Essays in Labor Economics

by

Tyler Williams

B.A., Middlebury College (2006)

Submitted to the Department of Economics
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2013

Signature of Author ....

Department of Economics
August 21, 2013

Certified by ....

Joshua Angrist
Ford Professor of Economics
Thesis Supervisor

Accepted by..

Michael Greenstone
3M Professor of Environmental Economics
Chairman, Departmental Committee on Graduate Studies

Essays in Labor Economics

by

Tyler Williams

Submitted to the Department of Economics
on August 21, 2013 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy in
Economics

ABSTRACT

I addressed three questions in Labor Economics, using experimental and quasi-experimental variation to determine causality. In the first chapter, I ask whether playing longer in the NFL increases mortality in retirement. I compared players with very short careers with those with long careers. I also examined mortality for replacement players used briefly during the 1987 players' strike. I find that mortality is 15 percent higher for players with longer careers. This difference is even larger for positions with a high risk of injury.

In the second chapter, we use a randomized experiment to evaluate the effects of academic achievement awards for first- and second-year college students studying at a Canadian commuter college. The award scheme offered linear cash incentives for course grades above 70. Awards were paid every term. Program participants also had access to peer advising by upperclassmen. Program engagement appears to have been high but overall treatment effects were small. The intervention increased the number of courses graded above 70 and points earned above 70 for second-year students, but generated no significant effect on overall GPA. Results are somewhat stronger for a subsample that correctly reproduced the program rules.

In the third chapter, we examine two questions: (1) What is the value of receiving the first draft pick in the National Basketball Association?, and (2) Do teams lose intentionally to secure higher draft positions? We answer the first question by adjusting for the probability of winning the lottery using a propensity score methodology. The estimates indicate that winning the draft lottery increases attendance by 6 percentage points during the five-year period following the draft. Receiving the first pick is also associated with a small increase in win percentage. To answer the second question, we use a fixed-effects methodology that compares games in which a team can potentially change its lottery odds to games at the end of the season in which these odds are fixed. Since 1968, playoff-eliminated teams have seen around a 5 percentage point increase in win percentage once their lottery odds are fixed. This difference has ballooned above 10 percentage points in more recent years.

Thesis Supervisor: Joshua Angrist
Title: Ford Professor of Economics

Acknowledgements

I am grateful to the MIT Labor/Public Finance Workshop, the MIT Labor Lunch, the Harvard Labor Economics Workshop, the MIT Sloan Sports Analytics Conference, and especially Christopher Walters, David Chan, Christopher Palmer, Amy Finkelstein, and David Autor for excellent feedback that helped shape this thesis. I am also indebted to the George and Obie Shultz Fund at MIT, the Higher Education Quality Council of Ontario, the Spencer Foundation, and the National Science Foundation for funding this research and my doctoral education. Most of all, I thank my advisors, Josh Angrist, Robert Stavins, and Heidi Williams, for their patience and valuable suggestions as this work evolved.

Chapter 1

Long-Term Mortality Effects of an NFL Career

Professional football is a high-profile dangerous industry. Around 30 percent of National Football League (NFL) players see at least one season end early due to tendon or ligament damage, broken bones, concussions, or other injuries. However, for many young men, the perceived benefits of an NFL career – fame, high salaries, medical and training services, and utility from playing football – outweigh the perceived injury risk. Using mortality data on all NFL players born between 1936 and 1976, I attempt to quantify the overall effects of time spent in the NFL on mortality. Baron et al. (2012) report that NFL retirees have lower mortality than the age-matched general public, but NFL players are positively selected on many genetic characteristics, smoking rates, and physical fitness history. To reduce positive selection bias, I compare NFL retirees with short careers to retirees with long careers. Cox proportional hazard and OLS models yield no significant linear relationship between career length and mortality. Career-shortening injury histories may increase mortality, however, biasing these estimates downwards. To address this bias, I examine the early retirement margin, before most players can accumulate substantial injury histories. I estimate that mortality for players with careers longer than three years is about 15 percent higher than mortality for players with three year careers, assuming a constant proportional effect across all ages. For positions with a high risk of injury, the mortality hazard for players with more than three years' experience is 25 percent higher. As an additional check, I compare 1,000 "replacement players" from the three-game 1987 players' strike to the NFL population. Though many replacement players are not yet 50 years old, traditional NFL players appear to have 50 to 100 percent higher mortality rates.

I. Introduction

Professional football is a high profile example of a dangerous industry. Each year, more than 200 players are forced onto the National Football League's (NFL) injured reserve list, ending their season prematurely. Around 30 percent of players land on the injured reserve list at least once in their career, due to concussions, ligament or tendon damage, broken bones, or other injuries.[1] However, for many young men, the perceived benefits of an NFL career – fame, high salaries, medical and training services, and utility from playing football, among other factors – outweigh the perceived injury risk. In this paper, I attempt to quantify the overall effect of time spent in the NFL on mortality.

Several survey-based studies suggest that NFL retirees have a poor quality of life. Forty percent of NFL retirees self-report arthritis pain compared with 11 percent of the similarly aged general population (Golightly et al., 2009). Concussions sustained while playing football are correlated with self-reported dementia and memory loss (Guskiewicz et al., 2005) and depression diagnoses (Guskiewicz et al., 2007), which are symptoms of Chronic Traumatic Encephalopathy (CTE). Known as "punch-drunk syndrome" in boxers, CTE causes brain degeneration, leading to memory loss, depression, impaired motor function, and, eventually, death. Numerous clinical studies (see, for example, Stern et al., 2011) have diagnosed CTE in deceased athletes with a history of concussion. While these studies have generated substantial interest in the health effects of contact sports, they do not prove that playing in the NFL reduces quality of life or life expectancy. First, NFL players differ from the general population in many ways. Second, establishing a causal link between head injuries and CTE is difficult, since diagnosis must be done post-mortem and few brains have been made available for study. Third, concussion risk is just one aspect of playing in the NFL that may affect mortality.

Direct mortality analysis for NFL players is limited. The National Institute for Occupational Safety and Health (NIOSH) has published two reports (Baron et al., 1994, Baron et al., 2012) showing that NFL players live longer than the age-matched general population. However, in line with evidence relating CTE to concussions, the NIOSH observed higher death rates from brain degenerative diseases (CTE is often diagnosed as Alzheimer's, Parkinson's, or ALS) for NFL players compared to the general population (Lehman et al., 2012).

These comparative studies likely suffer from omitted variables bias as well; NFL players have lower smoking rates, higher BMIs, better athletic ability and other genetic characteristics, and a stronger

[1] NFL team rosters are capped at 53 players. Placing a player on injured reserve ends his season in nearly all cases and opens up a roster spot for another player. I generated injured reserve statistics using weekly injury report tracking (available from 1994 to present) from the STATS database used throughout this study (described in the following section).

fitness history compared with the general population. This bias is an example of the "healthy worker effect": in most settings, employees have better health outcomes than the general population because a minimum health standard is required to maintain employment (see, for example, McMichael, Spirtas, and Kupper, 1974, McMichael, 1976, Carpenter, 1987). The NFL also provides health insurance to many retirees and disability payments to those who qualify. These benefits reduce the costs and increase the incentives, respectively, for NFL retirees to obtain various health diagnoses. Because of these issues, the mortality effects of an NFL career remain unknown.

I employ a different approach to measure mortality effects. Using data on game participation and mortality for all players born between October 1, 1936, and October 1, 1976, I compare mortality for players with shorter versus longer careers. Estimating the relationship between career length and mortality reduces the contribution of some omitted variables, but may introduce other biases. First, there may be an inframarginal healthy worker effect: players may have shorter careers because they lack physical ability or have a poor work ethic. These characteristics likely increase mortality for players with short careers, biasing estimated effects of career length on mortality downwards. Second, acute injuries shorten careers (Edgeworth Economics, 2010). These injuries may also increase mortality, biasing estimated effects downwards again. Third, players may leave the NFL because they have a strong outside option. The value of a player's best outside option is probably negatively correlated with his mortality, biasing the estimates upwards.

Cox proportional hazard models yield no significant linear relationship between career length and mortality. The precision in these estimates is sufficient to rule out a 5 percent difference in mortality for each additional year played for most subgroups studied. OLS estimates are similar, suggesting a slight decrease in mortality before age 40 for players with longer careers and no mortality differences at older ages. These results contrast with estimates for Major League Baseball players showing that the correlation between career length and mortality in a less risky environment is strongly negative (Abel and Kruger, 2006, Saint Onge, Rogers, and Krueger, 2008). Under the assumption that omitted variables biases are similar in professional baseball and football, the difference between sports suggests that the true mortality impact of a longer NFL career is positive, especially given additional potential downward bias from career-shortening injuries in the NFL.

To reduce the possible bias due to injuries and make a stronger case for a positive mortality effect, I investigate the early retirement margin, before most players can accumulate substantial injury histories. Specifically, I compare mortality for players who played less than three years against mortality for those who played more than three years. Using Cox hazard models with constant proportional effects at different ages, I find that players with careers longer than three years have about 15 percent higher mortality than players with shorter careers, stratifying on birth cohort and playing position and controlling for NFL draft selection number, college football program quality, height, and BMI. While these estimates may still be biased by differences in ability, effort, and outside options, estimates for positions with a high risk of injury buttress the results. For these positions, mortality for longer-tenured players is 25 percent higher. This increased hazard ratio suggests that a greater physical toll explains the higher mortality for players with careers longer than three years, rather than variation in outside options (for example).

Players with careers under three years are still exposed to substantial injury risk and may participate in professional football for longer than three years in some cases (on practice squads or in other short-lived leagues). To get closer to the full effects of an NFL career, I compare mortality rates between NFL players and around 1,000 "replacement players," who played up to three games during the 1987 NFL players' strike. These players were not quite good enough to play in the NFL in normal times and thus serve as a plausible control group with very low exposure to professional football. I find that NFL player mortality, especially for those with long careers, is 50 to 100 percent higher than replacement player mortality, although the small sample limits precision and mortality for replacement players is only around 2 percent (many are not yet 50 years old). Still, along with the results at the three year margin, these estimates improve on simple comparisons between the NFL and the general population and suggest that time spent in the NFL does increase mortality for at least some players.

The next section describes the NFL and my data, followed by a detailed presentation of results and concluding remarks.

II. Data

The NFL grew out of several smaller semi-professional and professional football outfits in the 1920s. League membership fluctuated through the 1940s before settling to 12 consistent teams in the 1950s. From 1947 through 1960, NFL teams played 12 regular season games and a varying number of preseason games. The rival American Football League (AFL) launched in 1960 with eight teams and a 14 game schedule, and the NFL itself added two teams from 1960 to 1961 while moving to a 14 game schedule. The AFL became competitive with the NFL fairly quickly, and the two leagues announced a merger in 1966 (consummated in 1970), keeping the NFL name. All AFL statistics from 1960 to 1969 are generally compiled together with NFL statistics, and I will refer to the AFL and NFL jointly throughout the paper as simply the NFL. The ten team expansion from 1960-1961 is by far the NFL's largest permanent expansion. The NFL has added a total of ten more teams since then and increased the number of regular season games to 16 in 1978, reducing the number of preseason games to compensate. It also went through brief player strikes in 1982 (seven games canceled) and 1987 (one game canceled and replacement players used for three games).

My sample includes every NFL player born between 10/1/1936 and 10/1/1976. Unless otherwise stated, all data originated from STATS, LLC, which maintains game statistics and other information for various professional sports. Table 1 reports that players average about 5.1 years from their first year to their last year in the league, generally entering between ages 22 and 24. This number is slightly lower for smaller, faster players (non-lineman) at 4.9 years and higher for bigger players (lineman) at 5.5 years. Many players do not play every game per season, due to injuries or coaching decisions. Players average 57.6 games over a career, which is equivalent to 3.6 sixteen game seasons or about 11 games per year.[2]

NFL players are large. Average BMI in my sample is 29.3, while the average for the U.S. adult male population increased from about 25 in 1960 to 28 in 2000 (Ogden et al., 2004).[3] A BMI over 30 is classified as obese, and, as shown in Figure 1, average BMI for more recent players is well over 30. It is important to note, however, that BMI does not account for body fat percentage or muscle mass. During their careers, NFL players are probably healthier than the average BMI-matched U.S. male or even the average U.S. male. Still, I will control for BMI in all regressions, since it is an important risk factor for many causes of death and no more sophisticated health measures are available.[4]

I gain an additional control variable from the NFL's annual amateur draft. About 67 percent of my sample was drafted into the NFL (the others signed contracts in the open market after the draft). During the draft, teams take turns selecting one college football player at a time. Teams have exclusive rights for one year to negotiate contracts with all of their selected players, and new players are not allowed to sign a contract with any team before the draft. Therefore, teams generally choose the best prospects first, which makes each player's selection number in the draft order a proxy for (perceived) ability. To account for undrafted players, I assign them the highest selection number plus one and include a dummy variable equal to one for undrafted players in all regressions. To account for years where the AFL held a separate draft, I take the draft number for the league in which the player started

[2] I calculate years in the league as (last year played – first year played + 1).
[3] Throughout the paper, I calculate BMI as (weight in pounds / height in inches^2) × 703. Though I do not know exactly when player height and weight are measured, these values are generally measured early in players' careers.
[4] Various studies investigate obesity as a risk factor for heart-related health problems among NFL players (see, for example, Chang et al., 2009).

his career (many players were selected in both drafts) and include a dummy variable equal to one for players who chose to play in the AFL.[5] Figure 2 shows that players drafted sooner have longer careers. This relationship confirms that draft number is a rough proxy for ability.

Using death dates provided by STATS, LLC, mortality for my sample and subgroups is between 5 and 10 percent (last row, Table 1). However, not all deaths appear in the STATS database. Matching a 10 percent sample of the data to the Social Security Death Master File (SSDMF) suggests that an additional 1 percent of the NFL retiree population has died.[6] Unsurprisingly, the missing deaths in the STATS database concentrate among players with careers shorter than three years. To correct this problem, I run simulations that add additional deaths in line with the tenure-specific SSDMF match percentage and repeat the main analysis of the paper. I describe this simulation procedure in more detail in the following section. I am in the process of matching the complete dataset to the SSDMF and to a listing of football retiree deaths from state and national administrative records maintained at www.oldestlivingprofootball.com.

Deceased players had slightly longer careers, as shown in Table 2 (5.3 years played compared to 5.1 for the whole sample), and were slightly more likely to be drafted. The last row shows that about 5 percent of the sample died while still playing football or retired due to a terminal illness or non-football injury. I exclude these players from all regressions, since they would introduce direct reverse causality from mortality to career length.[7]

In line with Baron et al. (1994) and Baron et al. (2012), I find that NFL players live longer than the general population. Figure 3 displays Kaplan-Meier survival curves by age for NFL players and survival curves for the general population based on 2007 cross-sectional mortality and longitudinal mortality for the 1950 birth cohort from the Social Security Administration. The NFL mortality pattern is broadly

---

[5] Alternatively, I could include comprehensive dummies for status in each draft, but for Cox hazard estimation, adding this set of dummies to my preferred specifications leads to failed convergence in many cases. Draft status controls have little effect on OLS point estimates of interest regardless of the specification chosen.

[6] For uncommon last names, I initially matched players to the SSDMF using birth month and last name and then looked within the matched set for close matches on birthday, first/middle name, initials, and state of social security number issuance (birth state in my data). If a player matched exactly on the first two conditions, I counted the player as deceased. If a player matched exactly on birthday and first/middle name were missing but initials matched, I counted the player as deceased only if the state of social security number issuance matched the birth state. Before finalizing any matches, I performed Google searches to ensure that players were not definitively still alive. For common last names, I initially matched by birth month, last name, and first initial, and then followed a similar procedure.

[7] There are few such individuals, so it is unlikely that their exclusion biases the results substantially. Also, slightly more than half of these players had careers longer than three years, and so including them in regressions with a dummy variable for careers longer than three years would only increase the estimated mortality effects of longer careers.

similar to that of the U.S. male population, but NFL mortality is lower at every age. NFL players with BMIs above 30 show survival rates closer to the U.S. male population up to age 60 and higher survival rates thereafter. Since BMI in this range is positively correlated with mortality at most ages and average U.S. adult male BMI has risen from 25 to 28 over the period, it is likely that NFL players outlive their BMI-matched general population counterparts by a substantial margin. However, these comparisons do not answer whether playing football decreases mortality, since NFL players are a highly selected group.

For easier reference, Table 3 presents raw NFL mortality rates at different ages. Each row (and all future age-specific mortality calculations) excludes cohorts that have not reached the stated age threshold, since more players may die in these cohorts before the threshold passes. About 1-2 percent of each subgroup has died by age 40. Mortality is slightly higher for bigger players and increases to 7-12 percent before age 60 and 18-24 percent before age 70.

Since the long-term effects of head trauma have received substantial attention recently, I collected data on cause of death for players with at least a three years' experience by conducting Google searches for death reports and obituaries, using individuals' full name, nickname(s), college attended, NFL teams played for, position played, place of birth and death, and other personal characteristics when available. Tables 4 and 5 list cause of death distributions at each age for all players with at least three years' experience and for the subset with BMI over 30, respectively. For comparison, the tables list cause of death breakdowns for the general population from 1999 to 2009, according to the Center for Disease Control.[8] Car accidents make up a relatively high share of NFL player deaths, especially before age 40 (20 percent). Heart attacks are also overrepresented among deaths before 40. Despite significant concern over recent high profile NFL suicides, NFL suicide rates are lower than general population rates, declining from 7 percent of NFL deaths before age 40 to 3 percent of deaths before age 70. However, deaths directly attributed to brain degenerative diseases (ALS, Parkinson's, CTE, Dementia, and Alzheimer's) make up 2 to 3 percent of NFL deaths at all ages, while these causes barely register in the general population. Unfortunately, I could not obtain cause of death for about 25 percent of deaths before ages 60, 65, and 70. Many of these missing causes are reported as "natural causes" and likely explain some heart and cancer death discrepancies compared to the general population at older ages. Missing data issues aside, these differences may not reflect causal effects of playing in the NFL due to possible selection biases and differences in accuracy of diagnosis.[9]

[8] I obtained these data from the WONDER system, accessible at http://wonder.cdc.gov/mcd.html.

[9] Since this is at best a descriptive exercise, I exclude players with one and two year careers. Initial searching uncovered very limited information on cause of death for these lesser known cohorts.

The next section presents OLS and Cox proportional hazard regression estimates exploring the relationship between NFL career length and mortality/cause of death. To help facilitate interpretation of these results, I will present effects separately for positions with relatively high and low injury propensities. If a positive estimated relationship between career length and mortality is causal, then it should be larger for positions at higher risk of injury. Table 6 reports position-specific injury rates from 2004 to 2009 from the NFL Injury Surveillance System (summarized in Edgeworth Economics, 2010). These data include all injuries noted on teams' weekly injury reports as well as some injuries prior to the season. Fast moving, frequently tackled offensive players (running backs, wide receivers, and tight ends) are injured most often, accruing about 1.5 reported injuries per player per season. These players also have the highest concussion rates (about 0.1 per season), along with fast-moving defensive players (defensive backs) and quarterbacks.[10] Larger, slow-moving players (offensive line, defensive line, and linebackers) suffer only 0.04 to 0.08 head injuries per season, though defensive linemen have a high overall injury rate (about 1.5 per season).

III. Results

A. Linear Career Length Specifications

I first estimate a linear relationship between career length and mortality. While these estimates may be biased by career-shortening injuries that also increase mortality, they provide a good starting point for comparison. Table 7 presents Cox proportional mortality hazard regressions according to the following constant-proportion specification:

$$h(t \mid X_i, C_i, s_i) = h_0(t \mid s_i)\exp(X_i'\beta_1 + C_i\beta_2),$$

where s_i is the stratum for individual i (defined, for example, by birth cohort), h_0 is the baseline hazard function for stratum s_i, X_i is a vector of personal characteristics (for example, BMI, height, draft status), C_i is a measure of career length, $h(t \mid X_i, C_i, s_i)$ is the hazard function for a given X_i, C_i, and s_i, and hence $\exp(\beta_2)$, averaged across all strata, is the hazard ratio of interest.[11] The first and third rows ("Years

[10] Serious injuries may include some head injuries. Also, head injuries are generally under-diagnosed (Guskiewicz et al., 2007). They may be reported more reliably for quarterbacks since hits to the quarterback's head are highly visible. The study does not report injury rates for kickers. Kickers generally do not engage in the contact portion of the game; anecdotal evidence suggests that they have a very low injury rate and an even lower severe/head injury rate compared to other positions.

[11] I stratify by birth cohort and position played rather than control for these covariates for two reasons: (1) stratifying improves optimization and precision, since the Cox proportional hazard model does not actually estimate the hazard function for each stratum (the hazard function can take any form in the Cox model), and (2) I am not especially interested in the estimated hazard ratios for these variables. Controlling for polynomials in birth cohort instead of stratifying yields similar career length hazard ratios. I also collected racial percentages by

Played") set C_i equal to (last year played) – (first year played) + 1 for player i. The second and third rows ("Season Equivalents") set C_i equal to (games played)/16. Panel A controls for demographics only (height, BMI, and birth cohort), while Panel B adds in controls related to the NFL (draft status and position). The column headings define the sample for analysis, and the table lists hazard ratios with 95 percent confidence intervals in square brackets.

There is little change in the mortality hazard for a one unit increase in career length, measured by years played or season equivalents. Hazard ratios for all specifications and samples are between 100 percent and 102 percent of baseline. The 95 percent confidence intervals for the full sample rule out any effect smaller than a 4 percent shift in mortality. The standard deviation for years played is about 4.1; therefore, these estimates are precise enough to rule out a 16 percent mortality change for a one standard deviation shift in career length.

OLS regressions estimating mortality differences at various ages give similar results. Table 8 presents these estimates using the following OLS specification:

$$D(A)_i = \alpha + X_i'\gamma_1 + C_i\gamma_2 + \varepsilon_i$$

where $D(A)_i$ is a dummy equal to one if player i died before age A, and X_i and C_i are defined as above. The sample excludes individuals who could not have reached age A by October 1, 2012, and all coefficients in this and later OLS tables are reported in percentage point units. Estimates for mortality differences by career length at ages 50, 60, 65, and 70 are not statistically significant and give a similar level of precision to the hazard models.[12] However, players with longer careers have measurably lower mortality at age 40. For every additional year played, the probability of death before age 40 is about 0.07 percentage points lower (significant at the 1 percent level). This difference is primarily driven by low injury risk positions, where age 40 mortality is about 0.10 percentage points lower for each additional year.[13] Although this point estimate is 7 percent of age 40 mortality, in absolute terms it is small. A longer NFL career may help high-BMI players stay fit and help all players avoid early accidental deaths, but only 1.4 percent of players die before age 40. More importantly, all of the results in table 8

common last names from the Census Bureau, which matched to about 75 percent of the sample. After estimating the fraction of black players in the NFL in each year, I calculated the probability that each player is black, given their last name. Controlling for this probability in the regression reduces sample size since the match was incomplete and does not meaningfully change most point estimates.

[12] For example, whole-sample mortality at age 50 is about 3.4 percent. The estimated standard errors for age 50 mortality differences by years played rule out any change less than 0.11 percentage points. The standard deviation of years played is about 4.1 years, so I can rule out a mortality difference of about 0.45 percentage points for a one standard deviation change. This difference is around 13 percent of observed age 50 mortality.

[13] Significant OLS estimates at just one age suggest that mortality differences by career length may violate the constant proportions assumption made for Cox hazard estimation. However, interactions between age and career length are not statistically significant when added to the Cox models.

could be driven by the biases discussed earlier. Some players with short careers may have retired due to severe injuries that increased their mortality risk at all ages, and some players who died before age 40 may have had fitness or substance abuse problems that caused both their retirement and early death. To learn more, I examine mortality differences between players with at three years' experience and players with less experience.

B. Three-Year Threshold Specifications

Players who retire after three years or less may retire due to injury, but they have not had much time to accumulate a long, mortality-increasing injury history. The results for this specification are quite different: players with careers longer than three years have higher mortality than players with shorter careers. Tables 9 and 10 repeat the Cox hazard and OLS analyses in Tables 7 and 8, respectively, replacing linear measures of career length with dummies for playing more than three years. Table 9 shows that, after controlling for personal characteristics, draft status, and position, players who last longer than three years have mortality rates over 25 percent higher than players with shorter careers (significant at the 1 percent level). Within positions with a high risk of injury, mortality is 37 percent higher for those with careers over three years. Although sample size shrinks at older ages, Kaplan-Meier survival curves presented in Figures 4 and 5 suggest that the mortality hazard ratios for the whole sample and for high-injury positions are similar across all ages. The mortality point estimates in Table 10 are somewhat consistent with these findings, though the effect appears to be concentrated at age 60 mortality.

These results suggest that NFL injuries increase mortality. However, missing deaths in the STATS database may be biasing the results. To test this claim, I matched a 10 percent sample of players with no recorded death date in the STATS data with the SSDMF (I describe the match procedure in the previous section). Among players in this sample who played for three years or less, about 1.4 percent appeared in the SSDMF. For players who played more than three years, the match rate was about 0.6 percent. I added in deaths to match these rates in 500 separate simulations, using the distribution of observed deaths in the STATS database to select age of death for the simulated data. After running the Cox hazard models on each of the 500 simulated versions, I averaged the point estimates and 95 percent confidence intervals to arrive at the numbers presented in table 11. Though smaller, these numbers suggest that players with more than three years' experience have 16 percent higher mortality rates (significant at the 10 percent level). As before, this difference grows to 27 percent for high injury positions. Without more accurate death data on hand, I take these to be my preferred estimates.

While this analysis may reduce bias due to career shortening injuries, it ignores extant selection biases. Draft selection number helps proxy for ability: the average selection number for players with three years of experience or less is about 174, while the average selection number for players with more than three years' experience is 116 (lower numbers imply earlier selection and therefore higher ability). However, remaining ability and effort deficiencies for shorter career players not captured by the draft could push mortality estimates lower. More worrying given the positive point estimate, these players may also have better outside options that induce early retirement. This omitted variable could explain the mortality increase for players with careers longer than three years. While these factors prevent clean interpretation of the results, the larger point estimates for positions with high injury risk suggest that injuries from time spent in the NFL cause at least some increase in mortality.

C. Replacement Players

The estimates above compare the results of a longer versus a shorter NFL career. Since players with one, two, or three year careers still face substantial injury risk, this comparison may not reflect the total effect of an NFL career, long or short. To get closer to the total effect, I could compare NFL players to college football players who did not play in the NFL. In the absence of comprehensive data on college player mortality, replacement players used for three games during the 1987 NFL players' strike are a similar control group. These players were primarily former college football players who were not good enough or chose not to play in the NFL, and their exposure to the NFL was very low. I identify replacement players as any player whose career started in 1987 and played three games or fewer, yielding 937 individuals. Some traditional players also satisfy these criteria (between 10 and 30 do so most years), and some replacement players earned contracts to stay in the NFL after the three game strike, but the vast majority of the 937 are replacement players.[14]

Table 12 presents estimates from Cox hazard models comparing replacement players with various subsets of the traditional NFL player population.[15] The column headings define the traditional player sample included in the hazard model. Precision is low due to the small number of replacement players, but the estimates generally suggest that traditional players have higher mortality rates than replacement players. Compared to the full sample in column 1, the difference is large (around 40 percent higher mortality for traditional players) but not statistically significant. However, traditional

[14] Unfortunately, the STATS data do not include specific game participation data for each player prior to 1991, so I cannot identify replacement players with 100 percent accuracy.

[15] Among the 10 percent random sample of players with no death date in the STATS data referenced earlier, no replacement players appeared in the SSDMF.

players with careers longer than three years (column 2) have 60 to 70 percent higher mortality than replacement players (significant at the 10 percent level with demographic controls only). The differences grow larger and increase in statistical significance when I restrict the sample to players born between 1958 and 1969, an 11 year window surrounding the most common birth year for 1987 NFL entrants.

Although these results are quite imprecise, on the whole they suggest that NFL players, especially those with longer careers, have 50 to 100 percent higher mortality than replacement players. Replacement players are mostly between 48 and 55 years old on 10/1/2012 with mortality around 2 percent, meaning that this large relative difference corresponds to a small absolute difference. Still, the results are consistent with the findings above suggesting that an NFL career increases mortality, especially for players with careers longer than three years.

D. Cause of Death and Concussions

Recent research suggests that repeated concussions may increase the probability of death due to advanced CTE, though these studies are not yet conclusive due to small samples and limited use of control populations (see, for example, Stern et al., 2011). If time in the NFL actually increases mortality, head injuries are one possible channel for the effect. CTE causes erratic behavior, irritability, depression, memory loss, and, eventually, a reduction in muscle control that mimics other diseases such as Parkinson's, ALS, and Alzheimer's, leading to misdiagnosis (McKee et al., 2009). As with other degenerative brain diseases, CTE itself may cause death, but CTE also increases risk factors for many other causes of death. To test for this channel, I analyze whether death rates for causes tied closest to CTE are higher for players with more than three years' experience than for players with exactly three years' experience.[16]

Specifically, I examine whether death due to degenerative brain diseases or suicide is more common for individuals with careers longer than three years. Though CTE linkages for other causes are weaker, I also address whether death due to alcohol and drug abuse is more common for these individuals. Finally, I repeat the analysis by position-specific concussion risk. Tables 13 through 15 present the relationship between career length and specific causes of death, controlling for personal characteristics, position, draft status, and number of NFL participants from each player's college. These tables list coefficients in percentage point terms from separate OLS regressions for each cause of death,

[16] Cause of death information available online becomes more and more limited for players with careers shorter than three years.

setting the dependent variable equal to one for the cause of interest and equal to zero for survivors and for all other causes of death.[17]

For the whole sample, mortality at age 40 due to degenerative brain diseases or suicide is about 0.09 percentage points higher for players with careers longer than three years versus three-year players (significant at the 5 percent level). While this difference seems small, overall mortality at age 40 due to these causes is only about 0.14 percent. Similarly, the difference at age 50 is about 0.16 percentage points, against an overall mortality rate of 0.25. Mentally-related mortality rates are not significantly different between groups at age 60 or 65, and though statistically significant at age 70, the difference is implausibly large. Differences for mortality rates including drug and alcohol abuse are a similarly large percentage of the overall mortality rate at ages 40 and 50.

Although these results are consistent with the claim that longer careers lead to more concussions, which lead to brain degeneration, there are three caveats that prevent this interpretation. First, cause of death is more likely to be missing for players with shorter careers, inflating estimates for all other causes of death. This is particularly evident at older ages in Table 13. Large, offsetting coefficients for the missing category balanced against the cardiovascular, cancer, transport-accident, and mentally-related mortality categories probably reflect data limitations for players with short careers rather than any real difference. Second, most mentally-related deaths under age 50 in the sample are due to suicide, which may be driven by depression or other life factors unrelated to concussions. Third, Tables 14 and 15 show that differences in mentally-related mortality rates are counterintuitively biggest for positions with the lowest risk of concussion. Given these challenges, I can draw no conclusion concerning what injuries generate the mortality increase observed for NFL players with careers longer than three years.

IV. Conclusions

Popular opinion holds that playing in the NFL increases mortality risk greatly, lowering life expectancy into the 50s (see, for example, Will, 2012). This view is greatly exaggerated; a recent NIOSH study (Baron et al., 2012) shows that NFL players actually live longer than the general population. However, NFL players are a highly selected group of athletically talented individuals, and a trove of survey and clinical work suggests that quality of life for NFL retirees is poor due to high rates of arthritis and depression and dementia (possibly linked to concussions). In this context, I analyze whether time

[17] I present OLS results only for this section, since the Cox hazard models are highly erratic and often fail to converge with full controls included for these low-mean dependent variables.

spent in the NFL increases mortality in retirement for players born between 10/1/1936 and 10/1/1976. This approach eliminates many omitted variables that plague earlier studies comparing NFL players to the general population. However, barriers to causal inference remain. Career-shortening injuries, ability/heredity, perseverance, and differing outside options may all bias estimates of the effect of career length on mortality, primarily downwards.

Naive analysis employing a linear measure of career length shows that longer careers are associated with slightly lower age 40 mortality, while career length is unrelated to mortality above age 40. This result suggests that the true effect could be positive. For comparison, Major League Baseball career length is strongly negatively correlated with mortality at all ages, likely due to income effects and the same negative biases relevant in football (Abel and Kruger, 2006, Saint Onge, Rogers, and Krueger, 2008). While the NFL's selection process and financial rewards may differ from Major League Baseball's, baseball has far fewer career-shortening and mortality-increasing injuries. These injuries may bias my NFL estimates downwards, yet I still estimate a less negative relationship in the NFL than the relationship reported for Major League Baseball.

To reduce injury bias and narrow in on the causal mortality effect of playing in the NFL, I compare players with more than three years of NFL experience to players with three years' experience or less, who may have shorter injury histories that have less effect on mortality in retirement. Players who persisted for longer than three years have 15 percent higher mortality, across all ages. This increase is driven in large part by 25 percent higher mortality within positions with a high risk of injury. Using replacement players from the 1987 NFL players' strike as the control group inflates the mortality difference to 50 to 100 percent, though replacement players are only around 50 years old today with mortality around 2 percent, and these estimates are not precise.

Although variation in players' outside options may explain some of these differences, these estimates provides the most direct evidence to date that time spent in the NFL increases mortality in retirement. These results reflect past playing conditions and player populations, which may lose relevance as the NFL evolves and salaries increase. Still, they are a step towards understanding the risks associated with contact sports, and, in fact, the mortality differences presented here increase for younger cohorts. In future analysis, I plan to investigate these changes over time. Also, I hope to test whether an instrumental variables approach with playoff participation instrumenting for total games played yields similar results, since playoff qualification and success depends on the efforts of the entire team.

Table 1. NFL Player Characteristics

	All	Non-Lineman	Lineman	BMI > 30
Years played	5.1	4.9	5.5	5.5
	(4.1)	(3.9)	(4.2)	(4.2)
Games played	57.6	54.1	62.9	63.7
	(57.3)	(53.5)	(60.8)	(60.6)
Season equivalents*	3.6	3.4	3.9	4.0
	(3.6)	(3.3)	(3.8)	(3.8)
Career starting year	1982	1982	1982	1985
	(11.1)	(11.0)	(11.4)	(10.8)
BMI	29.3	27.7	33.2	33.0
	(3.6)	(2.4)	(3.0)	(2.6)
Drafted	0.67	0.67	0.70	0.69
	(0.47)	(0.47)	(0.46)	(0.46)
Dead	0.07	0.06	0.08	0.07
	(0.25)	(0.23)	(0.28)	(0.26)
N	12,295	8,140	3,724	4,778

*Games played divided by 16

Note: Sample includes players born between 10/1/1936 and 10/1/1976. Standard deviations are in parentheses.

Table 2. Deceased NFL Player Characteristics

	All	Non-Lineman	Lineman	BMI > 30
Years played	5.3	4.9	6.1	6.1
Games played	61.6	54.2	72.6	73.0
Season equivalents*	3.8	3.4	4.5	4.6
Career starting year	1972	1971	1973	1974
BMI	29.5	27.6	32.4	32.5
Drafted	0.46	0.45	0.47	0.50
Death ended career	0.05	0.06	0.04	0.05
N	808	477	311	348

*Games played divided by 16

Note: Sample includes players born between 10/1/1936 and 10/1/1976.

Table 3. Percent Mortality by Age

	All	Non-Lineman	Lineman	BMI>30
Dead by 40	1.4	1.3	1.8	1.8
N	10,889	7,244	3,262	3,992
Dead by 50	3.4	3.2	4.0	4.5
N	7,093	4,704	2,132	2,190
Dead by 60	7.9	7.1	10.0	11.3
N	3,707	2,439	1,151	1,034
Dead by 65	12.5	12.1	13.1	15.5
N	2,449	1,606	776	699
Dead by 70	19.6	18.8	21.2	23.8
N	1,155	751	378	341

Note: Samples include players born between 10/1/1936 and 10/1/1976 who could have reached the specified age by 10/1/2012.

Table 4. Cause of Death Distributions by Age of Death

	Age < 40		Age < 50		Age < 60		Age < 65		Age < 70	
	NFL	Gen Pop	NFL	Gen Pop	NFL	Gen Pop	NFL	Gen Pop	NFL	Gen Pop
Cardiovascular	0.25	0.14	0.28	0.25	0.29	0.32	0.29	0.32	0.28	0.33
Cancer	0.14	0.09	0.16	0.17	0.16	0.29	0.18	0.31	0.18	0.33
Degenerative*	0.02	0.00	0.02	0.00	0.02	0.00	0.02	0.00	0.03	0.01
Other Internal	0.01	0.22	0.03	0.25	0.04	0.24	0.04	0.23	0.05	0.23
Transport Accident	0.20	0.13	0.12	0.11	0.09	0.06	0.08	0.04	0.07	0.04
Drugs/Alcohol	0.05	0.14	0.04	0.07	0.03	0.03	0.03	0.02	0.03	0.02
Suicide	0.07	0.12	0.05	0.07	0.04	0.03	0.03	0.03	0.03	0.02
Other External	0.19	0.15	0.12	0.07	0.09	0.03	0.08	0.03	0.07	0.02
Missing	0.08		0.16		0.26		0.26		0.27	
Total	**1.00**	**1.00**	**1.01**	**1.00**	**1.02**	**1.00**	**1.02**	**1.00**	**1.02**	**1.00**
N	106		226		380		447		517	

*ALS, Parkinson's, Chronic Traumatic Encephalopathy, Dementia, Alzheimer's

Note: The NFL population includes players with at least three years' experience born between 10/1/1936 and 10/1/1976. "Total" may be greater than one due to cases with multiple causes of death. General population death rates (from the Center for Disease Control) reflect all deaths over age 30 and under the indicated age.

Table 5. Cause of Death Distributions by Age of Death (BMI > 30)

	Age < 40		Age < 50		Age < 60		Age < 65		Age < 70	
	NFL	Gen Pop	NFL	Gen Pop	NFL	Gen Pop	NFL	Gen Pop	NFL	Gen Pop
Cardiovascular	0.34	0.14	0.39	0.25	0.35	0.32	0.35	0.32	0.33	0.33
Cancer	0.12	0.09	0.14	0.17	0.15	0.29	0.15	0.31	0.17	0.33
Degenerative*	0.03	0.00	0.03	0.00	0.02	0.00	0.01	0.00	0.03	0.01
Other Internal	0.00	0.22	0.02	0.25	0.03	0.24	0.03	0.23	0.05	0.23
Transport Accident	0.22	0.13	0.13	0.11	0.11	0.06	0.11	0.04	0.09	0.04
Drugs/Alcohol	0.03	0.14	0.03	0.07	0.03	0.03	0.02	0.02	0.02	0.02
Suicide	0.07	0.12	0.06	0.07	0.04	0.03	0.03	0.03	0.03	0.02
Other External	0.12	0.15	0.10	0.07	0.07	0.03	0.06	0.03	0.06	0.02
Missing	0.05		0.12		0.23		0.23		0.23	
Total	**1.00**	**1.00**	**1.02**	**1.00**	**1.02**	**1.00**	**1.01**	**1.00**	**1.01**	**1.00**
N	58		119		188		203		237	

*ALS, Parkinson's, Chronic Traumatic Encephalopathy, Dementia, Alzheimer's

Note: The NFL population includes players with at least three years' experience born between 10/1/1936 and 10/1/1976. "Total" may be greater than one due to cases with multiple causes of death. General population death rates (from the Center for Disease Control) reflect all deaths over age 30 and under the indicated age.

Table 6. Injuries per Player per Season by Position (2004-2009)

Position	Any	Severe	Head
Quarterback	0.86	0.20	0.11
Offensive Line	1.16	0.28	0.04
Linebacker	1.25	0.26	0.08
Defensive Back	1.37	0.29	0.10
Running Back	1.45	0.29	0.11
Wide Receiver	1.47	0.26	0.10
Defensive Line	1.47	0.29	0.04
Tight End	1.52	0.30	0.11

Note: Statistics from the NFL Injury Surveillance System as reported in Edgeworth Economics (2010).

Table 7. Career Length and Cox Mortality Hazard Ratios

	All	High Injury Position	Low Injury Position	High Concussion Position	Low Concussion Position
	Panel A. Demographic Controls				
Years Played	1.01	1.02	1.01	1.00	1.02
	[0.99 - 1.03]	[0.99 - 1.04]	[0.98 - 1.03]	[0.97 - 1.03]	[1.00 - 1.04]
Season Equivalents	1.01	1.02	1.00	0.99	1.02
	[0.99 - 1.03]	[0.99 - 1.05]	[0.97 - 1.03]	[0.96 - 1.03]	[0.99 - 1.05]
	Panel B. Full Controls				
Years Played	1.01	1.02	1.01	1.00	1.02
	[0.99 - 1.03]	[0.99 - 1.05]	[0.98 - 1.03]	[0.97 - 1.03]	[0.99 - 1.05]
Season Equivalents	1.01	1.03	1.00	0.99	1.02
	[0.99 - 1.03]	[0.99 - 1.06]	[0.97 - 1.03]	[0.96 - 1.03]	[0.99 - 1.05]
N	12,197	5,388	6,379	6,355	5,412

Note: The table reports mortality hazard ratios for Cox proportional hazard models with constant effects by age, stratified by birth cohort in panel A and stratified by birth cohort and position in panel B. The row labels denote the career length measure used and the column headings denote the group analyzed in each regression. Demographic controls (panel A) include BMI, BMI squared, height, and height squared. Panel B adds draft status controls, as described in the Data section of the paper, and the number of NFL participants from each player's college. High injury positions are tight ends, defensive lineman, wide receivers, and running backs; high concussion positions are quarterbacks, defensive backs, running backs, wide receivers, and tight ends. Low injury and low concussion positions are the respective complements, including kickers. I derive 95 percent confidence intervals (in square brackets) from robust standard errors.

Table 8. Career Length and Percent Mortality (OLS)

	All		High Injury Position		Low Injury Position	
	Years Played	Season Equivalents	Years Played	Season Equivalents	Years Played	Season Equivalents
Dead by 40	-0.072***	-0.079***	-0.025	-0.032	-0.103***	-0.111***
	(0.023)	(0.026)	(0.035)	(0.038)	(0.034)	(0.038)
N	10,852	10,852	4,787	4,787	5,682	5,682
Dead by 50	-0.073	-0.076	-0.088	-0.083	-0.072	-0.081
	(0.054)	(0.062)	(0.089)	(0.099)	(0.076)	(0.089)
N	7,070	7,070	3,084	3,084	3,729	3,729
Dead by 60	0.076	0.115	0.215	0.265	-0.004	0.042
	(0.114)	(0.134)	(0.211)	(0.243)	(0.139)	(0.169)
N	3,694	3,694	1,628	1,628	1,949	1,949
Dead by 65	-0.200	-0.255	-0.361	-0.372	-0.127	-0.220
	(0.168)	(0.200)	(0.292)	(0.341)	(0.207)	(0.251)
N	2,438	2,438	1,071	1,071	1,300	1,300
Dead by 70	-0.046	-0.107	-0.007	0.011	-0.024	-0.147
	(0.287)	(0.335)	(0.519)	(0.612)	(0.356)	(0.413)
N	1,148	1,148	493	493	629	629

Note: OLS regressions include position fixed effects, a third degree polynomial in date of birth, baseline BMI, BMI squared, height, height squared, draft status controls as described in the Data section of the paper, and the number of NFL participants from each player's college. High injury positions are tight ends, defensive lineman, wide receivers, and running backs; high concussion positions are quarterbacks, defensive backs, running backs, wide receivers, and tight ends. Low injury and low concussion positions are the respective complements, including kickers. Coefficients are measured in percentage points. Robust standard errors are in parentheses.

Table 9. Career Length and Cox Mortality Hazard Ratios

	All	High Injury Position	Low Injury Position	High Concussion Position	Low Concussion Position
	Panel A. Demographic Controls				
Years Played > 3	1.24***	1.34**	1.17	1.20	1.28**
	[1.07 - 1.45]	[1.07 - 1.67]	[0.94 - 1.46]	[0.96 - 1.51]	[1.03 - 1.59]
	Panel B. Full Controls				
Years Played > 3	1.26***	1.37***	1.18	1.24*	1.29**
	[1.07 - 1.48]	[1.09 - 1.72]	[0.94 - 1.50]	[0.98 - 1.57]	[1.02 - 1.62]
N	12,197	5,388	6,379	6,355	5,412

Note: The table reports mortality hazard ratios for Cox proportional hazard models with constant effects by age, stratified by birth cohort in panel A and stratified by birth cohort and position in panel B. The row labels denote the career length measure used and the column headings denote the group analyzed in each regression. Demographic controls (panel A) include BMI, BMI squared, height, and height squared. Panel B adds draft status controls, as described in the Data section of the paper, and the number of NFL participants from each player's college. High injury positions are tight ends, defensive lineman, wide receivers, and running backs; high concussion positions are quarterbacks, defensive backs, running backs, wide receivers, and tight ends. Low injury and low concussion positions are the respective complements, including kickers. I derive 95 percent confidence intervals (in square brackets) from robust standard errors.

Table 10. More than Three Years Played and Percent Mortality (OLS)

	All	High Injury Position	Low Injury Position
Dead by 40	-0.079	0.376	-0.418
	(0.215)	(0.324)	(0.304)
N	10,852	4,787	5,682
Dead by 50	0.174	0.276	0.061
	(0.438)	(0.677)	(0.597)
N	7,070	3,084	3,729
Dead by 60	2.112**	2.488	1.517
	(0.915)	(1.542)	(1.119)
N	3,694	1,628	1,949
Dead by 65	-1.097	-2.061	-0.315
	(1.404)	(2.303)	(1.726)
N	2,438	1,071	1,300
Dead by 70	0.179	0.717	-0.550
	(2.362)	(3.918)	(3.016)
N	1,148	493	629

Note: OLS regressions include position fixed effects, a third degree polynomial in date of birth, baseline BMI, BMI squared, height, height squared, draft status controls as described in the Data section of the paper, and the number of NFL participants from each player's college. High injury positions are tight ends, defensive lineman, wide receivers, and running backs; high concussion positions are quarterbacks, defensive backs, running backs, wide receivers, and tight ends. Low injury and low concussion positions are the respective complements, including kickers. Coefficients are measured in percentage points. Robust standard errors are in parentheses.

Table 11. Average Cox Mortality Hazard Ratios with Simulated Missing Deaths

	All	High Injury Position	Low Injury Position
	Panel A. Demographic Controls		
Years Played > 3	1.15*	1.24**	1.09
	[1.00 - 1.33]	[1.00 - 1.53]	[0.88 - 1.34]
	Panel B. Full Controls		
Years Played > 3	1.16*	1.27**	1.09
	[1.00 - 1.36]	[1.01 - 1.58]	[0.87 - 1.37]
N	12,197	5,388	6,379

Note: The table reports mean mortality hazard ratios for Cox proportional hazard models with constant effects by age, run on 500 simulations of data with added deaths to match the estimated missing rate as described in the paper. I stratify by birth cohort in panel A and by birth cohort and position in panel B. The row labels denote the career length measure used and the column headings denote the group analyzed in each regression. Demographic controls (panel A) include BMI, BMI squared, height, and height squared. Panel B adds draft status controls, as described in the Data section of the paper, and the number of NFL participants from each player's college. High injury positions are tight ends, defensive lineman, wide receivers, and running backs; high concussion positions are quarterbacks, defensive backs, running backs, wide receivers, and tight ends. Low injury and low concussion positions are the respective complements, including kickers. I derive 95 percent confidence intervals (in square brackets) from averaged robust standard errors across all 500 simulations.

Table 12. Cox Mortality Hazard Ratios, 1987 Replacement Players versus Traditional Players

	All	Years Played > 3	Born 1958-69	Years Played > 3, Born 1958-69	Entered 1987	Years Played > 3, Entered 1987
			Panel A. Demographic Controls			
Traditional Player	1.45	1.69*	1.55	1.73*	1.04	1.97
	[0.85 - 2.47]	[0.98 - 2.91]	[0.89 - 2.70]	[0.97 - 3.07]	[0.42 - 2.59]	[0.76 - 5.10]
			Panel B. Full Controls			
Traditional Player	1.38	1.60	1.67*	2.12**	1.85	5.27***
	[0.80 - 2.39]	[0.91 - 2.83]	[0.92 - 3.03]	[1.10 - 4.06]	[0.64 - 5.36]	[1.66 - 16.66]
N	12,197	7,582	4,308	3,016	1,318	1,148

Note: The table reports mortality hazard ratios for Cox proportional hazard models with constant effects by age, stratified by birth cohort in panel A and stratified by birth cohort and position in panel B. The row labels note that ratios compare traditional players to 1987 replacement players (see text for definition) and the column headings denote the subset of traditional players analyzed in each regression. Demographic controls (panel A) include BMI, BMI squared, height, and height squared. Panel B adds draft status controls, as described in the Data section of the paper, and the number of NFL participants from each player's college. High injury positions are tight ends, defensive lineman, wide receivers, and running backs; high concussion positions are quarterbacks, defensive backs, running backs, wide receivers, and tight ends. Low injury and low concussion positions are the respective complements, including kickers. I derive 95 percent confidence intervals (in square brackets) from robust standard errors.

Table 13. More than Three Years Played and Percent Mortality due to Selected Causes of Death

	Age < 40	Age < 50	Age < 60	Age < 65	Age < 70
Any	0.109	0.958	2.164	0.389	1.052
	(0.353)	(0.694)	(1.528)	(2.507)	(4.616)
Cardiovascular	0.174	0.504*	1.708***	1.614*	2.635
	(0.134)	(0.295)	(0.438)	(0.859)	(1.657)
Cancer	0.021	0.321	0.901	1.872**	3.811***
	(0.133)	(0.303)	(0.666)	(0.754)	(1.120)
Transport Accident	-0.224	-0.183	0.488***	0.569**	0.394
	(0.198)	(0.259)	(0.183)	(0.244)	(0.319)
Other External	0.114	-0.005	-0.618	-0.631	-0.746
	(0.132)	(0.267)	(0.566)	(0.787)	(1.221)
Missing	-0.115	-0.019	-0.484	-3.792*	-7.307*
	(0.175)	(0.420)	(1.149)	(2.089)	(4.024)
Degenerative*,	0.092**	0.156***	0.014	0.115	1.478**
Suicide	(0.042)	(0.059)	(0.300)	(0.530)	(0.628)
Degenerative*,	0.122***	0.311***	0.328	0.317	1.478**
Suicide, Drugs/Alcohol	(0.045)	(0.093)	(0.331)	(0.546)	(0.628)
N	6,665	4,549	2,383	1,558	750

*ALS, Parkinson's, Chronic Traumatic Encephalopathy, Dementia, Alzheimer's

Note: OLS regressions include position and NFL entry year fixed effects, a third degree polynomial in date of birth, baseline BMI, BMI squared, height, height squared, and draft status controls, as described in the Data section of the paper. Coefficients are measured in percentage points. Robust standard errors are in parentheses.

Table 14. More than Three Years Played and Percent Mortality due to Selected Causes of Death (High Concussion Risk)

	Age < 40	Age < 50	Age < 60	Age < 65	Age < 70
Any	-0.044	1.339*	2.755	1.270	4.418
	(0.452)	(0.777)	(1.730)	(3.191)	(5.519)
Cardiovascular	0.158*	0.444	1.113*	1.100	1.937
	(0.082)	(0.339)	(0.631)	(1.072)	(2.653)
Cancer	0.149*	0.676***	0.919	1.956	5.144***
	(0.087)	(0.237)	(0.951)	(1.258)	(1.750)
Transport Accident	-0.156	0.094	0.427*	0.602	--
	(0.222)	(0.064)	(0.258)	(0.381)	--
Other External	0.026	-0.017	-0.152	0.405	0.226
	(0.209)	(0.302)	(0.457)	(0.271)	(0.526)
Missing	-0.336	-0.241	-0.024	-3.302	-7.045
	(0.303)	(0.572)	(1.184)	(2.545)	(4.663)
Degenerative*,	0.064	0.090	-0.373	-0.368	1.302*
Suicide	(0.051)	(0.057)	(0.534)	(0.986)	(0.774)
Degenerative*,	0.074	0.304**	0.083	0.003	1.302*
Suicide, Drugs/Alcohol	(0.052)	(0.130)	(0.587)	(1.015)	(0.774)
N	3,421	2,316	1,198	772	377

*ALS, Parkinson's, Chronic Traumatic Encephalopathy, Dementia, Alzheimer's

Note: OLS regressions include position and NFL entry year fixed effects, a third degree polynomial in date of birth, baseline BMI, BMI squared, height, height squared, and draft status controls, as described in the Data section of the paper. Coefficients are measured in percentage points. Robust standard errors are in parentheses.

Table 15. More than Three Years Played and Percent Mortality due to Selected Causes of Death (Low Concussion Risk)

	Age < 40	Age < 50	Age < 60	Age < 65	Age < 70
Any	0.386	1.231	1.089	0.315	-0.026
	(0.634)	(1.299)	(2.863)	(4.098)	(8.077)
Cardiovascular	0.180	0.706	2.062***	2.986***	3.671*
	(0.334)	(0.548)	(0.599)	(0.976)	(2.043)
Cancer	-0.142	-0.004	1.268	2.301***	4.426**
	(0.305)	(0.734)	(0.964)	(0.835)	(1.978)
Transport Accident	-0.326	-0.122	0.598*	0.819*	0.922
	(0.396)	(0.450)	(0.332)	(0.477)	(0.749)
Other External	0.238*	0.031	-1.102	-1.863	-1.845
	(0.130)	(0.530)	(1.150)	(1.703)	(2.466)
Missing	0.230*	0.308	-1.544	-4.929	-7.112
	(0.135)	(0.698)	(2.297)	(3.613)	(7.166)
Degenerative*,	0.152*	0.316*	0.560**	0.764*	2.100*
Suicide	(0.091)	(0.170)	(0.278)	(0.391)	(1.141)
Degenerative*,	0.206**	0.396**	0.691**	0.764*	2.100*
Suicide, Drugs/Alcohol	(0.099)	(0.185)	(0.297)	(0.391)	(1.141)
N	3,006	2,062	1,108	740	354

*ALS, Parkinson's, Chronic Traumatic Encephalopathy, Dementia, Alzheimer's

Note: OLS regressions include position and NFL entry year fixed effects, a third degree polynomial in date of birth, baseline BMI, BMI squared, height, height squared, and draft status controls, as described in the Data section of the paper. Coefficients are measured in percentage points. Robust standard errors are in parentheses.

Figure 1. NFL Player BMI Over Time

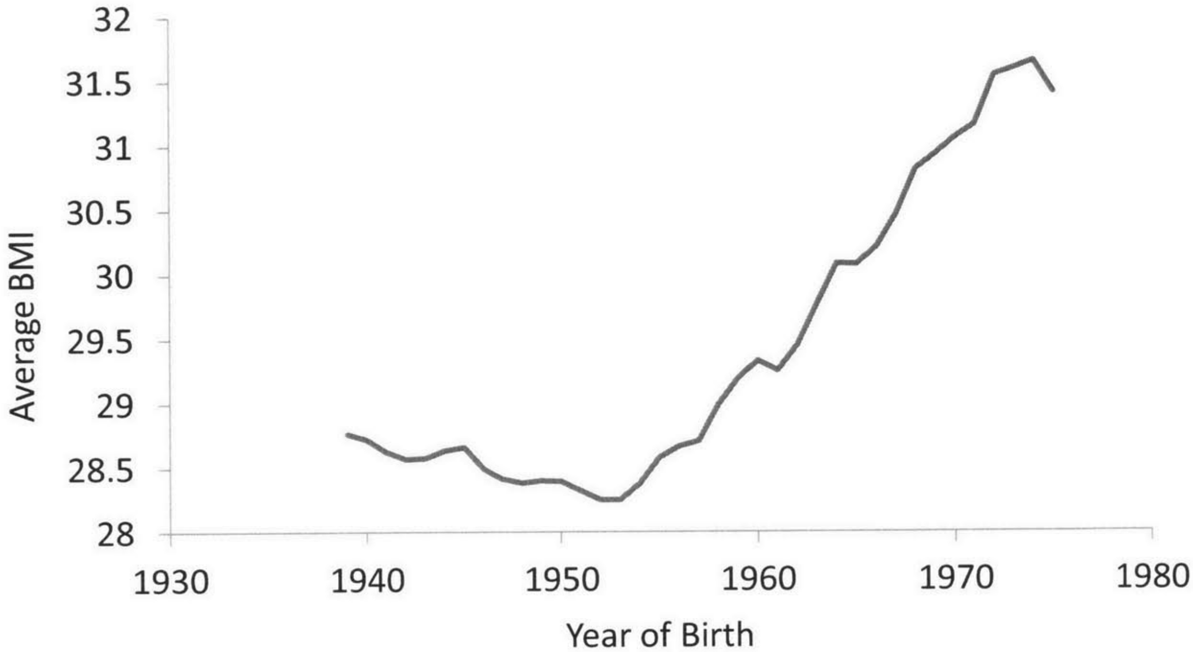


Figure 2. AFL/NFL Draft Selection Number and Years Played

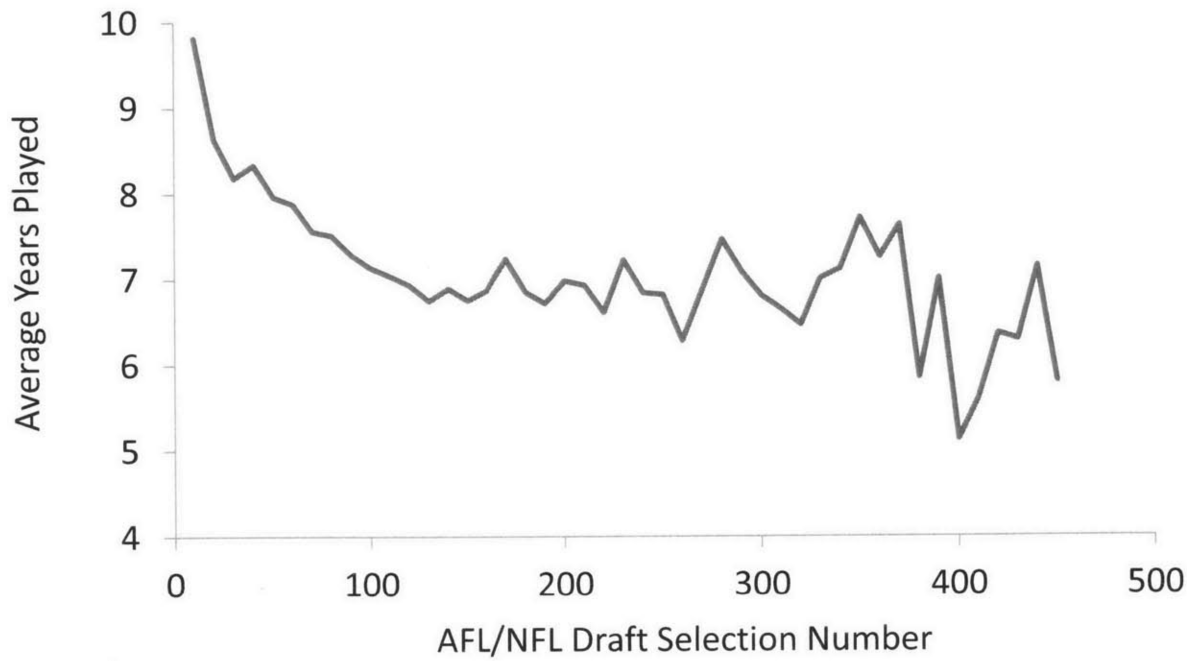


Figure 3. Kaplan-Meier Survival Curves for NFL Players and the General Population

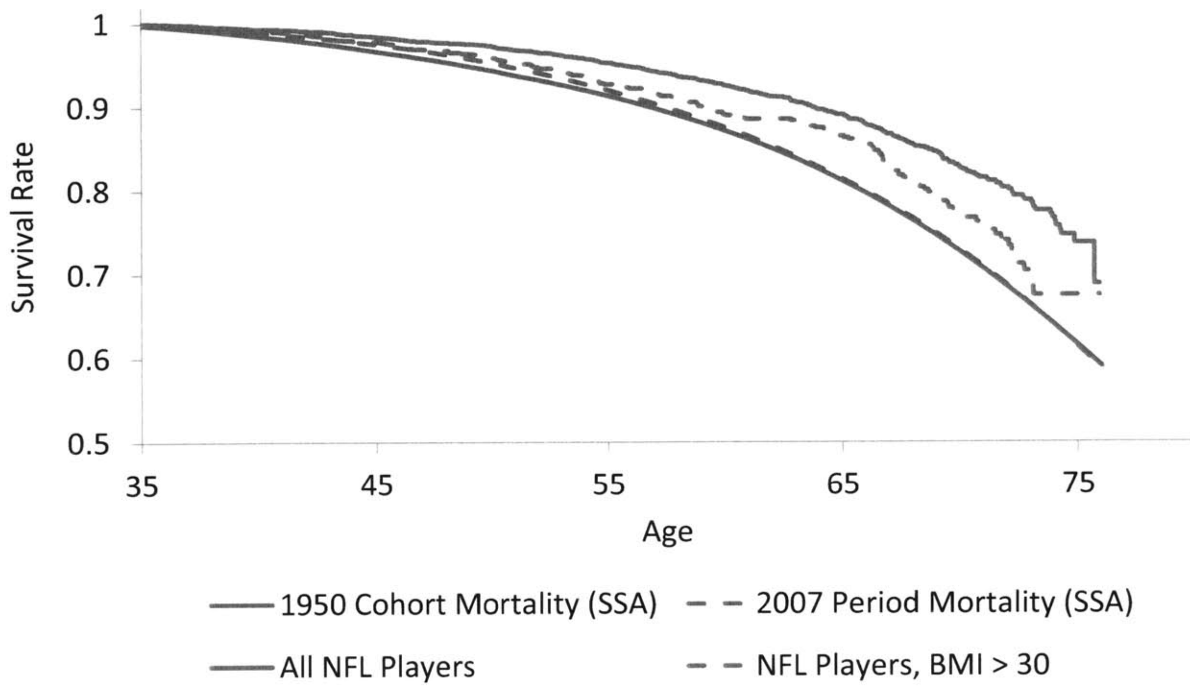


Figure 4. Kaplan-Meier Survival Curves by Career Length

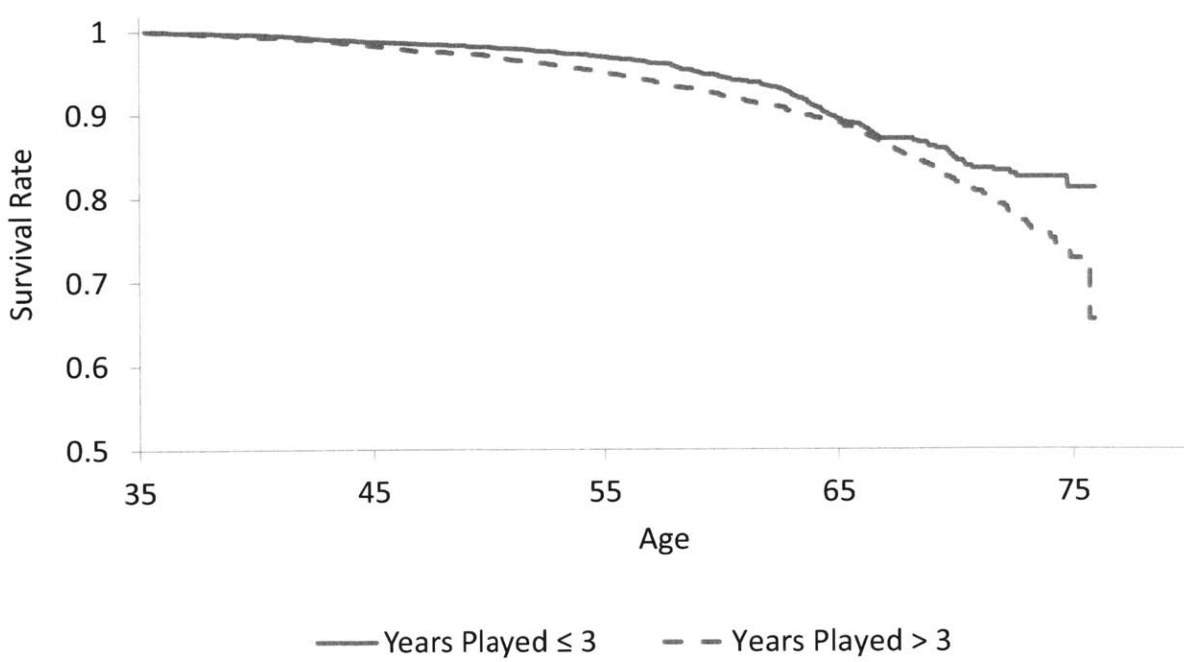
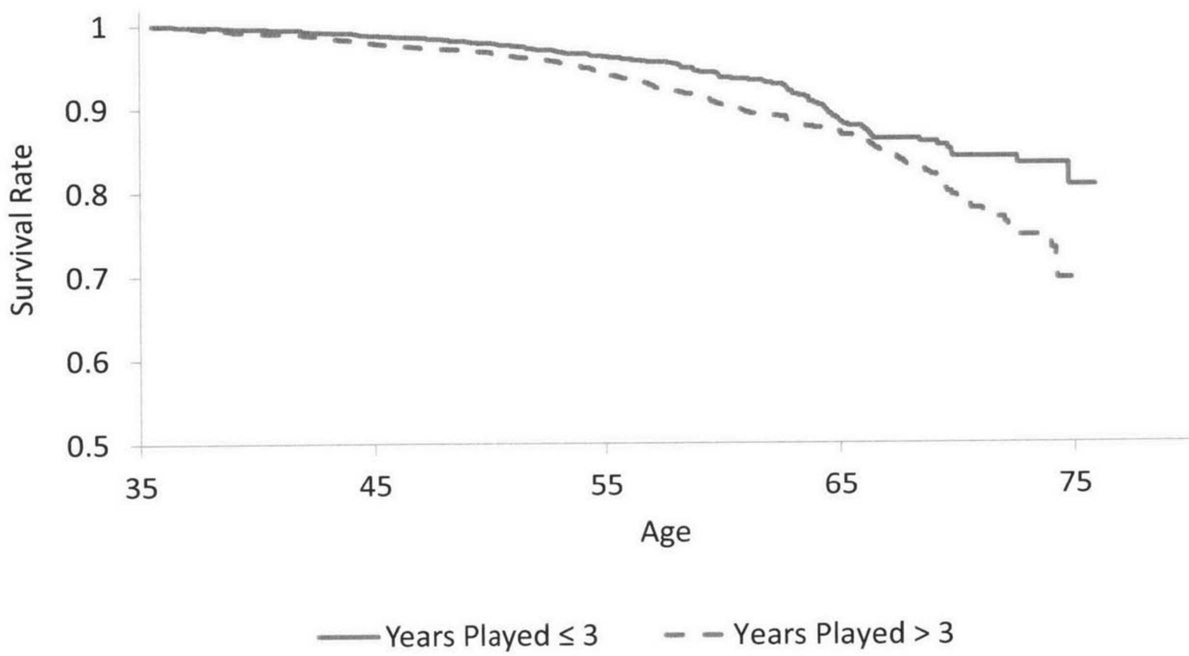


Figure 5. Kaplan-Meier Survival Curves by Career Length, High Injury Risk Positions



References

Abel, E.L., and M.L. Kruger. 2006. "The Healthy Worker Effect in Major League Baseball," Research in Sports Medicine 14(1): 83-87.

Baron, Sherry L., and R. Rinsky. 1994. "Health Hazard Evaluation Report, National Football League Players Mortality Study," Report No. HETA 88-085. Atlanta, GA: Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health.

Baron, Sherry L., Misty J. Hein, Everett Lehman, and Christine M. Gersic. 2012. "Body Mass Index, Playing Position, Race, and the Cardiovascular Mortality of Retired Professional Football Players," The American Journal of Cardiology 109(6): 889-896.

Carpenter, L.M. 1987. "Some Observations on the Healthy Worker Effect," British Journal of Industrial Medicine 44: 289-291.

Chang, Alice Y., Shannon J. FitzGerald, John Cannaday, Song Zhang, Amit Patel, M. Dean Palmer, Gautham P. Reddy, Karen G. Ordovas, Arthur E. Stillman, Warren Janowitz, Nina B. Radford, Arthur J. Roberts, and Benjamin D. Levine. 2009. "Cardiovascular Risk Factors and Coronary Atherosclerosis in Retired National Football League Players," The American Journal of Cardiology 104(6): 805-811.

Edgeworth Economics. 2010. "Dangers of the Game: Injuries in the NFL," Analysis for the NFLPA.

Golightly, Yvonne M., Stephen W. Marshall, Leigh F. Callahan, and Kevin Guskiewicz. 2009. "Early-Onset Arthritis in Retirement," Journal of Physical Activity and Health 6: 638-643.

Guskiewicz, Kevin M., Stephen W. Marshall, Julian Bailes, Michael McCrea, Robert C. Cantu, C. Randolph, and B.D. Jordan. 2005. "Association between Recurrent Concussion and Late-Life Cognitive Impairment in Retired Professional Football Players," Neurosurgery 57(4): 719-726.

Guskiewicz, Kevin M., Stephen W. Marshall, Julian Bailes, Michael McCrea, Herndon P. Harding, Jr., Amy Matthews, Johna Register Mihalik, and Robert C. Cantu. 2007. "Recurrent Concussion and Risk of Depression in Retired Professional Football Players," Medicine and Science in Sports and Exercise 39(6): 903-909.

Lehman, Everett J., Misty J. Hein, Sherry L. Baron, and Christine M. Gersic. 2012. "Neurodegenerative Causes of Death among Retired National Football League Players," Neurology 79: 1-5.

McKee, Ann C., Robert C. Cantu, Christopher J. Nowinski, E. Tessa Hedley-Whyte, Brandon E. Gavett, Andrew E. Budson, Veronica E. Santini, Hyo-Soon Lee, Caroline A. Kubilus, and Robert A. Stern. 2009. "Chronic Traumatic Encephalopathy in Athletes: Progressive Tauopathy following Repetitive Head Injury," Journal of Neuropathy and Experimental Neurology 68(7): 709-735.

McMichael, A.J. 1976. "Standardized Mortality Ratios and the 'Healthy Worker Effect': Scratching Beneath the Surface," Journal of Occupational Medicine 18(3): 165-168.

McMichael, A.J., R. Spirtas, and L.L. Kupper. 1974. "An Epidemiological Study of Mortality within a Cohort of Rubber Workers, 1964-1972," Journal of Occupational Medicine 16(7): 458-464.

Ogden, Cynthia L., Cheryl D. Fryar, Margaret D. Carroll, and Katherine M. Flegal. 2004. "Mean Body Weight, Height, and Body Mass Index, United States 1960-2000," Advance Data from Vital and Health Statistics 347: 1-18.

Saint Onge, Jarron M., Richard G. Rogers, and Patrick M. Krueger. 2008. "Major League Baseball Players' Life Expectancies," Social Science Quarterly 89(3): 817-830.

Stern, Robert A., David O. Riley, Daniel H. Daneshvar, Christopher J. Nowinski, Robert C. Cantu, Ann C. McKee. 2011. "Long-term Consequences of Repetitive Brain Trauma: Chronic Traumatic Encephalopathy," Physical Medicine and Rehabilitation 3(10S2): S460-S467.

Will, George F. 2012. "Football's Growing Killer Problem," New York Post, http://www.nypost.com/p/news/opinion/opedcolumnists/football_growing_killer_problem_1fnQKvhkFse7V6x4tG9iwN.

Chapter 2

When Opportunity Knocks, Who Answers? New Evidence on College Achievement Awards

with

Joshua Angrist
Ford Professor of Economics, MIT

and

Philip Oreopoulos
Professor of Economics, University of Toronto

We evaluate the effects of academic achievement awards for first and second-year college students studying at a Canadian commuter college. The award scheme offered linear cash incentives for course grades above 70. Awards were paid every term. Program participants also had access to peer advising by upperclassmen. Program engagement appears to have been high but overall treatment effects were small. The intervention increased the number of courses graded above 70 and points earned above 70 for second-year students, but generated no significant effect on overall GPA. Results are somewhat stronger for a subsample that correctly reproduced the program rules.

I. Introduction

As college enrollment rates have increased, so too have concerns about rates of college completion. Around 45 percent of United States college students and nearly 25 percent of Canadian college students fail to complete any college program within six years of postsecondary enrollment (Shaienks and Gluszynksi 2007; Shapiro et al. 2012). Those who do finish now take much longer than they used to (Turner 2004; Bound, Lovenheim, and Turner 2010; Babcock and Marks 2011). Delays and dropouts may be both privately and socially costly. Struggling college students often show little evidence of skill improvement (Arum and Roksa 2011). They pay a higher cost in foregone earnings than those who do finish, while losing the benefit of any possible "sheepskin effects" from degree completion. Time on campus is also subsidized at public colleges and universities, so repeated course failures and long completion times are costly for taxpayers. A recent analysis by Harris and Goldrick-Rab (2010) shows steadily declining degree-to-expenditure ratios in American public colleges, a trend generated by falling completion rates as well as increasing sticker prices.

In an effort to boost grades and on-time graduation rates, most universities deploy an array of support services. These efforts reflect a practical response to an important problem, but evidence that academic support services improve outcomes is mixed at best. A randomized trial discussed by Scrivener and Weiss (2009) finds that campus support services generate small improvements in grades and reduce student attrition, but Angrist, Lang, and Oreopoulos (2009) and MacDonald, Bernstein, and Price (2009) find virtually no effect from support services. Part of the problem seems to be that take-up rates for most support services are low. More pro-active programs that facilitate higher take-up and more intensive support have been found to be more successful than relatively passive interventions offering only "service availability" (Scrivener, Sommo, and Collado 2009; Bettinger and Baker 2011).

A parallel effort to boost college achievement and completion looks to financial incentives. Traditional need-based grant aid – which makes up the bulk of North American aid – flows to recipients in a manner that is mostly independent of academic performance, while embedding little incentive for timely degree completion. Merit-based aid, on the other hand, depends on academic achievement. Most merit awards go to top performing students, who can be expected to do reasonably well with or without support. Performance-based awards for students not already on top are a new but rapidly expanding policy development. If successful, such awards may improve academic outcomes, increase the rate of degree completion, and ultimately save both taxpayers and recipients money.

Georgia's Helping Outstanding Pupils Educationally (HOPE) program, introduced in 1993, is a pioneering effort in this direction. Funded by lottery ticket sales, HOPE covers tuition and fees at any Georgia public college or university for students who earned at least a 3.0 high school GPA. Students lose the HOPE scholarship if their college GPA dips below 3.0. Georgia HOPE has been a model for dozens of similar state programs. Accumulating empirical evidence suggests HOPE-like award schemes improve high school achievement (see, for example, Pallais 2009). On the other hand, such programs also appear to reduce recipients' college course loads (Cornwell et al. 2005), increase their automobile consumption (Cornwell and Mustard 2007), and reduce attendance at out-of-state colleges and college quality (Cornwell, Mustard, and Sridhar 2006; Cohodes and Goodman 2013).

Estimates of the effects of HOPE-style programs on college enrollment and completion are mixed. Dynarski (2008) reports large increases in Georgia and Arkansas's college-educated populations a few years after the introduction of HOPE and a similar Arkansas program, while Castleman (2013) estimates that Florida's HOPE-style public university scholarship boosted recipients' in-state public college completion rates. By contrast, recent analyses by Sjoquist and Winters (2012a; 2012b) find no effect when looking at a broader range of state programs with more recent data and updated clustered standard error estimation.

Most research on HOPE-style programs uses observational designs. Among the most credible of the HOPE-style evaluations, Scott-Clayton's (2011) regression discontinuity investigation of West Virginia's Providing Real Opportunities for Maximizing In-State Student Excellence (PROMISE) scholarship generates evidence of substantial increases in four and five-year graduation rates. Importantly, however, this study shows the PROMISE scholarship increased GPAs and credits earned during the first three years of college only, when students faced a minimum GPA requirement to maintain award eligibility. This suggests that the incentive effects of the scholarships are larger than the income effects resulting from greater financial aid.

Incentive experiments and quasi-experimental research designs in European universities have also produced mixed results. Using a regression-discontinuity design, Garibaldi et al. (2012) found that higher tuition induces faster degree completion by Italian women. De Paola, Scoppa, and Nistico (2012) also find substantial positive effects of a randomized financial award for business administration students in southern Italy. On the other hand, randomized evaluations of financial incentives offered to

Dutch university students generated little overall effect (Leuven, Oosterbeek, and van der Klaauw 2010; Leuven, et al. 2011).[18]

In an effort to encourage on-time completion and retention, a few incentive programs target college credits for those already enrolled. In a randomized evaluation managed by MDRC, Barrow et al. (2012) find significant effects on credit accumulation for a subsample of Louisiana community college students enrolled at least half time. Early results from a series of similar randomized evaluations show small but statistically significant increases in cumulative earned credits by the first or second term (Cha and Patel 2010; Miller et al. 2011; Richburg-Hayes, Sommo, and Welbeck 2011). Evaluating a Canadian community college retention program, MacDonald et al. (2009) report significant increases in GPA and retention; this program paid \$750 per semester for those with a GPA above 2.0, who maintained a full load *and* made use of academic services.

Motivated by the wide range of findings to date, we implemented a financial incentive demonstration program that builds on the lessons from earlier work, including ours. Overall academic performance in our study population was poor. Our merit aid therefore rewarded above-average performance for enrolled students. Specifically, the "Opportunity Knocks" (OK) experiment, piloted at a large Canadian commuter university, was designed to explore whether students who qualify for need aid can also be motivated by merit aid, and whether this improved performance would carry over into subsequent years. Incentivizing higher grades in one year could generate better subsequent performance through habit formation or learning by doing, even after incentives disappear.

OK was offered to first- and second-year students who applied for financial aid. Those who signed up were randomly assigned to treatment and control groups. In contrast to earlier programs that primarily rewarded students for achieving GPA thresholds, treated students earned \$100 for *each* class in which they attained a grade of 70 or better and an additional \$20 for each percentage point above 70 percent (roughly the average grade in the control group). A student with a full course load scoring 75 in every course qualified for \$2,000 over the course of the school year (10 × (\$100 + (5 × \$20))). Treated

[18] Randomized trials and quasi-experimental evaluations of financial incentives have been somewhat more encouraging for elementary and secondary students than for college students. Studies showing substantial positive effects on primary or secondary school students include Angrist et al. (2002), Henry and Rubinstein (2002), Kremer, Miguel, and Thornton (2009), Angrist and Lavy (2009), Dearden et al. (2009), Pallais (2009), and Dee (2011). Also in a primary or secondary context, Fryer (2012) reports large effects of aligned parent, teacher, and student incentives and Levitt et al. (2011) demonstrate some response to immediate rewards for test performance. Other recent experimental studies at this level have generated less reason for optimism. See, for example, Bettinger (2012), Rodriguez-Planas (2010), and Fryer (2011), who evaluate an array of award schemes for primary and middle school students in a variety of settings. For a general review of research on financial incentives, see Gneezy, Meier, and Rey-Biel (2011).

students also had the opportunity to interact with randomly assigned peer advisors. These were upper-class students who had been trained to provide advice about study strategies, time management, and university bureaucracy.

OK was developed in view of the findings from our earlier randomized evaluation on a similar campus. The Student Achievement and Retention (STAR) project (Angrist, et al. 2009) offered three interventions, the most successful of which combined financial incentives at widely spaced GPA thresholds with academic support services. OK provided an opportunity for replication and the chance to offer a more intense and perhaps even more successful treatment. By rewarding performance in each class and setting a low bar for the minimum payment, we hoped to make incentives stronger (92 percent of controls earned a grade of 70 percent or above in at least one class). This contrasts with STAR awards, which were paid out to only about 18 percent of eligible students. We opted for a partially linear payout scheme on theoretical grounds (see, for example, Holmstrom and Milgrom 1987).

OK awards were potentially more generous than those offered in STAR; high achievers could earn up to $700 *per class*.[19] The expected OK award among controls was $1,330, while the expected STAR award was only about $400. OK engendered more program engagement than STAR as well: Almost 90 percent of OK participants had some kind of interaction with peer advisors and/or the program website, in contrast with about 50 percent engagement in STAR.

OK had many novel and promising features: linear incentives at the class level, high reward levels, and high program engagement. It's therefore interesting, surprising, and somewhat disappointing that OK had only a modest impact on targeted outcomes. Treated second-year students earned about 13 percent more than expected based on the distribution of control-group grades, suggesting the program had an incentive effect. The strongest effects appear around the $100 award threshold, where completion of payment-qualifying courses increased, especially among students who appeared to understand the program well. OK also increased the number of second-year courses graded above 70 and grade points earned above 70, but these effects were not large enough to generate a significant increase in students' overall GPAs. OK generated no discernible impacts in the year after incentives were removed.

The following section describes the OK campus setting, program rules, and our random-assignment research design. Section III reports descriptive statistics and indicators of program engagement. Section IV discusses the experimental results while Section V reports on participants' impressions of the program as revealed in post-program surveys. The paper concludes in Section VI with

[19] Tuition at this university is around $5,000 per year.

a brief look at how our results fit in with other post-secondary incentive demonstrations. We also discuss possible explanations for differences between the findings reported here and those in our earlier study.

II. Background and Research Design

Motivated by the mixed results for college incentives to date, we developed an intervention meant to build on what we saw as the strongest features of the program discussed in Angrist, et al. (2009). The OK intervention combined incentives with academic support services; a combination of incentives and services appeared to be especially effective in the earlier STAR evaluation, which ran in a similar setting. The services delivered through STAR were more elaborate and expensive, however. STAR included the opportunity to participate in facilitated study groups as well as email-based peer mentoring, while OK services consisted of email-based peer mentoring only. We opted for email because the take-up rate for STAR's facilitated study groups was low. Also, because a number of STAR participants saw the awards as essentially out of reach, OK award rates were designed to be much higher. OK awards were also paid out more frequently, in this case, every term. Unlike STAR, the OK study population consisted only of students that had applied for financial aid prior to the start of the school year. This was partly in response to political constraints but it also seemed likely that aid recipients would be most responsive to the opportunity to earn additional awards.

Opportunity Knocks (OK) was piloted on an Ontario commuter campus affiliated with a large public university. The six-year completion rate on this campus is about 73 percent. There are about 2,500 students in an entering class. In late summer of 2008, we invited 1,056 first years and 1,073 second years to participate in OK. Eligible students are those who had requested financial aid, had an email address, had a high school GPA recorded in the university administrative information system, and who had enrolled for at least 1.5 credits for the upcoming fall term. Invitees who completed the intake survey and gave consent were eligible for random assignment. Of the 1,271 students who completed the survey and were eligible, 400 were treated. Treatment assignment was stratified by year (first and second) and sex, with 100 in each group. Within sex-year cells, assignment was stratified by high school GPA quartile, with 25 in each group (the analysis below controls for strata).

Previous studies have generally rewarded students for completing courses or reaching GPA thresholds (see, for example, Angrist et al. 2009, Cha and Patel 2010). In contrast, OK participants earned \$100 for each class in which they received at least a 70 percent grade, and an additional \$20 for each percentage point above 70.[20] For example, a student who earned a grade of 75 in each of five classes over one semester (five classes constitute a full load) would have received 5 × ($100 + (5 × $20)) = $1,000. We focused on grades near 70 because anything worse is typically seen as unsatisfactory and because awards for lower levels of achievement are likely to be prohibitively expensive (a GPA of at least C- is required for graduation; this translates to a percentage grade in the low 60s). Still, a grade of 70 is attainable for most students in at least one class, and the OK awards schedule provided incentives for above-average performance as well.

The services component of OK assigned treated students to (trained and paid) same-sex peer advisors. Peer advisors were enthusiastic upper-year students or recent graduates with good grades. Each peer advisor covered 50 participants. Advisors emailed advisees once every two to three weeks, whether or not the advisees responded. These emails offered advice on upcoming academic events and workshops and guidance relevant to key periods in the academic calendar, such as midterms and finals. Advisors also provided information about OK scholarships, including reminders of the scholarship calculation and payment schedules. Advisors frequently invited their clients to turn to them for help with any academic or personal issues that seemed relevant to academic success.

III. Descriptive Statistics and Program Response

The data for this study come primarily from the university records containing information on applicants, enrolled students, and course grades. We supplemented this with data from a baseline survey used to identify the population eligible for random assignment, as well as more descriptive focus-group style information collected from a few subjects after the experiment.

Table 1, which presents descriptive statistics, shows that OK participants were mostly college students of traditional age. Control group students had average grades around 82 percent in high school. Less than half of the control group spoke English as a first language, reflecting the relatively high proportion of immigrants on the OK host campus. About half of control group parents graduated from a postsecondary institution (44 percent of mothers and 53 percent of fathers), while nearly 80 percent of parents graduated from high school, a figure comparable to the Canadian average for college student parents. The OK scholarships were within reach for most participants: 92 percent of controls would have received an award under the OK scholarship formula. Table 1 also documents the fact that random assignment successfully balanced the background characteristics of those in the treatment and control groups (as evidenced by insignificant effects in the "Treatment Difference" columns). Although not

[20] Payoffs were doubled and issued in the spring for year-long courses.

documented in the table, student course selection and completion as measured by number of courses, difficulty, or subject area are also well balanced between treatment and control groups for the whole sample and within subgroups (random assignment occurred after students had pre-registered for courses).[21]

The OK intake survey, included in the packet describing the program to those eligible for random assignment, included two questions meant to gauge subjects' understanding of program award rules. The first asked students to calculate the award amount for one class, and the second asked them to calculate the total award amount from five classes. Two-thirds of the students answered the second question correctly (documented in Table 1), and over 80 percent answered the first question correctly. Those who responded incorrectly to either question received a clarification by email. In the program analysis, we look at treatment effects for the entire sample and for those who answered the second assessment question correctly to see if those who understood the scholarship formula also had a stronger program response.

Student involvement with OK was high. This can be seen in Table 2, which shows that about 73 percent of treated students checked their scholarship earnings on the program website. Women were nine points more likely to check than men. Only 38 percent of treated participants sent an email to their assigned peer advisor in the fall, but this number increased to 50 percent in the spring. By year's end, 70 percent had emailed an advisor at least once over the course of the year. First-year students and women were more likely to contact advisors than were second-year students and men. At least 86 percent of treated students made some kind of program contact: they emailed a peer advisor, checked scholarship earnings, or emailed program staff.

Following a presentation of intention-to-treat effects, we discuss two-stage least squares (2SLS) estimates of treatment effects using a dummy indicating any program contact as the endogenous variable. The idea here is that subjects who made no program contact of any kind, and did not even check their scholarship earnings, are unlikely to have been affected by either OK awards or advisor services. In other words, we think of a dummy indicating any contact as a good surrogate for program treatment status. 2SLS estimates treating program contact as an endogenous variable should therefore capture the effect of treatment on the treated for the subpopulation of active program participants (because endogenous compliance is one-sided, the local average treatment effect is the treatment on the treated effect; see Imbens and Angrist, 1994, for details).

---

[21] Attrition was also balanced between treatment and control (about 5 percent of OK participants dropped out during the study), and treatment and control group dropouts have similar characteristics (results are available upon request).

IV. Program Effects

A. Main Findings

A natural starting point for our analysis is a comparison of the amount earned by the experimental group with the earnings that students in the control group would have been entitled to had they been in the program. A large program effect should be reflected in larger-than expected earnings, where expected earnings are measured using the grade distribution in the control sample.[22] Our estimates of earnings and other effects come from regressions like this one:

$$y_{ij} = \alpha_j + \beta T_i + \delta' X_i + \varepsilon_{ij}, \tag{1}$$

where y_{ij} is the outcome for student i in stratum j, the α_j are strata effects, T_i is a treatment assignment indicator, and X_i is a vector of additional controls.[23] Causal effects of the OK program are captured by β. Since treatment is randomly assigned, covariates are unnecessary to reduce omitted variables bias in the estimated treatment effects. Models with covariates may, however, generate more precise estimates.

The OK program had no impact on earnings for first-year men and women, a result that can be seen in columns 1, 4, and 7 of Table 3. On the other hand, there is some evidence of higher-than-expected earnings for second-year treated students, especially second-year men. The estimated effect on second-year men in the spring term, reported in column 5, is a significant 170 dollars. Estimates over the course of the year are about 255 dollars for second-year men and 180 dollars for all second years.[24] Both of these estimates are significant at better than a 10 percent level and amount to 15-20 percent of a standard deviation of hypothetical control group earnings.

Our experimental design stratifies on sex, year of study, and high school GPA, mitigating concerns about mining for significant findings in subgroups. The analysis by sex and class is of substantive interest and a pre-specified feature of our research plan. Still, it's worth noting that under the null hypothesis of no treatment effect for all four sex by class subgroups, the probability that at least one observed full-year effect is significant at the 8 percent level is $1 - 0.92^4 = 0.28$ (assuming no

[22] Ashenfelter and Plant (1990) use a similar hypothetical payment outcome to measure the labor supply effects of exposure to a negative income tax.

[23] Additional controls include parental education, an indicator for English mother tongue, and indicators for students who answered scholarship formula questions correctly.

[24] Restricting the fall and spring samples to be the same as the full-year sample generates effects for the full year equal to the sum of the fall and spring effects. Estimated effects for the full year need not equal the sum (or average) of the two semester effects because the full-year sample differs slightly from the sample for either semester alone.

outcomes correlation across subgroups). The results in Table 3 emerge more strongly, however, when we limit the sample to students who understood the award formula well and are consistent with a marked response in grades around the 70 percent award threshold, as discussed below.

The question of whether the OK program caused more complex distributional shifts in hypothetical earnings is explored in Figure 1, which shows treatment and control earnings distributions in separate panels by sex and year. The only (marginally) significant distributional contrast in the figure is for second-year men (using a Kolmogorov-Smirnov test). On the other hand, the contrast by treatment status for second-year women looks similar to that for men. For both men and women, treatment seems to have shifted second-year earnings from below a level around 1,500 to more than 1,500 dollars. The shift emerges roughly one hundred dollars above mean earnings for controls.

The evidence for an effect on average grades (measured on a 0-100 scale) and GPA is weaker than that for earnings. The grades results appear in Table 4a and the GPA results appear in Table 4b. Average grades for second-year men increased by about 2.5 percentage points in the spring but this estimate is only marginally significant, and it's the only significant result in the table. The corresponding GPA effect amounts to about 0.27 GPA points, an estimate significant at the 5 percent level.[25] Power is not an issue with these comparisons. For the full sample, we are able to reject GPA and grade effects as small as 10 percent of the control standard deviation, meaning that our zeros are quite precise.

The earnings gains documented in Table 3 are necessarily explained by increases in the number of courses graded at least 70 and grade points over 70. Table 5 reports full-year program effects on each of these components of the scholarship award formula. Panel A shows effects on the number of courses in which a student earned a grade of at least 70. Treatment appears to have increased the number of over-70 grades awarded to second-year men by almost a full course. The number of over-70 courses increases by about half a course for all second years. These estimates are reasonably precise. On the other hand, the estimated effects on grade points earned over 70 are not estimated very precisely. The only (marginally) significant point gain is for all second years, an effect of 6.2 percentage points. It's also worth noting, however, that the magnitudes come out such that effects on total earnings are equally distributed between a threshold effect at 70 and awards for points over 70.

OK may have had a weaker effect on grades and GPA than on earnings because students substituted effort from classes with a grade above 70 to classes with a grade below 70. To test this claim and look for additional evidence of effects concentrated around the award threshold, we estimated

[25] GPA is not a linear transformation of average grades, so we expect slight differences in results. Effects on GPA should be more similar to effects on earnings, since GPA also jumps at 70 percent.

treatment effects on indicators for grade>g, where g runs from 60 to 80 (reported in Figure 2; these plots also show the control grade distribution). This investigation uncovers no negative treatment effects on courses above the higher thresholds, suggesting that students generally did not substitute effort from higher- to lower-graded courses.[26]

We found no evidence of an increased likelihood of crossing any threshold for first years. Treatment appears to have increased the likelihood that second-year women earned a grade of 72-74, a series of effects concentrated around the minimum award threshold. Effects concentrated around the threshold may be evidence of strategic grade-seeking behavior on the part of treated students. For example, students who expected a grade around 68 or 69 may have made a special effort (through negotiation or extra work) to clear 70. On the other hand, treatment appears to have boosted the grades of second-year men over a wide interval running from 60-75 percent. This pattern of effects weighs against a negotiation-based view of the incentive response, at least among men.

Although most students appeared to understand the OK program rules and award formula, a non-trivial minority did not. Those who misunderstood the formula linking grades and awards seem less likely to have been motivated by the awards. We therefore report estimates for a sample restricted to participants who correctly applied the OK earnings formula to an example in the baseline survey (information collected before random assignment). Two-thirds of the sample evaluated the example correctly.

Extrapolation from this selected subgroup is necessarily speculative, but if we assume that only those who understand the program change their behavior in response to OK incentives, average causal effects on those who understand program rules provide a measure of "theoretical effectiveness." Specifically, this parameter captures an upper bound for what the program might do when it becomes part of the routine. We'd expect to approach this bound over time, were schemes like OK a regular part of the college landscape. Estimates limited to the correct-responders sample are reported in Table 6.

Estimates for correct responders show larger program effects on earnings than the estimates computed using the full sample. Specifically, earnings gains are estimated to have been 370 for second-year men and 245 for all second years, both significant at the 5 percent level. On the other hand, neither GPA nor grade effects are significantly different from zero. The apparent difference in findings for grades and earnings is explained by the last two rows of Table 6, which reports estimates for the components of the award formula in the restricted sample. These estimates show reasonably clear

[26] Similar analysis on courses graded above thresholds from 80 to 100 percent demonstrates little difference between treatment and control students.

effects on the number of courses above 70 with weaker effects on points earned above. The shift in grades around the 70 percent threshold was apparently inadequate to boost overall GPA by a statistically significant amount.

Given the modest program effects observed during the treatment period, it seems unlikely that OK boosted achievement substantially in the longer-run. This conjecture is confirmed in Table 7, which reports full-sample treatment effects for fall 2009 (the semester after the program ended). The results in Table 7 show marginally significant positive effects on average grades and GPA for first-year women and in the pooled sample of first years (who are second years in the post-treatment period), but these effects are small. The post-program outcomes also offer a specification test for the analysis above, since we would not expect to see threshold effects around 70 percent in the post-program period. There is no evidence of a treatment effect on the number of fall 2009 courses graded at or above 70 percent.[27]

B. Subgroup Differences

The results presented differ by gender and year in school. First years do not appear to have responded to the OK program at all, while treated second years – particularly second-year men – showed some improvement in grades, especially in courses graded over 70. Although we cannot be sure why results differ by sex and class, we hypothesize that first-years did not respond as strongly because many first-year students have not yet developed successful study techniques, raising their costs of grade improvement beyond OK's marginal returns.

The impact range of OK's marginal incentives might also depend on how well students can target their grades. For example, a student with accurate grade knowledge may only respond to the $100 payment at 70 if she has a course currently graded just below 70. A student with inaccurate or imprecise grade knowledge may respond to the $100 payment even if his actual grades are well below or above 70. A possible explanation for the gender difference in our findings is a female advantage in effort targeting in response to the $100 payment. Figure 2 (discussed in detail above) shows localized positive treatment effects for second-year women around 72 to 73 percent, resulting in little effect on grades overall. Treated second-year men, however, increased courses graded above most thresholds from 60 to 75, contributing to stronger overall effects. It also seems likely that more-experienced second years could target grades better than first years, though high improvement costs for first years appear to have overwhelmed the marginal incentives.

[27] Roughly 100 program participants dropped out between the first and second years. Dropout rates were similar in the treatment and control groups.

C. Additional Results

We might expect OK incentives to be more powerful for financially constrained students. But treatment effects come out similar in subgroups defined by expected financial aid and whether students expressed concerns about funding. Effects are somewhat larger in the subsample of students whose parents had not been to college than among those with college-educated parents, but the gap by parents' schooling is not large or precisely estimated.

Effort substitution from easy to hard classes might also explain the small treatment effects. To maximize their award, OK participants should substitute effort from difficult classes to easy classes, where the financial return to effort is higher. However, treatment effects do not vary by class difficulty, as measured by the average class grade among control students (results available upon request). As noted above, course enrollment, difficulty, and completion are also unaffected, and students do not appear to substitute effort to focus solely on the larger incentive at 70 percent.

The effects of program assignment reported in Tables 3 to 7 are diluted by non-compliance, that is, by the fact that some of those assigned to treatment did not really participate in the program because they were unaware of their assignment or uninterested in the program offerings. It's therefore worth estimating the effect of the scholarship and advisor treatment on program participants. The decision to engage with the program is not randomly assigned; this is a choice made by those offered the opportunity to participate. However, we can use the randomly assigned offer of OK treatment as an instrument for program take-up. By virtue of random assignment the OK offer is unrelated to characteristics of eligible students. The OK offer is also highly correlated with participation status: As shown in Table 2, 86 percent of those offered OK were engaged in some way – either through program/advisor contact or through checking scholarship earnings – while no one in the control group had access to OK awards or services. We assume that those with no program engagement were unaware of and therefore unaffected by the OK awards and services. The overall first stage effect of OK offers on participation (awareness) is around 0.88, controlling for strata (see Table 8). Moreover, because no one in the control group participated, 2SLS estimates in this case capture the effect of treatment on the full sample of program participants, as described in Bloom (1984) and Imbens and Angrist (1994). Program participants are a self-selected group, but effects of OK on these students are of interest because they tell us how much achievement was boosted for those who were clearly aware of and responded to program opportunities in some measurable way.[28]

The first stage effect of OK offers on participation rates is between 0.84 and 0.9 in the full sample and between 0.86 and 0.92 in the subsample that appears to have understood OK program rules. The first-stage estimates appear in the first row of each panel in Table 8, which also reports 2SLS estimates of the effect of participation on participants. Adjusting reduced-form offer effects (the estimates of program effects reported in Tables 3-6) for non-compliance necessarily leads to somewhat larger treatment effects, in this case larger by about 10-20 percent.

The most impressive effects in Table 8 are for the number of courses in which students earned a grade above 70. Here, effects on second years in the full sample are on the order of two-thirds of a course, while the gains among those who understood the program well amount to almost a full course (an estimate of 0.91 with a standard error of 0.33, reported at the bottom of column 8). The last column of Table 8 shows a marginally significant effect on the number of courses in which students earned at least 70 among all students who understood the program well (pooling men and women, and first and second years). The effect for all men is also significant at the 5 percent level in this sample, with a marginally significant impact on second-year women. A robust and substantial impact on hypothetical earnings and points above 70 also emerges from the 2SLS estimates in Panel B. At the same time, neither the earnings effects nor the increase in the number of courses graded above 70 translated into higher overall average grades among participants.

V. Student Impressions

The OK sign-up survey asked students to predict their average grades in two scenarios, one as an OK participant and one as a non-participant. To encourage a thoughtful response to this question, we offered those who answered the opportunity to win a $500 prize to be given to the student whose predictions came closest to the mark. About 60 percent predicted the same grade either way and the average predicted effect on grades was about 2.2 points. This is considerably larger than most of the effects reported in Tables 6 and 8. It also seems noteworthy that those who predicted a positive response do not appear to have responded more strongly than those who predicted no effect.

[28] Some students may have been aware of the financial awards, even though they failed to check their earnings or otherwise engage with the program. In this case, the reported first stage effects on participation/awareness will be slightly too small, leading to inflated 2SLS estimates. Also, there is control noncompliance in the sense that control students have access to standard university support services. Therefore, the support services aspect of the OK treatment should be interpreted as a more engaging addition to a similar service, rather than a new program implemented in a vacuum (Heckman et al. 2000).

After the program ended, we asked students who predicted no effect in the intake survey why they had expected this. Of the 226 emails sent to treated participants predicting no effect, only 34 responded. Most of these respondents said they were planning to do as well as possible either way. For example, one said: "Before starting courses, I had already decided that I would do my best. And so, I felt a scholarship would be an added motivation, but fundamentally it came down to my own ability and commitment." Two thought the award was too remote, commenting: "I predicted the program would have no effect because it provides a long-term reward for regular short-term behavior (daily intense studying)." Only three respondents said the incentives were too small. One said OK was "not too catchy and/or something worth dying for." Another mentioned the 70 percent threshold: "I believe the cash reward for each course was not high enough per percentage point above 70 percent. If the cash reward was perhaps 30 or 40 dollars per percent point above 70 percent, I would've worked even harder."

We also surveyed a random sample of 50 students from the treatment group at the end of the school year (May 13, 2009), offering $25 movie gift certificates to those who responded. Among the 30 respondents to this survey, 27 said the scholarships motivated them. Some thought the program was very effective. For example, one respondent commented: "Every time I began to lose interest in a particular course, I would remind myself that I just need to well . . . keep with it; the rewards will be tremendous. A scholarship is one such reward . . . and it sure is helpful, as it lifts a lot of the financial burdens I'm faced with when it comes to paying tuition & other fees." Others saw the program was somewhat effective, as in this comment: "This scholarship did affect my motivation to study at some point . . ." Respondents often cited concerns about tuition and fees as a motivating factor that boosted their interest in OK.

Half of the post-program treated respondents felt the program led them to study more, though some felt their opportunity for more study time was limited. This comment was typical: "The program made me study more, but not much. I usually follow my schedule between work and school. So the amount of time I could have spent on study is somehow limited." Others felt the program helped them focus on schoolwork: "As someone who gets sidetracked easily, I kept it in mind that staying focused would pay off in more than one way, and so yes, it did affect the amount of time I devoted to studying." Another said, "I think what's great about the program is that when you feel like you're beginning to procrastinate, you think about the outcome of this program and want to get back to studying." On the other hand, one second-year student reporting feeling somewhat demoralized by OK: "I did abnormally poor this year compared to my usual standards and it just so happened to coincide with Opportunity Knocks. The money reminder just kind of made me feel 'worse' about myself."

Among those who responded to the post-program follow-up survey, almost all felt the program improved their academic performance. Some appreciated the opportunity to earn scholarships for good but not necessarily outstanding grades: "Personally, I don't find that [the university] offers as many scholarship opportunities as other [universities], so I think it was rewarding to know that my academic performance was acknowledged and rewarded." Some felt they increased performance out of financial concerns: "[E]specially now with the economic downfall, it is extremely difficult to muster up the finances to help pay for tuition without relying on OSAP [financial aid]. I kind of looked at Opportunity Knocks as my employer who gives me more money the better I performed in my studies." One student volunteered the view that the program would have a long-lasting effect on him/her: "The program had significantly improved my grades! And I cannot wait to see what I can accomplish next year."

Everyone we contacted afterwards reported that they received peer advisor e-mails about once or twice a month. All but one of the respondents said the advisor e-mails were helpful. One noted, "I think the advisor made good decisions between sending us important reminders and information without being redundant. It was especially important to receive the e-mails about the scholarship money quickly after marks were sent in." Another said, "I find it very useful that someone was actually helping me through school." All but one respondent felt the program was worth continuing. Virtually everyone seemed grateful for having being selected for OK. One respondent closed with this endorsement: "The OK Program has been an essential part of my student experience, and in many ways crucial to my academic performance. I think that having a peer advisor as opposed to just the regular counselors working in the University is very important. With all the stress that universities cause their students – financially or otherwise, it's really nice to know there is a program like Opportunity Knocks to help students every step of the way." Overall, this feedback leaves us feeling that most treated students were aware of and engaged with OK, and that a large minority expected some benefit. Others who thought the program would have little effect seem to feel this way because they were already anxious to succeed and willing to devote time to their studies.

VI. Summary and Conclusions

The OK program was popular with participants: sign-up rates and program engagement were high, and in follow-up focus group interviews many program participants were enthusiastic about their experiences. This enthusiasm probably reflects the high award rates for OK. It's therefore disappointing that, despite the introduction of substantial awards at almost every relevant level of achievement, overall program effects on achievement were modest. On the plus side, treated second-year students earned more in OK scholarship money than we would have expected based on the control-group grade distribution, increased the number of courses in which they earned a grade of 70, and gained a few grade points above 70. This localized response to the large program incentive to earn a grade of 70 percent did not translate into a substantial boost in overall achievement, though it was noticeably stronger in the subsample of students who appear to have understood the OK award scheme well.

The past decade has seen a growing number of randomized evaluations of pay-for-performance schemes for students at various levels and quasi-experimental studies looking at effects of state-based merit aid programs. Table 9 summarizes studies using randomized designs to look at financial incentives in college and Table 10 lists results from quasi-experimental studies of state-based merit scholarships.[29] A number of randomized evaluations show effects on credits earned in response to incentives for course completion and grade thresholds (see, for example, MacDonald et al. 2009; Cha and Patel 2010; Barrow et al. 2012). These results, along with the findings in Angrist et al. (2009) and those reported here suggest that students react to threshold targets more strongly than to marginal incentives beyond the initial target. Our linear incentive scheme came with a fairly forgiving target required to get payments started, a fact that may have induced a stronger threshold response. The OK program's novel linear incentive of $20 per percentage point provides a lower bound (in this context at least) on the marginal incentive needed to induce substantially higher student effort over a broad range of grades, especially for first years.

We were also surprised when the OK demonstration failed to replicate the strong positive results for women seen in the STAR experiment. Women may have shown a weaker, localized response to OK because they could strategically target and attain the class-specific OK minimum award standard; targeting the GPA award levels in STAR was likely harder. Men did not target their grades as accurately in OK, yet they did not respond to STAR's incentives. Perhaps the STAR GPA awards were simply too uncertain to motivate men, but at minimum, these results suggest that incentive size, targeting ability, and gender effects matter and interact.

Incentives seem to be more effective when combined with academic support services. On balance, however, the picture that emerges from Table 9 and from this study is one of mostly modest effects. In particular, overall GPA seems largely unaffected except in some subgroups, and Angrist et al.

[29] The studies listed in Table 9 use random assignment to evaluate financial incentives for college students. This list is the result of a citation search (that is, citing studies we were previous¬ly aware of), a keyword search (for "experiment, incentives, college") using Google Scholar, and helpful suggestions from anonymous referees. Table 10 was constructed similarly based on studies using difference in differences, regression discontinuity, event study designs to test impacts of state-based merit aid programs on college performance and completion.

(2009) is the only randomized evaluation to date to find college achievement effects persisting into the post-treatment period. Table 10 describes similarly discouraging results from studies of state-based merit aid programs. A few studies have found positive effects, most notably Scott-Clayton's (2011) evaluation of the West Virginia PROMISE. However, other positive results appear weaker in light of updated empirical work (Sjoquist and Winters 2012a, 2012b) and a better understanding of selection effects (Cohodes and Goodman 2013).

One general explanation for the muted effectiveness of merit scholarships may be that poor performing students have trouble developing effective study strategies. For example, Israeli high school students have easy access to test-focused study sessions in public school, a fact that may explain some of the stronger Angrist and Lavy (2009) results on achievement awards for high school girls. Indeed, second-year students may have responded more strongly in our study precisely because they have a better sense for how to improve their grades. Fryer (2011) similarly argues that incentives for learning (in his case, reading books) look more promising than pay for performance on achievement tests. These intriguing results come from elementary and secondary school settings. Investigation of the merits of as-yet-untried recipes combining learning incentives with academic support schemes seems a worthy priority for future research on college achievement.

Our study also indicates that program awareness and understanding could be important aspects of college incentive design. The positive effects of OK, though muted, are concentrated among students who understood the awards formula well. And, about 14 percent of students assigned to treatment did not engage with the program in any way, suggesting that treatment effects on those who were aware of the program were actually larger than the OLS estimates. These two subgroups are not representative, but their responses suggest that simple, high-profile programs may be more successful and that program effects may evolve over time as awareness and understanding increase.

There are potentially unlimited variations of financial incentives alone. In the context of past work, the Opportunity Knocks project suggests that future studies should consider all aspects of design to have a chance at success, including incentive size, targeting ability, excitement and awareness, simplicity, and gender considerations. In particular, students appear driven by large threshold payments that are hard to target. Creative designs such as lottery- or games-based payments or incentives leveraging social support or competition may be able to capitalize on this behavior, improve effort in new ways, and keep costs low.

Table 1. Descriptive Statistics and Covariate Balance by Gender

	Women				Men				All	
	First Years		Second Years		First Years		Second Years			
	Control Mean	Treatment Difference	Control Mean	Treatment Difference	Control Mean	Treatment Difference	Control Mean	Treatment Difference	Control Mean	Treatment Difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Age	18.2	-0.105	19.2	0.011	18.4	0.014	19.2	0.069	18.7	-0.012
	[0.608]	(0.056) *	[0.514]	(0.056)	[0.815]	(0.104)	[0.460]	(0.070)	[0.757]	(0.036)
High school	82.8	0.145	82.4	0.302	82.3	-0.344	82.1	-0.387	82.5	-0.024
grade average	[6.56]	(0.238)	[6.19]	(0.217)	[6.44]	(0.310)	[6.73]	(0.338)	[6.44]	(0.134)
1st language	0.404	0.057	0.426	-0.046	0.479	-0.060	0.333	0.097	0.416	0.009
is English	[0.491]	(0.056)	[0.495]	(0.057)	[0.501]	(0.065)	[0.474]	(0.069)	[0.493]	(0.031)
Mother a college	0.395	0.065	0.477	-0.016	0.479	0.050	0.424	-0.034	0.439	0.020
graduate	[0.490]	(0.056)	[0.500]	(0.058)	[0.501]	(0.065)	[0.497]	(0.070)	[0.496]	(0.031)
Father a college	0.479	0.051	0.581	0.009	0.603	0.047	0.475	0.105	0.532	0.049
graduate	[0.500]	(0.057)	[0.494]	(0.058)	[0.491]	(0.063)	[0.502]	(0.071)	[0.499]	(0.031)
Correctly answered harder	0.616	0.022	0.690	-0.010	0.719	-0.080	0.697	0.002	0.666	-0.014
question on scholarship formula	[0.487]	(0.053)	[0.464]	(0.054)	[0.451]	(0.061)	[0.462]	(0.065)	[0.472]	(0.029)
Controls who would have	0.883		0.968		0.908		0.978		0.923	
earned some scholarship money	[0.322]		[0.177]		[0.289]		[0.148]		[0.266]	
Hypothetical earnings for	1,240		1,390		1,430		1,400		1,330	
controls	[1,220]		[1,090]		[1,230]		[1,270]		[1,190]	
Observations	449		377		246		199		1,271	
F test for joint significance	1.11		0.453		0.858		1.43		0.515	
	{0.355}		{0.843}		{0.525}		{0.198}		{0.797}	

Notes: "Control Mean" columns report averages and standard deviations for variables in the left-most column, within the relevant gender-year subgroup. "Treatment Difference" columns report coefficients from regressions of each variable in the left-most column on a treatment dummy, with sampling strata controls (gender, year in school, and high school grade quartile). The last row presents within-column F tests of joint significance of all treatment differences. Control group standard deviations are in square brackets, robust standard errors are in parentheses, and p values for F tests are in curly braces. Some respondents did not answer the parents' education questions. They are coded as a separate category ("missing") and are not coded as high school or college graduates.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 2. Fraction of Treated Students Making Program-Related Contact by Gender and Year

	Women			Men			All		
Contact Type	First Years (1)	Second Years (2)	All (3)	First Years (4)	Second Years (5)	All (6)	First Years (7)	Second Years (8)	All (9)
Emailed advisor (Fall)	0.450	0.390	0.420	0.410	0.270	0.340	0.430	0.330	0.380
Emailed advisor (Spring)	0.520	0.440	0.480	0.660	0.380	0.520	0.590	0.410	0.500
Emailed advisor (Fall or Spring)	0.790	0.700	0.745	0.750	0.560	0.655	0.770	0.630	0.700
Checked scholarship earnings online	0.760	0.780	0.770	0.650	0.710	0.680	0.705	0.745	0.725
Emailed the program website	0.270	0.320	0.295	0.250	0.300	0.275	0.260	0.310	0.285
Any contact	0.900	0.870	0.885	0.840	0.840	0.840	0.870	0.855	0.863
Observations	100	100	200	100	100	200	200	200	400

Notes: This table shows the proportion making the indicated form of program-related contact.

Table 3. Effects on (Hypothetical) Program Earnings

	Women			Men			All		
	First Years	Second Years	All	First Years	Second Years	All	First Years	Second Years	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				Panel A. Fall					
Control	645	695	667	770	774	760	682	707	693
mean	[657]	[589]	[628]	[670]	[642]	[658]	[663]	[602]	[637]
Treatment	-18.8	99.7	39.9	33.9	49.2	11.9	-5.73	72.0	28.0
effect	(53.1)	(60.9)	(39.9)	(69.8)	(73.1)	(51.3)	(41.9)	(45.9)	(31.1)
N	444	374	818	246	195	441	690	569	1,259
				Panel B. Spring					
Control	589	711	640	644	655	649	605	696	642
mean	[608]	[598]	[606]	[600]	[683]	[633]	[606]	[622]	[614]
Treatment	-57.6	24.7	-19.1	-20.0	170	35.5	-52.5	77.3	4.47
effect	(49.4)	(66.4)	(39.6)	(59.5)	(80.7) **	(49.4)	(37.6)	(51.0)	(30.8)
N	441	340	781	242	183	425	683	523	1,206
				Panel C. Full Year					
Control	1,240	1,390	1,300	1,430	1,400	1,420	1,290	1,390	1,330
mean	[1,220]	[1,090]	[1,170]	[1,230]	[1,270]	[1,240]	[1,230]	[1,140]	[1,190]
Treatment	-80.2	165	33.0	7.01	255	54.8	-64.3	180	41.1
effect	(95.3)	(121)	(74.1)	(121)	(144)*	(95.2)	(74.3)	(91.3) **	(58.2)
N	441	339	780	242	181	423	683	520	1,203

Notes: "Control Mean" rows list averages and standard deviations of program earnings, within the relevant gender-year subgroup. "Treatment Effect" rows report coefficients from regressions of program earnings on a treatment dummy, with sampling strata controls (gender, year in school, and high school grade quartile) and controls for high school grade average, whether students' first language is English, parents' education, and whether students answered questions on program rules correctly. Control earnings are hypothetical; treated earnings are actual. Full year courses are double-weighted in the earnings calculation. The sample used for the full year estimates includes students with grades in fall and spring. The fall analysis omits full year courses. If we restrict the fall and spring samples to be the same as the full year sample, then the effects for the full year are the sum of the fall and spring effects (this is also true in later tables). Robust standard errors are in parentheses; standard deviations are in square brackets.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 4a. Effects on Average Grades

	Women			Men			All		
	First Years	Second Years	All	First Years	Second Years	All	First Years	Second Years	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				Panel A. Fall					
Control	68.1	71.0	69.4	70.7	72.4	71.4	68.9	71.4	70.0
mean	[11.6]	[8.40]	[10.4]	[10.9]	[8.39]	[10.0]	[11.4]	[8.41]	[10.3]
Treatment	0.424	0.420	0.461	0.452	-0.520	-0.496	0.236	0.064	0.076
effect	(0.945)	(0.947)	(0.662)	(1.18)	(1.07)	(0.827)	(0.740)	(0.694)	(0.515)
N	444	374	818	246	195	441	690	569	1,259
				Panel B. Spring					
Control	67.4	71.2	68.9	68.8	70.0	69.3	67.8	70.8	69.0
mean	[11.3]	[9.02]	[10.5]	[11.2]	[10.6]	[10.9]	[11.2]	[9.46]	[10.6]
Treatment	-0.814	-0.118	-0.471	-0.971	2.54	0.106	-0.966	0.727	-0.225
effect	(1.16)	(1.13)	(0.801)	(1.56)	(1.41)*	(1.03)	(0.901)	(0.901)	(0.634)
N	441	340	781	242	183	425	683	523	1,206
				Panel C. Full Year					
Control	67.9	71.1	69.2	69.9	71.5	70.5	68.4	71.2	69.6
mean	[10.7]	[7.77]	[9.69]	[10.3]	[8.59]	[9.70]	[10.6]	[7.99]	[9.70]
Treatment	-0.323	0.470	0.076	-0.233	1.17	-0.146	-0.458	0.614	-0.025
effect	(0.958)	(0.932)	(0.662)	(1.21)	(1.09)	(0.840)	(0.745)	(0.719)	(0.522)
N	441	339	780	242	181	423	683	520	1,203

Notes: "Control Mean" rows list averages and standard deviations of average grades, within the relevant gender-year subgroup. "Treatment Effect" rows report coefficients from regressions of average grades on a treatment dummy, with sampling strata controls (year in school, and high school grade quartile) and controls for high school grade average, whether students' first language is English, parents' education, and whether students answered questions on program rules correctly. Average grades are on a 100 point scale. Full year courses are double-weighted in the average grade calculation. The sample used for the full year estimates includes students with grades in fall and spring. The fall analysis omits full year courses. Robust standard errors are in parentheses; standard deviations are in square brackets.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 4b. Effects on GPA

	Women			Men			All		
	First Years (1)	Second Years (2)	All (3)	First Years (4)	Second Years (5)	All (6)	First Years (7)	Second Years (8)	All (9)
					Panel A. Fall				
Control mean	2.39	2.64	2.50	2.61	2.75	2.66	2.46	2.67	2.55
	[0.982]	[0.765]	[0.900]	[0.920]	[0.743]	[0.856]	[0.968]	[0.760]	[0.890]
Treatment effect	0.021	0.046	0.038	0.046	-0.039	-0.034	0.014	0.015	0.009
	(0.079)	(0.081)	(0.056)	(0.103)	(0.098)	(0.073)	(0.063)	(0.061)	(0.044)
N	444	374	818	246	195	441	690	569	1,259
					Panel B. Spring				
Control mean	2.34	2.64	2.47	2.47	2.54	2.50	2.38	2.61	2.48
	[0.916]	[0.783]	[0.875]	[0.935]	[0.880]	[0.912]	[0.922]	[0.810]	[0.885]
Treatment effect	-0.049	0.018	-0.016	-0.003	0.266	0.071	-0.037	0.102	0.022
	(0.081)	(0.090)	(0.059)	(0.106)	(0.119)**	(0.079)	(0.064)	(0.073)	(0.048)
N	441	340	781	242	183	425	683	523	1,206
					Panel C. Full Year				
Control mean	2.37	2.64	2.49	2.55	2.67	2.59	2.42	2.65	2.52
	[0.895]	[0.689]	[0.825]	[0.870]	[0.739]	[0.822]	[0.890]	[0.702]	[0.825]
Treatment effect	-0.021	0.055	0.018	0.019	0.126	0.021	-0.019	0.075	0.019
	(0.073)	(0.079)	(0.053)	(0.096)	(0.097)	(0.070)	(0.058)	(0.061)	(0.042)
N	441	339	780	242	181	423	683	520	1,203

Notes: "Control Mean" rows list averages and standard deviations of GPA, within the relevant gender-year subgroup. "Treatment Effect" rows report coefficients from regressions of GPA on a treatment dummy, with sampling strata controls (year in school, and high school grade quartile) and controls for high school grade average, whether students' first language is English, parents' education, and whether students answered questions on program rules correctly. GPA is on a four point scale. The sample used for the full year estimates includes students with grades in fall and spring. The fall analysis omits full year courses. Robust standard errors are in parentheses; standard deviations are in square brackets.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 5. Effects on Components of the OK Scholarship Formula

	Women			Men			All		
	First Years	Second Years	All	First Years	Second Years	All	First Years	Second Years	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Panel A. Number of Courses with Grade of At Least 70 Percent								
Control	4.58	5.22	4.85	5.18	5.01	5.11	4.75	5.16	4.92
mean	[3.35]	[2.84]	[3.16]	[3.17]	[2.96]	[3.08]	[3.30]	[2.87]	[3.14]
Treatment	-0.034	0.422	0.185	0.128	0.954	0.338	-0.010	0.572	0.239
effect	(0.260)	(0.335)	(0.205)	(0.356)	(0.405)**	(0.268)	(0.208)	(0.252)**	(0.161)
N	441	339	780	242	181	423	683	520	1,203
	Panel B. Total Grade Percentage Points Over 70 Percent								
Control	38.9	43.3	40.8	45.5	45.0	45.3	40.9	43.8	42.1
mean	[46.2]	[42.1]	[44.5]	[47.4]	[50.4]	[48.5]	[46.6]	[44.4]	[45.7]
Treatment	-3.84	6.16	0.726	-0.290	7.98	1.05	-3.17	6.15	0.861
effect	(3.76)	(4.64)	(2.88)	(4.57)	(5.49)	(3.62)	(2.87)	(3.52)*	(2.25)
N	441	339	780	242	181	423	683	520	1,203

Notes: The dependent variable in Panel A is the total number of courses in which the student received a grade at 70 percent or higher over both semesters. In Panel B, the dependent variable is the sum of the percentage points by which the student's grades exceeded 70 percent. "Control Mean" rows list averages and standard deviations, within the relevant gender-year subgroup. "Treatment Effect" rows report coefficients from regressions on a treatment dummy, with sampling strata controls (gender, year in school, and high school grade quartile) and controls for high school grade average, whether students' first language is English, parents' education, and whether students answered questions on program rules correctly. Full year courses are double-weighted in the calculation of both dependent variables. The sample used to make this table includes students with grades in fall and spring. Robust standard errors are in parentheses; standard deviations are in square brackets.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6. Full Year Effects (Students Who Calculated Awards Correctly)

	Women			Men			All		
	First Years (1)	Second Years (2)	All (3)	First Years (4)	Second Years (5)	All (6)	First Years (7)	Second Years (8)	All (9)
(Hypothetical) Program earnings	-218	219	-9.32	102	370	160	-80.4	245	63.7
	(130)*	(155)	(101)	(144)	(172)**	(111)	(97.2)	(114)**	(74.8)
Average grades	-1.23	0.999	-0.161	0.839	1.73	0.754	-0.351	1.03	0.219
	(1.10)	(1.12)	(0.779)	(1.51)	(1.31)	(1.00)	(0.913)	(0.879)	(0.634)
GPA	-0.107	0.112	-0.002	0.123	0.167	0.103	-0.008	0.117	0.044
	(0.088)	(0.095)	(0.064)	(0.118)	(0.117)	(0.083)	(0.072)	(0.074)	(0.052)
Number of courses with grade of at least 70 percent	-0.339	0.715	0.165	0.429	1.19	0.637	-0.008	0.813	0.353
	(0.333)	(0.410)*	(0.264)	(0.431)	(0.497)**	(0.323)**	(0.265)	(0.309) ***	(0.203)*
Total grade percentage points over 70 percent	-9.21	7.38	-1.29	2.97	12.6	4.82	-3.98	8.19	1.42
	(5.25)*	(5.98)	(3.96)	(5.37)	(6.49)*	(4.19)	(3.81)	(4.37)*	(2.91)
N	441	339	780	242	181	423	683	520	1,203

Notes: "Number of Courses with Grade of At Least 70 Percent" is the total number of courses in which the student received a grade at 70 percent or higher. "Total Grade Percentage Points Over 70 Percent" is the sum of the percentage points by which the student's grades exceeded 70 percent. Each row reports coefficients from regressions of the indicated variable on a treatment dummy, with sampling strata controls (gender, year in school, and high school grade quartile) and controls for high school grade average, whether students' first language is English, parents' education, and whether students answered questions on program rules correctly. Full year courses are double-weighted in the calculation of the dependent variables. The sample used to make this table includes students with grades in fall and spring. Robust standard errors are in parentheses.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 7. Effects in Fall 2009

	Women			Men			All		
	First Years (1)	Second Years (2)	All (3)	First Years (4)	Second Years (5)	All (6)	First Years (7)	Second Years (8)	All (9)
(Hypothetical) Program earnings	7.22	60.0	33.5	77.6	22.8	36.8	22.7	54.1	33.0
	(58.4)	(68.7)	(44.2)	(73.2)	(77.9)	(52.7)	(45.2)	(51.4)	(33.9)
Average grades	1.44	0.344	0.844	1.36	-2.16	-0.448	1.35	-0.618	0.299
	(0.917)	(1.17)	(0.736)	(1.49)	(1.46)	(1.06)	(0.803)*	(0.912)	(0.603)
GPA	0.148	0.019	0.082	0.083	-0.144	-0.037	0.119	-0.041	0.033
	(0.079)*	(0.096)	(0.062)	(0.127)	(0.122)	(0.088)	(0.068)*	(0.074)	(0.050)
Number of courses with grade of at least 70 percent	0.196	0.166	0.180	0.224	0.072	0.127	0.197	0.131	0.145
	(0.163)	(0.184)	(0.121)	(0.226)	(0.230)	(0.162)	(0.132)	(0.141)	(0.096)
Total grade percentage points over 70 percent	-0.620	2.17	0.776	2.76	0.782	1.21	0.152	2.05	0.921
	(2.32)	(2.69)	(1.75)	(2.74)	(3.02)	(1.99)	(1.75)	(2.02)	(1.32)
N	395	334	729	209	165	374	604	499	1,103

Notes: "Number of Courses with Grade of At Least 70 Percent" is the total number of courses in which the student received a grade at 70 percent or higher. "Total Grade Percentage Points Over 70 Percent" is the sum of the percentage points by which the student's grades exceeded 70 percent. Each row reports coefficients from regressions of the indicated variable on a treatment dummy, with sampling strata controls (gender, year in school, and high school grade quartile) and controls for high school grade average, whether students' first language is English, parents' education, and whether students answered questions on program rules correctly. Full year courses are excluded from the calculation of all five dependent variables. "First Year" and "Second Year" continue to refer to the students' standing during the program period. Robust standard errors are in parentheses.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 8. IV Estimates for Participants – Panel A: Full Sample

	Women			Men			All		
	First Years (1)	Second Years (2)	All (3)	First Years (4)	Second Years (5)	All (6)	First Years (7)	Second Years (8)	All (9)
First stage (any contact)	0.901	0.891	0.897	0.844	0.874	0.858	0.876	0.882	0.878
	(0.029) ***	(0.032) ***	(0.022) ***	(0.037) ***	(0.035)***	(0.025)***	(0.023)***	(0.024)***	(0.017)***
Second stages:									
(Hypothetical) Program earnings	-89.0	186	36.8	8.31	292	63.9	-73.4	204	46.8
	(104)	(131)	(81.3)	(139)	(156)*	(108)	(83.6)	(101)**	(65.4)
Average grades	-0.359	0.527	0.084	-0.276	1.34	-0.171	-0.523	0.696	-0.029
	(1.05)	(1.02)	(0.727)	(1.38)	(1.18)	(0.956)	(0.840)	(0.795)	(0.587)
GPA	-0.023	0.062	0.020	0.023	0.144	0.024	-0.022	0.084	0.021
	(0.079)	(0.086)	(0.058)	(0.110)	(0.105)	(0.080)	(0.065)	(0.068)	(0.047)
Number of courses with grade of at least 70 percent	-0.037	0.473	0.206	0.152	1.09	0.394	-0.011	0.648	0.272
	(0.283)	(0.362)	(0.225)	(0.407)	(0.437)**	(0.304)	(0.234)	(0.277)**	(0.180)
Total grade percentage points over 70 percent	-4.27	6.92	0.809	-0.344	9.14	1.22	-3.62	6.97	0.981
	(4.11)	(5.05)	(3.16)	(5.23)	(5.96)	(4.12)	(3.24)	(3.89)*	(2.53)
N	441	339	780	242	181	423	683	520	1,203

Table 8 (continued). IV Estimates for Participants – Panel B: Students Who Calculated Awards Correctly

	Women			Men			All		
	First Years (1)	Second Years (2)	All (3)	First Years (4)	Second Years (5)	All (6)	First Years (7)	Second Years (8)	All (9)
First stage (any contact)	0.922	0.907	0.915	0.863	0.900	0.875	0.896	0.895	0.895
	(0.033) ***	(0.035) ***	(0.024) ***	(0.043) ***	(0.037)***	(0.030)***	(0.027)***	(0.028)***	(0.019)***
Second stages:									
(Hypothetical) Program earnings	-237	241	-10.2	119	411	183	-89.8	274	71.2
	(139)*	(164)	(108)	(158)	(178)**	(123)	(106)	(123)**	(82.0)
Average grades	-1.34	1.10	-0.176	0.972	1.92	0.862	-0.392	1.15	0.245
	(1.16)	(1.19)	(0.835)	(1.66)	(1.35)	(1.10)	(0.997)	(0.950)	(0.696)
GPA	-0.116	0.123	-0.002	0.143	0.186	0.117	-0.008	0.130	0.049
	(0.094)	(0.101)	(0.069)	(0.129)	(0.120)	(0.091)	(0.079)	(0.080)	(0.057)
Number of courses with grade of at least 70 percent	-0.368	0.788	0.181	0.497	1.32	0.729	-0.009	0.908	0.394
	(0.353)	(0.432)*	(0.282)	(0.475)	(0.511)**	(0.356)**	(0.289)	(0.332)***	(0.222)*
Total grade percentage points over 70 percent	-9.99	8.13	-1.41	3.45	14.0	5.51	-4.44	9.15	1.59
	(5.58)*	(6.34)	(4.25)	(5.91)	(6.71)**	(4.62)	(4.16)	(4.73)*	(3.19)
N	274	236	510	166	127	293	440	163	803

Notes: "First Stage (Any Contact)" rows report coefficients from a regression of a dummy variable equal to one if the student made any program-related contact (see Table 2) on a treatment dummy. "Second Stage" rows report coefficients from IV regressions, instrumenting for the program contact dummy with the treatment dummy. All regressions include sampling strata controls (gender, year in school, and high school grade quartile) and controls for high school grade average, whether students' first language is English, parents' education, and whether students answered questions on program rules correctly. Full year courses are double-weighted in the calculation of second stage dependent variables. The sample used for this table includes students with grades in fall and spring. Standard errors are in parentheses.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 9. Randomized Evaluations of College Achievement Awards

Study (1)	Sample (2)	Treatment (3)	Outcome (4)	Effects: All (5)	Effects: Men (6)	Effects: Women (7)
1. Angrist, Lang, and Oreopoulos (2009) [The Student Achievement and Retention Project]	First year students at Canadian commuter university in 2005-2006, except for top HS grade quartile	$1,000 for C+ to B- first year performance, $5,000 for B+ to A performance (varies by HS grade)	First year GPA	0.010 (0.066) [1.805]	-0.110 (0.103) [1.908]	0.086 (0.084) [1.728]
			First year credits earned	-0.012 (0.064) [2.363]	-0.157 (0.106) [2.453]	0.084 (0.082) [2.298]
			Second year GPA	-0.018 (0.066) [2.040]	-0.081 (0.108) [2.084]	0.030 (0.085) [2.008]
			Second year credits earned	0.027 (0.108) [2.492]	0.155 (0.180) [2.468]	-0.024 (0.137) [2.509]
		Incentives and support services	First year GPA	0.210 (0.092)** [1.805]	0.084 (0.162) [1.908]	0.267 (0.117)** [1.728]
			First year credits earned	0.092 (0.087) [2.363]	-0.196 (0.015) [2.453]	0.269 (0.108)** [2.298]
			Second year GPA	0.072 (0.091) [2.040]	-0.170 (0.161) [2.084]	0.276 (0.106)*** [2.008]
			Second year credits earned	0.072 (0.130) [2.492]	-0.240 (0.206) [2.468]	0.280 (0.172) [2.509]

Table 9 (continued). Randomized Evaluations of College Achievement Awards

Study	Sample	Treatment	Outcome	Effects		
				All	Men	Women
(1)	(2)	(3)	(4)	(5)	(6)	(7)
2. Angrist, Oreopoulos, and Williams (2013) [Opportunity Knocks]	First year students on financial aid at Canadian commuter university in 2008-2009	Over two semesters and for each semester-long course, $100 for attaining at least 70 percent and $20 for each percentage point higher than this (full course load = 10 semester courses)	First year GPA	-0.019 (0.058) [2.42]	0.019 (0.096) [2.55]	-0.021 (0.073) [2.37]
			GPA, fall term of year after program	0.119 (0.068)* [2.60]	0.083 (0.127) [2.58]	0.148 (0.079)* [2.61]
	Second year students on financial aid at Canadian commuter University in 2008-2009		First year GPA	0.075 (0.061) [2.65]	0.126 (0.097) [2.67]	0.055 (0.079) [2.64]
			GPA, fall term of year after program	-0.014 (0.026) [2.83]	-0.144 (0.122) [2.79]	0.019 (0.096) [2.85]
3. Barrow et al. (2010) [Opening Doors Louisiana]	Low-income parents beginning community college in Louisiana between 2004 and 2005	For each of two semesters, $250 for at least half-time enrollment, $250 for C average or better at end of midterms, and $500 for maintaining a C average, plus optional Enhanced college counseling	First year GPA	0.068 (0.104) [2.171]		
			First year credits earned	3.345 (0.849)*** [7.623]	sample is mostly female	
			Registered one year after program year	0.053 (0.038)		
4. Cha and Patel (2010) [Ohio Performance-Based Scholarship Demonstration]	Low-income Ohio college students in 2008 with children and eligible for TANF	$1,800 for earning a grade of C or better in 12 or more credits, or $900 for a C or better in 6 to 11 credits, with payments at end of each semester	First year credits attempted	0.5 (0.8) [19.5]	sample is mostly female	
			First year credits earned	2.0 (0.5)*** [13.4]		

Table 9 (continued). Randomized Evaluations of College Achievement Awards

Study (1)	Sample (2)	Treatment (3)	Outcome (4)	Effects: All (5)	Effects: Men (6)	Effects: Women (7)
5. De Paola, Scoppa, and Nistico (2010)	First year business students at the University of Calabria in 2008-2009	$1,000 for students with the 30 highest cumulative scores on all exams	Cumulative exam score	6.023 (3.059)**	5.390 (4.615)	5.841 (4.061)
			Credits earned	2.335 (1.197)**	1.759 (1.854)	2.490 (1.518)*
		$350 for students with the 30 highest cumulative scores on all exams	Cumulative exam score	5.350 (3.164)*	2.354 (4.877)	6.157 (4.207)
			Credits earned	2.194 (1.266)*	0.714 (1.970)	2.766 (1.655)*
6. Leuven, Oosterbeek, and van der Klaauw (2010)	First year economics and business students at the University of Amsterdam in 2001-2002	$600 for completion of all first year requirements by start of new academic year	Met first year requirements	0.046 (0.065) [0.195]	not reported	
			Total "credit points" in first three years	-1.2 (9.8) [84.3]		
		$200 for completion of all first year requirements by start of new academic year	Met first year requirements	0.007 (0.062) [0.195]		
			Total "credit points" in first three years	-2.5 (9.6) [84.3]		

Table 9 (continued). Randomized Evaluations of College Achievement Awards

				Effects		
Study (1)	Sample (2)	Treatment (3)	Outcome (4)	All (5)	Men (6)	Women (7)
7. Leuven et al. (2011)	First year economics and business students at the University of Amsterdam in 2004-2005 and 2005-2006	\$1,250 for the student with the top microeconomics exam score	Microeconomics exam score	0.974 (0.877) [18.7]	not reported	
		\$3,750 for the student with the top microeconomics exam score	Microeconomics exam score	1.184 (0.617)* [18.9]		
		\$6,250 for the student with the top microeconomics exam score	Microeconomics exam score	-0.629 (0.644) [21.2]		
8. MacDonald et al. (2009) [Foundations for Success]	At-risk students beginning community college in Ontario, Canada, between 2007 and 2008	\$750 each of three semesters for 1) obtaining 2.0 GPA or higher, 2) eligible to continue in a full program the following semester, and 3) completing at least 12 hours of tutorial, case management, or career workshops	First semester GPA during program (missing imputed)	0.08 $p>0.1$ [2.11]	not reported	0.12 $p>0.1$ [2.20]
			Second semester GPA during program (missing imputed)	0.12 $p<0.05$** [1.88]		0.14 $p<0.05$** [2.04]
			Third semester GPA during program (missing imputed)	0.01 $p>0.1$ [2.01]		0.12 $p<0.05$** [2.16]
			Registered in fourth semester (after program)	0.02 $p>0.1$ [0.557]		0.014 $p>0.1$ [0.58]
9. Miller et al. (2011) [New Mexico Performance-Based Scholarship Demonstration]	Low-income students starting at the University of New Mexico in fall, 2008, and fall, 2009	\$1,000 each of four semesters for 1) obtaining 2.0 GPA or higher, 2) enrolling full time, and 3) completing two extra advisor meetings per semester	First semester credits earned	0.0 (0.2) [12.8]	not reported	
			Second semester credits earned	0.6 (0.3)* [11.1]		

Table 9 (continued). Randomized Evaluations of College Achievement Awards

Study (1)	Sample (2)	Treatment (3)	Outcome (4)	Effects: All (5)	Effects: Men (6)	Effects: Women (7)
10. Richburg-Hayes Sommo, and Welbeck (2011) [New York Performance-Based Scholarship Demonstration]	New York City community college students aged 22-35 who required remediation fall, 2008, through fall, 2009	Up to $1,300 each of two or three semesters, paid in installments for achieving 1) registration, 2) continued mid-semester enrollment, and 3) 2.0 GPA in at least six credits	First semester credits earned	0.6 (0.3)* [8.1]	not reported	
			Second semester credits earned	0.6 (0.4) [9.3]		

Notes: The table reports main baseline sample outcomes for grades, credits earned, and measures of persistence. Standard errors are shown in parentheses. Control means are shown in square brackets. See text for sources and more details

* significant at 10% level. ** significant at 5% level. *** significant at 1% level

Table 10. Quasi-Randomized Evaluations of Merit-Based College Scholarships

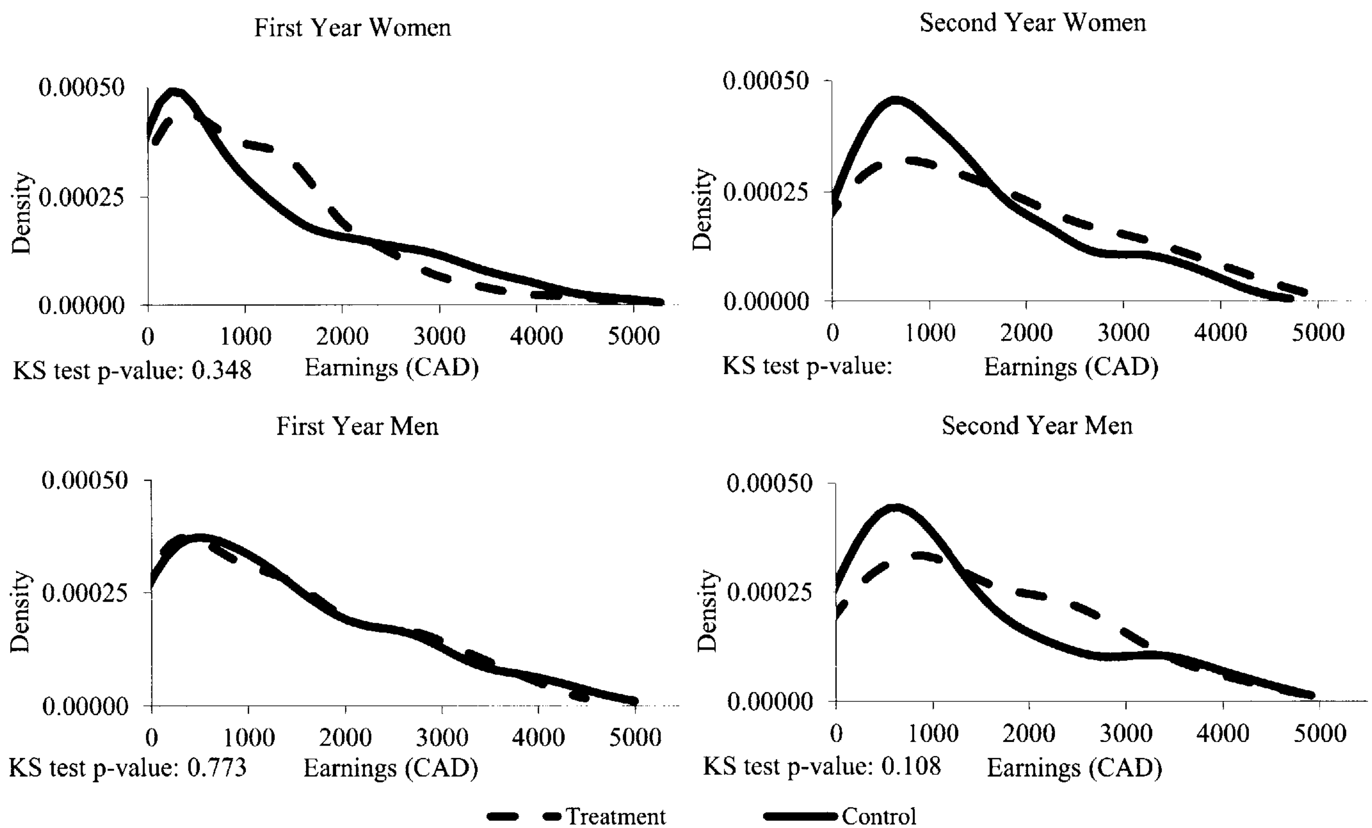
	Study (1)	Treatment (2)	Methodology (3)	Outcome (4)	Effects (5)
1.	Castleman (2013) [Bright Futures Scholarship, Florida Medallion Scholars (FMS) and Florida Academic Scholars (FAS)]	FMS: 75 percent of public college tuition and fees for students with a 3.0 high school GPA and at least 20 on the ACT or 970 on the SAT	Difference in differences, non-eligible students as controls	FL public college credits, four years	-0.634 (1.844)
				FL public college BA in four years	0.011 (0.019)
		FAS: 100 percent of public college tuition and fees for students with a 3.5 high school GPA and at least 28 on the ACT or 1270 on the SAT	Difference in differences, non-FMS students as controls	FL public college credits, four years	8.466 (1.744)***
				FL public college BA in four years	0.058 (0.021)***
2.	Cohodes and Goodman (2013) [John and Abigail Adams Scholarship Program (MA)]	MA public school tuition waived (excluding substantial fees) for students who score in the top 25th percentile of their school district and attain minimum absolute benchmarks on the statewide tenth grade test; must maintain 3.0 GPA in college	Regression discontinuity on tenth grade test score	Enrolled on time at a four-year college	0.018 (0.005)*** [0.710]
				Graduated in four years from a four-year college	-0.017 (0.004)*** [0.418]
3.	Cornwell, Lee, Mustard (2005) [Georgia HOPE]	Full tuition/fees at GA public colleges for students with a 3.0 high school GPA; must maintain 3.0 GPA in college	Differences in differences, non-GA-resident students as controls	Enrolled in full freshman course load at University of Georgia	-0.042 (0.016)*** [0.812]
				Completed full freshman course load at University of Georgia	-0.060 (0.019)*** [0.588]
4.	Dynarski (2008) [Georgia HOPE and Arkansas merit aid program]	$1,000 at inception (now $2,500) for tuition/fees at AR colleges for students with at least 19 on the ACT and a 2.5 core high school GPA; full tuition/fees at GA public colleges for students with a 3.0 high school GPA; for AR and GA, must maintain 3.0 GPA in college	Difference in differences, other state populations as controls	Fraction of age 22-34 population with a college degree	0.0298 (0.004)*** [0.337]

Table 10. Quasi-Randomized Evaluations of Merit-Based College Scholarships

Study (1)	Treatment (2)	Methodology (3)	Outcome (4)	Effects (5)
5. Scott-Clayton (2011) [West Virginia PROMISE]	Full tuition/fees at WV public colleges for students with a 3.0 overall and core high school GPA and at least 21 on the ACT or 1000 on the SAT	Regression discontinuity on ACT score	Cumulative four-year GPA for WV public college students	0.099 (0.045)**
			Earned BA in four years	0.094 (0.022)**
		Event study, program introduction (small sample T-distribution critical values)	Cumulative four-year GPA for WV public college students	0.039 (0.018)
			Earned BA in four years	0.067 (0.005)***
6. Sjoquist and Winters (2012a) [Georgia HOPE and Arkansas merit aid program]	$1,000 at inception (now $2,500) for tuition/ fees at AR colleges for students with at least 19 on the ACT and a 2.5 core high school GPA; full tuition/fees at GA public colleges for students with a 3.0 high school GPA; for AR and GA, must maintain 3.0 GPA in College	Difference in differences, other state populations as controls; increased sample and updated clustering compared with Dynarski (2008)	Fraction of age 22-34 population with a college degree	0.0091 *p*=0.216 [0.3567]
7. Sjoquist and Winters (2012a)	25 state merit aid programs with requirements on high school GPA, ACT/SAT scores, and college credit enrollment and GPA	Difference in differences, non-merit state populations as controls	Fraction of age 24-30 population with a college degree	0.0009 (0.0031) [0.388]
	Nine strongest state merit aid programs with requirements on high school GPA, ACT/SAT scores, and college credit enrollment and GPA		Fraction of age 24-30 population with a college degree	0.0011 (0.0037) [0.388]

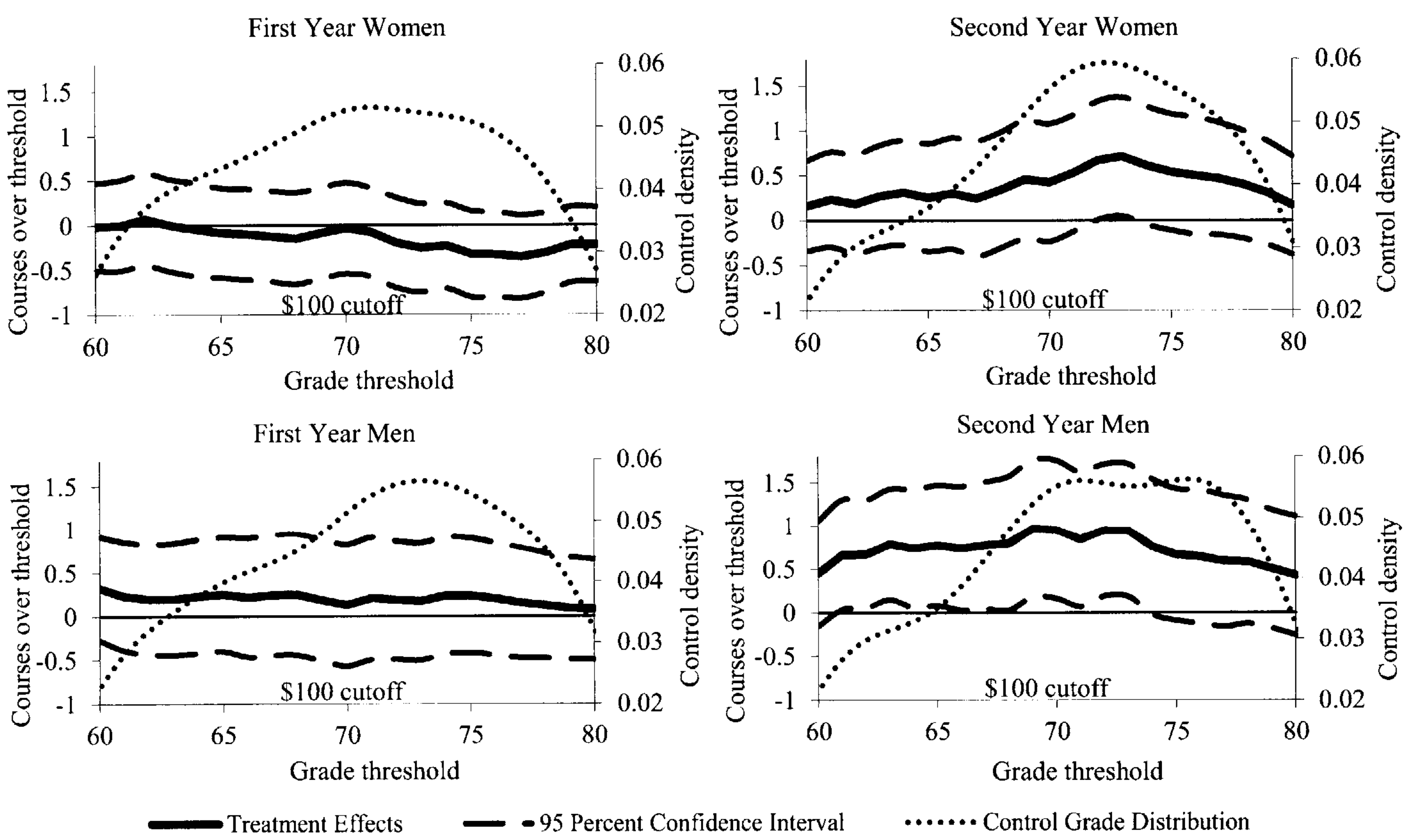
Notes: The table reports main baseline sample outcomes for grades and measures of persistence. Standard errors are shown in parentheses. Control means are shown in square brackets. See text for sources and more details.
* significant at 10% level. ** significant at 5% level. *** significant at 1% level

Figure 1. Densities of Full Year (Hypothetical) Earnings



Note: The figure plots the smoothed kernel densities of OK program earnings for the full year from fall 2008 through spring 2009. Control earnings are hypothetical; treated earnings are actual. Full-year courses are double-weighted in the earnings calculation. The sample used to make this figure includes students with grades in fall and spring.

Figure 2. Full Year Effects on Number of Courses over Grade Thresholds



Note: The figure shows treatment effects on the number of courses in which students earned a grade at or above a given threshold, where the thresholds are plotted on the *x*-axis. Control densities are kernel density plots of grades at the course level using a normal kernel, taking only grades between 60 and 80 percent (inclusive). Treatment effects were estimated using the same models as for Table 3.

References

Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer. 2002. "Vouchers for Private Schooling in Columbia: Evidence from a Randomized Natural Experiment." *American Economic Review* 92(5): 1535-58.

Angrist, Joshua, Daniel Lang, and Philip Oreopoulos. 2009. "Incentives and Services for College Achievement: Evidence from a Randomized Trial." *American Economic Journal: Applied Economics* 1(1): 136-63.

Angrist, Joshua D., and Victor Lavy. 2009. "The Effect of High School Matriculation Awards: Evidence from a Randomized Trial." *American Economic Review* 99(4): 1384-414.

Arum, Richard and Josipa Roksa. 2011. *Academically Adrift: Limited Learning on College Campuses*. Chicago, IL: University of Chicago Press.

Ashenfelter, Orley, and Mark W. Plant. 1990. "Nonparametric Estimates of the Labor-Supply Effects of Negative Income Tax Programs." *Journal of Labor Economics* 8(1, Part 2: Essays in Honor of Albert Rees): S396-415.

Babcock, Philip, and Mindy Marks. 2011. "The Failing Time Cost of College: Evidence from Half a Century of Time Use Data." *Review of Economics and Statistics* 93(2): 468-78.

Barrow, Lisa, Lashawn Richburg-Hayes, Cecilia Elena Rouse, and Thomas Brock. 2012. "Paying for Performance: The Educational Impacts of a Community College Scholarship Program for Low-Income Adults. Working Paper, Federal Reserve Bank of Chicago.

Bettinger, Eric. 2012. "Paying to Learn: The Effect of Financial Incentives on Elementary School Test Scores." *Review of Economics and Statistics* 94(3): 686-98.

Bettinger, Eric, and Rachel Baker. 2011. "The Effects of Student Coaching in College: An Evaluation of a Randomized Experiment in Student Mentoring." NBER Working Paper No. 16881. Cambridge, MA: National Bureau of Economic Research.

Bloom, Howard S. 1984. "Accounting for No-Shows in Experimental Evaluation Designs." *Evaluation Review* 8(2): 225-46.

Bound, John, Michael F. Lovenheim, and Sarah Turner. 2010. "Why Have College Completion Rates Declined? An Analysis of Changing Student Preparation and Collegiate Resources." *American Economics Journal: Applied Economics* 2(3): 129-57.

Castleman, Benjamin L. 2013. "All or Nothing: The Impact of Partial vs. Full Merit Scholarships on College Entry and Success." Working Paper, Harvard University.

Cha, Paulette, and Reshma Patel. 2010. "Rewarding Progress, Reducing Debt: Early Results from Ohio's Performance-Based Scholarship Demonstration for Low-Income Parents." MDRC Report, October. New York and Oakland: MDRC.

Cohodes, Sarah, and Joshua Goodman. 2013. "Merit Aid, College Quality and College Completion: Massachusetts' Adams Scholarship as an In-Kind Subsidy." Working Paper, Harvard University.

Cornwell, Christopher, Kyung Hee Lee, and David B. Mustard. 2005. "Student Responses to Merit Scholarship Retention Rules." *Journal of Human Resources* 40(4): 895-917.

Cornwell, Christopher, and David B. Mustard. 2007. "Merit-Based College Scholarships and Car Sales." *Education Finance and Policy* 2(2): 133-51.

Cornwell, Christopher, David B. Mustard, and Deepa J. Sridhar. 2006. "The Enrollment Effects of Merit-Based Financial Aid: Evidence from Georgia's HOPE Program." *Journal of Labor Economics* 24(4): 761-86.

De Paola, Maria, Vincenzo Scoppa, and Rosanna Nistico. 2012. "Monetary Incentives and Student Achievement in a Depressed Labor Market: Results from a Randomized Experiment." *Journal of Human Capital* 6(1): 56-85.

Dearden, Lorraine, Carl Emmerson, Christine Frayne, and Costas Meghir. 2009. "Conditional Cash Transfers and School Dropout Rates." *Journal of Human Resources* 44(4): 827-57.

Dee, Thomas. 2011. "Conditional Cash Penalties in Education: Evidence from the Learnfare Experiment." *Economics of Education Review* 30(5): 924-37.

Dynarski, Susan. 2008. "Building the Stock of College-Educated Labor." *Journal of Human .Resources* 43(3): 576-610.

Fryer, Roland G., Jr. 2011. "Financial Incentives and Student Achievement: Evidence from Randomized Trials." *Quarterly Journal of Economics* 126(4): 1755-98.

__________. 2012. "Aligning Student, Parent and Teacher Incentives: Evidence from Houston Public Schools." Harvard University, Working Paper.

Garibaldi, Pietro, Francesco Giavazzi, Andrea Ichino, and Enrico Retorre. 2012. "College Cost and Time to Complete a Degree: Evidence from Tuition Discontinuities." *Review of Economics and Statistics* 94(3): 699-711.

Gneezy, Uri, Stephan Meier, and Pedro Rey-Biel. 2011. "When and Why Incentives (Don't) Work to Modify Behavior." *Journal of Economic Perspectives* 25(4): 191-210.

Harris, Douglas N., and Sara Goldrick-Rab. 2010. "The (Un)Productivity of American Higher Education: From 'Cost Disease' to Cost-Effectiveness." University of Wisconsin, Working Paper.

Henry, Gary T., and Ross Rubinstein. 2002. "Paying for Grades: Impact of Merit-Based Financial Aid on Education Quality." *Journal of Policy Analysis and Management* 21(1): 93-109.

Heckman, James, Neil Hohmann, Jeffrey Smith, and Michael Khoo. 2000. "Substitution and Dropout Bias in Social Experiments: A Study of an Influential Social Experiment." *Quarterly Journal of Economics* 115(2): 651-94.

Holmstrom, Bengt, and Paul Milgrom. 1987. "Aggregation and Linearity in the Provision of Intertemporal Incentives." *Econometrica* 55(2): 303-28.

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62(2): 467-75.

Kremer, Michael, Edward Miguel, and Rebecca Thornton. 2009. "Incentives to Learn." *The Review of Economics and Statistics* 91(3): 437-56.

Leuven, Edwin, Hessel Oosterbeek, and Bas van der Klaauw. 2010. "The Effect of Financial Rewards on Students' Achievements: Evidence from a Randomized Experiment." *Journal of the European Economic Association* 8(6): 1243-65.

Leuven, Edwin, Hessel Oosterbeek, Joep Sonnemans, and Bas van der Klaauw. 2011. "Incentives Versus Sorting in Tournaments: Evidence from a Field Experiment." *Journal of Labor Economics* 29(3): 637-58.

Levitt, Steven D., John A. List, Susanne Neckermann, and Sally Sadoff. 2011. "The Impact of Short-Term Incentives on Student Performance." University of Chicago, Working Paper.

MacDonald, Heather, Lawrence Bernstein, and Cristofer Price. 2009. "Foundations for Success: Short-Term Impacts Report." Report to the Canada Millennium Scholarship Foundation. Ottawa, Canada: Canada Millennium Scholarship Foundation.

Miller, Cynthia, Melissa Binder, Vanessa Harris, and Kate Krause. 2011. "Staying on Track: Early Findings from a Performance-Based Scholarship Program at the University of Mexico." MDRC Report, August. New York and Oakland: MDRC.

Pallais, Amanda. 2009. "Taking a Chance on College: Is the Tennessee Education Lottery Scholarship Program a Winner?" *The Journal of Human Resources* 44(1): 199-222.

Richburg-Hayes, Lashawn, Colleen Sommo, Rashida Welbeck. 2011. "Promoting Full-Time Attendance among Adults in Community College: Early Impacts from the Performance-Based Scholarship Demonstration in New York." MDRC Report, May. New York and Oakland: MDRC.

Rodriguez-Planas, Nuria. 2010. "Longer-Term Impacts of Mentoring, Educational Incentives, and Incentives to Learn: Evidence from a Randomized Trial." IZA Discussion Paper No. 4754. Bonn, Germany: Institute for the Study of Labor.

Scrivener, Susan, Colleen Sommo, and Herbert Collado. 2009. "Getting Back on Track: Effects of a Community College Program for Probationary Students." MDRC Report, April.

Scrivener, Susan, and Michael J. Weiss. 2009. "More Guidance, Better Results? Three-Year Effects of an Enhanced Student Services Program at Two Community Colleges." MDRC report, August. New York and Oakland: MDRC.

Scott-Clayton, Judith. "On Money and Motivation: A Quasi-Experimental Analysis of Financial Incentives for College Achievement." *Journal of Human Resources* 46(3): 614-46.

Shaienks, Danielle, and Tomasz Gluszynski. 2007. "Participation in Postsecondary Education: Graduates, Continuers, and Dropouts, Results from YITS Cycle 4." Research Paper, Culture, Tourism and the Centre for Education Statistics. Ottawa, Canada: Statistics Canada.

Shapiro, Doug, Afet Dundar, Jin Chen, Mary Ziskin, Eunkyong Park, Vasti Torres, Yi-Chen Chiang. 2012. "Completing College: A National View of Student Attainment Rates." Signature Report 4, National Student Clearinghouse Research Center. Herndon, VA: National Student Clearinghouse.

Sjoquist, David L., and John V. Winters. 2012a. "Building the Stock of College-Educated Labor Revisited." *Journal of Human Resources* 47(1): 270-85.

__________. 2012b. "State Merit-Based Financial Aid Progams and College Attainment," IZA Discussion Paper No. 6801. Bonn, Germany: Institute for the Study of Labor.

Turner, Sarah E. 2004. "Going to College and Finishing College: Explaining Different Educational Outcomes." In *College choices: The Economics of Where to Go, When to Go, and How to Pay for It*, ed. Catherine M. Hoxby, 13-56. Chicago, IL: University of Chicago Press.

Chapter 3
To Tank or Not to Tank? Evidence from the National Basketball Association

with

Christopher Walters
Assistant Professor of Economics, University of California, Berkeley

High draft picks are a coveted commodity in the NBA. Teams are often accused of losing intentionally ("tanking") in order to increase their odds of receiving a favorable pick in the league's draft lottery. This paper answers two questions related to tanking in the NBA: (1) What is the value of receiving the first draft pick?, and (2) Do teams lose intentionally to secure higher draft positions? We answer the first question by adjusting for the probability of winning the lottery using a propensity score methodology. The estimates indicate that winning the draft lottery increases attendance by 6 percentage points during the five-year period following the draft. Receiving the first pick is also associated with a small increase in win percentage, though this effect is less precisely estimated. To answer the second question, we analyze games played by non-playoff teams near the end of the season. Using a fixed-effects methodology that compares games in which a team can potentially change its lottery odds to games at the end of the season in which these odds are fixed, we find evidence of tanking. Since 1968, playoff-eliminated teams have seen around a 5 percentage point increase in win percentage once their lottery odds are fixed. This difference has ballooned above 10 percentage points in more recent years.

I. Introduction

In the National Basketball Association (NBA), the selection order in the annual amateur draft is determined via lottery. Weaker teams receive more weight in the lottery, but all non-playoff teams have at least some chance to receive the first overall pick. Draft picks are a coveted commodity in NBA front offices, and NBA teams are sometimes accused of "tanking" (that is, intentionally losing games) late in the season in order to increase their chances in the lottery. In an April 2007 article titled "Tanks for Nothing, NBA," ESPN's Bill Simmons chronicles a late-season game between the Milwaukee Bucks and the Boston Celtics and argues that both teams were "desperate . . . to blow the game for lottery position." The resulting game was a "stink bomb" in which "every paying customer lost" (Simmons, 2007).

This paper explores two questions related to tanking in the NBA. First, we ask whether receiving a high draft pick actually has economic and competitive value. Despite the perceived benefits of a favorable lottery position, there is substantial uncertainty associated with the performance of top amateur prospects. Table 1 summarizes the careers of the players selected with the first overall pick in the NBA drafts held from 1985 to 2010. Some first picks, like Tim Duncan and Shaquille O'Neal, became superstars and won multiple league championships; others, like Michael Olowokandi and Kwame Brown, failed to qualify for a single NBA all-star team. Since the NBA's rookie salary scale requires teams to make significant financial commitments to high draft picks regardless of their performance, the true value of owning a high draft pick is theoretically unclear.

The first step of our analysis uses propensity score re-weighting to estimate the causal effects of winning the NBA draft lottery. Since bad teams are more likely to win the lottery, simple comparisons of subsequent outcomes for winners and losers would not capture true causal effects. However, the draft lottery setting is favorable, because the probability of winning the lottery for each team (the propensity score) is known *ex ante.* After adjusting for this known probability of assignment, the lottery outcome is random, and comparisons of winners and losers are unbiased. Our estimates are based on two weighting schemes that capture causal effects for different segments of the population. First, we weight observations to reflect the distribution of characteristics among lottery winners, so as to estimate the effect of treatment on the treated (TOT). Second, we weight observations to match the full population of teams, so as to estimate the average treatment effect (ATE). We use this methodology to estimate the effect of receiving the first pick on attendance and win percentage. The results indicate that receiving the first pick leads to a TOT attendance increase of about 6 percent in the five years following the lottery and a similar but shorter-lived ATE, although the ATE is less precisely estimated.

Our second question is whether teams recognize these benefits and tank to increase their odds of receiving a high draft pick. To answer this question, we focus on the performance of non-playoff teams near the end of the season. We use a fixed-effects methodology that compares games in which a team's lottery odds are already determined to games in which these odds would be improved by a loss. The results show that teams do tank in order to improve their draft prospects. Between 1968 and 2009, playoff-eliminated teams were around 5 percentage points more likely to win when doing so could not hurt their draft position. This effect is primarily driven by recent behavior; since 1989, win percentage for teams whose lottery odds are fixed is over 10 percentage points higher.

In a related study, Price et al. (2010) study tanking in the NBA. They argue that receiving a high draft pick increases gate revenue, and that teams tank to improve draft position. Our paper improves and extends their study in several important ways. First, their analysis of the value of the first pick does not account for unobservable differences between lottery winners and losers, and is therefore contaminated by selection bias. We solve this problem by explicitly adjusting for the probability of winning the lottery, thus eliminating this bias. Second, their methodology for identifying tanking is based on comparisons between playoff and non-playoff teams; there may be underlying differences in team quality that drive the results of such comparisons. We use a richer methodology that holds team quality fixed by comparing games in which the incentive to tank is turned on or off for a given non-playoff team in a particular season. Our approach isolates the tanking incentive and yields uniquely credible evidence of tanking in the NBA.

II. The Value of the First Pick

We begin our empirical analysis by estimating the economic and competitive effects of receiving the first pick in the NBA draft. This analysis is complicated by the potential for selection bias: Since teams with poor records receive disproportionate weight in the lottery's selection mechanism, lottery winners are likely to continue to perform poorly relative to lottery losers in future seasons regardless of whether they win the lottery. However, a useful feature of the NBA lottery context is that the weight determining each team's probability of winning can be directly calculated. In the language of Rosenbaum and Rubin (1983), the propensity score (probability of winning) is known. With appropriate re-weighting using this probability, the lottery is purely random. We utilize this insight to estimate the effects of winning the lottery using two weighting schemes (see, for example, Kline, 2011, for a derivation of these weights):

TOT weights: 1 for lottery winners, $\frac{p_{it}}{1-p_{it}}\frac{1-\pi}{\pi}$ for lottery losers,

$$\text{ATE weights:} \quad \frac{\pi}{p_{it}} \text{ for lottery winners, } \frac{1-\pi}{1-p_{it}} \text{ for lottery losers,}$$

where p_{it} is the known lottery odds for lottery team *i* in year *t*, and π is the fraction of lottery winners in the sample. Our estimating equation for analyzing the value of the first pick is then

$$Y_{it} = \alpha + \beta F_{it} + \varepsilon_{it}, \tag{1}$$

weighted by either the TOT or ATE weights, where Y_{it} is an outcome of interest for team *i* in year *t* (for example, attendance or win percentage in a subsequent year), and F_{it} is an indicator variable for receiving the first pick. Our coefficient of interest is β, which captures the causal effect of winning the first pick. We estimate equation (1) using data on participants in the 25 NBA draft lotteries held from 1985 to 2009, excluding teams that traded their picks prior to the lottery drawing. Outcomes are taken from the 1986 to 2010 NBA seasons.[30]

Following Florke and Ecker (2003), p_{it} is calculated using standard rules of probability along with a historical database of NBA lottery rules. We use heteroskedasticity-robust standard errors and cluster them at the franchise level due to the potential for correlation within a franchise over time. Before discussing our results, we perform a simple balance check to verify that weighting by the TOT or ATE weights eliminates observable differences between lottery winners and losers. Table 2 displays the relationship between pre-lottery team characteristics and receipt of the first pick. Column (1) shows coefficients from regressions of a variety of team attributes on a dummy for winning the lottery. While assignment of the first pick is not significantly related to previous win percentage, points per game, or all-star appearances, it is strongly negatively correlated with the previous season's attendance and scoring differential. Teams that win the first pick have average attendance 6 points lower than other lottery teams, and score 1.7 fewer points per game relative to their opponents. This means that regressions using a simple dummy variable for winning the lottery will give biased estimates; lottery winners are systematically different than lottery losers, due to the weighting in the lottery odds.

Columns (2) and (3) of Table 2 illustrate that weighting by the probability of winning the lottery successfully adjusts for these differences between winners and losers. The reported coefficients are estimates from versions of equation (1) using each pre-lottery characteristic as the dependent variable. Notably, the coefficients on the first pick dummy in the equations for win percentage and scoring differential are no longer statistically significant, and the estimates are close to zero in most cases. Since the ATE is the average causal effect across all lottery teams and teams with low lottery odds rarely win

[30] NBA lottery rules are available at http://www.nba.com/history/lottery_probabilities.html. Team records and other characteristics are available at http://basketballreference.com/stats_download.htm. Attendance information is available at http://www.apbr.org/attendance.html and http://espn.go.com/nba/attendance.

the lottery, estimates with these weights tend to have larger standard errors. The first pick coefficients in the regressions for other team characteristics remain insignificant. Weighting by the probability of winning the lottery as described above is sufficient to erase observable differences between lottery winners and losers. This lends credibility to our use of propensity score re-weighting to estimate the causal effects of winning the lottery.

We now turn to our causal estimates. Table 3 reports OLS estimates from equation (1) using the natural log of attendance and win percentage as dependent variables and TOT and ATE weights. Columns (1) and (2) show results for attendance for years 1 through 5 after the lottery and total log attendance over these years. The estimates show attendance gains close to 6 percent in most years following a lottery win for the treated sample, significant at the 10-percent level or lower in all years. Average effects for the whole sample of lottery teams are similar in years 1, 2, and 3 (though less precise) and fade in years 4 and 5. Given that the worst teams are overrepresented in the lottery sample and have farther to go to reach their attendance ceiling, higher persistence for the TOT is not surprising. Overall, winning the draft lottery appears to substantially increase attendance in the short term.

Columns (3) and (4) of Table 3 repeat the analysis for the TOT and ATE, respectively, with team win percentage as the dependent variable. The results show that winning the draft lottery weakly increases subsequent win percentage for lottery winners, with the gain peaking at 8 to 9 percentage points in the 4th year after the lottery. The 4th-year TOT estimates are highly statistically significant. Though the TOT estimates for other years are small and not significantly different from zero, all of them are positive. The ATE results are somewhat stronger. Win percentage increases by between 6 and 9 percentage points in years 1 through 4, statistically significant in all years except year 3. Columns (5) and (6) show similar results for win percentage, strongest in year 4. These results are consistent with the hypothesis that winning the lottery leads to small increases in win percentage over the next few years.[31]

The last two columns of Table 3 show the TOT and ATE on the number of All-Star selections, which designate the top players in the NBA each season. The results are broadly consistent with the effects on win percentage. Four years after winning the lottery, treated teams average nearly 0.5 additional All-Stars (significant at the 5 percent level). The ATE is slightly smaller, around 0.4 additional All-Stars. Even in year one, the TOT for All-Stars is around 0.3 and significant at the 10 percent level, though the ATE is not significant in earlier years.

[31] Estimates beyond year five (not reported) are not reliable since the sample size shrinks as teams are followed for additional years.

III. Do NBA Teams Tank?

The results in Table 3 show that winning a high draft pick yields economic and competitive benefits. Do teams tank to increase their chances of receiving these benefits? To answer this question, we analyze the relationship between game outcomes and the incentive to tank for teams eliminated from the playoffs from 1968 to 2009. Our preferred estimating equation in this context is

$$W_{hatgf} = \kappa_{ht} + \lambda_{at} + \mu_g + \nu_f + \theta_h C_{htg} + \theta_a C_{atf} + \gamma_h E_{htg} + \gamma_a E_{atf} + \eta_{hatgf}, \quad (2)$$

where each game is a unique observation, W_{hatgf} is a dummy variable equal to one if home team h wins against away team a in year t with g games remaining for the home team and f games remaining for the away team, κ_{ht} and λ_{at} are full sets of home and away team-by-season fixed effects, respectively, μ_g and ν_f are sets of fixed effects for the number of games remaining in the season for the home and away team, respectively, C_{htg} and C_{atf} are dummy variables equal to one if home team h and away team a are eliminated from the playoffs and can no longer change their lottery odds with g and f games remaining in year t, and E_{htg} and E_{atf} are dummy variables equal to one if the home and away team are eliminated from the playoffs. In years 1968 to 1983 and 1989 to 2009, C_{htg} and C_{atf} equal one if team h and team a, respectively, have clinched their final rank, so that they can no longer pass or be passed by another team. In years 1984 to 1988, all lottery-eligible teams faced even odds in the lottery, and so the clinching dummies are always one for playoff-eliminated teams in these years. Thus, in games where the clinching dummies are one, the teams have no incentive to tank for draft position, while when these dummies are zero, the teams have an incentive to tank.

The game fixed effects control for changes in teams' performance over the course of a season that are not due to tanking. For example, it is possible that all playoff-eliminated teams put in less effort as the season winds down, regardless of their tanking incentive. The team-season effects control for team quality in a particular season for a given team. The coefficients of interest, θ_h and θ_a, capture the effect of having an incentive to tank on team performance. We perform the analysis at the level of a game rather than at the level of each team in the game to avoid mechanical correlation between two observations from the same game. Standard errors for equation (2) are clustered at the home team-season level.

Tables 4 and 5 show the clinching progression for draft lottery ranking over the last five days of the 2007 and 2008 seasons. The cells with grey shading in each table represent days where the team with the given lottery rank had clinched its rank, setting its clinching dummy equal to one. In both tables, most teams did not clinch their spot before the end of the season. Based on the number of games remaining and the win totals for them and the nearest teams in the rankings, it was possible for their rank to change until the final day. However, three teams in 2007 and two teams in 2008 clinched their ranking before the final game. These teams had no explicit incentive to lose in the games played on the grey-shaded days.[32]

Our estimation strategy amounts to a difference in differences procedure, where the first difference is over time within teams (before and after clinching for teams that clinch) and the second difference is between teams. Before presenting estimates of tanking behavior in the NBA, we perform a standard check to determine whether clinching teams and non-clinching teams were on similar trends before the clinching teams locked in their ranking. To do this, in Table 6, we drop the games where C_{htg} equals one and recode C_{htg} equal to one in the two games prior to the actual clinching game. In other words, we test for treatment effects on the home team in the two games prior to clinching, when no treatment was actually present. Table 7 repeats the analysis for away teams.[33]

Columns (1) and (2) of Table 6 show pre-treatment estimates for home team win percentage, controlling for team-season and games-remaining fixed effects. Since the number of games remaining in the season for the home and away team entering a given game are highly correlated, we estimate regressions in column (1) using fixed effects for the home team's games remaining only, though this restriction has little effect on the point estimates and significance levels in this and future tables. Both columns show that win percentage for clinching home teams is about 7.5 percentage points worse in the two games preceding the clinching game (marginally significant at the 10 percent level). This result is problematic for our identification strategy, since it suggests that clinching home teams were on a downward trend relative to non-clinching teams prior to the clinching game.

We continue with the difference in differences strategy for three reasons. First, the same specification with point differential (home score minus away score) as the dependent variable shows no significant difference in trends for clinching and non-clinching home teams in columns (3) and (4). Point differential is a continuous outcome that may reflect team quality more accurately in small samples than win percentage, which is a discontinuous function of the point differential in each game. Second, the negative pre-trend for clinching teams may in fact reflect tanking behavior. Teams that clinch may do so by tanking more effectively than those who do not. If clinching teams tank more, our estimates will still reflect tanking behavior in general. Third, Table 7 shifts the focus to the away team, and all effects are

[32] In the event of a tie, the teams each receive the average of the odds for the two relevant ranks, so teams have an incentive to lose until ties are no longer possible.

[33] We maintain the actual treatment status for the opposing team in each table, so that the regression controls for the incentives of the opposing team properly.

far from statistical significance (though these tests are not very high powered). Still, we will keep the pre-trends in mind when interpreting results from the main specification.

Table 8 has the same structure as Tables 6 and 7, with C_{htg} and C_{atf} representing actual treatment instead of pre-treatment. The coefficients therefore estimate tanking behavior, under the assumptions of the difference in differences model. Columns (1) and (2) show that home team win percentage increases by about 6 percentage points when the home team has clinched (marginally statistically significant depending on the fixed effects specification), and decreases by about 5 percentage points when the away team has clinched (also significant at the 10 percent level). These nearly equal and opposite changes in home team win percentage are consistent with the hypothesis that team performance improves after clinching. We control for the playoff elimination status of each team in the regression as well, since eliminated teams may have a lower incentive to win. However, the point estimates for playoff elimination are not statistically significant and close to zero in all regressions, likely because most teams realize they will not make the playoffs long before they are officially eliminated.

Despite the consistent results on win percentage, the pre-treatment analysis above suggests that some of the difference is due to clinching home teams' declining win percentage before clinching. We turn to point differential to reduce these concerns somewhat. Columns (3) and (4) show similar effects on point differential, especially in response to the away team clinching. The home team loses more than 2 points of point differential over the away team once the away team has clinched (significant at the 1 percent level). While the point estimate for the home team's clinch effect is positive at nearly 1.5 points, the standard errors are slightly bigger and the effect is not statistically significant.

In Table 9, we explore whether tanking has increased or decreased over time. The regression specification is identical to the model employed in Table 8, but limited to the period before 1984 in the top panel and the period after 1988 in the top panel. We omit the seasons from 1984 to 1988, since lottery odds were the same for all playoff-eliminated teams in those years and therefore there was no incentive to tank in any games. The estimates in Table 8 suggest that tanking has increased substantially. In fact, while the point estimates for point differential in columns (3) and (4) have the expected sign before 1984, they are far from statistical significance. The point estimates for win percentage show no evidence of tanking either. The point estimates after 1988 suggest substantial tanking. Home team win percentage increases around 15 percentage points after the home team clinches and drops about 10 percentage points after the away team clinches (significant at the 5 and 10 percent levels, respectively). The effects on point differential mirror these effects, though significance levels are lower for the home-team clinch effect. Taken together, the results from before 1984 and after 1988 suggest that tanking is a relatively new phenomenon in the NBA.

IV. Conclusions

High picks in the NBA draft are perceived to benefit teams both competitively and financially. As a result, lottery-eligible teams have an incentive to tank at the end of the season in order to improve their chances of getting a high pick. This paper provides two new pieces of evidence on tanking in the NBA. First, weighting by the probability of winning the lottery to eliminate selection bias, we show that landing the first pick in the draft lottery increases attendance by about 6 percent over each of the subsequent five years for the treated sample (TOT) from 1989 to 2010. The average effect for all lottery-eligible teams (ATE) is similar but somewhat shorter lived. Winning the lottery also increases win percentage in the short run for treated teams, with the impact peaking at 8 to 9 percent in the 4th year following the lottery. The ATE on win percentage is more immediate and consistent. Overall, these results show that the first pick is competitively and economically valuable.

Second, we have uncovered new evidence that commentators and fans are right to blame teams for tanking. Among playoff-eliminated teams, those whose final rank is in doubt have worse performance. For the whole sample of teams, the probability of winning a home game increases by around 6 percentage points in response to the home team clinching its rank and decreases around 5 percentage points in response to the away team clinching. This difference is primarily driven by behavior since 1988. Our tanking estimates in that period balloon to 15 percentage points and -10 percentage points in response to the home and away team clinching, respectively. Point differential estimates are similar: home teams perform around 3 points better after clinching, while away teams perform almost 5 points better after clinching. Differential pre-trends for clinching versus non-clinching home team win percentage may muddy these results. However, given the consistency of our findings for point differential and for the away team, we believe our results are an improvement on past attempts to measure tanking. Importantly, our measurement of tanking is only a lower bound, since teams may make commitments to tanking that they cannot change upon clinching their lottery rank (for example, trading a star player or announcing that a player's season is over due to a mild injury). Other teams may struggle to improve their play after playing poorly for a long stretch of time.

Whether tanking is the right choice for teams remains an open question. Although we have measured clear economic benefits to tanking, our approach does not allow us to estimate lost revenue from this behavior due to decreased attendance in the long or short run. Similarly, we cannot comment on whether players, coaches, or management are responsible for tanking behavior, or whether tanking is bad for the NBA. The current draft order rules are meant to distribute new talent to the worst teams in the league, so as to increase parity between teams. Incentives to tank are a necessary part of any talent redistribution. If redistribution is valuable for the league, then there is a tradeoff between tanking and increased parity that the league must balance. This paper documents that the balance is nontrivial: measurable redistribution and tanking do occur under the current draft order rules.

Table 1: First Picks in the NBA Draft, 1985-2010

Lottery year	First Pick	Selecting team	Tenure with selecting team	All-star appearances	NBA Championship
(1)	(2)	(3)	(4)	(5)	(6)
1985	Patrick Ewing	New York Knicks	15	11	0
1986	Brad Daugherty	Cleveland Cavaliers	8	5	0
1987	David Robinson	San Antonio Spurs	16	10	2
1988	Danny Manning	Los Angeles Clippers	5	2	0
1989	Pervis Ellison	Sacramento Kings	1	0	0
1990	Derrick Coleman	New Jersey Nets	5	1	0
1991	Larry Johnson	Charlotte Hornets	5	2	0
1992	Shaquille O'Neal	Orlando Magic	4	15	4
1993	Chris Webber	Orlando Magic	0	5	0
1994	Glenn Robinson	Milwaukee Bucks	8	2	1
1995	Joe Smith	Golden State Warriors	2	0	0
1996	Allen Iverson	Philadelphia 76ers	10	11	0
1997	Tim Duncan	San Antonio Spurs	14	13	4
1998	Michael Olowokandi	Los Angeles Clippers	5	0	0
1999	Elton Brand	Chicago Bulls	2	2	0
2000	Kenyon Martin	New Jersey Nets	4	1	0
2001	Kwame Brown	Washington Wizards	4	0	0
2002	Yao Ming	Houston Rockets	8	8	0
2003	LeBron James	Cleveland Cavaliers	7	7	0
2004	Dwight Howard	Orlando Magic	7	5	0
2005	Andrew Bogut	Milwaukee Bucks	6	0	0
2006	Andrea Bargnani	Toronto Raptors	5	0	0
2007	Greg Oden	Portland Trailblazers	4	0	0
2008	Derrick Rose	Chicago Bulls	3	2	0
2009	Blake Griffin	Los Angeles Clippers	2	1	0
2010	John Wall	Washington Wizards	1	0	0

Notes: This table lists players selected with the first pick in the NBA draft from 1985-2010. Columns (4)-(6) list initial tenure with the selecting team, All-star appearances, and NBA championships won through the 2010-2011 NBA season.

Table 2: Covariate Balance

	Unweighted (1)	TOT weights (2)	ATE weights (3)
Log attendance	-0.060***	0.012	0.020
	(0.015)	(0.016)	(0.029)
Scoring differential	-1.660***	0.392	1.005
	(0.567)	(0.581)	(1.011)
Win percentage	-0.048	0.015	0.058
	(0.049)	(0.044)	(0.068)
Points per game	-0.687	-0.511	0.976
	(0.883)	(0.872)	(1.457)
Number of all-stars	0.033	0.158	0.104
	(0.111)	(0.108)	(0.173)
p-value from F-test	0.001	0.515	0.162
N		234	

Notes: This table shows coefficients from regressions of the pre-lottery variable in each row on a dummy for winning the draft lottery. Column (1) shows unweighted estimates. Column (2) uses weights equal to one for lottery winners and $(P/(1-P))*((1-\pi)/\pi)$ for lottery losers, where P is the propensity score and π is the fraction of teams winning the lottery. Column (3) uses weights equal to π/P for winners and $(1-\pi)/(1-P)$ for losers. Standard errors, clustered at the franchise level, are in parentheses. F-tests are for the hypothesis that receiving the first pick is unrelated to any of the displayed characteristics.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 3: Effects of Winning the NBA Draft Lottery on Attendance, Win Percentage, and All-Star Appearances

	Dep. var: Log attendance		Dep. var: Win percentage		Dep. var: Scoring differential		Dep. var: Number of All-Stars	
	TOT weights	ATE weights	TOT weights	ATE weights	TOT weights	ATE weights	TOT weights	ATE weights
Time period	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
After 1 year	0.066*	0.084	0.032	0.066*	0.268*	0.216	0.268*	0.216
	(0.038)	(0.058)	(0.029)	(0.037)	(0.153)	(0.182)	(0.153)	(0.182)
N	234		234		234		234	
After 2 years	0.061**	0.084*	0.023	0.087*	0.158	0.335	0.158	0.335
	(0.030)	(0.049)	(0.034)	(0.047)	(0.145)	(0.316)	(0.145)	(0.316)
N	221		221		221		221	
After 3 years	0.052*	0.049*	0.033	0.089	0.103	0.345	0.103	0.345
	(0.028)	(0.028)	(0.038)	(0.058)	(0.204)	(0.346)	(0.204)	(0.346)
N	207		207		207		207	
After 4 years	0.066**	0.033	0.087***	0.062**	0.459**	0.382*	0.459**	0.382*
	(0.028)	(0.023)	(0.032)	(0.026)	(0.226)	(0.213)	(0.226)	(0.213)
N			195					
After 5 years	0.064**	0.025	0.047	0.033	0.254	0.186	0.254	0.186
	(0.032)	(0.026)	(0.035)	(0.022)	(0.199)	(0.156)	(0.199)	(0.156)
N	182		182		182		182	
Total, next 5 years	0.052**	0.021	0.043	0.067*	0.266*	0.373*	1.328*	1.865*
	(0.023)	(0.021)	(0.029)	(0.040)	(0.137)	(0.198)	(0.683)	(0.991)
N	182		182		182		182	

Notes: This table shows regressions of subsequent log attendance, winning percentage, and All-Star appearances on a dummy for winning the NBA draft lottery. Columns (1), (3), (5) and (7) estimate the effect of treatment on the treated (TOT) by using weights equal to one for lottery winners and $(P/(1-P))*((1-\pi)/\pi)$ for lottery losers, where P is the propensity score and π is the fraction of teams winning the lottery. Columns (2), (4), (6) and (8) estimate the average treatment effect (ATE) by using weights equal to π/P for winners and $(1-\pi)/(1-P)$ for losers. Standard errors, clustered at the franchise level, are in parentheses.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 4. Clinching Progression Over the Last Five Days of the 2007 Season

Lotto Rank	Days Left: 5 Wins	5 Games Left	4 Wins	4 Games Left	3 Wins	3 Games Left	2 Wins	2 Games Left	1 Wins	1 Games Left
1	14	3	14	3	14	2	14	1	14	1
2	18	2	18	2	19	1	19	1	19	1
3	20	3	21	2	21	2	21	2	21	1
4	22	3	22	2	22	2	22	2	22	1
5	23	3	23	2	23	2	23	2	23	1
6	23	2	23	2	23	2	23	1	23	1
7	26	3	26	2	26	2	26	1	26	1
8	30	3	31	3	31	2	31	2	31	1
9	31	3	31	2	31	2	32	1	32	1
10	32	3	33	2	33	2	33	2	34	1

Note: Shaded columns indicate that the team with the corresponding rank in the left-most column had clinched their lottery rank at that point in the season.

Table 5. Clinching Progression Over the Last Five Days of the 2008 Season

Lotto Rank	Days Left: 5 Wins	5 Games Left	4 Wins	4 Games Left	3 Wins	3 Games Left	2 Wins	2 Games Left	1 Wins	1 Games Left
1	16	3	16	3	16	2	16	1	16	1
2	19	2	19	2	19	2	19	1	19	1
3	19	3	19	2	19	2	19	1	19	1
4	22	3	22	2	22	2	22	1	22	1
5	23	3	23	3	23	2	23	1	23	1
6	24	3	24	2	24	2	24	1	24	1
7	28	3	29	2	29	2	29	1	29	1
8	30	3	30	3	31	2	31	1	31	1
9	31	2	31	2	31	1	32	1	32	1
10	32	3	33	2	33	2	34	1	34	1

Note: Shaded columns indicate that the team with the corresponding rank in the left-most column had clinched their lottery rank at that point in the season.

Table 6. Pretrend Test for Home Teams Within Three Games of Clinching Draft Odds

	Dep. Var: Home Win Percentage		Dep. Var: Home Point Differential	
	(1)	(2)	(3)	(4)
Pre-Treatment (Home)	-0.074	-0.077	-0.957	-0.931
	(0.042)*	(0.041)*	(1.150)	(1.147)
Clinched (Away)	-0.047	-0.047	-2.722	-2.759
	(0.028)*	(0.028)*	(0.925)***	(0.926)***
Playoff Eliminated (Home)	-0.002	-0.001	-0.287	-0.261
	(0.016)	(0.016)	(0.409)	(0.410)
Playoff Eliminated (Away)	-0.011	-0.012	-0.251	-0.273
	(0.012)	(0.012)	(0.350)	(0.352)
N	38,061			
Games Left Controls	Home FEs	Home & Away FEs	Home FEs	Home & Away FEs

Notes: The table reports coefficients from OLS regressions of a dummy for a home team win (columns 1 and 2) and the home team score minus the away team score (columns 3 and 4) on a pre-treatment dummy equal to one if the home team clinched its draft odds within two games of the current game, a treatment dummy equal to one if the away team had already clinched its draft odds, and dummies for whether each team had been eliminated from the playoffs. The "Games Left Controls" row notes the controls included for the number of games remaining in the season (fixed effects for just the home team or for both teams). All regressions include home team by season and away team by season fixed effects. Standard errors clustered at the team-season level are in parentheses.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 7. Pretrend Test for Away Teams Within Three Games of Clinching Draft Odds

	Dep. Var: Home Win Percentage		Dep. Var: Home Point Differential	
	(1)	(2)	(3)	(4)
Clinched (Home)	0.055	0.050	1.015	0.912
	(0.038)	(0.038)	(0.980)	(0.987)
Pre-Treatment (Away)	-0.040	-0.034	-0.303	-0.096
	(0.044)	(0.044)	(1.208)	(1.212)
Playoff Eliminated (Home)	-0.004	-0.004	-0.257	-0.236
	(0.015)	(0.015)	(0.410)	(0.410)
Playoff Eliminated (Away)	-0.010	-0.011	-0.275	-0.311
	(0.013)	(0.013)	(0.352)	(0.354)
N	38,126			
Games Left Controls	Home FEs	Home & Away FEs	Home FEs	Home & Away FEs

Notes: The table reports coefficients from OLS regressions of a dummy for a home team win (columns 1 and 2) and the home team score minus the away team score (columns 3 and 4) on a pre-treatment dummy equal to one if the away team clinched its draft odds within two games of the current game, a treatment dummy equal to one if the home team had already clinched its draft odds, and dummies for whether each team had been eliminated from the playoffs. The "Games Left Controls" row notes the controls included for the number of games remaining in the season (fixed effects for just the home team or for both teams). All regressions include home team by season and away team by season fixed effects. Standard errors clustered at the team-season level are in parentheses.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 8. Effect of Clinching Draft Odds on Performance for Playoff-Eliminated Teams

	Dep. Var: Home Win Percentage		Dep. Var: Home Point Differential	
	(1)	(2)	(3)	
Clinched (Home)	0.062	0.058	1.489	1.368
	(0.035)*	(0.035)	(0.932)	(0.939)
Clinched (Away)	-0.048	-0.047	-2.307	-2.323
	(0.027)*	(0.027)*	(0.854)***	(0.855)***
Playoff Eliminated (Home)	-0.004	-0.004	-0.265	-0.236
	(0.015)	(0.015)	(0.398)	(0.398)
Playoff Eliminated (Away)	-0.008	-0.010	-0.181	-0.199
	(0.012)	(0.012)	(0.342)	(0.344)
N		40,083		
Games Left Controls	Home FEs	Home & Away FEs	Home FEs	Home & Away FEs

Notes: The table reports coefficients from OLS regressions of a dummy for a home team win (columns 1 and 2) and the home team score minus the away team score (columns 3 and 4) on treatment dummies for whether the home and away team had clinched their draft odds and dummies for whether each team had been eliminated from the playoffs. The "Games Left Controls" row notes the controls included for the number of games remaining in the season (fixed effects for just the home team or for both teams). All regressions include home team by season and away team by season fixed effects. Standard errors clustered at the team-season level are in parentheses.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 9. Effect of Clinching Draft Odds on Performance for Playoff-Eliminated Teams by Era

	Dep. Var: Home Win Percentage		Dep. Var: Home Point Differential	
	(1)		(3)	(4)
		Panel A. Before 1984		
Clinched (Home)	-0.019	-0.037	0.724	0.501
	(0.062)	(0.063)	(1.705)	(1.717)
Clinched (Away)	-0.049	-0.045	-1.844	-1.874
	(0.057)	(0.057)	(1.767)	(1.787)
N		12,982		
		Panel B. After 1988		
Clinched (Home)	0.149	0.153	3.095	3.277
	(0.072)**	(0.071)**	(1.887)	(1.853)*
Clinched (Away)	-0.105	-0.105	-4.898	-4.938
	(0.057)*	(0.058)*	(1.555)***	(1.551)***
N		22,304		
Games Left Controls	Home FEs	Home & Away FEs	Home FEs	Home & Away FEs

Notes: The table reports coefficients from OLS regressions of a dummy for a home team win (columns 1 and 2) and the home team score minus the away team score (columns 3 and 4) on treatment dummies for whether the home and away team had clinched their draft odds and dummies for whether each team had been eliminated from the playoffs. Panel A shows effects for the period before the even odds lottery began in 1984 and Panel B shows effects for the period after it ended in 1988. The "Games Left Controls" row notes the controls included for the number of games remaining in the season (fixed effects for just the home team or for both teams). All regressions include home team by season and away team by season fixed effects. Standard errors clustered at the team-season level are in parentheses.
* significant at 10%; ** significant at 5%; *** significant at 1%

References

Florke, C.R., and M.D. Ecker. 2003. "NBA Draft Lottery Probabilities," American Journal of Undergraduate Research 2(3) 19-29.

Kline, Patrick. 2011. "Oaxaca-Blinder as a Reweighting Estimator," American Economic Review 101(3): 532-537.

Price, J., B.P. Soebbing, D. Berri, and B.R. Humphreys. 2010. "Tournament Incentives, League Policy, and NBA Team Performance Revisited," Journal of Sports Economics 11(2): 117-135.

Rosenbaum, P.R., and D.B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies of Causal Effects," Biometrika 70: 41-55.

Simmons, William J. 2007. "Tanks for Nothing, NBA," ESPN The Magazine, April 2007.