work and broadly sharing that information." (IOM 2007, p. 410). It also notes that "All types of evaluation can make an important contribution to the evidence base upon which to design policies, programs, and interventions." (Ibid, p. 4). This paper makes a contribution to that effort by documenting attrition and weight loss in a large program that offers financial incentives for weight loss.

The program studied is of particular interest because it is a real-world intervention, not a pilot program designed and monitored by researchers. As a result, the data are informative about how such interventions work in the real-world. However, because it is a real-world intervention, it suffers the limitations of selection by employers of incentive schedule, and a relatively small control group (129 out of a total sample of 2,407).

We study the two outcomes recommended by the NIH for evaluating weight loss interventions: attrition and weight loss. We find higher attrition (up to 76.4% after one year) than virtually all previous studies (see Appendix Table 2.1 and Paul-Ebhohimhen and Avenell, 2007). Another recent study of real-world wagers on own weight loss also found 80% failure (Burger and Lynham, 2008).

We find that the financial rewards in this program are associated with modest changes in weight. After one year, those in the modified incentives group lose 1.9 pounds more than those in the control group, while the weight loss of those in the standard incentives group is not statistically distinguishable from that of the control group. The NIH considers a loss of 10% of baseline weight in 6 months to one year to be good progress for an obese individual (NHLBI 2000). By this standard, very few participants in this program achieve good progress toward weight loss: just 2.4% of the standard incentives group and 6.5% of the modified incentives group lost 10% of their starting weight in 12 months. By most measures, participants in the modified incentives group had 12-month weight loss that was greater than those in the standard incentives group, but it is not clear how much of this is due to selection and how much is due to bonds, controlling for selection.

The weight loss associated with the program we examine is generally smaller than that documented in the previous literature. For example, Volpp et al. (2008) estimate mean 16-week weight loss to be 13.1 lbs. when rewards take the form of a lottery with a daily expected value of $3, and 14.0 lbs. when the rewards take the form of deposit contracts or bonds, whose amount is chosen by the enrollee but can vary between $0 and $3 per day and is matched 1:1 if the weight loss goal is achieved.

Our findings are closer to those of Finkelstein et al. (2007), who find modest weight loss (between 2.0 and 4.7 lbs.) at three months, but no significant weight loss at six months, associated with financial rewards that varied between $7 and $14 per percentage point of weight lost after six months. Likewise, Butsch et al. (2007) find no significant difference in 12-week weight loss between a treatment group offered a $150 refund of their enrollment fee if they lost 6% of their initial weight, and a control group which was not eligible for such a refund.

Overall, our findings regarding attrition and weight loss suggest that the experience of pilot programs (such as those described in Appendix Table 2.1) may be overoptimistic about what can be achieved on a larger scale.

To put our findings in a the broader context of what works to promote weight loss, a literature review (Douketis et al. 2005) found that dietary and lifestyle therapy tends to result in less than 5 kg weight loss after 2-4 years, pharmacologic therapy results in 5-10 kg weight loss after 1-2 years, and surgical therapy results in 25-75 kg weight loss after 2-4 years. At this point, financial rewards remain a promising method for weight loss but it remains to be seen whether they can be as effective as traditional medical approaches.

This paper presents the basic patterns in the data. Our follow-up work will estimate hazard models of attrition and estimate regression models of weight loss to measure the change in weight associated with the incentive schedules, controlling for the observable characteristics of participants. Future research in this area should also focus on the optimal design of financial incentives for maximizing loss of excess weight, finding ways to decrease attrition, whether offering extrinsic rewards decreases intrinsic motivation, and whether weight loss is maintained after financial incentives for weight loss are removed.

APPENDIX

Appendix Table 2.1: Previous Literature on Financial Incentives for Weight Loss

Study	Study Design	Intervention and Incentives	Sample Size and Population	Duration	Weight Loss	Attrition
Volpp et al. (2008)	Randomized controlled trial	3 groups: Deposits contract of $0-$3 / day matched 1:1. Lottery for daily prize with E[V]=$3. Self-reported daily weight. $20 for monthly weigh- in, unconditionally	N=57 (19 in each of 3 groups) Patients at Philadelphia Veterans Affairs Medical Center with BMI 30-40	16 weeks	Mean weight loss: Lottery: 13.1 lbs Deposit contracts: 14.0 Control: 3.9 lbs.	8.8%
Burger and Lynham (2008) working paper	Opportunistic data from William Hill betting agency for 1993-2006	Maximum bet of $65. William Hill offered odds ranging from 5:1 to 50:1; potential payoff averaged $1,926. Average duration of bet is 8 months, weight to be lost ranges from 28-168 lbs. Each bettor weighed at start and end of bet by physician. No control group.	N=51 Self-selected members of British population.	Average of 8 months	Approximately 80% of people betting on their weight loss lose the bet	Approximately 80% of people betting on their weight loss lose the bet
Finkelstein et al. (2007)	Randomized trial, no control group	Three groups: Back loaded: $0 at 3 months, $14 per % point lost at 6 months Front loaded: $14 per % point lost at 3	N=207 (72 in Back Loaded, 64 in Front Loaded, 71 in Steady Payment)	6 months	Mean weight loss 3 months: 2 lbs for Back Loaded, 4.7 lbs for Front Loaded, 3 lbs for	54% in Back Loaded, 45% in Front Loaded, 31% in Steady

		months, $0 at 6 months Steady payment: $7 per % point lost at both 3 months and 6 months Weigh-ins at 3 months and 6 months. Incentives only up to 10% weight loss ($140)	Overweight and obese employees at one university and 3 community colleges in NC		Steady Payment Mean weight loss at 6 months not significantly different from zero	Payment
Butsch et al. (2007)	Sequential control-intervention, not randomized	Treatment group eligible for 50% reimbursement of enrollment fee ($150 of $300) if lose 6% of initial weight and attend 10 of 12 group sessions Control group was not eligible for reimbursement	N=401 (241 intervention of which 59 enrolled, 160 control of which 40 enrolled) Participants in Univ. Alabama at Birmingham EatRight Lifestyle Program BMI 30 and over.	12 weeks	Mean weight loss: 2.25% in control group, 3.27% in intervention group; difference not statistically significant	Not stated
Hubbert et al. (2003)	Propensity score matching of 4 controls to each member of intervention group	Treatment group eligible for 50% of cost of program fees ($150 of $300) if lose 6% of initial weight and attend 10 of 12 group sessions. Control group was not eligible for reimbursement	N=125: 25 in intervention group, 100 in control group Participants in Univ. Alabama at Birmingham EatRight Lifestyle Program and members of UAB-owned HMO BMI 30 and over.	12 weeks	Mean weight loss: 7.3 kg (6.1%) in intervention group, 4.0 kg (3.9%) in control group; both differences are statistically significant	Not stated
Jeffery, Forster, et	Block-randomized	Worksites divided evenly between treatment and control groups.	32 worksites in Minneapolis / St. Paul	2 years	No treatment effect was found for weight.	No attrition of worksites.

al. (1993)	controlled experiment (worksites randomized)	Treatment (Healthy Worker Project) consisted of health education classes and payroll deductions that served as bonds – refunded if achieve weight loss goals or donated to charity otherwise. Goals chosen by employee and ranged from minimum of 0 lb and maximum of 1% body weight loss each week. Participants chose amount of payroll deduction (minimum of $5 biweekly). 200 employees surveyed at baseline and again after 2 years (cohort). Another 200 employees surveyed after 2 years (cross-section). Weight self-reported but corrected for reporting error.	metropolitan area. Of 10,000 employees in treatment worksites, 2,041 employees participated in weight control program.		In cohort survey, average change in BMI was 0.08 units for control group, -0.02 units for treatment group; not statistically significant. In cross-sectional survey, average change in BMI was -0.05 in both the treatment and control groups.	
Jeffery, Wing, et al. (1993)	Randomized controlled experiment	Five groups: 1) control; 2) standard behavioral therapy (SBT); 3) SBT plus food provision; 4) SBT plus incentives; 5) SBT plus food provision plus incentives. Weekly incentives: $0 if gained weight, $2.50 if did not gain weight; $12.50 if weight loss was 50% of goal, $25 if weight loss reached goal. Weight-loss goals could be either 14, 18, or 23 kg during course of program. Weight measured at baseline, 6, 12,	N=202 men and women from Pittsburgh and Minneapolis-St. Paul, of which 40 to 41 were in each of the 5 groups. Had to be 14-32 kg overweight	18 months	No effect of financial incentives or the interaction of financial incentives with food provision	11% attrition at 6 months, 13% at 12 months, 15% at 18 months

		and 18 months. There were also optional weekly weigh-ins.				
Jeffery et al. (1990)	Randomized experiment	Two groups: 1) offered a weight control newsletter program for price of $5; 2) offered the same program for free but requiring a $60 deposit that would be refunded based on (proportional to) success in weight loss. Individuals chose weight loss goals of not more than 4 lb a month. Weight self-reported (questionnaire, telephone survey). For subset of respondents, validation of self-report through measurement of weight.	N=1,304 residents of Bloomington Minnesota: 1,190 in the $5 newsletter program group and 114 in newsletter plus incentive program group	6 months	Weight loss averaged about 4 lbs for $5 program and 8 lbs for incentive program.	3.8% did not return survey
Kramer et al. (1986)	Randomized controlled experiment	Three groups: 1) monthly financial contingencies for weight maintenance; 2) monthly financial contingencies for participation in training sessions to solidify behavioral changes; 3) no treatment. $120 deposit. For each of 12 sessions not attended, participant forfeited $10. Refund also withheld if weighed more than “baseline” (post-first-treatment) weight. Withheld refunds (forfeited moneys) were distributed among those who were at or below “baseline” weight at final session. Weight measured at “baseline” and at one year.	N=85 individuals who had already lost 10% or more of their body weight through a 15-week weight-loss program.	1 year	Incentives had no impact on weight maintenance / amount of weight regained. Average weight regain: 10.3 lbs in control group, 11.9 lbs. in group with incentives.	6 of 28 (21%) of the incentives group refused to attend final weigh-in. They self-reported weight, and 5 lbs was added to account for under-reporting.
Jeffery et al. (1984)	Randomized controlled	Three groups: 1) regular contract, 2) difficulty-grade contract; 3) no contract	N=113	15 weeks	Average weight loss: 26.2 lbs (12.8%) in	11 subjects (10%)

	experiment	(control) All deposited $150. Immediately refunded to control group. Regular contract group received $30 for each 5-lb. increment of weight loss Difficult-grade contract group received $5 for first 5 lbs lost, $10 for second, $20 for third, $40 for fourth, and $75 for fifth.	Roughly half recruited from population sample and the other half from newspaper advertisements		difficulty-grade contract 21.7 lbs (10.8%) in regular contract 17.7 lbs (8.5%) in control group	refused to attend final weigh-in. They self-reported weight, and 5 lbs was added to account for under-reporting.
Jeffery et al. (1983)	Randomized experiment	Six treatment groups: 3 levels of deposit ($30, $150, $300) times two types of payoff criteria: individual weight loss or mean group weight loss. All received 15-week behaviorally oriented program. Goal was 30 pounds lost. Cash refunds per week at rate of $1, $5, or $10 per pound up to 2 pounds per week. Monies not refunded for weight loss by end of program were distributed equally among those who achieved the 30-pound weight loss goal. Participants were weighed weekly.	N=89 Men in the Minneapolis area with self-reported weight at least 30 pounds above the ideal.	15 weeks	Individuals rewarded for group performance lost on average 5 lbs. more weight. This difference was maintained over 1 year follow-up. No significant effects of contract size.	None
Coates et al. (1982)	Randomized experiment	Four treatment groups: 2 incentivized behaviors (weight loss or decrease in calorie consumption) by 2 frequencies of therapeutic contact (5 times or 1 time per week)	N=36 Adolescents at least 10% above average weight-for-height.	15 weeks	The treatment group receiving rewards for weight loss and coming to the clinic 5 times per week was the only group to	None

		Deposits were equal to 15 weeks' allowance or 50% of earnings from part-time work; amounts varied from \$15-\$240 (mean=\$67.75). Source of payment: parents (51.5%), subjects (39.4%), shared (9.1%). Weight loss goal was 1 lb. per week, or caloric reduction necessary to lose 1 lb. per week. Monetary reward was delivered either once per week or once per week at treatment center. Weighed at each clinic visit. Food records checked			significantly reduce the percent overweight. Treatment effects maintained over a 6-month follow-up period. Significant correlation between initial monetary deposit and percent overweight lost. No significant difference based on whether parents or subject paid the deposit	
Jeffery et al. (1978)	Randomized controlled experiment.	Three treatment groups: deposits were returned contingent on either attendance, calorie restriction, or weight loss. Also a control group. Each of the three treatment groups deposited \$200. One group paid \$20 for losing 2 lbs. per week. Another paid \$20 for calorie restrictions calculated to cause loss of 2 lbs. per week. Third group paid \$20 for weekly attendance.	N=31 Respondents to newspaper advertisement for people who need to lose 50 lbs. or more.	10 weeks	Groups rewarded for weight loss or calorie reductions lost an average of 20 lbs, significantly more weight loss than either the group rewarded for attendance (8.6 lbs) or the control group (12.4 lbs).	4 of 7 in control group quit.
Mann (1972)	Single-subject reversal design	Subjects deposited a large number of valuables (e.g. money, jewelry, medals) with the researcher and signed a Contingency Contract allowing the researcher to switch them from	N=8 Respondents to newspaper advertisement. All	Durations of treatments varied; total study ran at least 400	Average weekly weight loss of 1.6 to 1.7 pounds during treatment, regain of 1.4 pounds per week	None

		treatment to control conditions, with the treatment being valuables being either returned or forfeited based on weight loss. One valuable was returned for each 2 lb weight loss over a 2-week period. Subjects weighed every Monday, Wednesday, and Friday.	agreed to lose 25 pounds or more and had physician approval.	days	when incentives removed.	

REFERENCES

Ainslie, G. (1975). "Specious Reward: A Behavioral Theory of Impulsiveness and Impulse Control." *Psychological Bulletin*. 82(4): 463-496.

Bastiat, M. F. (1850). *That Which is Seen, and That Which is Not Seen: The Unintended Consequences of Government Spending*. (West Valley City, Utah: Waking Lion Press), reprinted 2006.

Beunckens, C. (2005). "Direct Likelihood Analysis versus Simple Forms of Imputation for Missing Data in Randomized Clinical Trials." *Clinical Trials*. 2(5): 379-386.

Bish, C.L., Blanck, H.M., Serdula, M.K., et al. (2005). "Diet and Physical Activity Behaviors Among Americans Trying to Lose Weight: 2000 Behavioral Risk Factor Surveillance System." *Obesity Research*. 13(3): 596-607.

Burger, N. and Lynham, J. (2008). "Betting on Weight Loss … and Losing: Personal Gambles as Commitment Mechanisms." Working paper, University of California Santa Barbara.

Butsch, W.S., Ard, J.D., Allison, D.B., Patki, A., Henson, C.S., Rueger, M.M., Hubbert, K.A., Glandon, G.L. and Heimburger, D.C. (2007). "Effects of a Reimbursement Incentive on Enrollment in a Weight Control Program." *Obesity*, 15(11): 2733-2738.

Camerer. C. (2005). "Three Cheers-Psychological, Theoretical, Empirical—For Loss Aversion." *Journal of Marketing Research.* 42(5): 129-133.

Cawley, J. (2004) "The Impact of Obesity on Wages." *Journal of Human Resources.* Spring, 39(2): 451-474.

Cawley, J, Rizzo, J.A. and Haas, K. (2007) "Occupation-Specific Absenteeism Costs Associated with Obesity and Morbid Obesity." *Journal of Occupational and Environmental Medicine*. December, 49(12): 1317-1324.

Coates, T.J., Jeffery, R.W., Slinkard, L.E., Killen, J.D. and Danaher, B.G. (1982). "Frequency of Contact and Monetary Reward in Weight Loss, Lipid Change, and Blood Pressure Reduction with Adolescents." *Behavior Therapy.* 13: 175-185.

Crawford, D., Jeffery, R.W. and French, S.A. (2000). "Can Anyone Successfully Control Their Weight? Findings of a Three Year Community-Based Study of Men and Women." *International Journal of Obesity.* 24: 1107-1110.

Deci, E.L. (1971). “Effects of Externally Mediated Rewards on Intrinsic Motivation.” *Journal of Personality and Social Psychology*. 18(1): 105-115.

Douketis, J.D., Macie, C., Thabane, L and Williamson, D.F. (2005). “Systematic Review of Long0Term Weight Loss Studies in Obese Adults: Clinical Significance and Applicability to Clinical Practice.” *International Journal of Obesity.* 29: 1153-1167.

Finkelstein, E.A., Fiebelkorn, I.C., and Wang, G. (2003) “National Medical Spending Attributable to Overweight and Obesity: How Much, and Who’s Paying?” *Health Affairs.* Web Exclusive, W3-219.

Finkelstein, E.A., Linnan, L.A., Tate, D.F., Birken, B.E. (2007) “A Pilot Study Testing the Effect of Different Levels of Financial Incentives on Weight Loss Among Overweight Employees.” *Journal of Occupational and Environmental Medicine.* 49(9): 981-989.

Fishbach, A. and Trope, Y. (2005). “The Substitutability of External Control and Self-Control.” *Journal of Experimental Social Psychology*. 41(3): 256-270.

Ford, E.S., Moirarty,D.G., .G, Zack, M.M., Mokdad, A.H. and Chapman, D.P. (2001). “Self-Reported Body Mass Index and Health-Related Quality of Life: Findings from the Behavioral Risk Factor Surveillance System.” *Obesity Research*, 9(1): 21-31.

Gadbury, G.L., Coffey, C.S. and Allison, D.B. (2003). “Modern Statistical Methods for Handling Missing Repeated Measurements in Obesity Trial Data: Beyond LOCF.” *Obesity.* 4: 175-184.

Higgins, S.T., Alessi, S.M. and Dantona, R.L. (2002). “Voucher-Based Incentives: A Substance Abuse Treatment Innovation.” *Addictive Behaviors*. 27: 887-910.

Hubbert, K.A., Bussey, B.F., Allison, D.B., Beasley, T.M., Henson, C.S. and Heimburger, D.C. (2003). “Effects of Outcome-Driven Insurance Reimbursement on Short-Term Weight Control.” *International Journal of Obesity.* 27, 1423-1429.

Institute of Medicine. (2007). *Progress in Preventing Childhood Obesity: How do we Measure Up*? (National Academies Press: Washington, D.C.).

Jeffery, R.W., Bjornson-Benson, W.M., Rosenthal, B.S., Kurth, C.L., Dunn, M.M. (1984). “Effectiveness of Monetary Contracts with Two Repayment Schedules on Weight Reduction in Men and Women from Self-Referred and Population Samples.” *Behavior Therapy.* 15: 273-279.

Jeffery, et al. (1998).“Use of Personal Trainers and Financial Incentives to Increase Exercise in a Behavioral Weight-Loss Program” *Journal of Consulting and Clinical Psychology*. 66(5): 777-783.

Jeffery, R.W., Forster, J.L., French, S.A., Kelder, S.H., Lando, H.A., McGovern, P.G., Jacobs, D.R. and Baxter, J.E. (1993). “The Healthy Worker Project: A Work-Site Intervention for Weight Control and Smoking Cessation.” *Journal of Public Health.* 83(3): 395-401.

Jeffery, R.W., Gerber, W.M., Rosenthal, B.S. and Lindquist, R.A. (1983). “Monetary Contracts in Weight Control: Effectiveness of Group and Individual Contracts of Varying Size.” *Journal of Consulting and Clinical Psychology.* 51(2): 242-248.

Jeffery, R.W., Hellerstedt, W.L. and Schmid, T.L. (1990). “Correspondence Programs for Smoking Cessation and Weight Control: A Comparison of Two Strategies in the Minnesota Heart Health Program.” *Health Psychology.* 9(5): 585-598.

Jeffery, R.W., Thompson, P.D. and Wing, R.R. (1978). “Effects on Weight Reduction of Strong Monetary Contracts for Calorie Restriction or Weight Loss.” *Behavior Research and Therapy.* 16: 363-369.

Jeffery, R.W., Wing, R.R., Thorson, C., Burton, L.R., Raether, C., Harvey, J. and Mullen, M. (1993). “Strengthening Behavioral Interventions for Weight Loss: A Randomized Trial of Food Provision and Monetary Incentives.” *Journal of Consulting and Clinical Psychology*. 61: 1038-1045.

Kahneman, D. and Tversky, A. (1979). “Prospect Theory: An Analysis of Decision Under Risk. *Econometrica.* 47(2): 263-292.

Kane, R.L., Johnson, P.E., Town, R.J. and Butler, M. (2004). “A Structured Review of the Effect of Economic Incentives on Consumers’ Preventive Behavior.” *American Journal of Preventive Medicine.* 27(4): 327-352.

Kramer, et al. (1986). “Maintenance of Successful Weight Loss Over One Year: Effects of Financial Contracts for Weight Maintenance or Participation in Skills Training.” *Behavioral Therapy*. 17: 295-301.

Laibson, D. (1997). “Golden Eggs and Hyperbolic Discounting.” *Quarterly Journal of Economics.* 112(5): 443-477.

Lepper, M.R., Greene, D., Nisbett, R.E. (1973). “Journal of Personality and Social Psychology.” 28(1): 129-137.

Loewenstein, G., Weber, E.U., Hsee, C.K., et al. (2001). “ Risk as feelings.” *Psychological Bulletin*. 127: 267-286.

Lussier, J.P., Heil, S.H., Mongeon, J.A., et al. (2006). “A Meta-Analysis of Voucher-Based Reinforcement Therapy for Substance Use Disorders.” *Addiction*. 101: 192-203.

Mann, R.A. (1972) “The Behavior-Therapeutic Use of Contingency Contracting to Control and Adult Behavior Problem: Weight Control.” *Journal of Applied Behavior Analysis.* 5(2): 99-109.

NIH Technology Assessment Conference Panel. (1993). “Methods for Voluntary Weight Loss and Control.” *Annals of Internal Medicine*. 119(7, Part 2): 764-770.

O’Donoghue, T. and Rabin, M. (2001). “Choice and Procrastination.” *Quarterly Journal of Economics.* 116(1): 121-160.

Ogden, C.L., Carroll, M.D., Curtin, L.R., et al. (2006).“Prevalence of Overweight and Obesity in the United States, 1999-2004.” *Journal of the American Medical Association.* 295: 1549-1555.

Paul-Ebhohimhen, V. and Avenell, A. (2007). “Systematic Review of the Use of Financial Incentives in Treatments for Obesity and Overweight.” *Obesity Reviews*. 9: 355-367.

Petty, N.M. and Martin, B. (2002). “Low-Cost Contingency Management for Treating Cocaine- and Opiod-Abusing Metadone Patients.” *Journal of Consulting and Clinical Psychology*. 70(2): 398-405.

Read, D., Loewenstein, G. and Rabin, M. (1999). “Choice Bracketing.” *Journal of Risk and Uncertainty.* 19:1-3; 171-197.

Samuelson, P. A. (1937). “A note on measurement of utility.” *Review of Economic Studies*, 4: 155-161.

Serdula, M.K., Mokdad, A.H., Williamson, D.F., Galuska, D.A., Mendlein, J.M. and Heath, G.W. (1999). “Prevalence of Attempting Weight Loss and Strategies for Controlling Weight.” *Journal of the American Medical Association.* 282(14): 1353-1358.

Sindelar, J.L. (2008). “Paying for Performance: The Power of Incentives Over Habits.” *Health Economics*. 17: 449-451.

Strotz, R.H. (1955-1956). “Myopia and Inconsistency in Dynamic Utility Maximization.” *Review of Economic Studies.* 23(3): 165-180.

Stunkard, A. and McLaren-Hume, M. (1959). “The Results of Treatment for Obesity.” *Archives of Internal Medicine*. 103, 79–85.

Thayer, R.H. and Shefrin, H.M. (1981). “An Economic Theory of Self-Control.” *Journal of Political Economy.* 89(2): 392-406.

Tversky, A. and Kahneman, D. (1991). “Loss Aversion in Riskless Choice: A Reference-Dependent Model.” *Quarterly Journal of Economics.* 106: 1039-61.

U.S. Department of Health and Human Services. (2000). “The Practical Guide: Identification, Evaluation, and Treatment of Overweight and Obesity in Adults.” *National Institutes of Health #00-4084.* 80 pages.

Vidal, J. (2002). “Updated Review on the Benefits of Weight Loss.” *International Journal of Obesity*. 26 (Suppl 4): S25-S28.

Volpp, K.G., John, L.K., Troxel, A.B., et al. (2008). “Financial Incentive Based Approaches for Weight Loss: A Randomized Trial.” *Journal of the American Medical Association.* 300(22): 2631-2637.

Ware, J.H. (2003). “Interpreting Incomplete Data in Studies of Diet and Weight Loss.” *New England Journal of Medicine*. 348(21): 2136-2137.

Wing, R.R. and Phelan, S. (2005). “Long-Term Weight Loss Maintenance.” *American Journal of Clinical Nutrition.* 82 (supplement): 222S-225S.

CHAPTER 3

THE EFFECTS OF HIGHER ADMISSION STANDARDS ON NCAA STUDENT-ATHLETES: AN ANALYSIS OF PROPOSITION 16

Abstract

This study examines the effect of an increase in minimum admissions standards on college enrollment and graduation rates of student-athletes. In 1996, the National Collegiate Athletic Association (NCAA) enacted Proposition 16, which increased admission standards for freshmen student-athletes at Division I schools, in an effort to improve graduation rates. Results indicate that Proposition 16 increased graduation rates significantly for black student-athletes, and had no significant impact on graduation rates for white student-athletes. Results also indicate that graduation rates declined for black student-athletes at Division II schools, which may have been driven by students transferring to Division I. As a result of the higher admission standards, Division I schools changed recruiting patterns and relied less on freshmen student-athletes, particularly black student-athletes, to fill scholarships. Even though fewer black freshmen student-athletes enrolled in Division I schools, the overall number of black student-athletes did not change, suggesting that a greater proportion of transfer students into Division I schools were black.

Introduction

In January 1989, Georgetown was scheduled to play Boston College in a men's basketball game. Right before tip-off, future Hall of Fame coach John Thompson, in protest of a National Collegiate Athletic Association (NCAA) proposal, walked off the court and into the locker room. The proposal called for the use of standardized test scores to determine eligibility for student-athletes. John Thompson felt that relying heavily on standardized test scores for athletic eligibility and athletic financial aid would disadvantage black student-athletes. His protest raised the question of whether NCAA policies which rely on standardized test score to determine initial eligibility have differential impacts by race.

The National Collegiate Athletic Association (NCAA) asserts the pursuit of academic excellence as one if its core values and objectives. In accordance with this, the NCAA has passed several legislative policies that established eligibility requirements for first-time freshmen and continuing student-athletes[16]. In the early 1980s, the NCAA began to use standardized test scores, in addition to high school GPAs, as a requirement for freshmen eligibility. In the late 1980s, the NCAA began discussion about increasing the initial eligibility standard through requiring a higher test score, GPA, and number of core classes taken in high school. It continued until the early 1990s, when the NCAA raised its admission standard, requiring either higher test scores or a higher GPA for freshmen student-athletes. This again raised the debate of whether relying on these measures of ability disadvantaged minority students more than non-minority student-athletes.

In this study, I examine the effects of changes in the NCAA's policy from the viewpoint of universities and colleges. The NCAA policy raised the admission standards for student-athletes who attended schools that were a member institution of

[16] For a comprehensive review of NCAA eligibility policies see Covell & Barr (2001).

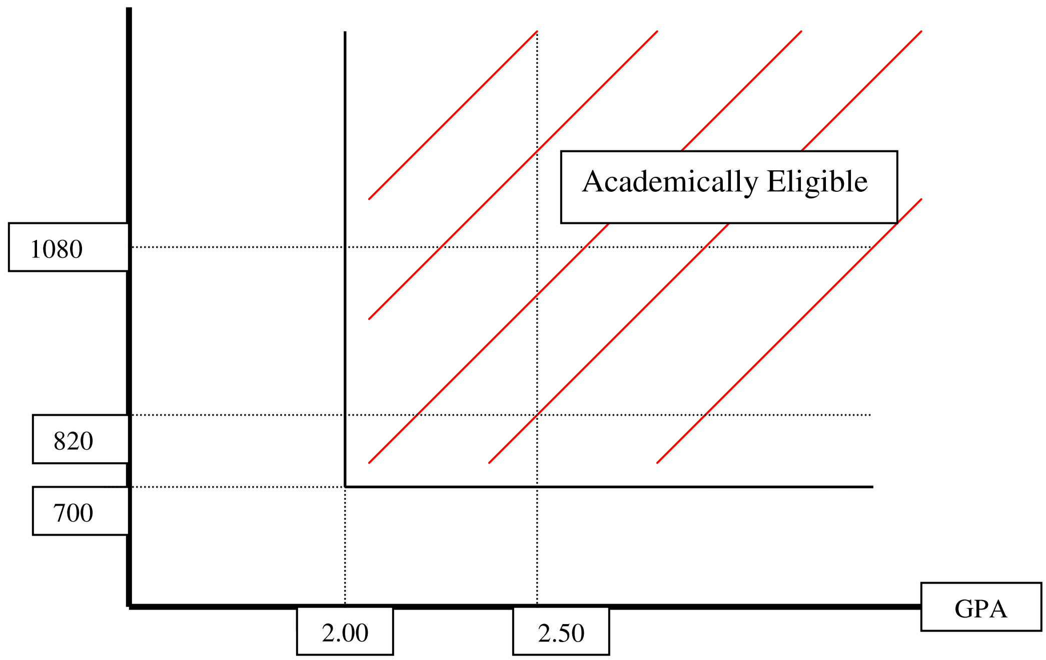
the NCAA. Student-athletes must abide by the NCAA standards to participate in inter-collegiate athletics, regardless of their school's admission standard. The question that I address is how increasing the admission standard changes the type of students that those schools admit. I also test whether Proposition 16 was effective in its goal of increasing the graduation rate of student-athletes.

Proposition 48

The first policy enacted by the NCAA that altered admission requirements for student-athletes was Proposition 48. Previously, only a 2.00 high school GPA was required for student eligibility to participate in inter-collegiate athletics. Proposition 48 included standardized test scores to determine initial eligibility in order to set a national standard to "level the playing field" of academic recruiting standards. Beginning with Division I schools in 1986, potential student-athletes needed to achieve a 700 combined score on the Standard Aptitude Test (SAT) (17 ACT) in addition to a 2.00 high school GPA (See Figure 3.1, top panel). By meeting these standards, students became eligible to practice, compete, and receive athletic financial aid. Proposition 48 was later implemented among Division II schools beginning with the 1988–1989 academic school year.

The NCAA released a series of reports analyzing the effects of Proposition 48 on six-year graduation rates (*NCAA Research Report Series 01, 1990*). The findings indicate that in the three years prior to the implementation of the policy, graduation rates averaged 52% for all student-athletes at Division I schools. For the first cohort subject to the regulations of Proposition 48, graduation rates increased to 57% and

Prior to Proposition 16



Sliding Scale of Proposition 16

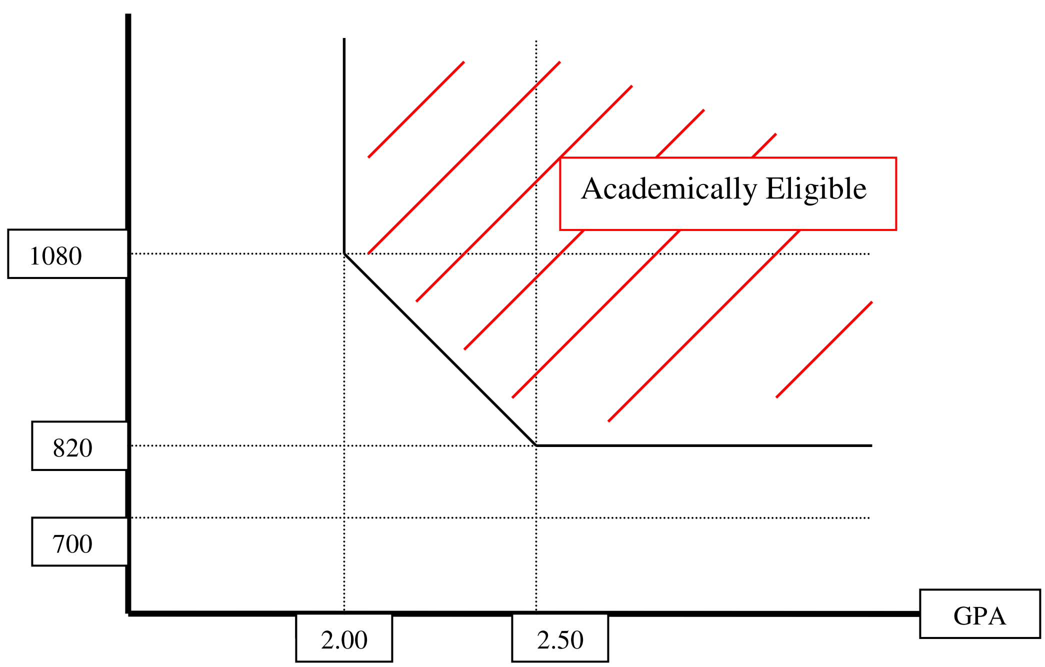


Figure 3.1 Eligibility Scale for Potential Student-Athletes

continued to rise to 59% for the entering cohort of 1988. Some of the largest increases in completion rates were found among minority students. Graduation rates rose 8 percentage points for African-American students and 11 percentage points for Hispanic students, up to 44 and 50% respectively. From these results, the NCAA concluded that increasing admission standards led to significantly higher graduation rates.

While the focus of Proposition 48 was to establish a national standard of admissions based on indicators of college success, requiring a benchmark score on the SAT might have created a disadvantage for minorities. The total number of African-American freshmen student-athletes went from 3,724 in 1985 to 3,041 in the first year after Proposition 48, an 18.3% decrease. In spite of this significant decrease, the total number of African-Americans admitted in 1986 who graduated within six years was virtually identical to the 1985 cohort (1,334 compared to 1,337 respectively). Examining recruiting behavior of Division IA football programs, Heck and Takahashi (2006) find that freshmen enrollment decreased after Proposition 48 and that there was an increase in the number of transfer students.

Overall, previous analyses of Proposition 48 indicate that establishing an SAT score cut-off in conjunction with a required high school GPA increased graduation rates of those admitted to Division I schools. African-Americans and Hispanics saw the greatest increase in six-year graduation rates. However, as a result of the policy, the proportion of minority freshmen student-athletes decreased dramatically.

Proposition 16

With continued concern toward admitting students who were not prepared to succeed in college and graduate, the NCAA enacted Proposition 16, which took effect in the fall of 1996 among Division I schools, and did not apply to Division II schools.

Similar to its predecessor, Proposition 16 increased the admission requirement of high school GPA and standardized test scores, but differed by using a sliding scale. The sliding scale required a student who earned a 2.00 GPA to achieve at least a 1080 on the SAT, or it required a student who scored an 820 on the SAT to receive a 2.50 GPA, or a linear combination of the two (see Figure 3.1, bottom panel)[17]. Those who did not meet the new sliding scale standard but still met the prior requirements were still eligible for athletic scholarships from Division II schools, but no longer from Division I.

Critics of Proposition 16 claimed that this policy would have differential effects on specific racial and ethnic groups, because it placed more weight on the SAT for admission into a Division I school. A National Center for Educational Statistics (NCES) study showed that, among high school graduates who graduated in 1992 and who applied to college and took the SAT (or ACT), only 46.4% of black students and 54.1% of Hispanic students meet the standard set by Proposition 16, compared to 67% of white and Asian students (NCES, 1995). Other studies have shown that black students score significantly lower on the SAT (Hoffman & Lowitzki, 2005; Fleming, 2002; Camara & Schmidt, 1999). In the book, *Black-White Test Score Gap* (Jencks & Phillips, 1998), Vars and Bowen find evidence that the SAT is a poorer predictor of college success for black students compared to white students. Fleming (2002) asserts that while the SAT is a better predictor for black students who attend Historically Black Colleges and Universities (HBCU), it is a poor predictor of college success for Black males who attend a non-HBCU. Sellers (1992) finds that high school grades are a good predictor of college GPA for black student-athletes. In contrast, high school grades and standardized test scores are good predictor of college

[17] Beginning in 1996 the SAT rescaled the scoring of the verbal and math section such that a 700 in 1995 was equivalent to an 820 in 1996.

grades for white student-athletes.

Due to the differential performance between minorities and non-minorities on standardized tests and lower predictive power of college performance, opponents of Proposition 16 brought a lawsuit against the NCAA, claiming that Proposition 16 violated the Civil Rights Act of 1964 (Title VI) by discriminating against minorities (Cureton v. NCAA). While the courts originally ruled Proposition 16 to violate Title VI, the ruling was later overturned in Circuit Court because, according to the ruling, the NCAA was not subject to Title VI's regulations since the NCAA does not receive federal funding.

The NCAA's own research examining the effects of Proposition 16 focused primarily on the increases in graduation rates, comparing the year before the policy to the first year after the policy. Their study found that student-athletes matriculating under the guidelines set by Proposition 16 experienced a two percentage point increase in graduation rates, to an all-time high of 62% (NCAA News 2003). They break down the effect and show the number of black students in men's basketball increased from 35 to 41%. Division I-A black football players also had an increase in graduation rates, from 46 to 49. White student athletes in men's basketball and football were one percentage point lower than the 1995 graduation rate. In this report they also note that as a result of the policy, the proportion of black student-athletes in the freshmen class fell, particularly in men's basketball and football, by 2.9 and 3.7 percentage points respectively. A managing director of research for the NCAA said, "though the research indicates fewer black student-athletes overall, those in the system graduated at a higher rate than previous cohorts" (NCAA News 2003). Yet comparisons across just two years do not account for trends that may drive the findings. Therefore, a more rigorous analysis is needed to identify the true effects of the change in the policy.

Hypotheses

There are two specific hypotheses that I test in this study. First, higher admission standards based on a standardized test score and GPA will cause Division I schools to admit fewer freshmen student-athletes, particularly minorities, and enroll more transfer student-athletes. Second, the higher admission standard will increase graduation rates for all student-athletes at Division I schools, with minority-student athletes experiencing the greatest increase in graduation rates.

When admitting freshmen student-athletes, athletic departments seek to maximize a combination of athletic and academic ability of student-athletes. The specific combination is decided by each school, since it is certainly the case that some schools place more weight on the academic ability of student-athletes. However, schools face two constraints. First, student-athletes must meet a school minimum athletic ability to be offered an athletic scholarship. Second, student-athletes must meet the higher of two academic eligibility standards. The first is the standard set by the institution that all students must meet in order to be admitted. This varies widely by institution and is not always an explicit benchmark publicized by individual schools. The second is the standard set by the NCAA that all student-athletes must meet to be eligible. Many schools admit student-athletes through a special admission process, and often these students have significantly lower test scores than the overall student body population. An investigation by the Atlanta Journal Constitution reported that at Clemson University, UCLA, Rutgers University, Texas A&M University, Louisiana State University, and the University of Georgia more than half of all student-athletes are special admits, with University of Georgia having 73.5% of all student-athletes admitted this way (Knobler, 2008). Thus, it very well may be that even at schools with higher admission standards, the NCAA academic requirement is the binding constraint due to special admissions. When admitting transfer student-athletes, schools face a different admission standard based on grades in college courses and not on high school GPA and test scores. Between freshmen and upper class athletes, schools try to fill all scholarships available in a given year.

For a Division I school to fill their allotted number of scholarships, they could have either reduced the athletic ability required to be offered a scholarship or they could have given scholarship offers to non-freshmen student-athletes. It is unlikely that athletic departments would reduce the athletic ability required for scholarship recipients, due to possible returns to the athletic program and the University for having a successful athletic program. These returns include higher publicity, increases in SAT scores sent to the school (Pope & Pope, 2006), a better student environment, and increases in alumni donations to the school (Tucker, 2004; Baade & Sundberg, 1996; Brooker & Klastorin, 1981). Rather than admit student-athletes with lower athletic ability, it is expected that athletic departments would turn to transfer students to fill the scholarships available[18]. Thus, it is hypothesized that Division I schools changed recruiting patterns by relying more on transfer student-athletes and less on first time freshmen student-athletes to fill scholarships after the implementation of the policy.

Furthermore, it is hypothesized that with higher admission standards in place, graduation rates should have increased at Division I schools. This hypothesis relies on the assumption that with higher standards, better prepared students were admitted, and as a result of having better prepared students, graduation rates would have increased. This would correspond with the NCAA's study examining the effect of Proposition 48 (*NCAA Research Report Series 01, 1990)*, as well as with other research stating that high school GPAs and standardized test scores are predictors of academic success at college (Burton & Ramist, 2001; Fleming, 2002; Betts & Morrell, 1999).

[18] In addition to transfer students, athletic departments could also offer scholarships to walk-ons.

However, the hypothesized effect of the policy on graduation rates for Division II schools is ambiguous. Graduation rates would increase if the student-athletes who no longer qualified for Division I schools academically were admitted to Division II schools and had higher probabilities of graduating than other student-athletes at Division II schools. Or it may be the case that these individuals were at the bottom of the ability distribution of Division I schools, and when they enrolled in Division II schools, they remained at the bottom of the distribution, which would decrease the graduation rate at their Division II schools. It is possible that although these individuals did not achieve the required SAT score or GPA for Division I eligibility they still had the ability to be successful at college, and they possessed the athletic ability to compete in Division I athletics. Thus, they may spend enough time at a Division II school to meet the requirements to transfer to a Division I school, in which they would then be counted as a dropout, and graduation rates would decrease at the Division II schools.

Data

The data for the present analysis comes from the National Collegiate Athletic Association's *Graduation Report*, which publishes enrollment and graduation rates for each member institution as mandated by the *Student Right to Know Act*. Individual institutions submit the data to the NCAA where it is reviewed by research staff. After a review by the NCAA, a report is published and each institution verifies the report. The *Graduation Report* is found on the NCAA website (www.ncaa.org) and was converted for the present study into a panel data set that consists of institutional level data. Five entering freshmen cohorts were used for this analysis, three cohorts prior to Proposition 16 (1993–1995) and two cohorts after Proposition 16 (1996–1997). In 1993 institutional-level data was made available to the public.

Although data is available for cohorts past 1998, the NCAA implemented a program known as Academic Progress Reports, which required student-athletes to complete a specific fraction of their degree by each year. Punishments were also established for schools who received poor Academic Progress ratings, and such punishments could result in loss of allotted scholarships. Thus, schools might change the effort put forth to aid student-athletes towards completing a degree and graduation. While this policy did not take effect until 2005, any changes that institutions implemented in response to this policy would affect the six-year graduation rate of the 1998 entering cohort. In addition to this policy change, the NCAA was also subjected to stricter confidentiality requirements by the Department of Education which caused them to suppress more data when cell sizes were small. Therefore, using more than two years after the policy in the analysis might not allow the study to cleanly identify the effects of Proposition 16 from the effects of other policies.

Proposition 16 only applied to those student-athletes who participated in Division I schools and not to those who participated in Division II. The distinction between divisions is made by criteria set by the NCAA based on number of scholarship athletes, number of sports offered, and other factors. The data contains 306 schools in Division I and 288 schools in Division II. Furthermore, the policy only affected student-athletes participating in intercollegiate athletics and defines student-athletes as full-time students who receive athletic aid[19]. Therefore, those schools who do not offer athletic aid (Division III institutions, Ivy League schools, and U.S. Military Academies) are excluded from this analysis.

To protect the privacy of individuals, the NCAA suppresses data when the number of enrolled student-athletes or number of student-athletes who graduate is one

[19] Athletic aid is defined as any grant, scholarship, tuition waiver, or any other financial assistance from the college or university based on a student's athletic ability.

or two. If the number of enrolled student-athletes is suppressed, then the graduation rate is also not reported. When this is the case, I impute the number of enrollees equal to one and the graduation rate is still reported as missing. However, if the number enrolled is greater than two and the number of graduates is one or two, the graduation rate is not reported, but the number of enrollees is still provided. When this is the case, an imputed graduation rate is used such that the number of student-athlete graduates is equal to one[20]. By using this imputation method, the results on enrollment and graduation rate should be seen as a lower bound.

The first key variable of interest is the number of first time freshmen student-athletes who enroll at a given institution. As shown on Figure 3.2, the average number of freshmen student-athletes at Division I and II schools increased between 1993 and 1997. These increases over the years are also present for white student-athletes at both divisions. However, the average number of black student-athletes remained fairly constant but decreased slightly at Division I schools and slightly increased at Division II schools over the time span. Division I averages more student-athletes per school than Division II, and both types of schools experienced an increase in the average number of enrolled student-athletes since 1993. The total number of student-athletes at a school can increase for one of two reasons, the number of allotted scholarships can be increased by the NCAA or schools can divide a scholarship between multiple individuals.

It may be thought that it is the “big-time” sports that would be most affected by the stricter admission policies, namely football and basketball. Thus, examining individual sports would be ideal to see if the higher admission standard effected sports differently. The data does not provide a continuous measure of the number of

[20] Results do not change significantly when imputing the graduation rate assuming the unreported number is two and imputing the number of enrollees to be equal to two.

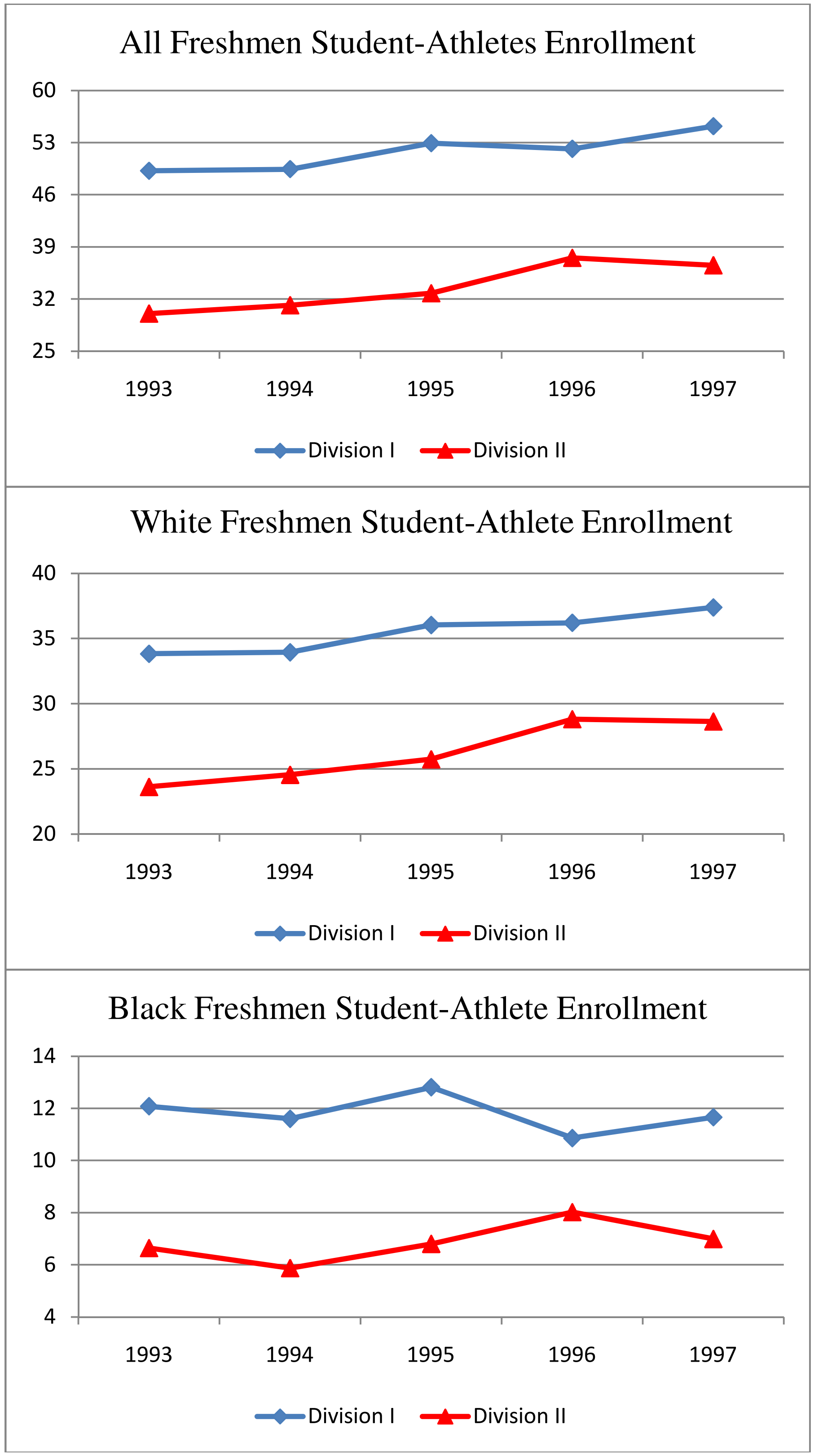


Figure 3.2 Enrollment Trends in the Number of Freshmen Student-Athletes

freshmen student-athletes by sport. What the data does provided is enrollment by groups (1-5, 6-10, 11-15, 16-20, and more than 20) for certain sports and then clumps all other sports into one category. Because of the size of the groups, men's football is the only big-time sport that experiences variation between the different categories due to the number of football scholarships allotted. Basketball admits so few freshmen student athletes each year that any change in recruiting will not be detected by the bins that the data provides. Due to the clumping of other sports in one category[21], a comparison between football and other non-big-time sports can be made. Even though many student-athletes may receive special admissions, it is believed that football players are the greatest benefactors of special admissions. This leads to the hypothesis that higher admission standards should affect the enrollment of football participants more than other sport participants.

The other key outcome of interest is six-year graduation rates, which is defined as the number who graduate within six years divided by the size of the entering cohort at a given institution. For example, the graduation rate for the entering cohort of 1995 measures those that graduate by 2002. This measure of graduation rate is limited by the treatment of transfer students. Students who transfer out of a school while in good academic standing and graduate at another institution are considered dropouts for the initial school. Furthermore, students who transfer into a school are not accounted for with this measure of graduation rate, regardless of whether they graduate or not. One of the strengths of using graduation rates is that it is measured in the same way for student-athletes and students at each institution, thus comparisons can be made between the two types of students Graduation rates for student-athletes at Division I schools are higher than student-athletes at Division II schools. Compared to the

[21] For men, other sports include all sports except baseball, basketball, football, and track/cross country and for women other sports include all sports except basketball and track/cross country.

general student body population, student-athletes in both Division I and Division II schools experience higher graduation rates (see Figure 3.3 and Figure 3.4), and rates have been increasing between 1993 and 1997.

Methods

Prior to Proposition 16, schools in both Division I and II had the same academic requirements for student-athletes. The treatment in this analysis is the change in admission standards caused by the policy change of Proposition 16. With the implementation of Proposition 16, only Division I schools experienced an increase in admissions requirements. Prior to the treatment, Division I and II schools experienced similar trends in enrollment of student-athletes (see Figure 3.2). Due to these similar trends prior to the policy change, I am able to identify the effect of the higher admission standards on enrollment of freshmen by using a difference-in-differences (DID) approach, using Division II as a control group. The first difference accounts for changes that occurred within each division. The second difference is between differences in changes of enrollment at Division I and Division II schools[22]. The reason for using the natural log of enrollment as the dependent variable is to be able to interpret the estimated coefficients as a percent change in enrollment. Additionally, standard errors are clustered by individual institutions.

With the large incentives that schools have for competing in "big time" sports, there is a perception that participants in these sports are most at risk of being affected by a higher admission standard. While the data does not allow for an analysis for number of freshmen student athletes, a discrete choice model will be used to estimate the effects of Proposition 16 on freshmen enrollment in football and other sports. The

[22] To account for individual institutional policies that may have affected enrollment of student-athletes, a third difference was taken between student-athletes and non-athletes. Results are both quantitatively and qualitatively similar.

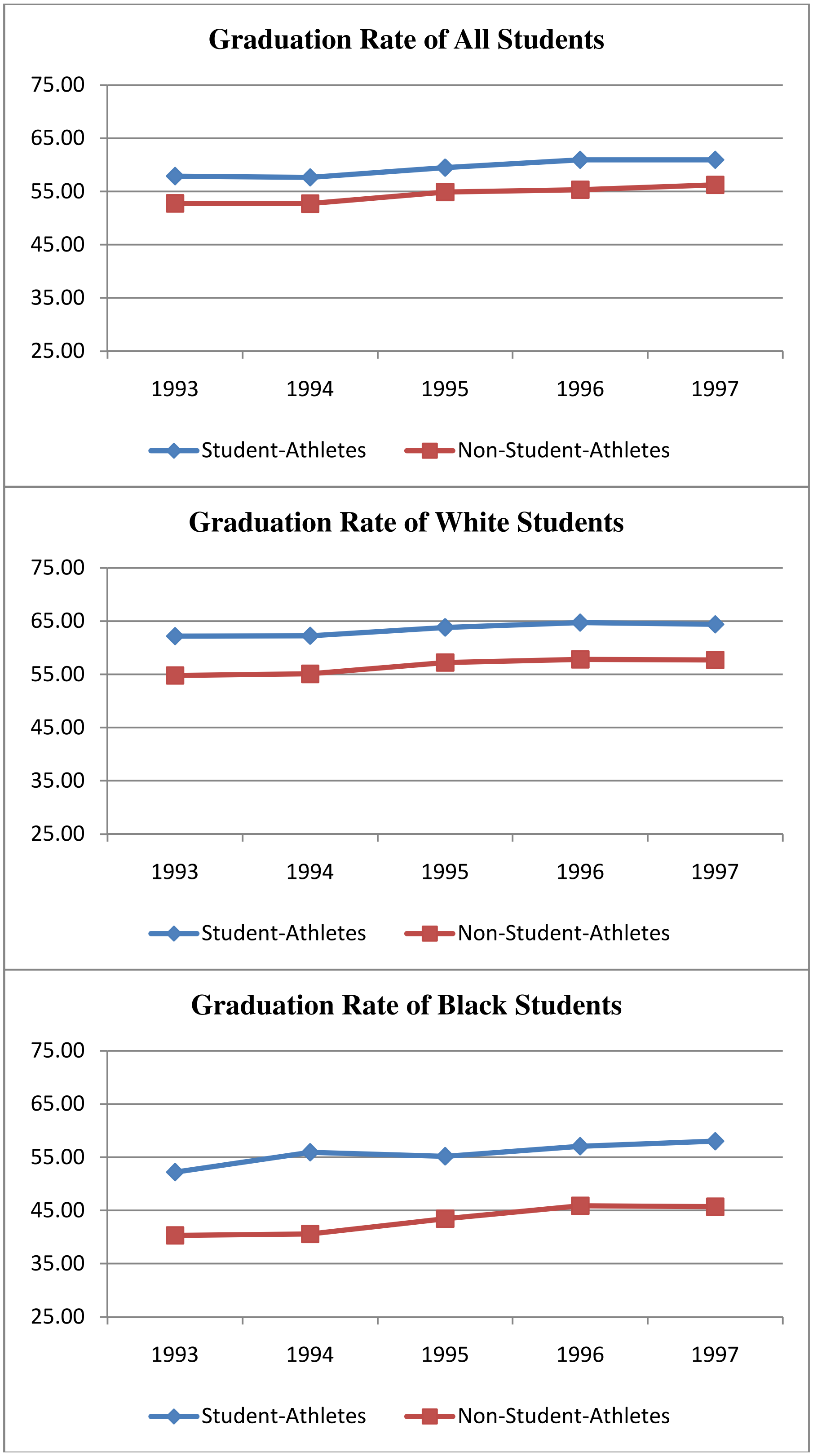


Figure 3.3 Division I Graduation Rate Trends

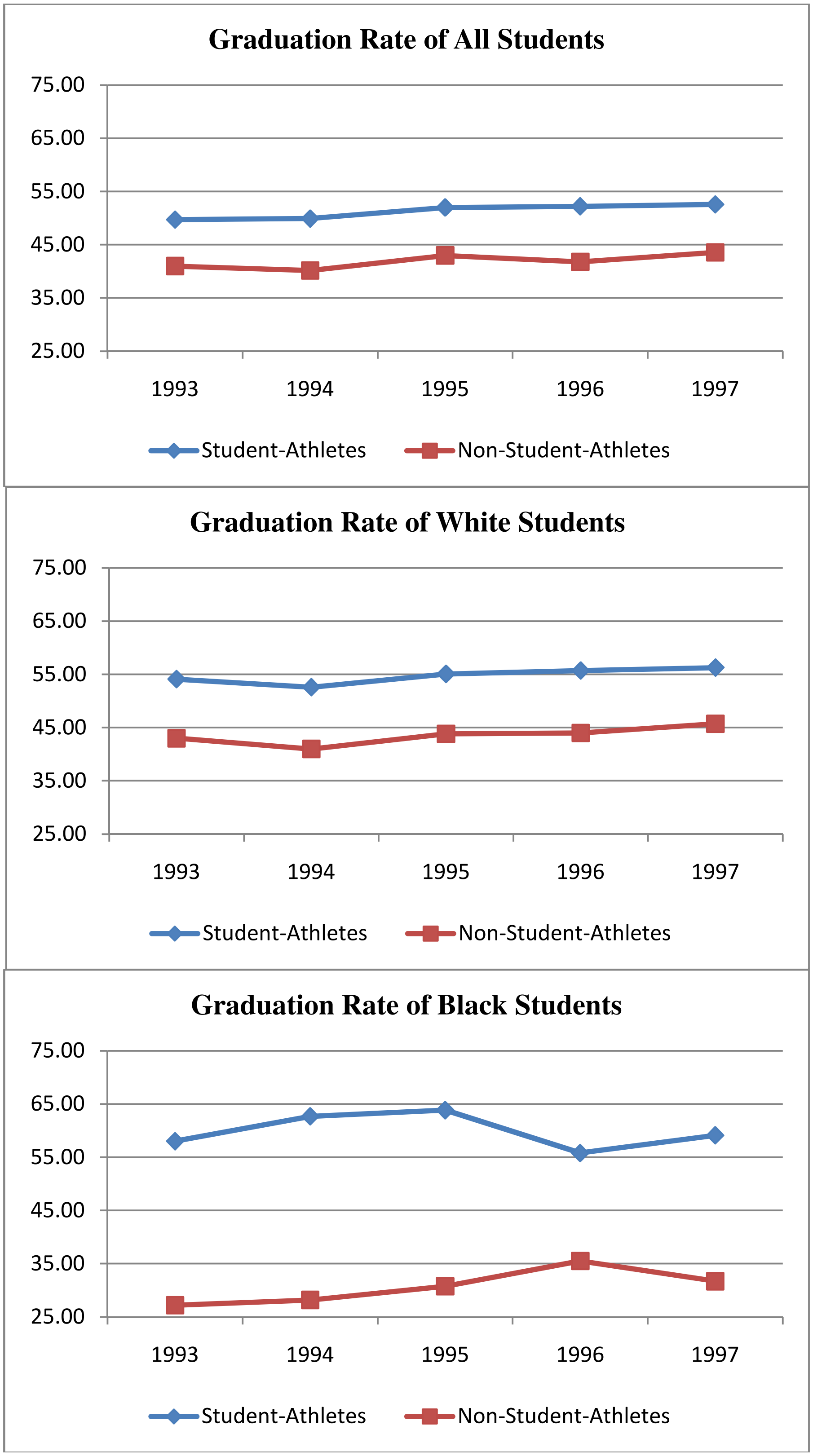


Figure 3.4 Division II Graduation Rate Trends

outcome of interest is enrollment of freshmen student-athletes in the sport and takes on the values of 1 if there are 1–5 recruits, 2 if there are 6–10 recruits, 3 if there are 11–15 recruits, 4 if there are 16–20 recruits, and 5 if there are more than 20 recruits. An ordered probit model will be used because of the ordinal property of the dependent variable. Due to the small cell size of observations when dividing the sample by race, this will only be estimated for all freshmen student-athletes.

A DID approach is also used to examine the effects of Proposition 16 on graduation rates. However, if Proposition 16 changed the types of students who attend Division I and Division II schools, then Division II schools should also be seen as a treatment group and not as a true control. Thus when examining graduation rates, I will compare graduation rates of student-athletes to graduation rates of non-student-athletes at Division I and Division II schools. By using non-student-athletes as the control group, I can separate the effect of Proposition 16 and any other policies that individual universities implemented that affect the graduation rates of the entire student body. For example, a university may increase student services which could increase the graduation rate of all the student body, including student-athletes. Since Division I and Division II experienced two separate treatments, I estimate the DID model separately by Division. Figure 3.3 shows the trends of graduation rates for Division I and Figure 3.4 shows the trends for Division II. Prior to the implementation of Proposition 16, the treatment and control groups experienced similar trends, lending support for the use of a DID approach.

As shown in Figures 3.2, 3. 3, and 3.4, there are general upward trends in both enrollment and graduation rate prior to 1996. A time-trend is included in all specifications to account for these upward trends[23]. It might be thought that there may

[23] Additionally, to account for changes in the slope after the policy, a time-trend interacted with a dummy variable for post is also included. The results do not change qualitatively when using this specification.

be different effects among Division IA compared to other Division I schools or between Bowl Championship Series (BCS) and non-BCS schools. I estimate the models based on these specifications, and the results indicate there are no significant differences among differing classifications within Division I[24].

Results

It was hypothesized that with higher admission standards for first-time freshmen Division I schools would decrease the enrollment of freshmen in the years following the implementation of Proposition 16. Results indicate that the average enrollment of all freshmen student-athletes at Division I schools decreased by 9.9% following the policy change (Table 3.1). Analyzing the results by race show that the average enrollment of white and black freshmen student-athletes decreased by 12.2 and 20.1%, respectively[25]. The average enrollment decrease for black student-athletes was nearly twice as large as the effect on white student-athletes. This result provides support for the claim that black student-athletes were more affected by this policy. To check the robustness of these findings, a third difference is taken between student-athletes and non-student-athletes, and these results indicate that after the policy change black student-athletes enrollment decreased by over 20% at Division I schools.

To further examine the effect of the NCAA policy on enrollment of freshmen student-athletes, the fraction of student-athletes who are freshmen is used as the dependent variable. During the years of analysis, the number of students with athletic scholarships increased, which reflects the ways schools disperse partial scholarships, which may cause the increase in the number of freshmen student-athletes to also increase. If this were the case, then the fraction of student-athletes who were freshmen

[24] These results are available from the author upon request.
[25] A similar analysis is not done for other races or ethnicities due to small sample size of these groups at individual institutions.

Table 3.1 Effect of Proposition 16 on Enrollment

Dependent Variable is log (enrollment)			
Difference-in-Differences			
Student-Athletes	Total	White	Black
Div I * Post	-0.099**	-0.122**	-0.201**
	[0.036]	[0.046]	[0.063]
Div I	0.611**	0.478**	0.915**
	[0.044]	[0.060]	[0.082]
Post	0.062	0.075	0.133
	[0.039]	[0.051]	[0.070]
Constant	3.143**	2.829**	1.185**
	[0.044]	[0.058]	[0.072]
time trend	yes	yes	yes
Observations	2790	2626	2505
Difference-in-Difference-in-Differences			
	Total	White	Black
Ath * Div I * Post	-0.150**	-0.099*	-0.245**
	[0.043]	[0.049]	[0.079]
Ath * Div I	-0.532**	-0.608**	-0.135
	[0.065]	[0.075]	[0.100]
Ath * Post	0.178**	0.179**	0.118
	[0.040]	[0.041]	[0.070]
Div I * Post	0.051	-0.023	0.044
	[0.026]	[0.048]	[0.063]
Div I	1.143**	1.087**	1.050**
	[0.064]	[0.087]	[0.121]
Post	-0.088**	-0.092*	0
	[0.024]	[0.042]	[0.062]
Ath	-2.807**	-2.761**	-2.191**
	[0.056]	[0.063]	[0.078]
Constant	5.972**	5.599**	3.364**
	[0.051]	[0.070]	[0.098]
time trend	yes	yes	yes
Observations	5555	5222	4944
Robust standard errors in brackets			
* significant at 5%; ** significant at 1%			

should not change. As presented in Table 3.2, Division I schools experienced a decrease of 2.8% in the fraction of freshmen student-athletes. The effect is being driven by the decreases in the fraction of black student-athletes who are freshmen, a 3.2% decrease. Taking a third difference with the fraction of the student-body who are freshmen shows that these results are robust. These findings suggest that even if the number of scholarships is increasing over time, the fraction of student-athletes who are freshmen is decreasing after Proposition 16.

There are two groups that should not be affected by this policy and an analysis of these groups can serve as falsification tests. The first of these groups is the total number of student-athletes. Each school is allotted a specific number of scholarships, and a DID estimation should show no change in the total number of scholarship student-athletes. Table 3.3 shows that there were no significant changes in the total number of student-athletes, particularly no change in the total number of black student-athletes. The measure of total number of student-athletes includes freshmen, thus for the total number of black student-athletes to remain unchanged, there must be an increase in the number of black transfer students to account for the decrease in the number of black freshmen. The other group that should not be affected by this policy is freshmen non-student-athletes. As shown in Table 3.4, there is no significant changes at Division I schools after the policy change. These two examples indicate that the policy did not have an effect on groups which were not subject to its requirements.

Focusing on individual sports, football, which has more student-athletes than any other sport, shows similar patterns of reductions in the number of recruits at Division I schools. Average marginal effects from an ordered logit estimation indicate that most of the change occurred from decreases in the probability of enrolling more freshmen recruits at Division I schools (Table 3.5). As hypothesized, the policy had no

Table 3.2 Effect of Proposition on Enrollment

Dependent Variable is Fraction of Students who are Freshmen			
Difference-in-Differences			
	Total	White	Black
Div I * Post	-0.028*	-0.011	-0.038*
	[0.013]	[0.015]	[0.017]
Div I	-0.01	-0.017	0.005
	[0.011]	[0.011]	[0.013]
Post	-0.01	-0.016	-0.017
	[0.016]	[0.017]	[0.020]
Constant	0.236**	0.253**	0.236**
	[0.011]	[0.012]	[0.013]
time trend	yes	yes	yes
Observations	2695	2530	2400
Difference-in-Difference-in-Differences			
	Total	White	Black
Ath * Div I * Post	-0.040*	-0.018	-0.070**
	[0.016]	[0.017]	[0.020]
Ath * Div I	-0.001	-0.011	0.018
	[0.012]	[0.013]	[0.015]
Ath * Post	0.008	0.004	0.019
	[0.014]	[0.015]	[0.018]
Div I * Post	0.012	0.006	0.032**
	[0.008]	[0.008]	[0.011]
Div I	-0.009	-0.006	-0.013
	[0.005]	[0.006]	[0.008]
Post	-0.014	-0.015	-0.034**
	[0.008]	[0.008]	[0.010]
Ath	0.054**	0.073**	0.039**
	[0.010]	[0.010]	[0.012]
Constant	0.186**	0.184**	0.199**
	[0.006]	[0.007]	[0.009]
time trend	yes	yes	yes
Observations	5267	4943	4672
Robust standard errors in brackets			
* significant at 5%; ** significant at 1%			

Table 3.3 Enrollment of All Student-Athletes

Dependent Variable is log (enrollment)			
Difference-in-Differences			
	Total	White	Black
Div I * Post	-0.032	-0.022	-0.008
	[0.035]	[0.046]	[0.046]
Div I	0.590**	0.471**	0.880**
	[0.044]	[0.077]	[0.078]
Post	0.05	0.065	0.041
	[0.041]	[0.054]	[0.050]
Constant	4.698**	4.257**	2.667**
	[0.040]	[0.061]	[0.063]
time trend	yes	yes	yes
Observations	2855	2765	2790
Robust standard errors in brackets			
* significant at 5%; ** significant at 1%			

Table 3.4 Freshmen Enrollment of Non-Student-Athletes

Dependent Variable is log (enrollment)			
Difference-in-Differences			
	Total	White	Black
Div I * Post	0.051	-0.023	0.044
	[0.026]	[0.048]	[0.063]
Div I	1.143**	1.087**	1.050**
	[0.064]	[0.087]	[0.121]
Post	-0.060*	-0.08	-0.016
	[0.024]	[0.044]	[0.065]
Constant	5.995**	5.608**	3.351**
	[0.052]	[0.072]	[0.100]
time trend	yes	yes	yes
Observations	2765	2596	2439
Robust standard errors in brackets			
* significant at 5%; ** significant at 1%			

Table 3.5 Effect of Proposition 16 on Enrollment by Sport

Average Marginal Effects of an Ordered Logit				
Enroll 1 to 5 Freshmen		Football	Male Other Sports	Female Other Sports
	Division I * Post	0.044**	-0.016	0
		[0.016]	[0.018]	[0.010]
	Division I	-0.124**	0.015**	-0.251**
		[0.017]	[0.010]	[0.002]
	Post	-0.03	0.001	0.008*
		[0.024]	[0.022]	[0.011]
Enroll 6 to 10 freshmen		Football	Male Other Sports	Female Other Sports
	Division I * Post	0.034*	0.001	-0.001
		[0.016]	[0.004]	[0.001]
	Division I	-0.111**	-0.195**	0.012**
		[0.009]	[0.001]	[0.001]
	Post	-0.021	0.001	-0.038*
		[0.013]	[0.014]	[0.004]
Enroll 11 to 15 Freshmen		Football	Male Other Sports	Female Other Sports
	Division I * Post	-0.026**	0.01	0
		[0.009]	[0.001]	[0.001]
	Division I	0.082**	0.124**	0.019**
		[0.011]	[0.002]	[0.006]
	Post	0.019	-0.008	0.024*
		[0.015]	[0.002]	[0.001]
Enroll 15 to 20 Freshmen		Football	Male Other Sports	Female Other Sports
	Division I * Post	-0.033*	0.004	-0.002
		[0.014]	[0.001]	[0.016]
	Division I	0.100**	0.042**	0.168**
		[0.007]	[0.007]	[0.001]
	Post	0.021	0.002	0.002*
		[0.014]	[0.002]	[0.017]
Enroll more than 20 Freshı		Football	Male Other Sports	Female Other Sports
	Division I * Post	-0.019*	0.001	0.003
		[0.009]	[0.011]	[0.004]
	Division I	0.052**	0.014**	0.052**
		[0.004]	[0.001]	[0.008]
	Post	0.011	0.005	0.003*
		[0.007]	[0.005]	[0.001]
Controlled for time trend		yes	yes	yes
Observations		3208	4666	5256
Standard errors in brackets,				
* significant at 5%; ** significant at 1%				

measurable effect on admission of other athletic recruits. These results come with limitations; variation does not come from a change in the number, but only changes across bins.   For example, a change in going from 16 to 15 recruits is measured the same as going from 20 to 11 recruits, and a change from 20 to 15 is not picked up by the measurement. Therefore, even though the results suggest a statistically significant change between the categories of football recruits, it is not possible to measure the actual change in number of recruits.

Taken as a whole, these results provide evidence that, as a result of higher initial eligibility standards, Division I schools changed their recruiting patterns. They relied less on freshmen to fill scholarships, particularly black freshmen, and they relied more on transfer students.

The objective of Proposition 16 was to admit students who were academically prepared to succeed at the institutions they attended. The measure of success that was of primary importance was graduation rate. The difference-in-differences approach indicates that overall graduation rate for student-athletes at Division I schools were not significantly different than from non-student-athletes (See Table 3.6). This indicates that the changes in the mean graduation rate from before the policy to after the policy is due to institutional policies and not Proposition 16. However, Proposition 16 had significant effects on graduation rates at Division II schools. Compared to black non-student-athletes at Division II schools, black student-athletes experienced an 8.8 percentage point decline in graduation rates as a result of Proposition 16.

Ex ante, it was unclear what effect the policy would have on the graduation rate of Division II student athletes. It was proposed that it depended on where the non-qualifying students fit in the distribution of the Division II schools they attended. This number might suggest that student-athletes who were deemed unprepared to succeed at Division I schools, as s result of their test score and GPA, were unprepared to

Table 3.6 Effect of Proposition 16 on Graduation Rates

Dependent Variable is graduation rate * 100				
Difference-in-Differences				
	Division I Schools	Total	White	Black
	Student-Athlete * Post	0.294	0.555	-0.208
		[0.504]	[0.935]	[0.702]
	Student-Athlete	4.872**	9.156**	7.021**
		[0.593]	[0.723]	[0.760]
	Post	0.225	-1.404	0.083
		[0.459]	[0.908]	[0.624]
	Constant	51.782**	39.222**	54.171**
		[1.132]	[1.126]	[1.178]
	Observations	3007	2536	2896
	Control for time trend	yes	yes	yes
	Division II Schools	Total	White	Black
	Student-Athlete * Post	-0.18	-1.233	-8.779**
		[1.131]	[1.289]	[2.330]
	Student-Athlete	11.315**	13.031**	32.822**
		[0.787]	[1.149]	[1.920]
	Post	0.751	1.027	0.019
		[1.074]	[1.168]	[2.203]
	Constant	41.399**	38.824**	24.939**
		[1.220]	[1.336]	[1.835]
	Observations	2304	2903	1979
	Control for time trend	yes	yes	yes
Robust standard errors in brackets				
* significant at 5%; ** significant at 1%				

succeed academically at a Division II school. While this may be the case, an alternative explanation for these results is that non-qualifying students who attended Division II schools possessed the athletic ability to compete at Division I schools. Thus, once they completed an academic year of college, they were no longer subject to initial eligibility requirements and could transfer into a Division I school based on college performance, and not standardized test scores. These transfer students would then count as non-graduators and, as a result, the graduation rate would decrease at Division II schools. A back of the envelope calculation shows that the 8% decrease in black student-athletes can be caused by an average of 1 student for every two Division II schools to transfer to a Division I school (an increase of 144 transfer students). This calculation shows that a small number of students can cause a significant change, and lends further support that Division I schools rely more on transfer students after the increase in freshmen eligibility requirements.

Conclusion

The objective of Proposition 16 was to increase graduation rates at Division I schools by increasing the required grade point average and test scores of entering freshmen student-athletes. The results of this study indicate that these higher standards had no real effect on the average graduation rates for Division I student-athletes. However, due to higher academic standards required for eligibility, athletic departments relied less on freshmen and more on transfer students to fill scholarships. This is evident in the decline in the number of freshmen enrolled as well as the fraction of student athletes who are freshmen at Division I schools but no change in the total number of student-athletes.

Opponents of Proposition 16 argued that by relying more on standardized test scores for admittance to Division I schools, minority student-athletes would be affected more so than non-minority student-athletes. This study indicates that enrollment of black freshmen student-athletes decreased as a result of the policy. However, the total number of black student-athletes did not decrease, indicating that more black transfer students were recipients of athletic scholarship after Proposition 16.

One question that this study cannot answer, but attempts to shed light on: are potential student-athletes better off as a result of this policy? Although this study reports changes in school level enrollment and graduation rates, its inability to track students over time does not allow it to examine choices by individuals. With individual-level data, an analysis can be conducted to examine how Proposition 16 affected the choices of the individual with regards to enrollment, persistence, and graduation.

The other question that needs to be addressed is what standard can or should the NCAA use to determine initial eligibility of freshmen student-athletes? Standardized test scores are a low-cost option that has limitations, but what other alternatives are there? It was suggested that the NCAA determine eligibility on a case by case basis, but due to the number of student-athletes this is not at all practical. Recently, many colleges have begun to rely less on the SAT/ACT and some have even dropped it all together (most notably Wake Forest University). The admission process at these schools may serve as useful examples of ways in which a standard could be set without requiring standardized test scores.

All in all, the policies established by the NCAA have attempted to emphasize the academic portion of the student-athlete equation. Graduation rates have continued to rise for student-athletes over the years and are significantly higher than overall student-body graduation rates. However, caution should be used when using only graduation rate as a measure of policy success. As seen with Proposition 16, when only graduation rates are examined, many other aspects of the educational process that play an integral role in evaluating the effectiveness of policies may be overlooked.

REFERENCES

Baade, Robert A. and Jeffrey O. Sundberg, 1996. “Fourth Down and Gold to Go? Assessing the Link between Athletics and Alumni Giving.” *Social Science Quarterly*. 77(4): 789-803.

Betts, Julian R. and Darlene Morell, 1999. “The Determinants of Undergraduate Grade Point Average.” *Journal of Human Resources.* 34(2): 268-293.

Brooker, George and T.D. Klastorin, 1981. “To the Victors Belong the Spoils? College Athletics and Alumni Giving” *Social Science Quarterly.* 62(4): 721-734.

Burton, Nancy and Leonard Ramist, 2001. “Predicting Success in College: SAT Studies of Classes Graduating Since 1980.” *College Board Research Report No. 2001-2.* College Entrance Examination Board, New York.

Camara, Wayne J.; and Amy Elizabeth Schmidt, 1999. “Group Differences in Standardized Testing and Social Stratification.” *College Board Report No. 99-5.* College Entrance Examination Board, New York.

Covell, Dan & Carol A. Barr, 2001. “The Ties that Bind.” *The Journal of Higher Education.* 72(4):414-452.

Fleming, Jacqueline, 2002. “Who Will Succeed in College? When the SAT Predicts Black Student’s Performance.” *The Review of Higher Education.* 25(3):281-296.

Heck, Ronald H. and Robin Takahashi, 2006. “Examining the Impact of Proposition 48 on Graduation Rates in Division 1A Football and Program Recruiting Behavior.” *Educational Policy*. 20(4): 587-614.

Hoffman, John L. and Katie E. Lowitzki, 2005. “Predicting College Success with High School Grades and Test Scores: Limitations for Minority Students.” *The Review of Higher Education.* 28(4): 455-474

Jenks, Christopher and Meredith Phillips, 1995. The *Black-White Test Score Gap.* Brookings Institute Press, 1998. Fredrick E. Vars. and William G. Bowen, Chapter 13. “Scholastic Aptitude Test Scores, Race, and Academic Performance in Selective Colleges and Universities.”

Knobler, Mike, 2008. "AJC Investigation: Many Athletes Lag Far Behind on SAT Scores." The Atlanta Journal Consitution. December 28, 2008.

NCAA Research Report Series 01. 1990-1997 http://www.ncaa.org/library/research/01Series/index.html

NCAA News (Sep. 1, 2003). "Athletic Graduation Rates Continue Climb." Retrieved from http://www.ncaa.org/wps/ncaa?key=/ncaa/ncaa/ncaa+news/ncaa+news+online/2003/division+i/athlete+graduation+rates+continue+climb+-+9-1-03

NCES, 1995. "Who Can Play? An Examination of NCAA's Proposition 16." *National Center for Education Statistics.* 95-763

Pope, Devin G. and Jaren C. Pope, 2006. "Understanding College Choice Decisions: How Sports Success Garners Attention and Provides Information." *Working Paper*

Profiles of American Colleges, 1997. *Baron's Educational Series, Inc. 21st Edition.* New York.

Sellers, Robert M. 1992. "Racial Differences in the Predictors for Academic Achievement of Student-Athletes in Division I Revenue Producing Sports. *Sociology of Sport Journal.* 9(1): 48-59.

Temp, George, 1971. "Validity of the SAT for Blacks and White in Thirteen Integrated Institutions." *Journal of Educational Measurement.* 8(4):245-251.

Tucker, Irvin B, 2004. "A Reexamination of the Effect of Big-Time Football and Basketball Success on Graduation Rates and Alumni Giving Rates." *Economics of Education.* 23(6):655-661.

CONCLUSION AND EXTENSIONS

In conclusion, this dissertation contains three distinct empirical chapters. Each chapter focuses on an issue pertaining to the economics of education or health. For each chapter, I plan on extending my current analyses through the following research outlined below.

Chapter 1 examines the effect of instructor race and gender on persistence in STEM fields. I plan on extending my current analysis on this topic in several different ways. First, the current study focuses on first semester courses and transitions between majors after the first semester and first year. I plan to examine the effect of courses taken after the first semester and transitions that occur within the first two years of college. I also plan to model the full decision process that students face between persisting in a STEM field, changing major, or dropping out. Second, I plan to analyze the type of selection that occurs when students are able to select their own courses. To do this, I will use data from a large public university which randomly assigned students to sections of courses and then changed its assignment policy by allowing students to select their own schedule of courses. Utilizing this change of policy, I will be able to identify the type of selection that occurs based on observable characteristics of faculty. Finally, I plan to work with Doug Weber and Ron Ehrenberg on an analysis of how expenditures other than instructional expenditures affect graduation and persistence rates.

Chapter 2 analyzes a program that offers financial incentives for weight loss. With coauthor John Cawley, I will be extending the analysis to further examine the outcomes of attrition and weight loss. We are planning on implementing a hazard model to estimate factors that predict attrition. We will also be using regression analyses to control for observable characteristics that are correlated with weight loss.

In our current analysis we focus on the first year that the program was implemented. However, multiple worksites continued using the program past the first year, which will allow us to examine longer term outcomes of the programs in future research Chapter 3 analyzes the effects of the NCAA's Proposition 16 on enrollment and graduation rates of student-athletes. This policy focused on initial eligibility requirements, subsequent NCAA policies, most notably the Progress-Toward-Degree Rule, have focused on increasing retention and graduation rates of student-athletes. I hypothesize that student-athletes will sort into majors that make it easier for institutions to comply with graduation benchmarks established by the NCAA. I plan to use data from multiple institutions to examine the distribution of student-athletes across majors and test this hypothesis.