Undergraduate grant employment and persistence

Margaret Levenstein, Daniela Morar and Jason Owen-Smith *

October 18, 2019

Preliminary draft- do not cite

Abstract

We study the impact of grant employment on undergraduate graduation rates. We use a unique dataset created by linking course-level student record data with transaction-level data on federal grant expenditures on personnel at a major public university. Our paper uses a selection on observables strategy and finds that a paid research experience on a federally funded grant increases graduation rates by 10.1 percent in STEM majors and by 5.5 percent across all fields of study. We find that the race and gender gaps are not narrowed by much and that the positive impact of grant employment is largest for male students and whites. Even though minorities and female students benefit from grant employment, the benefits are much lower across all measures used. Furthermore, we find that for students receiving who qualify for Pell grants and thus have more financial constraints, working with a faculty member helps increase their graduation rates significantly. Our results indicate potential benefits to students of matriculating in more research-intensive environments and the need of higher intensity interventions to improve the representativeness of STEM population.

JEL-Classification: I20, I23, J16, J15
Keywords: Higher education, teaching assistants, STEM persistence

*We are grateful for helpful comments from John Bound, Charlie Brown, Kevin Stange, Jeff Smith, Morris Hamilton, Nancy Calvin-Naylor, Natsuko Nicholls, and Sang Teck Oh. We acknowledge the NSF SMA-1262447 grant for generous research support. This research has been determined exempt from human subjects control under exemption 1 of the 45 CFR 46.101 (b) by the U.M. Institutional Research Board (HUM00085505).

1 Introduction

The United States' strength in terms of productivity, competitiveness and economic growth has been linked to technological development [Xie and Killewald, 2012, Goldin and Katz, 2008, Augustine, 2007]. Although a leader in global technology and economy, the U.S. lags behind other developed countries in the number of STEM graduates [Chen, 2013]. The National Science Foundation [NSB, 2016] finds that only around 9 percent of global university STEM degrees were conferred to U.S. students, while U.S. students hold about 20 percent of the world bachelor's degrees [Ryan and Bauman, 2016, Group, 2014]. There is considerable research that shows that a critical approach to stimulate the country's economic growth is helping students succeed and graduate in STEM disciplines [Ehrenberg, 2010, 2005].

The low number degrees conferred in STEM fields could be partially attributed to the low STEM participation of women and minorities [Hernandez et al., 2018, US Department of Health and Human Services, 2015, Olson and Riordan, 2012]. Women having outnumbered men in college enrollment, having earned about 57 percent of all Bachelor's degrees awarded since the late 1990s [NCSES, 2017]. Despite this fact, here still exists significant attainment gender gap in STEM degrees [Gayles and Ampaw, 2014, NCSES, 2017], with little change since the 1980s [DiPrete and Buchmann, 2013, England et al., 2007, England and Li, 2006, Mann and DiPrete, 2013]. More specifically, while women earned more bachelor's degrees in Psychology, biosciences, and social sciences (except for Economics) compared to men, they earned considerably fewer degrees in computer science, engineering and mathematics [NCSES, 2017]. Besides the gender STEM gap, previous studies also document an existent racial gap. While white and Asian American students are consistently well represented in STEM disciplines compared to their overall enrollment [Herrera and Hurtado, 2011, Goyette and Xie, 1999], African-American and Hispanic are underrepresented [NCSES, 2017].

Significant evidence shows that participation in faculty-mentored undergraduate research [Hu et al., 2007] helps improve the under-representation of students in STEM fields since interactions between students and faculty members have been directly linked to persistence in college [Terenzini and Pascarella, 1977, Pascarella and Terenzini, 1979, Tinto, 1993]. We hypothesize that working with a faculty member improves persistence of students in STEM majors by providing students with three different types of capital: human capital, social capital and financial capital. The first channel we consider is the accumulation of both generic and specific human capital through undergraduate research employment. Being employed by a professor has been linked with taking more advanced courses [Bauer and Bennett, 2003] and graduating at a higher rate [Kim et al., 2003, Gregerman et al., 1998]. Studies that examine undergraduate research experience have also found positive effects on student outcomes,[1] such as a higher knowledge and comprehension of science [Sabatini, 1997] that surpassed the knowledge achieved in ordinary science classes [Ward et al., 2003]. In addition, students who engaged in undergraduate research report increased oral communication and research skills [Bauer and Bennett, 2003, Seymour et al., 2004, Hunter et al., 2007, Kardash, 2000].

Another channel through which undergraduate research impacts persistence is through providing students with social capital which would further provide students with a sense of STEM identity. Akerlof and Kranton [2002] develop a model of identity where people make economic decisions based on their individual identities and social norms. If we believe that there are associations between students' social economic characteristics and the decision to pursue a STEM major, then this model could explain the low STEM participation of underrepresented groups. Thus, undergraduate grant employment was proposed as a solution to narrow these participation gaps. Working with a faculty member was linked to improvement in students' navigation through their STEM major [Cole and Espinoza, 2008, Ullah and Wilson, 2007], where students are more likely to leave the field due to a "chilly" climate [Seymour and Hewitt, 1997]. Students who engaged in undergraduate research display improved confidence in their science skills [Grandy, 1998, Graham et al., 2013] and are more likely to identify themselves as people who "do science" [Seymour et al., 2004, Hunter et al., 2007, Kardash, 2000, Russell et al., 2007]. Integration has been shown to be even more important in the persistence of URM students (i.e. Black/African American, Latino/a or Native American) in STEM majors, who gained a stronger science identities and higher self-efficacy [Chang et al., 2014, Cole and Espinoza, 2008, Hurtado et al., 2009, Hyde et al., 1990]. Furthermore, accumulation of social capital could also manifest itself through the opportunity to faculty members and familiarize themselves with research. Students involved in projects with faculty members were also more likely to attend graduate school [Barlow and Villarejo, 2004, Bauer and Bennett, 2003, Pender et al., 2010, Russell et al., 2007, Hunter et al., 2007, Hathaway et al., 2002]. In addition, undergraduate research provided students with a clearer understanding of the type of work involved in a scientific career [Lopatto, 2010], which increased their desire to pursue STEM-related careers [Bauer and Bennett, 2003, Russell et al., 2007, Zydney et al., 2002].

Last, but not least, undergraduate employment provides the financial capital for students to support themselves while acquiring a degree. Scott-Clayton and Minaya [2016] and Scott-Clayton and Zhou [2017] suggest that the cash channel is not as important as other channels. In fact, low

[1] A summary of the literature on undergraduate research experience is offered in Table 1.

income students benefit more from federal work study, but they attribute this result to the fact that lower-income students would have been employed even in the absence of federal work study and might have had less desirable jobs in the absence of campus employment. Thus, they conclude that participants in federal work study benefit from their employment because they have a job more related to their major and not through the cash channel. This result is also consistent to the broader result showing that on-campus employment, as opposed to off-campus employment, has a positive impact on persistence in college [Pascarella and Terenzini, 2005, Hossler et al., 2009].

In this study, we evaluate the impact of undergraduate research employment on persistence in STEM, where undergraduate research employment is defined as having been employed on a federally funded grant at a large public university. Despite a vast empirical literature on undergraduate research, previous studies rely heavily on surveys and fail to support causal claims [Mervis, 2006, Linn et al., 2015, Sadler et al., 2010]. In an ideal world, we would like to be able to randomly assign undergraduate employment in a controlled environment to correctly identify and estimate its mean impact on student outcomes. In the context where randomization is not possible, quasi-experimental designs can be used to identify a comparison group that is as similar as possible to the treatment group in terms of pre-treatment characteristics. These quasi-experimental designs help reduce the effect of selection bias, which is important in the absence of random assignment of researcher experience. In our case, students and faculty engage in a search and match process that selects students for research experiences who are more likely to graduate [Rosenbaum and Rubin, 1983, West et al., 2008]. We employ a selection on observables strategy, which allows us to construct our counter-factual, the mean outcome of students who worked on a grant had they not worked on a grant and we consider both general and STEM graduation rates. Since working with a faculty member gives students a glimpse of the type of work that a scientific career entails [Lopatto, 2010, Kinkead, 2003], grant employment should influence STEM graduation rates more than general graduation rates. Our results are consistent with this hypothesis and our preferred matching estimators show a positive and significant impact of approximately 10 percentage points of employment on STEM graduation rates and an impact of approximately 6 percentage points on general graduation rates.

Given that most previous papers have focused on examining small, short-term research programs [Gregerman et al., 1998], we extend the existing literature by examining all paid research positions within a large public university. We also employ performance-based evidence (grades, declaring a major, graduating), as opposed to the student self-reported, retrospective accounts of research experience [Hathaway et al., 2002] used by previous papers, which have been shown to be inconsistent with performance-based evidence [Bowman, 2010, Dunning et al., 2003, Feldon et al., 2015]. We further contribute to the existing literature by using an innovative dataset that combines administrative student transcript data with longitudinal administrative data on research funding. This unique data on research funding, the UMETRICS dataset, contains information on expenditures made on federally funded grants since 2001. It provides researchers with comprehensive information on employees' salaries, as well as payments to vendors and subcontractors made from federally funded grants. Thus, linking the UMETRICS data with student transcript data allows us to track all federally funded employment and course-taking history for the students who attended this large university over a period of thirteen years (2001-2014). In doing so, we demonstrate some of the value of linked administrative data from universities for understanding both the pathways by which research investments yield returns and the role of high-impact non-classroom experiences in shaping educational outcomes.

Our paper also disentangles the heterogeneous effects of undergraduate employment based on gender, race and financial status. Given the large differences in STEM persistence rates across students with different socio-demographic backgrounds, it's crucial to test whether undergraduate research employment helps alleviate these gaps. This paper considers two undergraduate students' outcomes: the likelihood of graduating with a degree in any field and the likelihood of graduation with a degree in a STEM field, conditional on graduation. Our results suggest that while undergraduate research employment has a positive impact on persistence, it does not narrow the gender, race and financial gaps in graduation rates. Finally, this study provides estimates of research employment for all types of grant employment, as well as more research intensive positions. In particular, we divide research employment by the amount of research intensity. We define research jobs as the jobs that are related to the student's science career, based on the job description of the university's HR department. Our findings show that research jobs increase STEM graduation rates, but that once again, this effect is higher for men and well represented racial groups.

The remainder of this paper is organized as follows. Section 2 discusses our data and variable construction. Section 3 outlines our identification strategy, Section 4 shows our results and Section 5 concludes.

2 Data

In this section, we describe the data that we use.

2.1 Institutional context and data

We study the impact of grant employment on graduation rates at a large research university. This large university is home to many colleges, among which are the College of Arts and Sciences (the largest one, making up 60 percent of total undergraduates enrolled) and the College of Engineering. Our dataset combines information from two different sources: student records data and administrative data on all the federal grants received by the faculty at the university. Both datasets contain a unique individual identifier that allows us to combine the two sources of data to get a complete history of each student's employment and course history.

The first data source contains administrative student data from a public Midwestern institution that contains all undergraduate students taking classes between Fall 2001 and Winter 2014. This data contains information on students regarding their demographic characteristics, financial aid status, course outcomes, and degree attainment. The demographic information includes each student's race, gender, and state and country of residency. The data also provide information about the courses taken by the students in each semester. We have access to the course subject and number, the credit and the grade obtained in the course. We have additional data on Advanced Placement exams and information about the last high school attended by the student, such as grade point average and the name of the high school (identified by its College Entrance Examination Board (CEEB) code). Data on intended major prior to attending the university is also available, and we define STEM fields as those designated by the U.S. Immigration and Customs Enforcement (ICE), as explained in the following subsection.[2]

The UMETRICS program is a university-specific program that builds upon the federally supported Science and Technology for America's Reinvestment: Measuring the Effect of Research on Innovation, Competitiveness, and Science (STAR METRICS) program.[3] The UMETRICS data are longitudinal administrative data on research funding from 19 major research institutions that contain information on expenditures made on federal awards: payments to individual people, as well as purchases to vendors and sub-contractors. More specifically, the data includes information on each of the grants received by individuals at the university, the number of people employed on each grant, their occupational status and the full-time equivalent (FTE). This comprehensive data contain information on all the employees working on federal grants between 2001 and 2014 at this

[2] In contrast, the National Science Foundation (NSF) uses a broader definition for STEM fields in which social sciences are also included.

[3] The STAR METRICS project was initiated in 2009 as a partnership between U.S. federal agencies and research universities to measure the impact of federally funded research.

public institution.[4] In our data, we cannot identify which positions are work-study positions. [5]

2.2 Dataset construction

This section informs the reader on the sample creation and variable construction.

2.2.1 Sample creation

We restrict the sample to students who are admitted as Freshmen and remove any transfer students, whose course taking behavior might vary due to past college experience. We only consider students who are enrolled for courses during the time period considered (Fall 2001-Winter 2014). In order to allow students to graduate in 5 years, we restrict the sample to students entering before the Fall of 2010. The resulting dataset of $35,720$ unique students provides a rich source for the analysis of our key questions.

2.2.2 Variable construction

Out of the various definitions used by previous literature to define STEM fields, we choose the one designated by the U.S. Immigration and Customs Enforcement agency on April 2008 when the extension for the Optional Practical Training (OPT) was introduced.[6] We do not take into account the additions made to this list in 2011 and 2012 (which include fields like psychology, agriculture, etc.). With this definition in mind, we use the CIP (Classification of Instructional Programs) codes that include all the disciplines offered in academic institutions in the United States to map the majors offered in the university to STEM fields.

These CIP codes are also used to identify students' intent to major in a STEM discipline. We collect information about intended major prior to attending college from three different sources: the Common Application, the SAT, and the ACT. The Common Application asks students to list their areas of interest in college, with no required upper limit for the answers provided. The SAT

[4]These data are created and maintained by the Institute for Research on Innovation and Science, which also makes them available for research use through a virtual data enclave. The full data documentation for the 2017 UMETRICS data release can be found at `https://doi.org/10.21987/R7MQ0S`. A newer version of the documentation for the 2018 data release can be found at `https://doi.org/10.21987/R7GW89`.

[5]Most of the standard work-study positions are under the umbrella of the Department of Education, but they do not have a clear code to allow us to identify them.

[6]The Optional Practical Training (OPT) is a period during which undergraduate and graduate students on a student visa are allowed to work for one year. More information about the OPT can be found at: `https://www.ice.gov/doclib/sevis/pdf/nces_cip_codes_rule_09252008.pdf`

exam contains a questionnaire on the choice of major, allowing up to three answers. The ACT asks students to list the college major they plan to have, with only one answer allowed.

Our list of covariates also includes Advanced Placement (AP) tests. Given our interest in STEM outcomes, we only select the science and math AP tests: Biology (BY), Chemistry (CH), Physics (Physics B (PHYSB), Physics C: Electricity and Magnetism (PHYSE), Physics C: Mechanics (PHYSM)), Computer Science (Computer Science A (CSA), Computer Science AB (CSAB)), Statistics (STAT), and Calculus (Calculus AB (CALAB), Calculus BC (CALBC)). In addition to AP test scores, we also have access to high school grade point average (GPA). The university recalculated all the high school GPAs on a 4.0 scale. One caveat is that before 2009, the university included only the courses taken in grades 9-11 for calculating the GPA. After 2009, the university considered all high school courses taken for all grades. However, we do not believe that this would be a major issue for our analysis. We also use composite ACT scores as a covariate, converting the SAT scores of the students who did not take the ACT into ACT scores using the concordance tables provided by the College Board.[7]

Another important factor for college persistence is parental education and income [Hellerstein and Morrill, 2011]. Unfortunately, data on parental education and income acquired from the admission office contains a very large number of missing observations (over 40 percent for parental income and over 20 percent for parental education) and we cannot use multiple imputation methods due to the non-randomness of the missing data. Instead, we use need-based grant eligibility as a proxy for parental income. Need-based grants have been showed to be good indicators of both the probability that students enroll in college [Deming and Dynarski, 2009], as well as persistence in college [Deming and Dynarski, 2009, Bettinger et al., 2009]. The largest of the need-based financial grants is the federal Pell Grant, a need-based grant that assists low-income students who are attending universities and other accredited secondary institutions. We create a binary Pell grant variable that identifies students who have received one (or more) of the following grants: Pell grant, Academic Competitiveness Grant (ACG), Supplemental Educational Opportunity Grant (SEOG) or SMART grant.

Student demographic characteristics also play an important role in their educational attainment. We have information on each student's gender (binary male/female), race (white, black, Hispanic, Asian, and other: native American, not indicated, Hawaiian and two or more) and country/county of residency. We use the information about each student's residence at the time of submitting their college application to define both in-state/out of state status, as well as international student status

[7] The concordance tables can be found at: `https://research.collegeboard.org/sites/default/files/publications/2012/7/researchnote-2009-40-act-sat-concordance-tables.pdf`

(for students with their country of residency outside of the United States).

For comparison purposes, we consider two measures of grant employment. The first measure of research experience considers all the grants from all federal sources listed in the UMETRICS dataset. One problem with this measure is that it includes administrative jobs that might not contribute much to the development of STEM-specific skills. Thus, we also consider an alternative measure of grant employment that includes only jobs that are related to the student's science career, denoted "research jobs". We select these jobs based on the job description from the university's HR department. For example, we categorize positions such as Research Associate, Assistant in Research, and Laboratory Assistant as research jobs, and positions such as Clerk, Library Assistant, and Secretary as non-research jobs.

3 Methodology

3.1 Treatment and outcome

We estimate the impact of research experience on persistence at a large public institution. Our outcome measure, graduation rate, is calculated based on a five-year window, starting with the first semester of courses attended at the university as a Freshman. Our outcomes are general graduation and STEM graduation rates.

Our treatment variable is undergraduate research experience, a binary variable that measures having been employed on a federally funded grant while attending courses at the university.[8] As seen in Table 1, the students at this university are slightly more likely to be working in their later years, but the majority (70 percent) start working in their first year of college.[9] Since we are concerned that students who work on a grant in their senior year are also more likely to graduate, we remove all instances of employment that happen in the student's senior year.[10]

Table 2 presents the descriptive statistics of the variables used in our analysis for the full sample, as well as for the treatment and control groups. Based on this table, the full sample contains about 52 percent women, with the sample employed on a grant having slightly more women. Our full sample contains about 6 percent black students, 5 percent Hispanic students and 11 percent Asian students. 64 percent of the population comes from the state where the university is located, 19 percent are ever recipients of Pell grants and the average ACT composite score is 28.1. Students

[8]We exclude all grant employment that takes place before the first semester and after the last semester enrolled for courses.

[9]The density of the number of months employed is shown in Table 2.

[10]The Sensitivity Analysis section talks more about this bias.

who are employed on a grant are more likely to have higher high school GPA and ACT composite scores and be Pell grant recipients. Furthermore, the AP scores for science and Math are higher for the students with job experience than for those without it.

The last two columns of Table 2 show that more female students who are employed have a research position, as compared to male students. Furthermore, while Hispanics and blacks working on a grant are less likely to have a research job, Asian students are more likely. On average, students who hold research positions are also more likely to graduate. This suggests a positive selection on observable characteristics for employment.

When estimating the impact of grant employment, one must account for a two-sided selection process. First, the student decides whether he or she wants to gain research experience and applies for a job. In the next step, each professor selects one or more applicants to be hired from the pool of all applicants. Because of the richness of our data and the possibility to control for pre-college interest in STEM (AP exams, high school GPA, intent to major in STEM) selection on observables is a plausible assumption. The basic regression we would want to estimate is one where we regress our two outcomes on the treatment status. In an ideal world, we would have a random assignment of treatment and we would not have to add any covariates to the equation. Absent such a set-up, we use matching techniques [Rosenbaum and Rubin, 1983] to compare treatment and control groups in the presence of selection. We adopt a "selection on observables strategy" (a term adopted from Heckman and Robb [1985]) to estimate the average treatment effect on the treated or the ATET. Formally, we denote Y_1 as the potential outcome for the students employed on a grant, Y_0 as the potential outcome for the students not employed on a grant, and T as the treatment (i.e. grant employment). The observed outcome is $Y = TY_1 + (1 - T)Y_0$ and we want to estimate:

$$\Delta_{TT} = E(Y_1 - Y_0|T = 1) = E(Y_1|T = 1) - E(Y_0|T = 1), \tag{1}$$

where $E(Y_1|T = 1)$ is the expected grant employment outcome conditional on grant employment and $E(Y_0|T = 1)$ is the expected non-grant employment outcome conditional on grant employment. Identifying $E(Y_0|T = 1)$ is challenging since it is unobservable: we cannot observe the outcomes for the students who did not work on a grant in a world where they had.

3.2 Identification

Given the selection issue, we employ the propensity score matching (PSM) technique which allows us to compare our treated groups with our controls even in the absence of a large treatment group. Our matching technique requires the three main conditions. The first one, the conditional independence assumption (CIA) states that once we control for all observable variables, the potential outcomes are independent of the treatment assignment (or that $(Y_0, Y_1) \perp T|X$). Rosenbaum and Rubin [1983] show that estimation doesn't require the CIA, but a weaker assumption called the conditional mean independence (CMI), also called the "balancing property". The CMI assumption implies that once we control for the covariates X, the treatment does not affect the conditional mean of each potential outcome:

$$E(Y_0|X, T = 1) = E(Y_0|X, T = 0) = E(Y_0|X) \tag{2}$$

Another assumption we need is the common support assumption, namely that for each value of X, there is a positive probability of participation given X, which translates to $0 < P(X) < 1$ for all X. We call this probability the propensity score: $P(X) = Pr(T = 1|X)$.

With this definition of the propensity score, Rosenbaum and Rubin [1983] show that if the CIA assumption holds for X, it also holds for $P(X)$ so that $Y_0 \perp T|P(X)$, and thus:

$$E(Y_0|P(X), T = 1) = E(Y_0|P(X), T = 0) = E(Y_0|P(X)) \tag{3}$$

The last assumption is the "Stable Unit Treatment Value Assumption" (SUTVA), which requires that the students who are not employed on a grant are not affected by the treatment. This assumption fails if there are general equilibrium effects generated by spillovers. While we cannot test for the existence of spillover effects, we believe that they are negligible, especially as the treated students are only 10 percent of the total. This assumption also fails if the students are employed on grants from outside of the university considered. Furthermore, if we assume that there are positive spillovers of being employed on a grant, then our estimates would biased downwards.

3.3 Estimation

Given these assumptions, there are many appropriate estimators we could use to calculate the ATET. We begin by considering two different estimators, with pros and cons of using each one of them. The first one, the inverse probability weighting (IPW) estimator, uses weighted averages of the observed outcome variable to estimate means of the potential outcomes.[11] In the estimation process, each weight is the inverse of the estimated probability that an individual receives a

[11] We estimate the treatment effects using the STATA command teffects. This command, unlike psmatch2, calculates the standard errors based on Abadie and Imbens [2012] and takes into account that the propensity score is estimated prior to the matching step.

treatment level and it is calculated using the following formula:

$$\hat{\Delta}_{TT} = \frac{1}{N^T}\sum_{i=1}^{N} Y_i T_i - \frac{1}{N^U}\sum_{i=1}^{N}\left(\frac{1}{N^U}\sum_{i=1}^{N}\frac{\hat{P}(X)(1-T_i)}{\hat{P}(X)}\right)^{-1}, \tag{4}$$

where N^T is the number of treated units and N^U is the number of untreated units.

Huber et al. [2013] and Busso et al. [2014] show that the IPW estimator has very good finite-sample properties when compared to other estimators and it requires no assumptions about the functional form of the outcome model. However, IPW can be problematic since it is very sensitive to extreme values of the propensity score and also to small misspecifications.

Therefore, we employ a second estimator, which is the nearest neighbor with replacement. Monte Carlo simulations [Frölich, 2004, Huber et al., 2013, Busso et al., 2014] show that the nearest neighbor with replacement estimator performs very poorly in comparison with other estimators in terms of mean squared error, due to the very high variance of the estimator. The high variance of the estimator is caused by the estimator ignoring all the observations close to the treated units, but not the closest ones. Despite this problem, the estimator exhibits a low bias and it is not sensitive to extreme values of the propensity score, which makes it a good alternative estimator. By increasing the number of neighbors, we increase bias (since we use matches that are farther away), but decrease variance (since we use more untreated units as points of comparison). In our estimation, we use the nearest neighbor with replacement estimator with one neighbor.

3.4 Propensity score specification

We first investigate the selection of students based on observable characteristics to explore the common support assumption. In addition to this, this procedure is informative for our sensitivity analysis. We consider different sets of conditioning variables that determine participation in research experience. Covariate selection is a relatively complicated task and there are benefits and costs to increasing the number of covariates. Not including all the important covariates can increase the bias of the estimates, as shown by Heckman et al. [1997]. However, including too many covariates can also be problematic. First of all, we only include the conditioning variables that affect both the treatment and the outcome and exclude the variables that are affected by the treatment. Bryson et al. [2002] and Augurzky and Schmidt [2001] find that including too many covariates that are not significant can increase the variance, even though they will not bias the estimates and it will not make them inconsistent.[12] Moreover, they show that including extraneous variables could lead to a violation of the common support condition.[13] We follow Stuart [2010] and include all the variables previously discussed in order to reduce the bias caused by not including the relevant variables.

We use institutional knowledge, previous empirical findings, and economic theory to determine the best choice of covariates. Previous research focused mainly on student characteristics and experiences (both from high-school and college) to disentangle the factors that influence persistence in STEM. Since high school class rank has been shown to be among the most important determinants of college success in STEM [Ellington, 2006], we use high school GPA quantiles[14] in our analysis. Furthermore, we also include ACT scores,[15] and a vector of different AP tests with scores of at least 3. Because interest in math and science are strong indicators of persistence in STEM [Maple and Stage, 1991, Mau, 2003, Tai et al., 2006, Maltese and Tai, 2010, 2011], we only use AP science and math test scores. Another pre-college variable that we include is a binary variable for interest in a STEM major. By including both AP tests and intent to major in STEM, we control for possible selection into STEM research jobs.

We also include an indicator for receiving a Pell grant (and Pell grant gender interaction) to account for the large differences in STEM persistence among students from different SES backgrounds [Schneider et al., 1998, Miller and Kimmel, 2012]. The propensity score specification further includes demographic characteristics (gender, race, gender-race interactions), in-state status (as measured by the address of the student at the time of enrollment), international student status and cohort fixed effects.

Tables 3-4 display the results from logistic regressions of undergraduate grant employment participation. The first column of each table shows estimates from a logistic regression, while the second column presents marginal effects estimated at the mean of observable characteristics. The regressions show that gender is a statistically significant predictor of research experience. Being female, Asian, and a Pell grant recipient significantly increase the likelihood of participating in research experience. In addition, students with higher high schools GPAs, ACT composite scores, and students from the state where the university is located are also more likely to be employed on a grant. Having taken an AP science or math test is also a very strong predictor of research experience. More precisely, students who have taken the AP tests in Biology (BY), Chemistry

[12] This is true in any multivariate setting, not just propensity score estimation.

[13] One test suggested for solving the selection of covariates for the propensity score problem is to start with a simple model and keep adding variables. Then the variables are kept if they are statistically significant and if they increase the prediction rate [Black et al., 2005].

[14] The university converts all the high school GPAs received on a 4.0 scale.

[15] We consider ACT composite scores (we convert SAT scores as explained above).

(CH), Physics B (PHYSB),[16] Physics C: Mechanics (PHYSM) and Calculus AB (CALAB)[17] are more likely to be employed on a federally funded grant. Interestingly, black and other race students are more likely to have been employed on a grant, but not more likely to hold a research-intensive position.

3.5 Balancing tests

We perform balancing tests to check that, at each value of the propensity score, the covariates chosen have the same distribution for the treatment and control group.[18] We use the standardized differences test [Rosenbaum and Rubin, 1985] to examine the balance, defined as:

$$\text{STD}_{\text{diff}} = 100\frac{(\bar{x}_t - \bar{x}_c)}{\sqrt{\frac{s_t^2+s_c^2}{2}}}, \tag{5}$$

where $\hat{x}_t$ and $\hat{x}_c$ are the sample means for a particular covariate in the treated and control groups, respectively and s_t^2 and s_c^2 are the sample variances.

This standardized difference is computed for each covariate used in the matching procedure. Since there is no formal criterion for how large this difference should be, we follow Rosenbaum and Rubin [1985] in considering 20 to be the allowed upper bound. Table 5 shows the standardized differences both before and after our matching analysis (using Inverse Probability Weighting as explained below). The results are calculated from the formula above and dividing the outcome by 100. Table 5 shows that our preferred matching estimator has standardized differences of the important variables all lower than 0.2,[19] considered small in the literature. Table 5 shows the standardized differences for grant employment as treatment. The other standardized (Tables 13-15) differences exhibit similar patterns.

Figure 3 shows the distribution of both the treated and non-treated students. We can visually see the significant overlap between the two populations. Given this, we can proceed to estimate the

[16]The AP Physics B, discontinued in 2014, is the equivalent of an introductory Physics college course. It was, later on, replaces by AP Physics 1 and 2. The AP Physics C: Mechanics test studies Newtonian mechanics, while the AP Physics C: Electricity and Magnetism test studies electricity and magnetism.

[17]The AP Calculus AB is the equivalent of the first Calculus course taken in college and includes topics such as limits, derivatives, definite integrals, and the Fundamental Theorem of Calculus. In comparison, the AP Calculus BC (CALBC) is a more advanced test and in addition to the topics included in the AP Calculus AB test it includes topics such as sequences and series.

[18]The balancing property is not equivalent to the CIA and vice-versa [Smith and Todd, 2005].

[19]This threshold was calculated by dividing 20 by 100.

ATET for our two outcomes and two treatments considered.

4 Results

Table 6 shows the estimates for the impact of grant employment on both general and STEM graduation rates. The estimates from Table 6 show an impact of about 10.1 percentage points of grant employment on STEM graduation rates and an impact of 5.5 percentage points on graduation rates. The nearest neighbor with replacement results with one neighbor are shown in Table 7 and they are very similar to the IPW results. For the remainder of the paper, we only focus on the IPW results, since this similarity is maintained for throughout our different analyses.

For an easier analysis of the results from Table 6 we combine all the results for the full sample in Figure 4. Our estimates suggest while the nature of the grant employment doesn't matter much for general graduation rates, it makes a big difference for STEM graduation rates. Research jobs have a positive impact of 13.6 percentage points on STEM graduation rates, much higher than the effect of all grant employment on STEM graduation rates. These results suggest that while all types of employment on a federally funded grant improve both graduation and STEM graduation rates, the research intensity of employment also matters and future research should also take this into account (Figure 4).Furthermore, the impact of grant employment is much higher for STEM graduation rates than general graduation rates. One possible explanation of this is the STEM specific human capital that research intensive positions provide and that other positions do not. Working in a laboratory, as opposed to working as an administrative person would facilitate the acquisition of STEM specific skills and knowledge. Another possible explanation is that more research intensive job provide the students with a higher level of social capital. Thus, the connections built with a faculty member might be closer for a job more related to their own research which would help the students navigate through their major in an easier way.

We are also interested in the heterogeneity of the treatment effect. When breaking down the analysis by the gender of the students, we can see from Table 6 that research employment improves persistence in college for both men and women, but that men benefit more from grant employment jobs than women. Figures 5 and 6 show the graphical representation of our results for gender, broken down by type of student employment. In fact, the impact for female students of any grant employment is almost identical for graduation rates and STEM graduation rates, while for male students the impact is much higher for STEM graduation rates (Figure 5). Even when just considering the research intensive jobs, female students do not gain much more in terms of STEM graduation rates as compared to general graduation rates, while male students do (Figure 6). While the gender of the student doesn't seem to create a big divide for graduation rates based on the type of job the student held, it makes does matter for STEM graduation rates (Figures 7 and 8). Research intensive jobs are have a higher positive impact on STEM graduation rates, and the effect is much higher for men than woman. The fact that women do not gain as much from grant employment as men in terms of persistence could be linked to the lower accumulation of human and social capital needed to steer women into STEM majors. Perhaps a stronger intervention is needed to increase female participation in STEM majors. This result can be tied to the stereotype threat literature, which states that women might not be able to act in accordance to their abilities in fear they might reinforce the negative stereotypes associated with their identity [Steele and Aronson, 1995]. In this context, having a research-intensive job might help women succeed in STEM more by giving them more confidence in their ability to be a STEM major. However, a research intensive job might not be enough to narrow the gender gap in STEM graduation rates.

Another result of interest is the heterogeneous effect of research experience based on the race of the student (also shown in Table 6). First, we are interested in whether the research intensity of the job impacted the outcomes differently for different races. Figures 9 and 10 show that white and Asian students benefit the most from working with a faculty member on a research grant. For blacks and Hispanics, working with a faculty member doesn't increase graduation rates as much as for the other races. Furthermore, blacks and Hispanics benefit more from working with a faculty member in terms of general graduation rates as compared to STEM graduation rates. Figures 11 and 12 show the results broken down by race and research intensity of grant employment. The type of job a student holds makes little difference for graduation rates, but not for STEM graduation rates. In fact, not all races benefit more from a research intensive job. A research intensive job has a smaller effect on graduation rates than all jobs for blacks and Hispanics. Interestingly, Hispanics benefit more from all jobs than research intensive jobs for STEM graduation. However, blacks benefit much more from research jobs for STEM graduation. This analysis suggests that the gains from grant employment are highly dependent on the race of the student. It is possible that black and Hispanic students are more financially constraint and thus benefit more from any type of job as opposed to a more research intensive job. Another explanation of the results is connected to the theory of identity invoked above. Perhaps, once again, student employment is not a sufficient force to move more underrepresented minority students into STEM majors. It is possible that the social networks and social support acquired through working with a faculty member vary vastly for students of various races.

We are also interested in the degree of financial capital accumulated and the impact of this on various outcomes. We can see from Table 6 that grant employment increases graduation rates for students with Pell grants by 7.2 percentage points and it increases STEM graduation rates by 6.3 percentage points. To compare these effects with the effects for the full sample, we focus on Figures 13 - 15. We conclude from Figures 15 and 16 that having a job increases graduation rates for Pell grant recipients by more than for the full sample, but the opposite is true for STEM graduation. The fact that we get a different result for STEM graduation suggests that, although having a job decreases the cost of education, there might be other factors involved in getting a STEM graduation that we are not accounting for.

Figures 13 and 14 show that Pell grant recipients benefit more than the full sample from grant employment in terms of increased general graduation rates, but less in terms of STEM graduation rates. One explanation is the fact that the financial channel is more salient than other channels for financially constraints students. Thus, holding a grant funded position helps students persist in college, but doesn't necessarily cause students to pursue a STEM education. Research jobs seem to be more important for Pell graduates for fostering higher STEM graduation rates than general graduation rates. Thus, we do find evidence of STEM specific capital accumulation, but it seems that income matters absolutely and relatively more to Pell students.

4.1 Inverse-probability-weighted regression adjustment estimator

We employ one last matching estimator for our analysis to check the robustness of our results to the estimator choice. The inverse-probability-weighted regression adjustment (IPWRA) estimator uses inverse probability weighting (IPW) weights when performing regression adjustment. It combines models for both the outcome and the treatment status. Wooldridge [2010] shows that this estimator is doubly robust meaning that the estimates of the effects will be consistent if either the treatment model or the outcome model, but not both, are misspecified.[20] Table 7 shows the results using the IPWRA estimator for grant employment. The results are very similar to the results obtained with the IPW estimator and they show an average treatment of the treated effect of grant employment on STEM graduation of 10.4 percentage points and on general graduation of 5.5 percentage points. For research intensive jobs, the results show an effect of 5.6 percentage points and of 14.0 percentage points on STEM graduation rates.

4.2 Sensitivity analysis

We are also interested in how robust our estimates are to different sources of biases. One source of bias that could arise is due to the timing of research employment. Since the students who are

[20]We could not find any studies that test the double robust property when both of the models are misspecified.

doing research later in their academic career are also more likely to graduate, we are concerned that our estimates might be biased. As noted before, the majority of the students considered (70 percent) have their first grant employment opportunity in their first year of college. We also remove all instances of grant employment in the students' senior year since these students are very likely to graduate.

Another issue that arises is that, by using a propensity matching technique, the selection problem is not necessarily fixed. The credibility of the matching procedure used relies entirely on the conditional independence assumption, which assumes that we observe all the variables that have an impact on research experience and graduation rates. We are interested in the sensitivity of our estimates with respect to confounding factors, i.e., unobserved variables that might influence both assignment to research experience and the likelihood of graduating in STEM.

While not possible to estimate the magnitude of selection bias without experimental data, it is possible to check the sensitivity of the estimated results with respect to deviations from the conditional independence assumption [Aakvik, 2001]. Thus, we are interested in how unobserved covariates that affect both our treatments and our outcomes would alter our conclusions. Such unobserved variables could be motivation, future career aspirations, or any other factor that affects both the probability of taking part in undergraduate research experience, as well as the probability of graduation in general or in STEM). For example, a student who shows more motivation towards his or her studies might have a higher chance of getting hired to work with a faculty member, and also might have a higher chance of graduating from college. This section informs us how our results would change if we had such unobservable characteristics that would affect both our treatments and outcomes.

As a first step, we assume that the relationship between treatment and observable is related to the relationship between treatment and unobservables [Altonji et al., 2005]. Our study includes a broad set of variables that cover socioeconomic characteristics, as well as data on the courses the students take and their high explanatory power suggests that the observable characteristics could provide useful information about the unobservable characteristics.

In our analysis, we control for pre-interest in STEM fields by including AP tests and intent to major in STEM. Given that AP tests are highly predictive of research experience, we re-run our propensity score matching procedure by excluding AP tests, as shown in Table 8. The base estimates are the IPW estimates including AP tests, while the sensitivity analysis estimates are the ones excluding AP tests. The results show that excluding AP tests increases our estimates of the ATET effect of all types of grant employment on graduation rates from 5.5 percentage points to 5.8 percentage points. We get a similar increase in the ATET effect of all grant employment on STEM graduation rates, where by excluding the AP tests controls, our IPW estimates increase from 10.1 percentage points to 11 percentage points.

In general, these results suggest that if there were other factors such as motivation, that were equally as important in determining both employment and the outcomes in question as having an AP test, the exclusion of these other factors would bias upward our estimated effects, but by a relatively small magnitude.

Another approach to examine if our estimates are sensitive to biases is to calculate bounds for the degree to which the unobserved variables affect selection into treatment [Rosenbaum, 2002]. With the notation used before, we have our binary outcome variable Y, our binary treatment T and a covariate vector X. Following the model in Aakvik [2001] and Becker and Caliendo [2007], we define probability of receiving treatment $\pi_i = Pr(x_i, u_i) = Pr(T_i = 1|x_i, u_i) = F(\beta x_i + \gamma u_i)$, where x_i are the observed characteristics for person i, while u_i are the unobserved characteristics. Here, γ is the effect that unobserved characteristics have on the treatment. In the case of no hidden bias, γ is zero, but in the presence of hidden bias, individuals with the same exact observed characteristics will have different probabilities of receiving treatment.

Assuming a matched pair of people i and j, and a logistic regression F, the odds that the individuals receive treatment are $\frac{\pi_i}{1-\pi_i}$ and $\frac{\pi_j}{1-\pi_j}$, respectively and the odds ratio is:

$$\frac{\frac{\pi_i}{1-\pi_i}}{\frac{\pi_j}{1-\pi_j}} = \frac{\pi_i(1-\pi_j)}{\pi_j(1-\pi_i)} = \frac{exp(\beta x_i + \gamma u_i)}{exp(\beta x_j + \gamma u_j)} \tag{6}$$

The matching procedure implies that $x_i = x_j$, so we have:

$$\frac{exp(\beta x_i + \gamma u_i)}{exp(\beta x_j + \gamma u_j)} = exp(\gamma(u_i - u_j)) \tag{7}$$

When the odds ratio is one, there is no hidden bias. This happens when either the unobserved variables are the same ($u_i = u_j$), or when the unobserved variables do not influence selection into treatment ($\gamma = 0$). Assuming that the unobserved variable is a binary variable, we get the following bounds for the odds ratio [Rosenbaum, 2002]:

$$\frac{1}{e^\gamma} \leq \frac{\pi_i(1-\pi_j)}{\pi_j(1-\pi_i)} \leq e^\gamma \tag{8}$$

If $e^\gamma = 1$, then both individuals i and j have the same probability of being assigned the treatment. Here, e^γ measures the amount of departure from the case where we have no hidden bias [Rosenbaum, 2002]. The equation above states that if two individuals differ in terms of unobservables, then their probability of receiving treatment depends on γ and the difference in u.

Thus, for a fixed $e^{\gamma} \geq 1$ and $u \in \{0, 1\}$, it can be shown that the test statistic Q_{MH} can be bounded by two known distributions [Rosenbaum, 2002].[21] In the case when $e^{\gamma} = 1$, the bounds are equal to the value of the test statistic. For values of e^{γ} greater than 1, the upper and lower bounds diverge as a consequence of unobserved selection bias creating uncertainty about the test statistics.

There are two STATA commands that have been developed for calculating these bounds: one for continuous outcomes, rbounds [DiPrete and Gangl, 2004], and one for binary outcomes, mhbounds [Becker and Caliendo, 2007]. Since this paper deals with dichotomous outcomes, we employ the latter STATA command which uses the Mantel and Haenszel test statistic. The test can be used to test for no treatment effect both within different strata of the sample and as a weighted average between the strata. In our example, we use a weighted average of the strata.

Following Aakvik [2001], we introduce additional notation to be able to apply this test. The outcome y is observed for both treatment and control groups, which under the null-hypothesis has a hypergeometric distribution. Furthermore, we define N_{1s} to be the number of treated individuals in stratum S and N_{0s} to be the number of untreated individuals in stratum s, so that $N_S = N_{1s} + N_{0s}$. In addition, Y_{1s} is the number of successful participants and Y_{0s} is the number of unsuccessful participants, which gives Y_s as the number of total successes in stratum s. With this notation in mind, the test statistic Q_{MH} is:[22]

$$\begin{aligned} Q_{MH} &= \frac{\mid Y_1 - \sum_{s=1}^{S} E(Y_{1s}) \mid -0.5}{\sqrt{\sum_{s=1}^{S} Var(Y_{1s})}} \\ &= \frac{\mid Y_1 - \sum_{s=1}^{S} \left(\frac{N_{1s} Y_S}{N_S} \right) \mid -0.5}{\sqrt{\sum_{s=1}^{S} \frac{N_{1s} N_{0s} Y_S (N_S - Y_S)}{N_S^2 (N_s - 1)}}} \end{aligned} \tag{9}$$

We denote Q^{+}_{MH} to be the test statistic in the case where we overestimated the treatment effect

[21] This test statistics can only be used after a matching procedure is performed since the individuals in the treatment and control groups have to be very similar to each other.

[22] This test statistic follows asymptotically the standard normal distribution.

and Q^-_{MH} to be the test statistic in the case where we underestimated the treatment effect. Thus, Q^+_{MH} statistic adjusts the MH test statistic downward for positive unobserved selection, while the Q^-_{MH} statistic adjusts the MH statistic upwards for negative unobserved bias. Then we have:

$$Q^+_{MH} = \frac{|Y_1 - \sum_{s=1}^{S} \tilde{E}^+_S| - 0.5}{\sqrt{\sum_{s=1}^{S} Var(\tilde{E}^+_S)}} \tag{10}$$

and

$$Q^-_{MH} = \frac{|Y_1 - \sum_{s=1}^{S} \tilde{E}^-_S| - 0.5}{\sqrt{\sum_{s=1}^{S} Var(\tilde{E}^-_S)}} \tag{11}$$

where $\tilde{E}_S$ and $Var(\tilde{E}_S)$ are the large-sample approximations to the expectation and variance of the number of participants in treatment.

To implement this procedure, we re-estimate the treatment effects using one-nearest-neighbor matching,[23] for different values of γ. The sensitivity analysis performed informs us how biases might influence our estimates, but does not inform us if biases exist.

Table 2 shows that the students are positively selected into treatment based on observed characteristics. More specifically, students with higher high school GPAs, students who took AP science courses and students who have a higher ACT composite score are more likely to be treated. Positive observed selection into treatment does not imply positive unobserved selection into treatment, but it does inform us that some unobserved factors would confound our estimated effect. Given that our estimated effect of research experience is positive, we are worried about overestimating our treatment effect and thus we are only interested in the bias related to overestimation of the treatment effect.

Tables 9-12 show the sensitivity of the test statistic for e^γ, as well as for the test statistic $\Gamma = e^\gamma = 1$, the case with no hidden bias. All p-values are based on one-sided significance tests. The first column in the table represents $\Gamma = e^\gamma \geq 1$ for which the sensitivity analysis is carried out.

For the analysis of all grant employment on graduation rates, Table 9 shows that the test statistic

[23]The estimates for the treatment effect using this procedure are slightly different from the estimates in the results section, due to the use of a different matching algorithm.

becomes not significant at the 5 percent level when the relative odds of working on any federally funded grant with a faculty member are 3. Thus, we would need an unobserved variable u that increases treatment by 30 percent to make the relationship non-significant at a 5 percent significance level.

To interpret these results, we can compare the estimates from our sensitivity analysis with the results from logit models predicting treatment, the ones we estimated for our propensity score analysis. To put our sensitivity analysis results into perspective, Table 3 shows that the movement from not being eligible for a Pell grant to being eligible shifts the relative odds ratio by 1.4. To completely get rid of the effect of all grant employment on graduation rates, an unobserved variable would have to have almost as large of an effect as Pell grant eligibility, net of all the control variables we already include. Thus, this test provides evidence that a high amount of selection into unobservables is needed to eliminate the treatment effects. The treatment effect stays positive for values of $\Gamma \leq 1.4$, after which point it becomes negative due to a large positive unobserved characteristics. Furthermore, we see from the same table that for values of $\Gamma \geq 1.6$, the test statistic becomes significant again at the standard 5 percent level.

Table 10 shows the sensitivity analysis results for the effect of research-intensive jobs on graduation rates. The effect at $\Gamma = 1$ is significant and stays significant until $\Gamma = 1.2$, when it is not even significant at 10 percent significance level. For $\Gamma = 1.2$, two students with the same observable characteristics differ in the odds of participating in research experience by a factor of 1.2, which is a very large number considering we have already adjusted for important student characteristics. Thus, the unobservable characteristics would have to increase the probability of receiving a research intensive job by 20 percent in order for the effects of research-intensive grant employment on graduation rates to stop being significant. This corresponds to a significant negative treatment effect, caused by large positive unobserved characteristics. Table 9 also shows that $\Gamma = 1.35$ is the point where the treatment effect changes signs, going from positive to negative. In addition to this, the test statistic becomes significant again at 5 percent level once Γ exceeds 1.45, but the treatment effect becomes negative.

We again perform a back-of-the-envelope calculation to see which covariates from our propensity score regressions produce coefficients similar to the ones from the sensitivity analysis. In a similar way, we see from Table 4 that the "weakest" unobservable characteristics we would need to have to render the effect of research intensive jobs on graduation rates as not significant would have to be at least as large as the effect of taking the AP Biology test. Again, the unobservable characteristics would have to be variables that are not already included in our analysis.

Tables 11-12 present the Mantel-Haenszel bounds for the case where we consider STEM graduation as the outcome. For both tables, even the relative odds of getting treatment are 1.5, the test statistic p^+_{MH} is still statistically significant at a 5 percent significance level. When considering all grant employment as the treatment variable, Table 11 shows that the test statistic p^+_{MH} is statistically significant at 5 significance level for all values considered, except for the values for Γ between 1.55 and 1.8. Thus, for values of the odds ratio between 1.55 and 1.8, our inferences would become not significant at 5 percent significance level. In addition to this, for $\Gamma = 1.7$, the treatment effect changes signs, going from positive to negative. Table 12 shows that our estimates of the effect of research-intensive jobs on STEM graduation would become statistically not significant at a 5 percent significance level for values of Γ between 1.7 and 2. The treatment effect at $\Gamma = 1.85$, although insignificant, remains positive. For values of $\Gamma \geq 1.85$, the treatment effect becomes negative.

These results are reassuring that while we cannot rule out the possibility of hidden biases, we can reassure ourselves that the unobservable characteristics would have to be quite large to make our results not significant. In both cases, it is quite unlikely to have unobservable characteristics that would switch the relative odds of receiving treatment by 1.55, and 1.7, respectively, given that switching the race or the gender of the student explains a much smaller change in the odds ratio. The results for considering STEM graduation rate as the outcome are less sensitive to unobservable characteristics than the results for general graduation rate as the outcome.

Mantel-Haenszel bounds for all the other subsamples are provided in Tables 16-19. One way to read our results is by looking at the smallest value of Γ for which the effect of our treatment variables stops being statistically significant at 5 percent significance level, which we define Γ'. We also define Γ'' as the highest value of $\Gamma = e^\gamma$ for which the test statistic p^+_{MH} is still not significant at a 5 percent significance level. Thus, in the Table 12 from before, Γ' is equal to 1.7 and Γ'' is equal to 2.

The results are fairly robust to the possible presence of unobservable characteristics. For example, Table 16 shows that for the effect of all grant employment on graduation rates for female students to go away, we would need an unobserved variable that would multiply the odds of treatment by 1.25. As explained in the analysis above, this effect is fairly large given that it is net of all the controls we included. In general, the tables suggest that the estimates for Black and Hispanic students have larger intervals $[\Gamma',\Gamma'']$ where the test statistic is not significant at 5 percent level. However, all of the subgroups considered are partially robust to selection bias. This implies that our estimates for these two groups are more sensitive to nonobservable characteristics, a result due in part to the lower representation of these groups in the overall student population. Another result is that the subgroup analysis is less robust to unobservables for general graduation rates than it is for STEM graduation rates. Thus, we need to be more careful when interpreting the estimates of our matching procedure when considering general graduation rates.

5 Conclusion and future research

The factors that impact persistence in STEM are of importance to the society due to the role of STEM graduates in technological advancement. This paper provides insights into the policy implications of research productivity by analyzing the role of grant employment on persistence in STEM. We use a unique dataset that combines administrative student transcript data with longitudinal administrative data on research funding at a public research institution. This innovative data allows us to track all employment and courses history for the students who attended this large university over 2001-2014.

Using this data, we quantify the impact of undergraduate employment, defined as having been employed on a federally funded grant at a large public university, on persistence in STEM. Using our preferred IPW estimator, we find a positive and significant impact of approximately 10.1 percentage points of grant employment on STEM graduation rates and an impact of approximately 5.5 percentage points on general graduation rates. In addition, this study evaluates the effect of grant employment on different subgroups. Given the large disparities among the students who select to major in STEM, it is important to study how research experience affects these different groups.

We actually find that the race and gender gaps are not narrowed by much and that the positive impact of grant employment is largest for male students and whites. Even though minorities and female students benefit from grant employment, the benefits are much lower across all measures used. Furthermore, we find that for students receiving who qualify for Pell grants and thus have more financial constraints, working with a faculty member helps increase their graduation rates significantly.

Finally, this study provides estimates of research employment for all types of grant employment, as well as more research-intensive positions. In particular, we divide research employment by the amount of research intensity. We define research jobs as the jobs that are related to the student's science career, based on the job description of the university's HR department. Our findings show that the research intensity matters for persistence, with more research intensive jobs having a larger impact on STEM graduation rates.

Given the overall positive effects of research experience, we suggest the implementation of policies that would increase the research opportunities of undergraduate students at academic institutions. We suggest that future research should focus primarily on the mechanisms that explain this positive impact of research experience and the reasons behind the subgroup differences in benefits.

While more employment opportunities should be available to female students and minority students together, this policy should be coupled with other policies aimed at improving the experience of underrepresented groups in STEM fields. In the future, we plan to extend our analysis to take into account the gender and race composition of the research teams that the students are part of. We envision extending our current research to investigate research collaborations in more depth, given that we have access to rich information on all the federally funded grants and their recipients. We also plan to investigate the effects of undergraduate research on longer-term student outcomes, such as graduate school attendance and labor market outcomes.

References

A. Aakvik. Bounding a matching estimator: the case of a Norwegian training program. *Oxford bulletin of economics and statistics*, 63(1):115–143, 2001.

A. Abadie and G.W. Imbens. A martingale representation for matching estimators. *Journal of the American Statistical Association*, 107(498):833–843, 2012.

G.A. Akerlof and R.E. Kranton. Identity and schooling: Some lessons for the economics of education. *Journal of Economic Literature*, 40(4):1167–1201, 2002.

J.G. Altonji, T.E. Elder, and C.R. Taber. An evaluation of instrumental variable strategies for estimating the effects of Catholic schooling. *Journal of Human Resources*, 40(4):791–821, 2005.

B. Augurzky and C.M. Schmidt. The propensity score: A means to an end. IZA Discussion Paper No. 271, 2001.

N.R. Augustine. Rising above the gathering storm: Energizing and employing America for a brighter economic future. *Washington DC: the National Academies Press*, 19:2007, 2007.

A.E.L. Barlow and M. Villarejo. Making a difference for minorities: Evaluation of an educational enrichment program. *Journal of Research in Science Teaching*, 41(9):861–881, 2004.

K.W. Bauer and J.S. Bennett. Alumni perceptions used to assess undergraduate research experience. *The Journal of Higher Education*, 74:210–230, 2003.

S.O. Becker and M. Caliendo. Sensitivity analysis for average treatment effects. *The Stata Journal*, 7(1):71–83, 2007.

E.P. Bettinger, B.T. Long, P. Oreopoulos, and L. Sanbonmatsu. The role of simplification and information in college decisions: Results from the H&R Block FAFSA experiment. Technical report, National Bureau of Economic Research, 2009.

D. Black, K. Daniel, and J. Smith. College quality and wages in the United States. *German Economic Review*, 6(3):415–443, 2005.

N.A. Bowman. Can 1st-year college students accurately report their learning and development? *American Educational Research Journal*, 47(2):466–496, 2010.

A. Bryson, R. Dorsett, and S. Purdon. The use of propensity score matching in the evaluation of labour market policies. Working Paper No. 4, Department for Work and Pensions, 2002.

M. Busso, J. DiNardo, and J. McCrary. New evidence on the finite sample properties of propensity score reweighting and matching estimators. *Review of Economics and Statistics*, 96(5):885–897, 2014.

M.J. Chang, J. Sharkness, S. Hurtado, and C.B. Newman. What matters in college for retaining aspiring scientists and engineers from underrepresented racial groups. *Journal of Research in Science Teaching*, 51(5):555–580, 2014.

X. Chen. STEM Attrition: College students' paths into and out of STEM Fields (NCES 2014-001). *Washington, Dc: US. Department of Education and National Center for Education Statistics*, 2013.

D. Cole and A. Espinoza. Examining the academic success of latino students in science technology engineering and mathematics (STEM) majors. *Journal of College Student Development*, 49(4): 285–300, 2008.

D. Deming and S. Dynarski. Into college, out of poverty? Policies to increase the postsecondary attainment of the poor. Technical report, National Bureau of Economic Research, 2009.

T.A. DiPrete and C. Buchmann. *The rise of women: The growing gender gap in education and what it means for American schools*. Russell Sage Foundation, New York, 2013.

T.A. DiPrete and M. Gangl. Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments. *Sociological Methodology*, 34(1):271–310, 2004.

D. Dunning, K. Johnson, J. Ehrlinger, and J. Kruger. Why people fail to recognize their own incompetence. *Current directions in Psychological Science*, 12(3):83–87, 2003.

R.G. Ehrenberg. Graduate education, innovation and federal responsibility. *Communicator*, 38: 1–8, 2005.

R.G. Ehrenberg. Analyzing the factors that influence persistence rates in STEM field, majors: Introduction to the symposium. *Economics of Education Review*, 29(6):888–891, 2010.

R. Ellington. *Having their say: Eight high-achieving African-American undergraduate mathematics majors discuss their success and persistence in mathematics*. PhD thesis, University of Maryland, College Park, 2006.

P. England and S. Li. Desegregation stalled the changing gender composition of college majors, 1971-2002. *Gender & Society*, 20(5):657–677, 2006.

P. England, P. Allison, S. Li, N. Mark, J. Thompson, M.J. Budig, and H. Sun. Why are some academic fields tipping toward female? The sex composition of US fields of doctoral degree receipt, 1971-2002. *Sociology of Education*, 80(1):23–42, 2007.

D.F. Feldon, M.A. Maher, M. Hurst, and B. Timmerman. Faculty mentors, graduate students, and performance-based assessments of students research skill development. *American Educational Research Journal*, 52(2):334–370, 2015.

M. Frölich. Finite sample properties of propensity-score matching and weighting estimators. *Review of Economics*, 86:77–90, 2004.

J.G. Gayles and F. Ampaw. The impact of college experiences on degree completion in STEM fields at four-year institutions: Does gender matter? *The Journal of Higher Education*, 85(4): 439–468, 2014.

C. Goldin and L.F. Katz. Transitions: Career and family life cycles of the educational elite. *The American Economic Review*, 98(2):363–369, 2008.

K. Goyette and Y. Xie. Educational expectations of Asian American youths: Determinants and ethnic differences. *Sociology of Education*, pages 22–36, 1999.

M.J. Graham, A. Byars-Winston, A.B. Hunter, and J. Handelsman. Science education: Increasing persistence of college students in STEM. *Science*, 341:1455–1456, 2013.

J. Grandy. Persistence in science of high-ability minority students: Results of a longitudinal study. *The Journal of Higher Education*, 69(6):589–620, 1998.

S.R. Gregerman, J.S. Lerner, W. von Hippel, J. Jonides, and B.A. Nagda. Undergraduate student-faculty research partnerships affect student retention. *The Review of Higher Education*, 22(1): 55–72, 1998.

World Bank Group. *World Development Indicators 2014*. World Bank Publications, 2014.

R.S. Hathaway, B.A. Nagda, and S.R. Gregerman. The relationship of undergraduate research participation to graduate and professional education pursuit: An empirical study. *Journal of College Student Development*, 43(5):614–631, 2002.

J. Heckman and R. Robb. *Longitudinal Analysis of Labor Market Data*, chapter Alternative Methods for Evaluating the Impact of Interventions, pages 156–246. New York: Cambridge University Press, 1985.

J. Heckman, H. Ichimura, and P. Todd. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies*, 64(4):605–654, 1997.

J.K. Hellerstein and M.S. Morrill. Dads and daughters the changing impact of fathers on women's occupational choices. *Journal of Human Resources*, 46(2):333–372, 2011.

P.R. Hernandez, A. Woodcock, M. Estrada, and P.W. Schultz. Undergraduate research experiences broaden diversity in the scientific workforce. *BioScience*, 68(3):204–211, 2018.

F.A. Herrera and S. Hurtado. Maintaining initial interests: Developing science, technology, engineering, and mathematics (STEM) career aspirations among underrepresented racial minority students. In *Association for Educational Research Annual Meeting, New Orleans, LA*, 2011.

D. Hossler, M. Ziskin, J.P.K. Gross, S. Kim, and O. Cekic. Student aid and its role in encouraging persistence. In *Higher education: Handbook of Theory and Research*, pages 389–425. 2009.

S. Hu, G.D. Kuh, and J.G. Gayles. Engaging undergraduate students in research activities: Are research universities doing a better job? *Innovative Higher Education*, 32(3):167–177, 2007.

M. Huber, M. Lechner, and C. Wunsch. The performance of estimators based on the propensity score. *Journal of Econometrics*, 175:1–21, 2013.

A.B. Hunter, S.L. Laursen, and E. Seymour. Becoming a scientist: The role of undergraduate research in students' cognitive, personal, and professional development. *Science Education*, 91 (1):36–74, 2007.

S. Hurtado, N.L. Cabrera, M.H. Lin, L. Arellano, and L.L. Espinosa. Diversifying science: Underrepresented student experiences in structured research programs. *Research in Higher Education*, 50(2):189–214, 2009.

J.S. Hyde, E. Fennema, and S.J. Lamon. Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, 107(2):139–155, 1990.

C.M. Kardash. Evaluation of an undergraduate research experience: Perceptions of undergraduate interns and their faculty mentors. *Journal of Educational Psychology*, 92(1):191–201, 2000.

M.M. Kim, G. Rhoades, and D.B. Woodard Jr. Sponsored research versus graduating students? Intervening variables and unanticipated findings in public research universities. *Research in Higher Education*, 44(1):51–81, 2003.

J. Kinkead. Learning through inquiry: An overview of undergraduate research. *New Directions for Teaching and Learning*, 93:5–17, 2003.

M.C. Linn, E. Palmer, A. Baranger, E. Gerard, and E. Stone. Undergraduate research experiences: Impacts and opportunities. *Science*, 347(6222), 2015. art. 1261757.

D. Lopatto. Survey of undergraduate research experiences (sure): First findings. *Cell Biology Education*, 3(4):270–277, 2004.

D. Lopatto. Undergraduate research as a high-impact student experience. *Peer Review*, 12(2): 27–30, 2010.

A.V. Maltese and R.H. Tai. Eyeballs in the fridge: Sources of early interest in science. *International Journal of Science Education*, 32(5):669–685, 2010.

A.V. Maltese and R.H. Tai. Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among US students. *Science Education*, 95(5):877–907, 2011.

A. Mann and T.A. DiPrete. Trends in gender segregation in the choice of science and engineering majors. *Social science research*, 42(6):1519–1541, 2013.

S.A. Maple and F.K. Stage. Influences on the choice of math/science major by gender and ethnicity. *American Educational Research Journal*, 28(1):37–60, 1991.

W.C. Mau. Factors that influence persistence in science and engineering career aspirations. *The Career Development Quarterly*, 51(3):234–243, 2003.

J. Mervis. Biomedical training programs: NIH told to get serious about giving minorities a hand. *Science*, 311:328–329, 2006.

J.D. Miller and L.G. Kimmel. Pathways to a STEMM profession. *Peabody Journal of Education*, 87(1):26–45, 2012.

NCSES. Women, minorities, and persons with disabilities in science and engineering: 2017, 2017. URL www.nsf.gov/statistics/wmpd.

NSB. *Science and engineering indicators 2016 (NSB-2016-1)*. Arlington, VA: National Science Foundation, 2016.

S. Olson and D.G. Riordan. Engage to excel: Producing one million additional college graduates with degrees in Science, Technology, Engineering, and Mathematics. Report to the President. *Executive Office of the President*, 2012.

E.T. Pascarella and P.T. Terenzini. Interaction effects in Spady and Tinto's conceptual models of college attrition. *Sociology of Education*, 52:197–210, 1979.

E.T. Pascarella and P.T. Terenzini. *How college affects students: A third decade of research (Vol. 2)*. San Francisco: Jossey-Bass, 2005.

M. Pender, D.E. Marcotte, M.R.S. Domingo, and K.I. Maton. The STEM pipeline: The role of summer research experience in minority students' Ph.D. aspirations. *Education Policy Analysis Archives*, 18(30):1–19, 2010.

P. Rosenbaum and D. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

P.R. Rosenbaum. Observational studies. In *Observational Studies*, pages 1–17. Springer, second edition, 2002.

P.R. Rosenbaum and D.B. Rubin. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39:33–38, 1985.

S.H. Russell, M.P. Hancock, and J. McCullough. The Pipeline: Benefits of undergraduate research experiences. *Science*, 316(5824):548–549, 2007.

C.L. Ryan and K. Bauman. Educational attainment in the united states: 2015, 2016. URL https://www.census.gov/content/dam/Census/library/publications/2016/demo/p20-578.pdf.

D.A. Sabatini. Teaching and research synergism: The undergraduate research experience. *Journal of Professional Issues in Engineering Education and Practice*, 123(3):98–102, 1997.

T. Sadler, S. Burgin, L. McKinney, and L. Ponjuan. Learning science through research apprenticeships: A critical review of the literature. *Journal of Research in Science Teaching*, 47:235–256, 2010.

B. Schneider, C.B. Swanson, and C. Riegle-Crumb. Opportunities for learning: Course sequences and positional advantages. *Social Psychology of Education*, 2:25–53, 1998.

J.E. Scott-Clayton and V. Minaya. Should student employment be subsidized? Conditional counterfactuals and the outcomes of work-study participation. *Economics of Education Review*, 52: 1–18, 2016.

J.E. Scott-Clayton and Y. Zhou. Does the federal work-study program really work–and for whom? 2017. URL `https://ccrc.tc.columbia.edu/media/k2/attachments/capsee-does-fws-program-really-work.pdf`.

E. Seymour and N.M. Hewitt. *Talking about Leaving: Why Undergraduates Leave the Sciences*. Westview Press Boulder, Colorado, 1997.

E. Seymour, A.B. Hunter, S.L. Laursen, and T. DeAntoni. Establishing the benefits of research experiences for undergraduates in the sciences: first findings from a three-year study. *Science Education*, 88(4):493–534, 2004.

J. Smith and P. Todd. Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics*, 125(1–2):305–353, 2005.

C.M. Steele and J. Aronson. Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5):797–811, 1995.

E.A. Stuart. Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25(1):1, 2010.

R.T. Tai, C.Q. Liu, A. Maltese, and X.T. Fan. Planning early for careers in science. *Science*, 312 (5777):1143–1144, 2006.

P.T. Terenzini and E.T. Pascarella. Voluntary freshman attrition and patterns of social and academic integration in a university: A test of a conceptual model. *Research in Higher Education*, 6(1): 25–43, 1977.

V. Tinto. *Leaving College: Rethinking the Causes and Cures of Student Attrition*. University of Chicago Press, Chicago, IL, 2nd edition, 1993.

H. Ullah and M.A. Wilson. Students' academic success and its association to student involvement with learning and relationships with faculty and peers. *College Student Journal*, 141(4):1192–1202, 2007.

US Department of Health and Human Services. National institute of general medical sciences 5-year strategic plan, 2015. URL `https://publications.nigms.nih.gov/strategicplan/NIGMS-strategic-plan.pdf`.

C. Ward, J.S. Bennett, and K.W. Bauer. Content analysis of undergraduate research student evaluations. Retrieved June 2015, 2003. URL `https://www.udel.edu/RAIRE/Content.pdf`.

S.G West, N. Duan, W. Pequegnat, P. Gaist, D.C. Des Jarlais, D. Holtgrave, J. Szapocznik, M. Fishbein, B. Rapkin, and M. Clatts. Alternatives to the randomized controlled trial. *American journal of public health*, 98(8):1359–1366, 2008.

J.M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. MIT press, 2010.

Y. Xie and A.A. Killewald. *Is American Science in Decline?* Cambridge: Harvard University Press., 2012.

A.L. Zydney, J.S. Bennett, A. Shahid, and K.W. Bauer. Impact of undergraduate research experience in engineering. *Journal of Engineering Education*, 91(2):151–157, 2002.

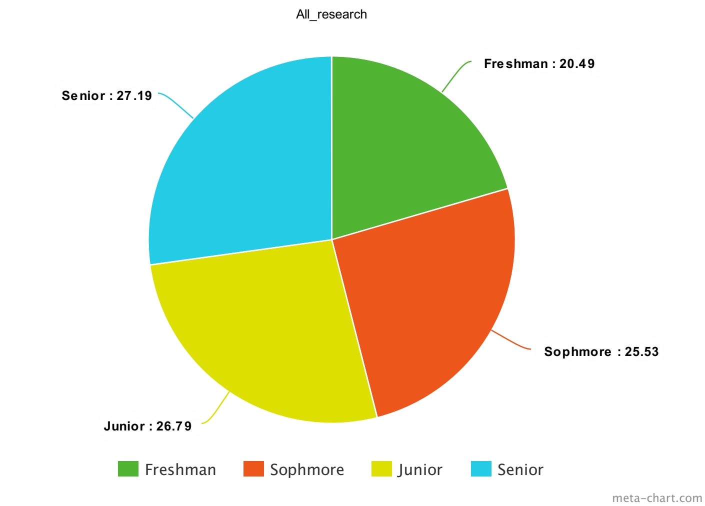
A Tables and Figures

Table 1: Literature review on undergraduate research experience

Study	Data	Type of research experience
Hurtado et al. [2009]	Focus groups (four research universities)	Structured science research programs
Barlow and Villarejo [2004]	University of California, Davis	Biology Undergraduate Scholars Program
Bauer and Bennett [2003]	Survey from University of Delaware	Undergraduate Research Program
Chang et al. [2014]	The Freshman Survey and College Senior Survey	Grants from NIH and NSF
Seymour et al. [2004]	Student interviews from four liberal arts colleges	Undergraduate research experience
Hunter et al. [2007]	Ethnographic study at four liberal arts colleges	Summer science undergraduate research experiences
Kardash [2000]	Midwestern, Carnegie Research I university	Undergraduate science research
Russell et al. [2007]	Web-based surveys	Undergraduate research opportunity (NSF)
Kim et al. [2003]	22 public research universities	R&D expenditures from NSF's CASPAR data
Gregerman et al. [1998]	University of Michigan	Undergraduate Research Opportunity Program
Pender et al. [2010]	University of Maryland Baltimore County	Meyerhoff Scholarship Program
Hathaway et al. [2002]	Survey from University of Michigan	Undergraduate Research Opportunity Program
Lopatto [2004]	Online survey of undergraduates from 41 institutions	Summer undergraduate research programs
Zydney et al. [2002]	Survey at the University of Delaware	Undergraduate science research experience

Figure 1: Timing of research experience

(a) All research experience



(b) First time research experience

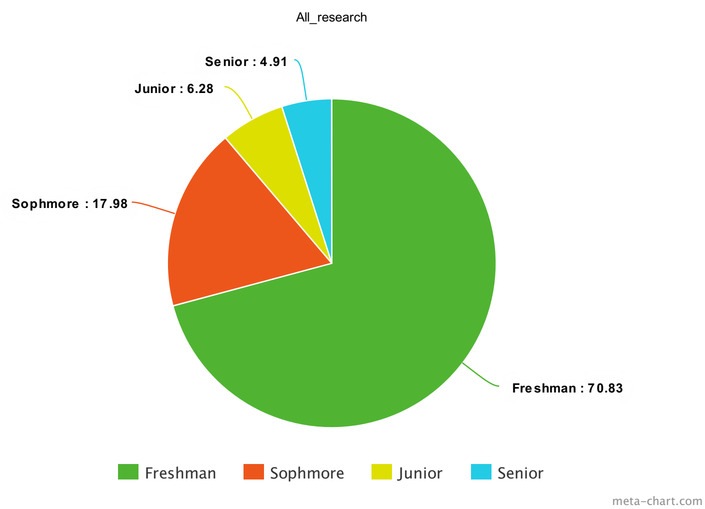


Figure 2: Number of months employed

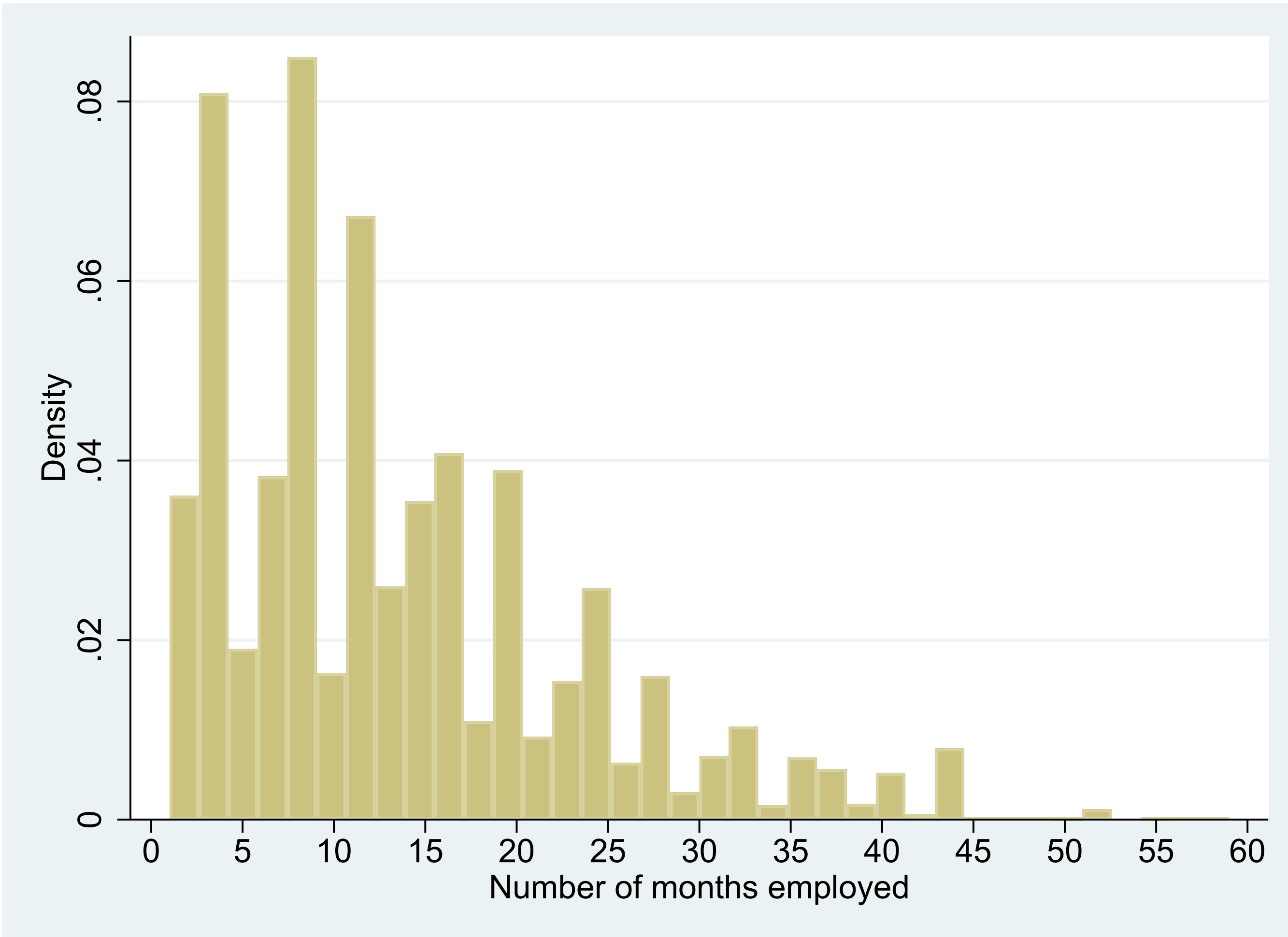


Table 2: Summary Statistics

	Full Sample		All jobs		Research jobs	
	Mean	SD	Mean	SD	Mean	SD
Female	0.529	0.499	0.594	0.491	0.568	0.495
Black	0.064	0.246	0.086	0.280	0.061	0.240
Hispanic	0.053	0.226	0.050	0.218	0.046	0.210
Asian	0.117	0.322	0.163	0.370	0.173	0.379
Other race	0.035	0.184	0.039	0.194	0.040	0.198
In state	0.645	0.479	0.771	0.420	0.780	0.415
High school GPA	3.723	0.292	3.781	0.251	3.801	0.232
International student	0.036	0.187	0.024	0.154	0.026	0.161
Pell grant	0.195	0.396	0.300	0.437	0.257	0.437
ACT composite score	28.10	3.445	28.34	3.351	28,71	3.35
Graduate	0.806	0.395	0.852	0.355	0.863	0.344
Graduate STEM	0.205	0.404	0.355	0.479	0.418	0.493
Graduate social science	0.285	0.451	0.223	0.416	0.177	0.382
AP BY score	0.626	1.467	0.919	1.741	1.057	1.844
AP CALAB score	1.140	1.816	1.351	1.940	1.477	1.997
AP CALBC score	0.453	1.307	0.609	1.503	0.694	1.593
AP CH score	0.490	1.260	0.758	1.534	0.863	1.620
AP CSA score	0.045	0.423	0.057	0.480	0.062	0.499
AP CSAB score	0.017	0.248	0.018	0.244	0.019	0.259
AP PHYSB score	0.166	0.747	0.209	0.864	0.244	0.938
AP PHYSE score	0.074	0.513	0.099	0.595	0.112	0.636
AP PHYSM score	0.201	0.853	0.303	1.042	0.348	1.115
AP STAT score	0.329	1.087	0.321	1.095	0.353	1.150
AP CALAB	0.253	0.435	0.302	0.459	0.329	0.470
AP CALBC	0.102	0.303	0.137	0.343	0.155	0.362
AP CALSB	0.109	0.312	0.145	0.352	0.165	0.371
AP BY	0.143	0.350	0.208	0.406	0.237	0.425
AP CH	0.112	0.316	0.175	0.380	0.200	0.400
AP CSA	0.010	0.100	0.013	0.114	0.014	0.118
AP CSAB	0.003	0.0611	0.004	0.064	0.004	0.066
AP PHYSB	0.039	0.195	0.048	0.215	0.057	0.233
AP PHYSE	0.015	0.122	0.020	0.142	0.022	0.150
AP PHYSM	0.045	0.209	0.068	0.253	0.079	0.270
AP STAT	0.078	0.268	0.074	0.263	0.081	0.274
Observations	35720	.	3653	.	2896	.

Notes: Five-year graduation rates reported.
Both AP scores and an indicator for taking AP tests are included.
Science AP tests included: Biology (BY), Chemistry (CH), Physics (Physics B (PHYSB), Physics C: Electricity and Magnetism (PHYSE), Physics C: Mechanics (PHYSM)), Computer Science (Computer Science A (CSA), Computer Science AB (CSAB)), Statistics (STAT), and Calculus (Calculus AB (CALAB), Calculus BC (CALBC)).

Table 3: Estimates of propensity score with any grant employment as the outcome, logit model

VARIABLES	Selection	Marginal effects
Female	0.195***	0.032***
	(0.023)	(0.003)
Black	0.154*	0.026*
	(0.063)	(0.010)
Hispanic	-0.028	-0.004
	(0.067)	(0.011)
Asian	0.181***	0.030***
	(0.041)	(0.007)
Other race	-0.023	-0.003
	(0.078)	(0.013)
In state	0.291***	0.048***
	(0.022)	(0.003)
HS GPA	0.047***	0.008***
	(0.008)	(0.001)
International student	0.073	0.012
	(0.059)	(0.010)
Pell grant	0.338***	0.056***
	(0.022)	(0.003)
ACT composite score	-0.042**	-0.007**
	(0.014)	(0.002)
ACT composite sq.	0.001***	0.0001***
	(0.000)	(0.000)
AP CALAB	0.063**	0.010**
	(0.033)	(0.003)
AP CALBC	0.048	0.008
	(0.033)	(0.005)
AP BY	0.144***	0.024***
	(0.025)	(0.004)
AP CH	0.123***	0.020***
	(0.028)	(0.004)
AP CSA	0.046	0.007
	(0.040)	(0.005)
AP CSAB	-0.003	-0.000
	(0.148)	(0.024)
AP PHYSB	0.097*	0.0163*
	(0.046)	(0.007)
AP PHYSE	-0.048	-0.008
	(0.081)	(0.013)
AP PHYSM	0.145**	0.024**
	(0.049)	(0.008)
AP STAT	-0.058	-0.009
	(0.035)	(0.006)
Constant	-1.829***	
	(0.195)	
Observations	35,720	35,720
Pseudo R-squared	0.058	0.058

Notes: All regressions contain race-gender interaction terms.
Selection equation estimated using a logistic regression. Marginal effects estimated at the mean of observable characteristics.
Standard errors in parentheses.
*** Statistical significance at the 1 percent level.
** Statistical significance at the 5 percent level.
* Statistical significance at the 10 percent level.

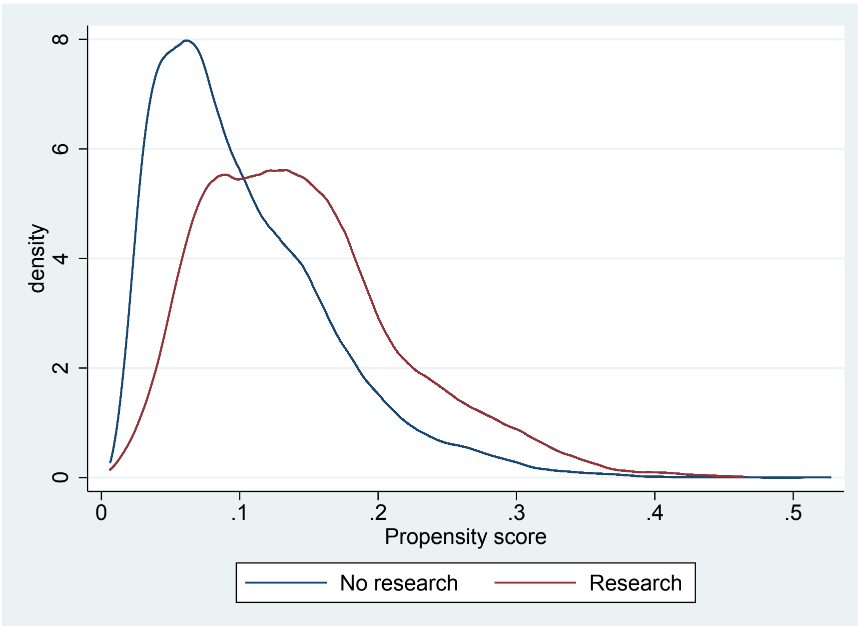
Table 4: Estimates of propensity score with research employment as the outcome, logit model

VARIABLES	Selection	Marginal effects
Female	0.160***	0.022***
	(0.024)	(0.003)
Other race female	0.126	0.017
	(0.106)	(0.014)
Black	0.032	0.004
	(0.074)	(0.010)
Hispanic	-0.0161	-0.002
	(0.0730)	(0.010)
Asian	0.176***	0.024***
	(0.043)	(0.006)
Other race	-0.050	-0.007
	(0.080)	(0.011)
In state	0.304***	0.042***
	(0.024)	(0.003)
HS GPA	0.057***	0.008***
	(0.008)	(0.001)
International student	0.144*	0.020*
	(0.063)	(0.008)
Pell grant	0.221***	0.031***
	(0.024)	(0.003)
ACT composite	-0.032	-0.004
	(0.017)	(0.025)
ACT composite score sq.	0.001**	-0.000**
	(0.000)	(0.000)
AP CALAB	0.074*	0.010*
	(0.024)	(0.004)
AP CALBC	0.050	0.007
	(0.034)	(0.004)
AP BY	0.182***	0.025***
	(0.026)	(0.003)
AP CH	0.140***	0.096***
	(0.029)	(0.004)
AP CSA	0.007	0.001
	(0.096)	(0.012)
AP CSAB	-0.045	-0.006
	(0.156)	(0.021)
AP PHYSB	0.143**	0.020**
	(0.047)	(0.006)
AP PHYSE	-0.068	-0.009
	(0.083)	(0.011)
AP PHYSM	0.164**	0.022**
	(0.050)	(0.007)
AP STAT	-0.035	-0.005
	(0.037)	(0.005)
Constant	-2.224***	
	(0.235)	
Observations	35,720	35,720
Pseudo R-squared	0.070	0.070

Notes: All regressions contain race-gender interaction terms.
Selection equation estimated using a logistic regression. Marginal effects estimated at the mean of observable characteristics.
Standard errors in parentheses.
*** Statistical significance at the 1 percent level.
** Statistical significance at the 5 percent level.
* Statistical significance at the 10 percent level.

Figure 3: Kernel density of probability of getting the treatment

(a) All grant employment



(b) Research jobs

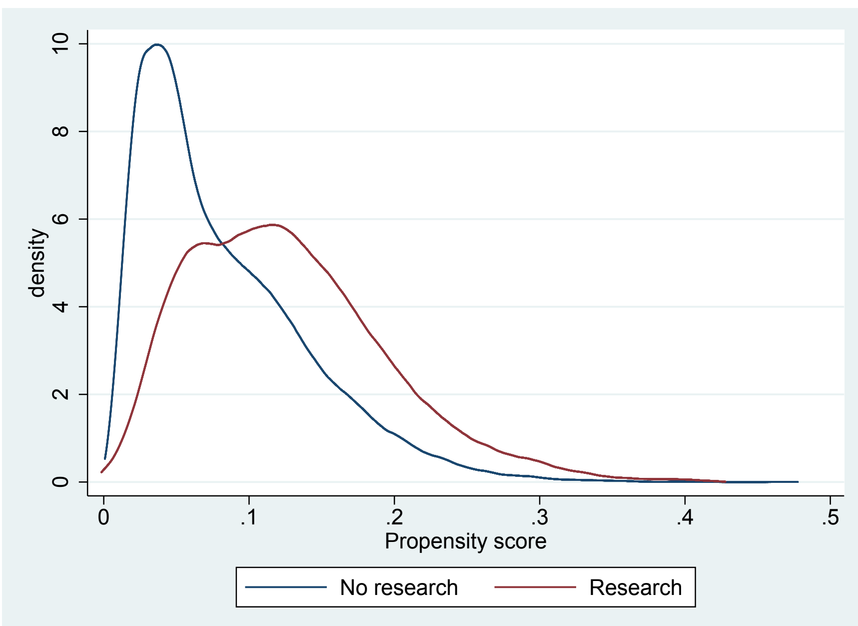


Table 5: Standardized differences for grant employment as treatment and graduation as outcome, inverse probability weighting

	Std diff Raw	Std diff Weighted	Var ratio Raw	Var ratio Weighted
Female	.147	-.004	.966	1.001
Black female	.117	-.005	1.648	.978
Hispanic female	.037	.001	1.230	1.008
Asian female	.137	-.005	1.570	.984
Other race female	.055	-.006	1.446	.962
Black	.090	-.008	1.348	.975
Hispanic	-.018	.003	.928	1.012
Asian	.148	-.005	1.371	.990
Other race	.024	-.003	1.125	.984
In state	.309	-.010	.758	1.014
Intent STEM	.395	-.008	.864	1.007
HS GPA	.208	-.007	.963	1.015
International student	-.077	.003	.654	1.020
Pell grant	.275	-.013	1.403	.988
Pell grant * female	.270	-.009	1.760	.985
ACT composite score	.078	.001	1.088	.978
ACT composite score sq.	.090	.002	1.108	1.018
AP CALAB	.121	-.000	1.131	.999
AP CALBC	.119	-.001	1.331	.996
AP BY	.191	.001	1.402	1.002
AP CH	.201	-.009	1.532	.983
AP CSA	.031	.000	1.341	1.008
AP CSAB	.006	-.000	1.106	.998
AP PHYSB	.050	.001	1.255	1.006
AP PHYSE	.046	-.005	1.410	.962
AP PHYSM	.111	-.002	1.552	.992
AP STAT	-.014	.002	.956	1.007

Notes: Standardized differences between the treatment and comparison groups, calculated based on the formula from Equation 5 divided by 100.

Table 6: IPW estimation results

		Graduation (percentage points)	STEM graduation (percentage points)
Overall	All jobs	0.055***	0.101***
		(0.006)	(0.008)
	Research jobs	0.056***	0.136***
		(0.006)	(0.009)
Female	All jobs	0.048***	0.0653***
		(0.008)	(0.010)
	Research jobs	0.049***	0.102***
		(0.009)	(0.012)
Male	All jobs	0.065***	0.147***
		(0.010)	(0.013)
	Research jobs	0.067***	0.176***
		(0.010)	(0.013)
White	All jobs	0.045***	0.108***
		(0.007)	(0.010)
	Research jobs	0.049***	0.146***
		(0.008)	(0.011)
Black	All jobs	0.101***	0.018
		(0.027)	(0.026)
	Research jobs	0.083**	0.075**
		(0.035)	(0.037)
Hispanic	All jobs	0.107***	0.094***
		(0.031)	(0.033)
	Research jobs	0.093***	0.084**
		(0.035)	(0.040)
Asian	All jobs	0.046***	0.109***
		(0.016)	(0.021)
	Research jobs	0.054***	0.126***
		(0.017)	(0.024)
Pell grant	All jobs	0.072***	0.063***
		(0.013)	(0.015)
	Research jobs	0.080***	0.110***
		(0.015)	(0.019)

Notes: Five-year graduation rates reported.
ATET results shown using the IPW estimator. All jobs refers to all types of federal grant employment, while research jobs refers to more research oriented types of federal grant employment.

Figure 4: IPW estimation results for grant employment by graduation type

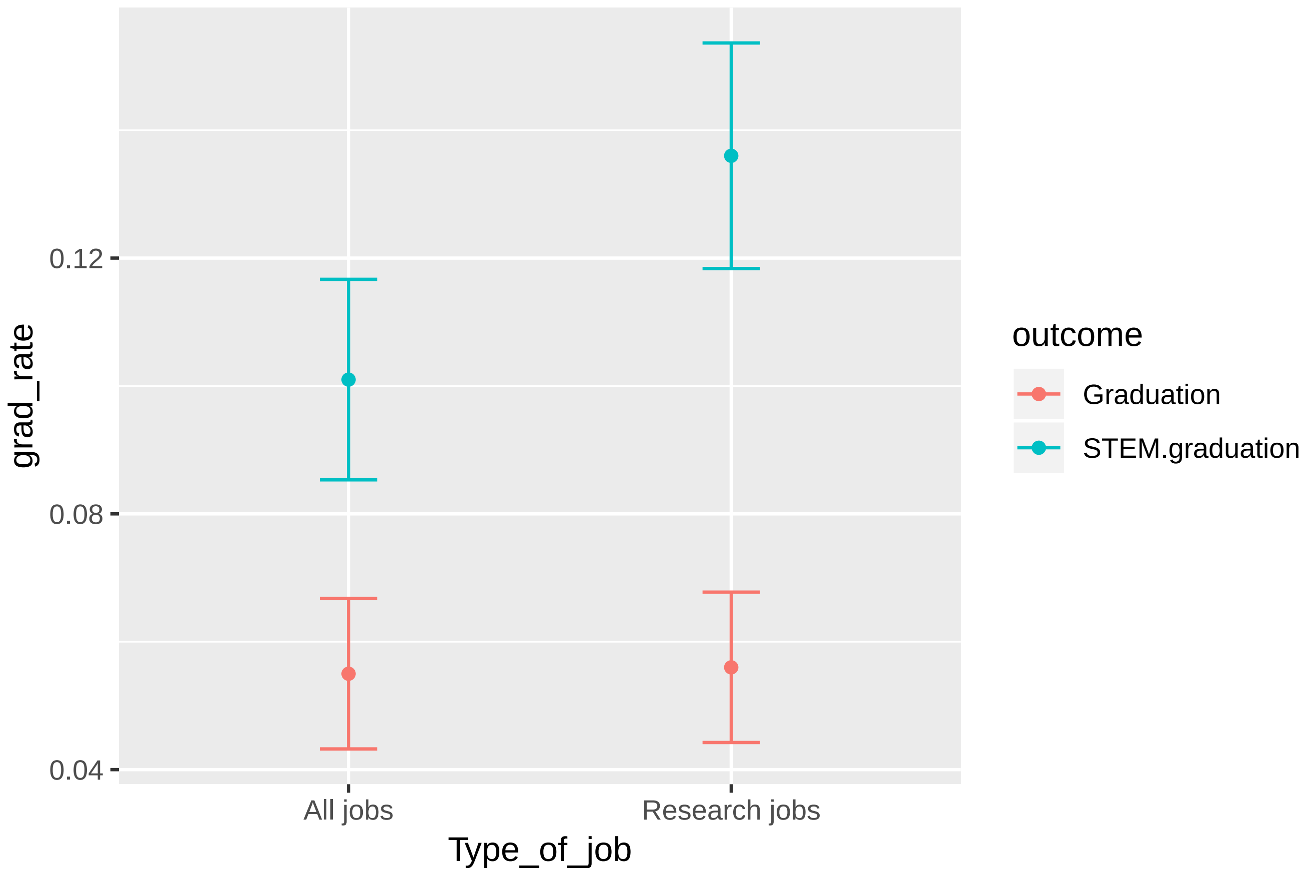


Figure 5: IPW estimation results for gender for all jobs by graduation type

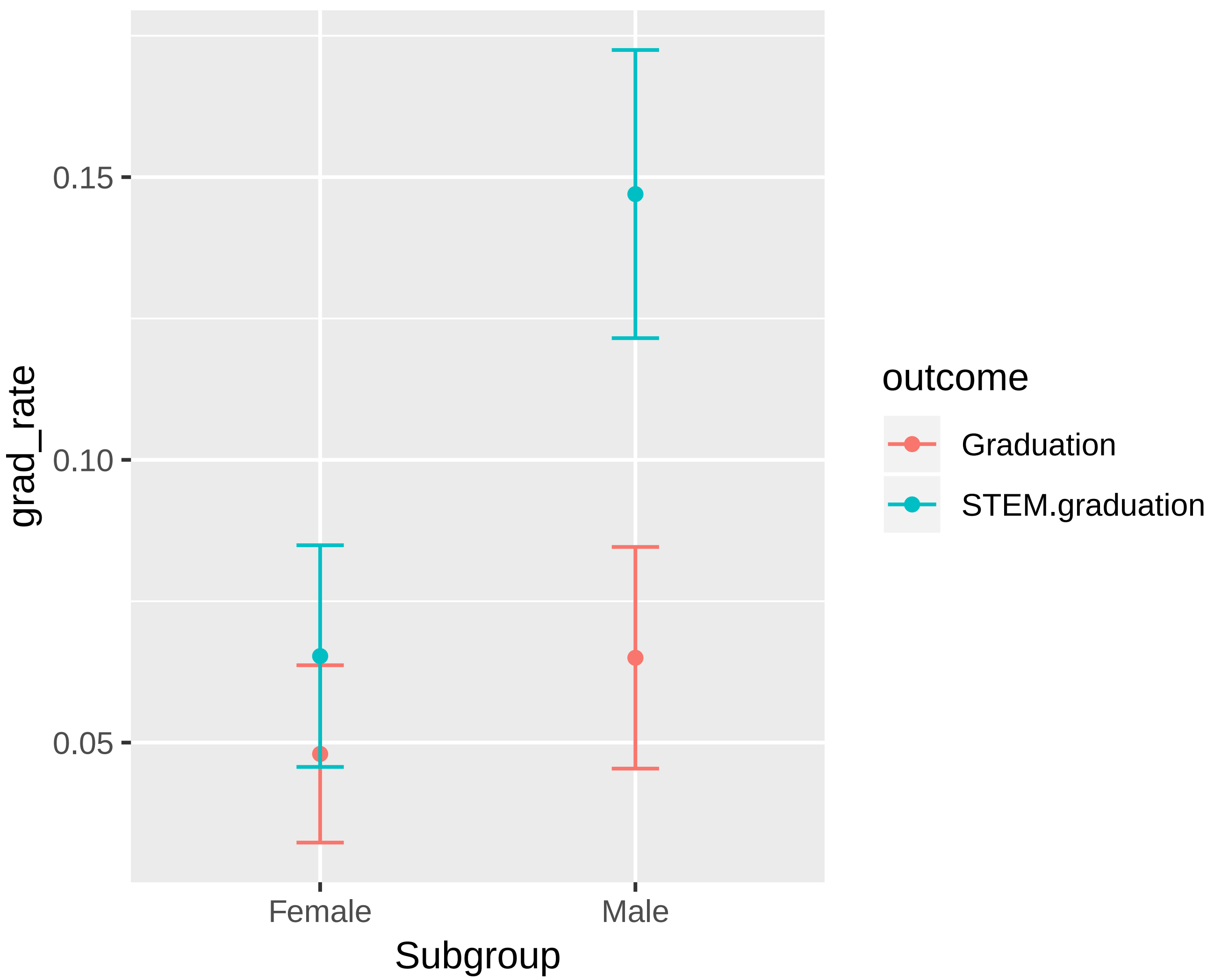


Figure 6: IPW estimation results for gender for research jobs by graduation type

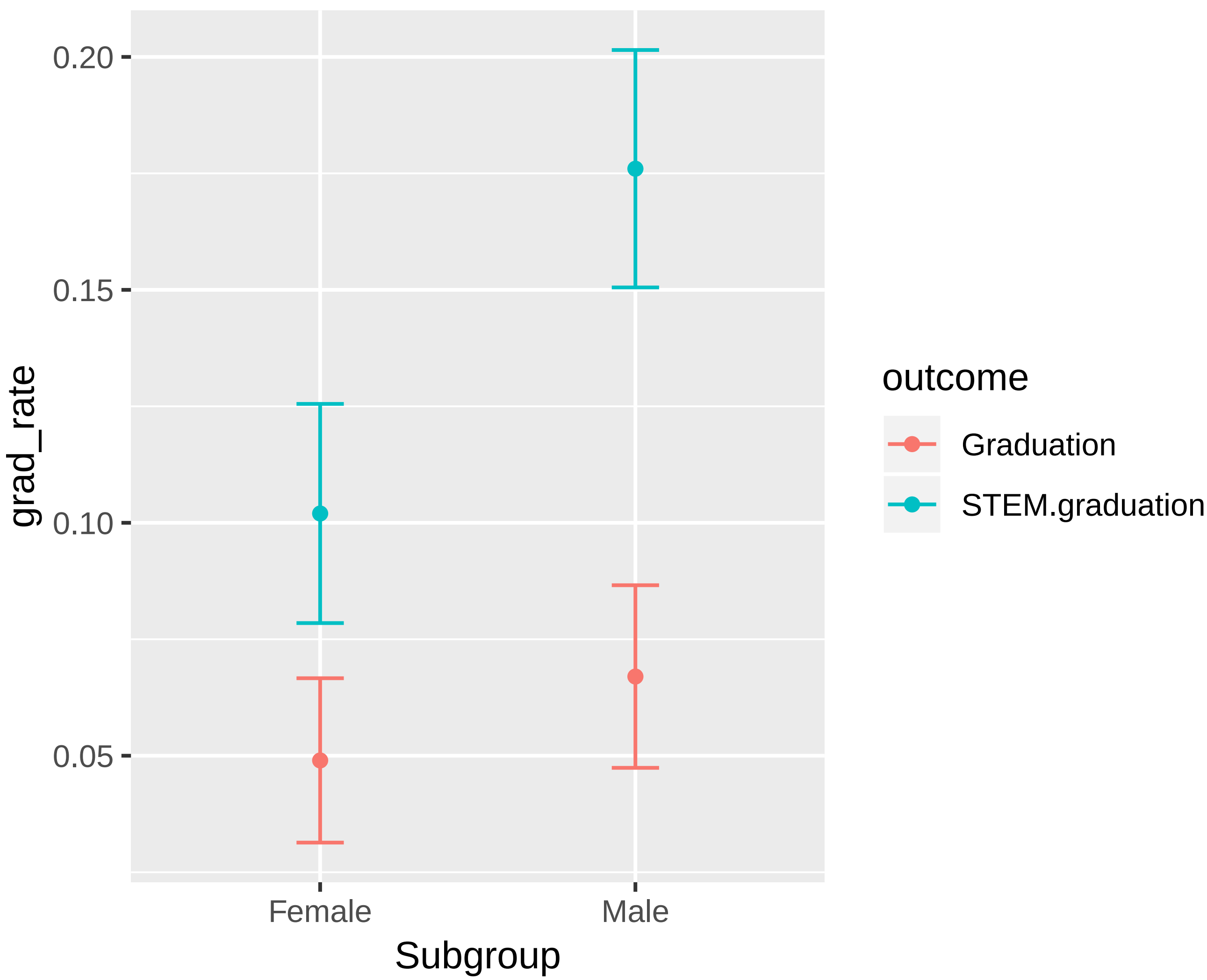


Figure 7: IPW estimation results for gender for general graduation rates by type of job

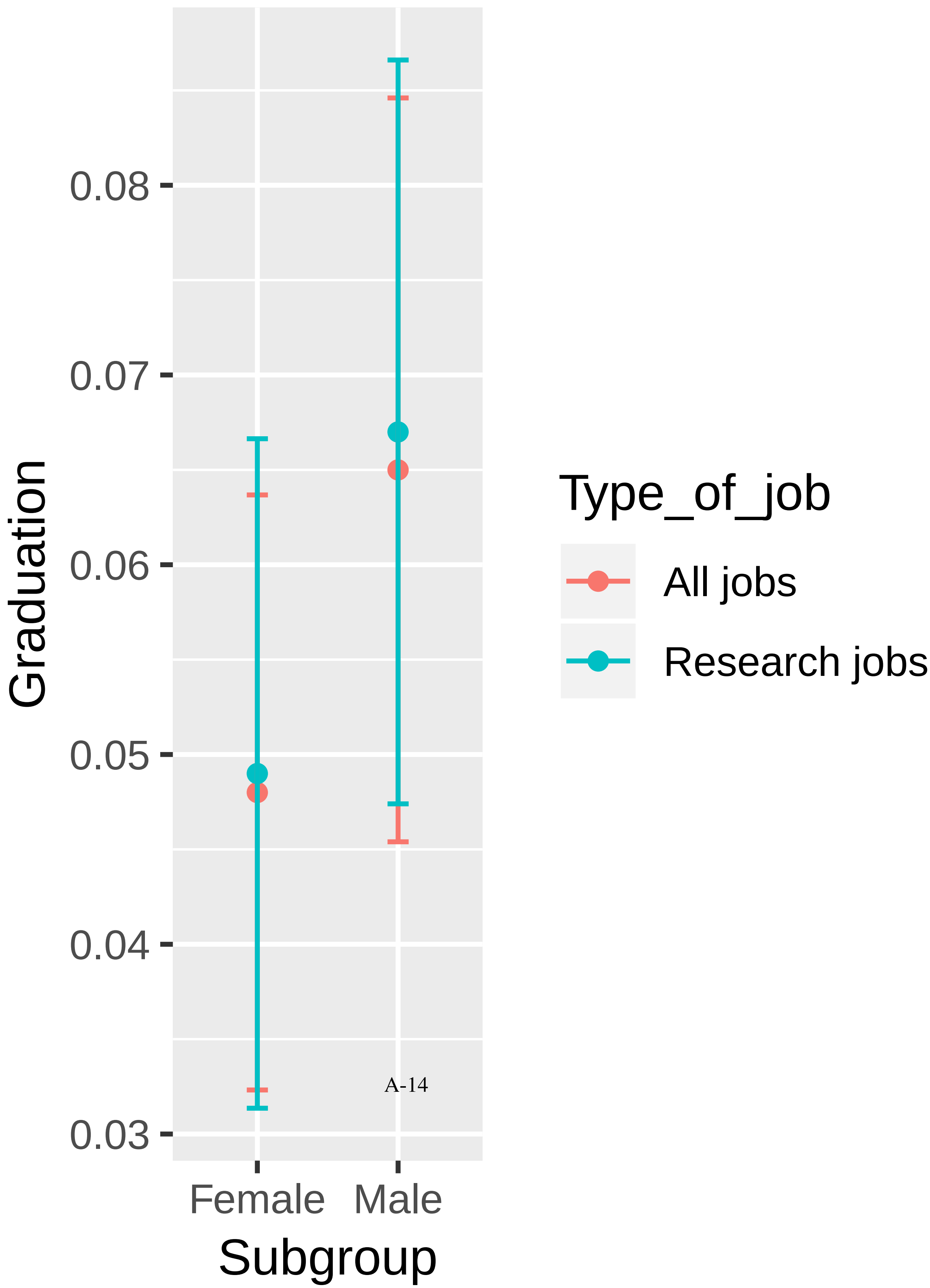


Figure 8: IPW estimation results for gender for STEM graduation rates by type of job

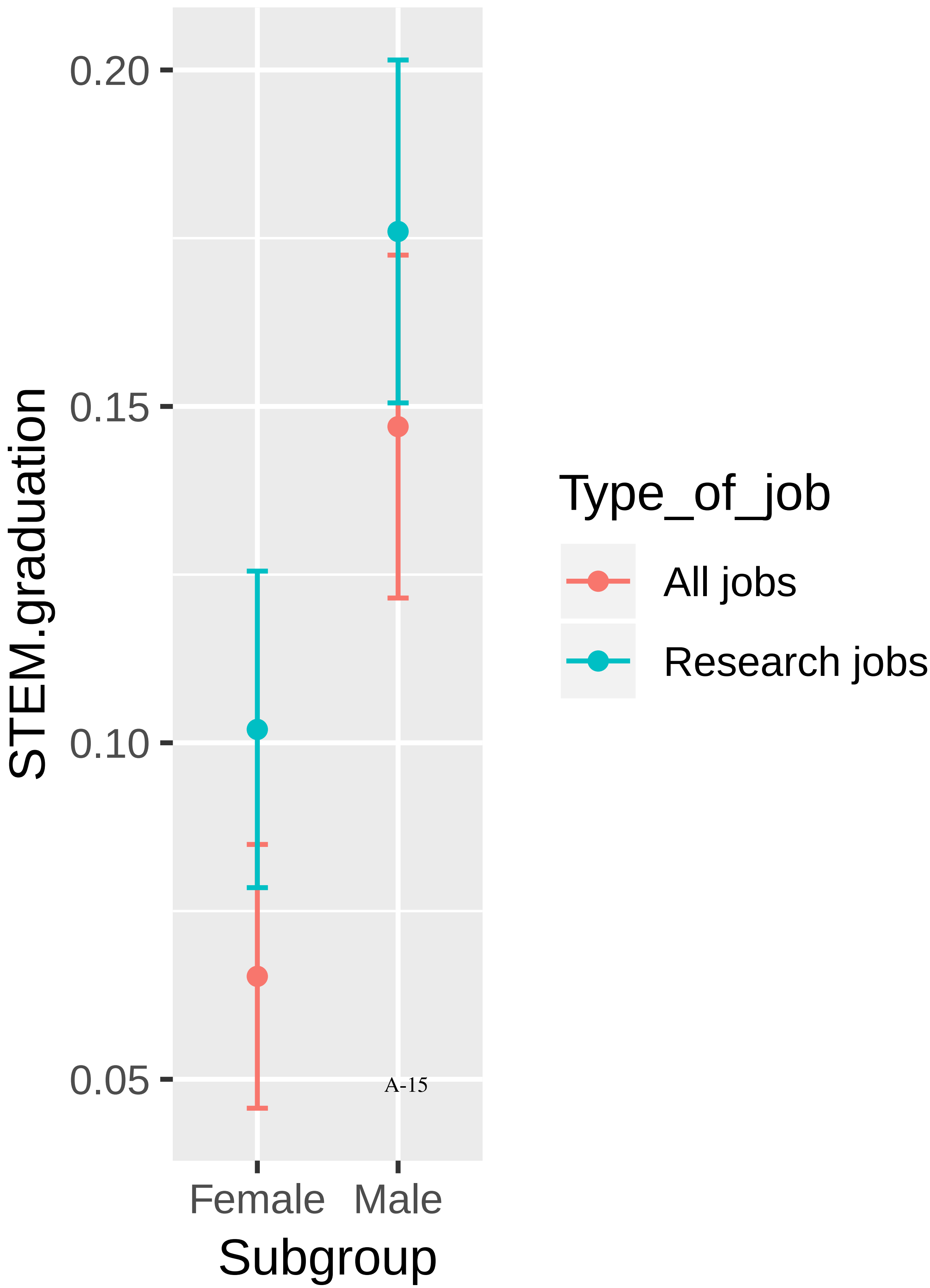


Figure 9: IPW estimation results for race for all jobs by graduation type

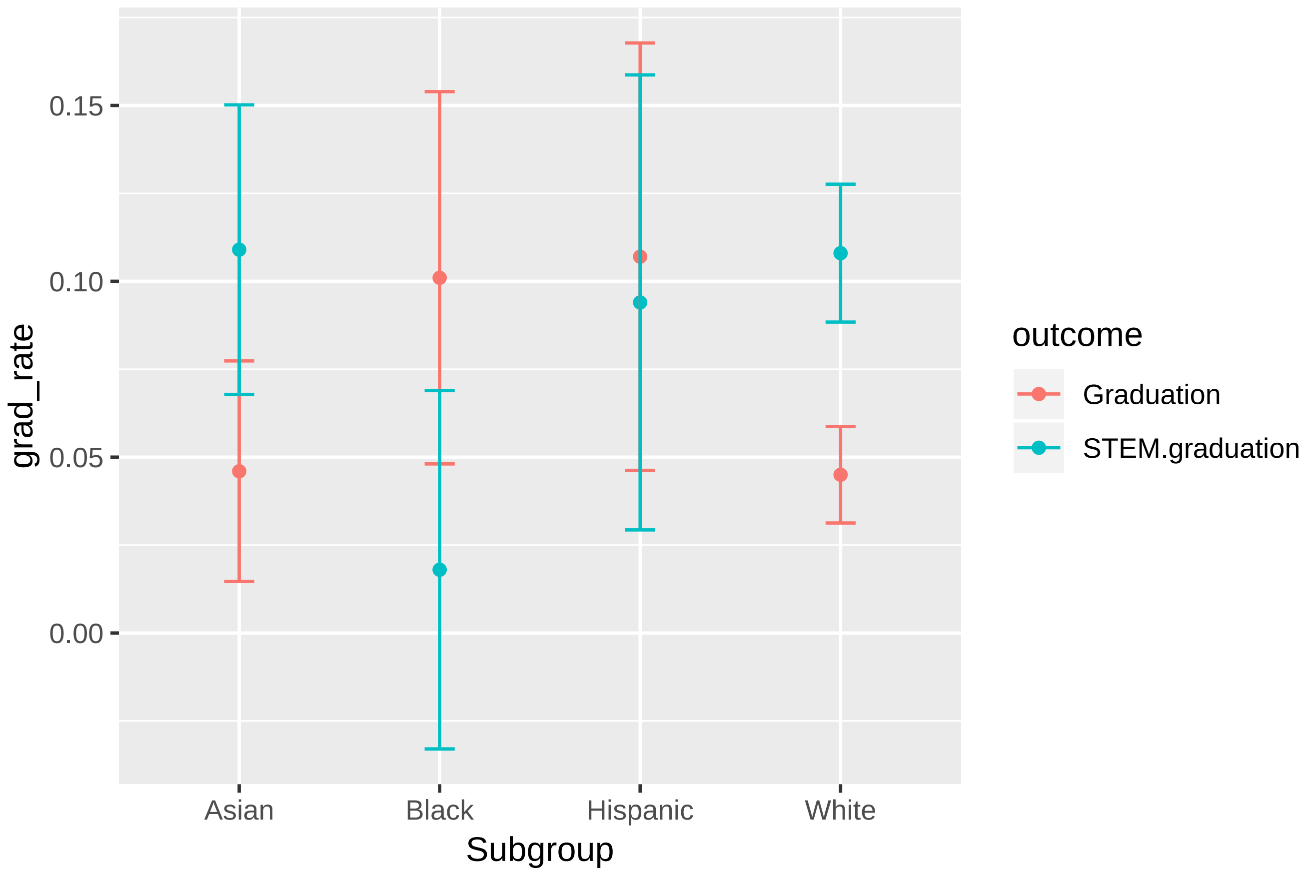


Figure 10: IPW estimation results for race for research jobs by graduation type

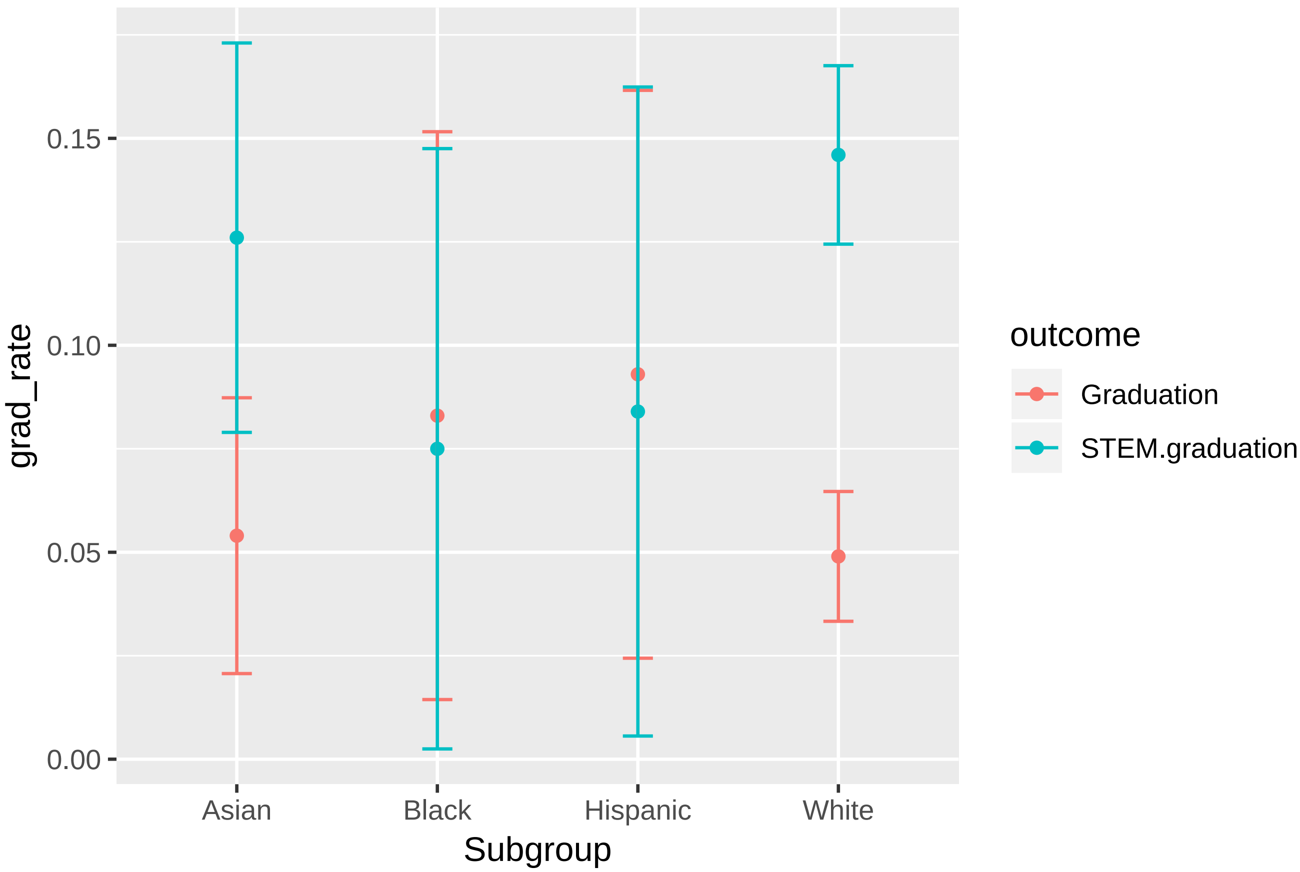


Figure 11: IPW estimation results for race for general graduation rates by type of job

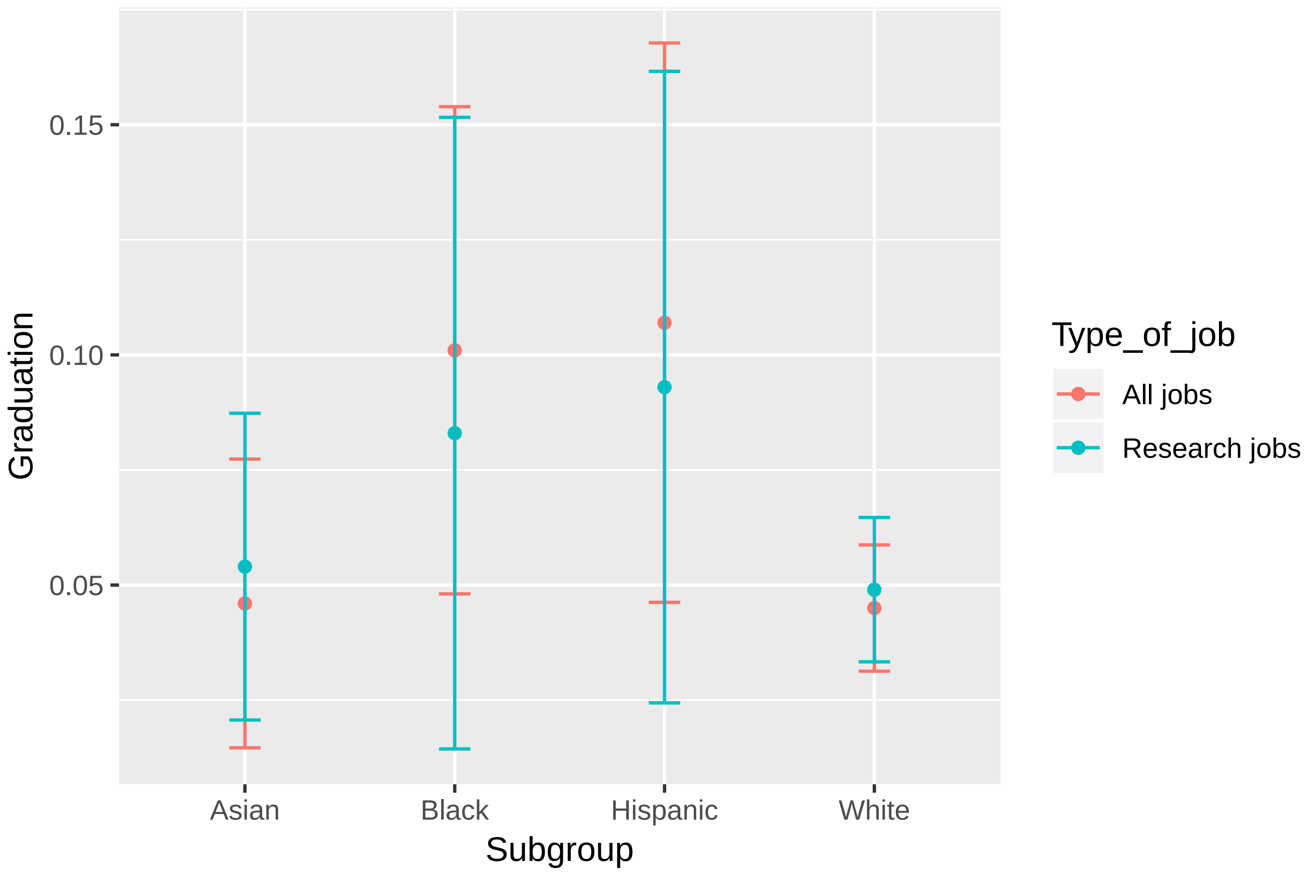


Figure 12: IPW estimation results for race for STEM graduation rates by type of job

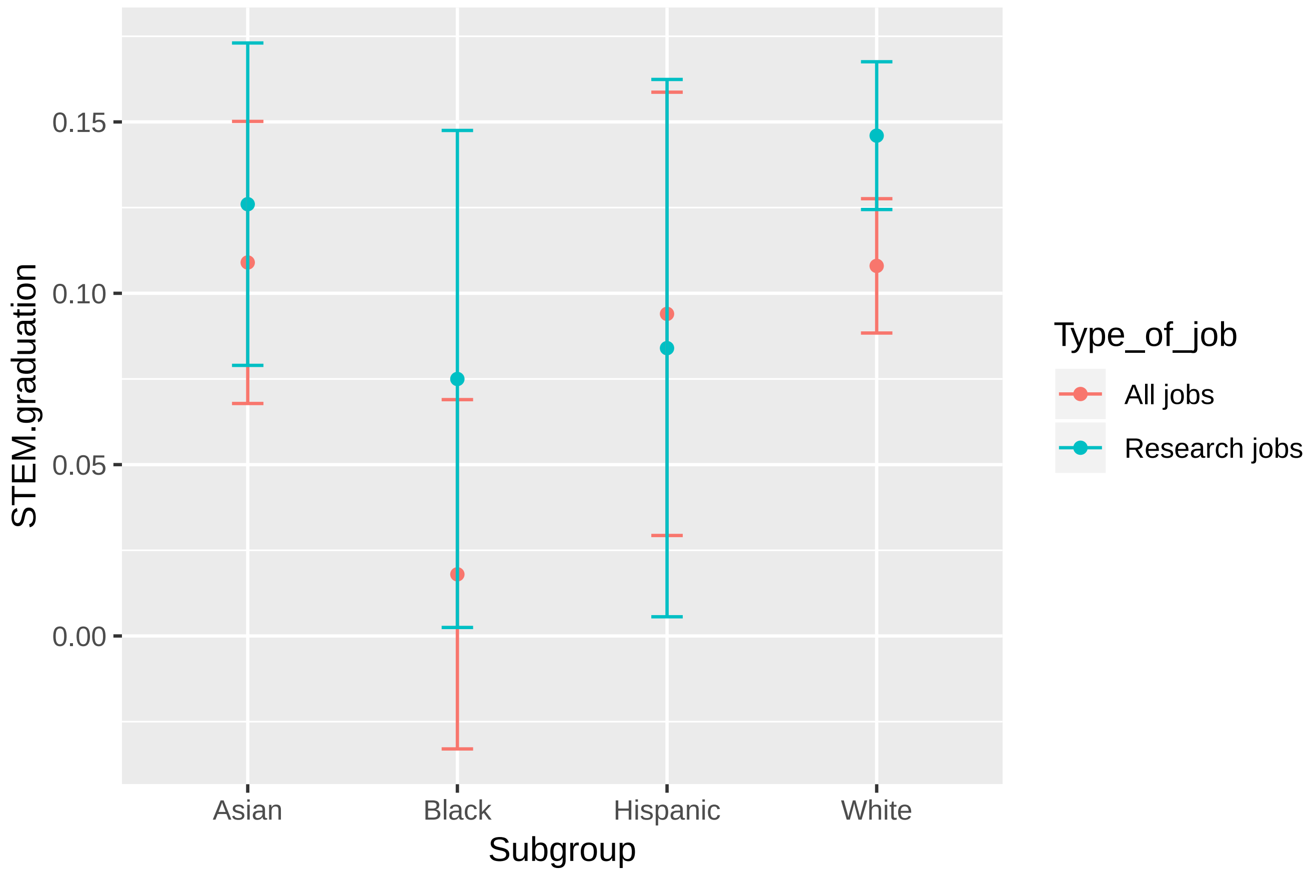


Figure 13: IPW estimation results for general graduation rates by type of job

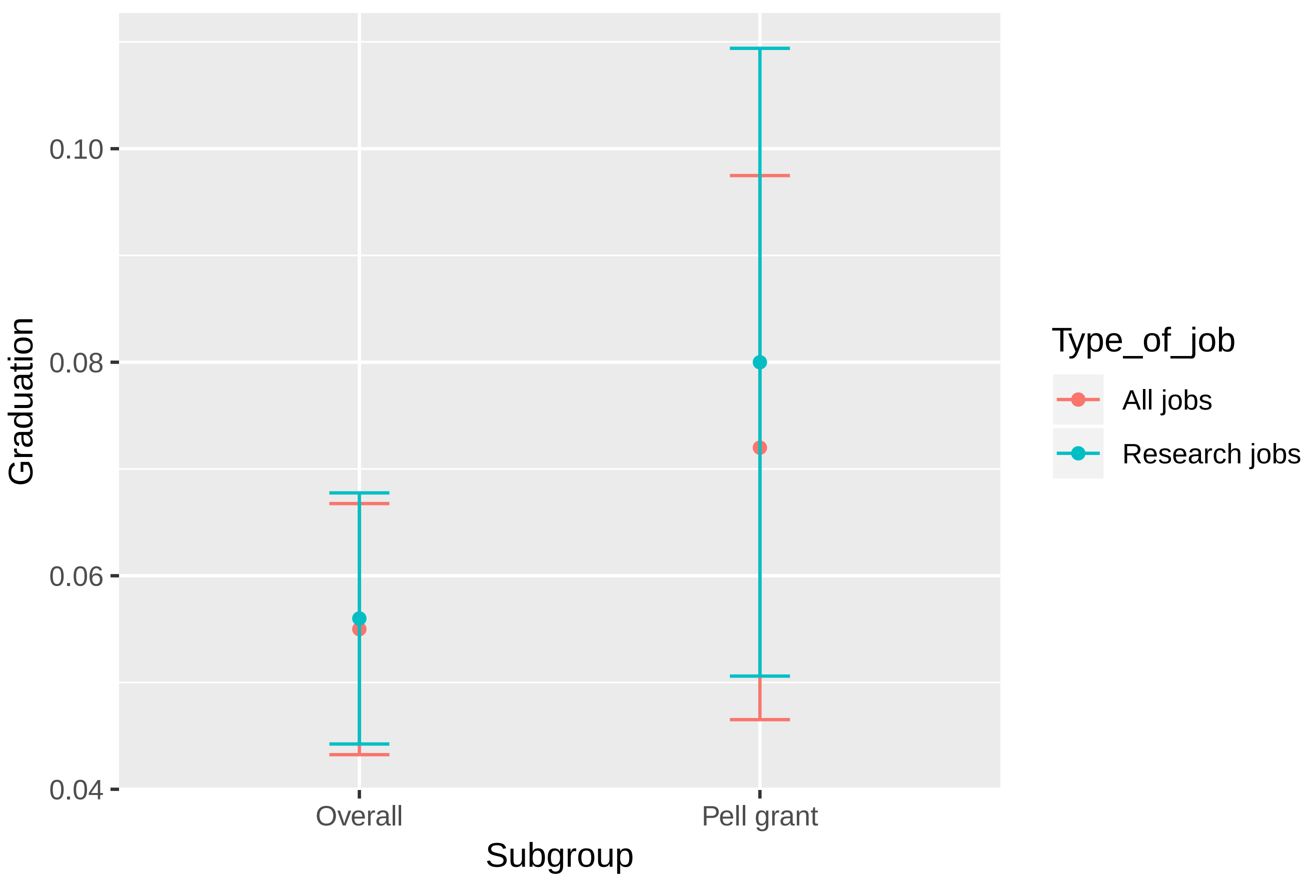


Figure 14: IPW estimation results for STEM graduation rates by type of job

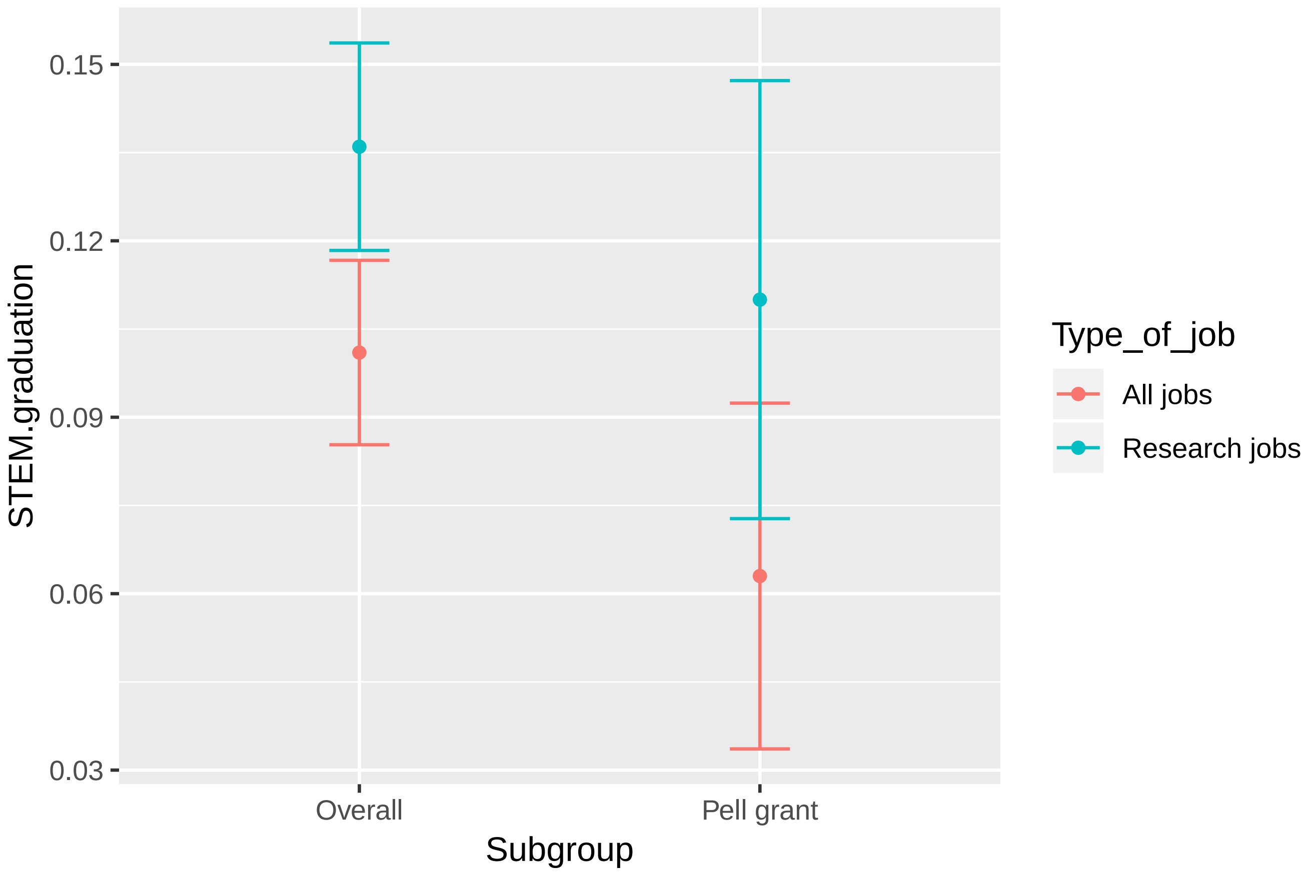


Figure 15: IPW estimation results for Pell recepients for all jobs by graduation type

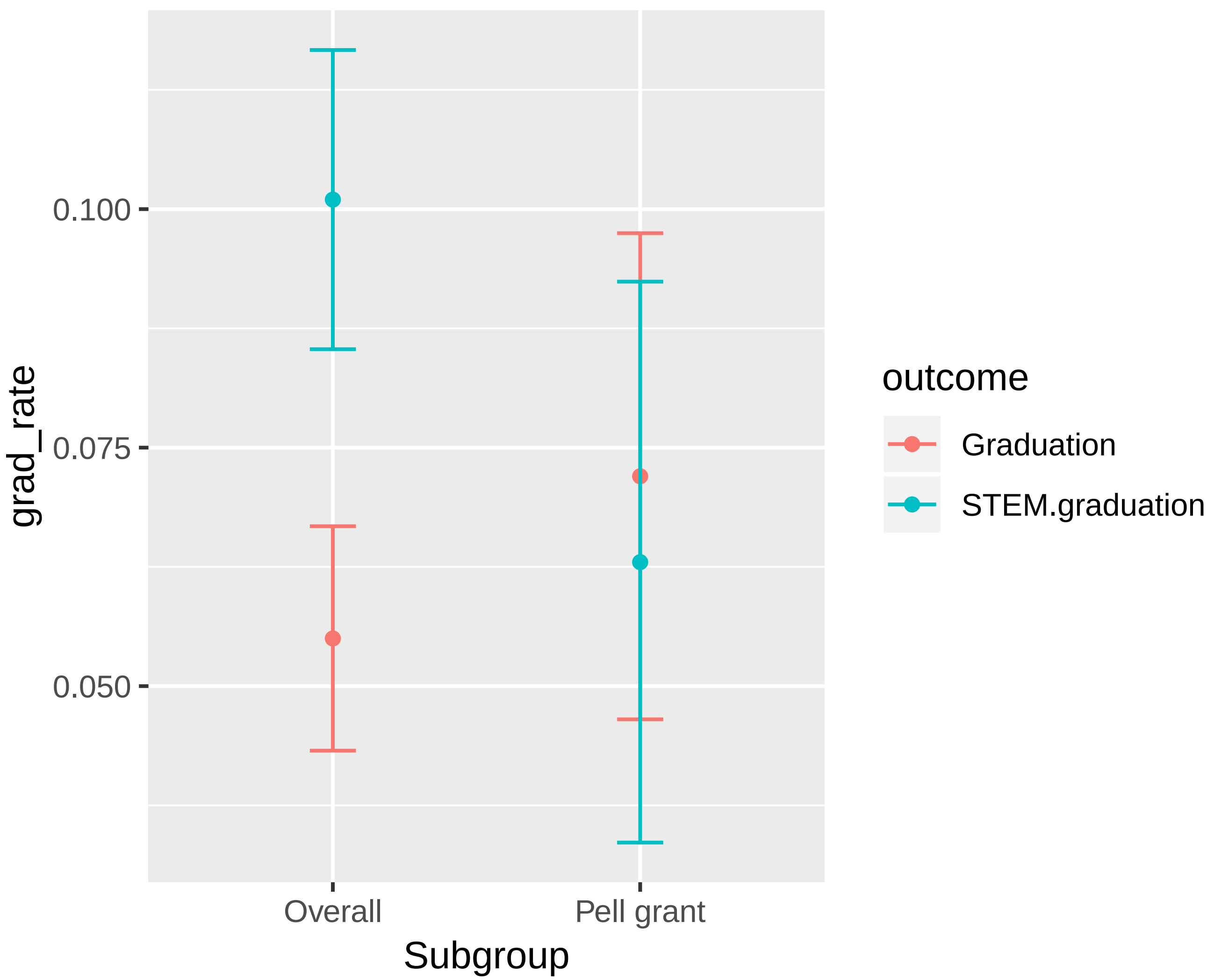


Figure 16: IPW estimation results for Pell recepients for research jobs by graduation type

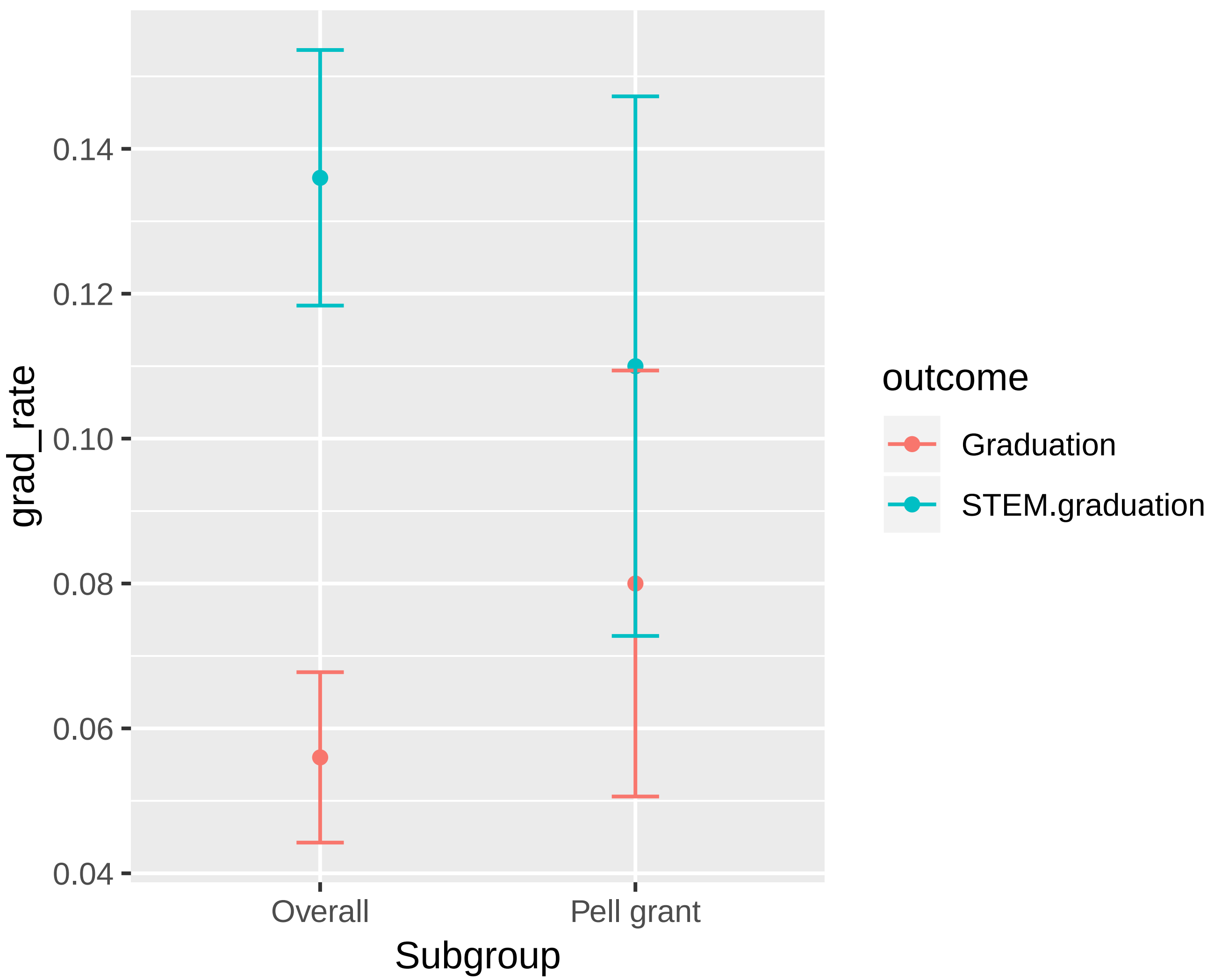


Table 7: Nearest neighbor (1) and IPWRA estimation results

		Graduation (percentage points)	STEM graduation (percentage points)
IPW	All jobs	0.055***	0.101***
		(0.006)	(0.008)
	Research jobs	0.056***	0.136***
		(0.006)	(0.009)
NN(1)	All jobs	0.052***	0.101***
		(0.008)	(0.010)
	Research jobs	0.049***	0.134***
		(0.009)	(0.012)
IPWRA	All jobs	0.055***	0.104***
		(0.006)	(0.00808)
	Research jobs	0.056***	0.140***
		(0.006)	(0.009)

Notes: Five-year graduation rates reported.
The column base estimates presents results with all the controls included. They are the same results using the IPW estimator from Table **??**. The column sensitivity analysis presents results using the same controls, excluding AP test controls, using the IPW estimator.

Table 8: Sensitivity analysis of the IPW estimation results

		Graduation (percentage points)		STEM graduation (percentage points)	
		Base estimates	Sensitivity analysis	Base estimates	Sensitivity analysis
Overall	All jobs	0.055***	0.058***	0.101***	0.110***
		(0.006)	(0.006)	(0.008)	(0.008)
	Research jobs	0.056***	0.060***	0.136***	0.148***
		(0.006)	(0.006)	(0.009)	(0.009)
Female	All jobs	0.048***	0.051***	0.0653***	0.073***
		(0.008)	(0.008)	(0.010)	(0.010)
	Research jobs	0.049***	0.051***	0.102***	0.114***
		(0.009)	(0.009)	(0.012)	(0.012)
Male	All jobs	0.065***	0.069***	0.147***	0.158***
		(0.010)	(0.010)	(0.013)	(0.013)
	Research jobs	0.067***	0.072***	0.176***	0.189***
		(0.010)	(0.010)	(0.013)	(0.014)
White	All jobs	0.045***	0.047***	0.108***	0.117***
		(0.007)	(0.007)	(0.010)	(0.010)
	Research jobs	0.049***	0.052***	0.146***	0.159***
		(0.008)	(0.008)	(0.011)	(0.011)
Black	All jobs	0.101***	0.099***	0.018	0.028
		(0.027)	(0.027)	(0.026)	(0.027)
	Research jobs	0.083**	0.086**	0.075**	0.093**
		(0.035)	(0.035)	(0.037)	(0.038)
Hispanic	All jobs	0.107***	0.107***	0.094***	0.108***
		(0.031)	(0.031)	(0.033)	(0.034)
	Research jobs	0.093***	0.094***	0.084**	0.097**
		(0.035)	(0.035)	(0.040)	(0.041)
Asian	All jobs	0.046***	0.049***	0.109***	0.121***
		(0.016)	(0.016)	(0.021)	(0.022)
	Research jobs	0.054***	0.056***	0.126***	0.136***
		(0.017)	(0.017)	(0.024)	(0.0244)
Pell grant	All jobs	0.072***	0.074***	0.063***	0.067***
		(0.013)	(0.013)	(0.015)	(0.015)
	Research jobs	0.080***	0.084***	0.110***	0.119***
		(0.015)	(0.015)	(0.019)	(0.019)

Notes: Five-year graduation rates reported.
the column base estimates presents results with all the controls included. They are the same results using the IPW estimator from Table **??**. The column sensitivity analysis presents results using the same controls, excluding AP test controls, using the IPW estimator.

Table 9: Sensitivity analysis using Mantel-Haenszel bounds for grant employment and graduation

$\Gamma = e^{\gamma}$	Q^{+}_{MH}	Q^{-}_{MH}	p^{+}_{MH}	p^{-}_{MH}
1.00	5.630	5.630	<0.001	<0.001
1.05	4.844	6.420	<0.001	<0.001
1.10	4.094	7.17	<0.001	<0.001
1.15	3.379	7.898	<0.001	<0.001
1.20	2.695	8.593	.003	<0.001
1.25	2.040	9.261	.020	<0.001
1.30	1.411	9.905	.079	<0.001
1.35	.806	10.527	.209	<0.001
1.40	.224	11.129	.411	<0.001
1.45	.275	11.711	.391	<0.001
1.50	.818	12.275	.206	<0.001
1.55	1.343	12.823	.089	<0.001
1.60	1.852	13.356	.031	<0.001

$\Gamma = e^{\gamma}$: odds of differential assignment due to unobserved factors
Q^{+}_{MH}: Mantel-Haenszel statistic (assumption: overestimation of treatment effect)
Q^{-}_{MH}: Mantel-Haenszel statistic (assumption: underestimation of treatment effect)
p^{+}_{MH}: significance level (assumption: overestimation of treatment effect)
p^{-}_{MH}: significance level (assumption: underestimation of treatment effect)

Table 10: Sensitivity analysis using Mantel-Haenszel bounds for research job and graduation

$\Gamma = e^{\gamma}$	Q^{+}_{MH}	Q^{-}_{MH}	p^{+}_{MH}	p^{-}_{MH}
1.00	3.773	3.773	<0.001	<0.001
1.05	3.102	4.446	<0.001	<0.001
1.10	2.462	5.090	.006	<0.001
1.15	1.852	5.706	.031	<0.001
1.20	1.268	6.297	.102	<0.001
1.25	.709	6.866	.239	<0.001
1.30	.172	7.415	.431	<0.001
1.35	.271	7.944	.392	<0.001
1.40	.770	8.456	.220	<0.001
1.45	1.250	8.951	.105	<0.001
1.50	1.715	9.431	.043	<0.001

$\Gamma = e^{\gamma}$: odds of differential assignment due to unobserved factors
Q^{+}_{MH}: Mantel-Haenszel statistic (assumption: overestimation of treatment effect)
Q^{-}_{MH}: Mantel-Haenszel statistic (assumption: underestimation of treatment effect)
p^{+}_{MH}: significance level (assumption: overestimation of treatment effect)
p^{-}_{MH}: significance level (assumption: underestimation of treatment effect)

Table 11: Sensitivity analysis using Mantel-Haenszel bounds for grant employment and STEM graduation

$\Gamma = e^{\gamma}$	Q^{+}_{MH}	Q^{-}_{MH}	p^{+}_{MH}	p^{-}_{MH}
1.00	9.430	9.430	<0.001	<0.001
1.05	8.508	10.355	<0.001	<0.001
1.10	7.629	11.238	<0.001	<0.001
1.15	6.791	12.083	<0.001	<0.001
1.20	5.990	12.895	<0.001	<0.001
1.25	5.222	13.675	<0.001	<0.001
1.30	4.485	14.426	<0.001	<0.001
1.35	3.776	15.150	<0.001	<0.001
1.40	3.094	15.850	<0.001	<0.001
1.45	2.436	16.526	.007	<0.001
1.50	1.800	17.181	.035	<0.001
1.55	1.185	17.816	.117	<0.001
1.60	.590	18.433	.277	<0.001
1.65	.0136	19.032	.494	<0.001
1.70	.492	19.614	.311	<0.001
1.75	1.035	20.181	.150	<0.001
1.80	1.563	20.733	.589	<0.001
1.85	2.077	21.271	.018	<0.001

$\Gamma = e^{\gamma}$: odds of differential assignment due to unobserved factors
Q^{+}_{MH}: Mantel-Haenszel statistic (assumption: overestimation of treatment effect)
Q^{-}_{MH}: Mantel-Haenszel statistic (assumption: underestimation of treatment effect)
p^{+}_{MH}: significance level (assumption: overestimation of treatment effect)
p^{-}_{MH}: significance level (assumption: underestimation of treatment effect)

Table 12: Sensitivity analysis using Mantel-Haenszel bounds for research job and STEM graduation

$\Gamma = e^{\gamma}$	Q^{+}_{MH}	Q^{-}_{MH}	p^{+}_{MH}	p^{-}_{MH}
1.00	10.402	10.402	<0.001	<0.001
1.05	9.555	11.253	<0.001	<0.001
1.10	8.748	12.065	<0.001	<0.001
1.15	7.978	12.842	<0.001	<0.001
1.20	7.242	13.588	<0.001	<0.001
1.25	6.536	14.305	<0.001	<0.001
1.30	5.859	14.996	<0.001	<0.001
1.35	5.208	15.662	<0.001	<0.001
1.40	4.582	16.305	<0.001	<0.001
1.45	3.977	16.927	<0.001	<0.001
1.50	3.394	17.529	<0.001	<0.001
1.55	2.830	18.113	.002	<0.001
1.60	2.284	18.680	.0111	<0.001
1.65	1.755	19.230	.039	<0.001
1.70	1.242	19.765	.107	<0.001
1.75	.744	20.286	.228	<0.001
1.80	.260	20.793	.397	<0.001
1.85	.152	21.288	.439	<0.001
1.90	.610	21.770	.270	<0.001
1.95	1.057	22.241	.145	<0.001
2.00	1.492	22.701	.067	<0.001
2.05	1.916	23.150	.027	<0.001

$\Gamma = e^{\gamma}$: odds of differential assignment due to unobserved factors
Q^{+}_{MH}: Mantel-Haenszel statistic (assumption: overestimation of treatment effect)
Q^{-}_{MH}: Mantel-Haenszel statistic (assumption: underestimation of treatment effect)
p^{+}_{MH}: significance level (assumption: overestimation of treatment effect)
p^{-}_{MH}: significance level (assumption: underestimation of treatment effect)

Table 13: Standardized differences for grant employment as treatment and graduation as outcome, inverse probability weighting

	Std diff Raw	Std diff Weighted	Var ratio Raw	Var ratio Weighted
Female	.085	-.002	.984	1.000
Black female	.026	-.003	1.131	.983
Hispanic female	.011	.001	1.067	1.007
Asian female	.145	-.003	1.602	.989
Other race female	.048	-.004	1.378	.971
Black	-.013	-.006	.951	.977
Hispanic	-.037	.002	.857	1.010
Asian	.174	-.004	1.436	.992
Other race	.032	-.003	1.171	.984
In state	.325	-.009	.739	1.013
Intent STEM	.538	-.006	.759	1.007
HS GPA	.279	-.005	.924	1.013
International student	-.060	.000	.722	1.005
Pell grant	.162	-.006	1.243	.992
Pell grant * female	.151	-.003	1.406	.993
ACT composite score	.196	.001	.944	.990
ACT composite score sq.	.205	.002	1.027	1.032
AP CALAB	.183	-.000	1.189	.999
AP CALBC	.172	-.002	1.486	.995
AP BY	.263	.002	1.547	1.003
AP CH	.268	-.010	1.709	.984
AP CSA	.040	.000	1.446	1.002
AP CSAB	.012	.001	1.217	1.022
AP PHYSB	.093	.001	1.493	1.004
AP PHYSE	.064	-.007	1.595	.954
AP PHYSM	.153	-.004	1.787	.986
AP STAT	.015	-.000	1.048	.997

Notes: Standardized differences between the treatment and comparison groups.

Table 14: Standardized differences for research jobs as treatment and graduation as outcome, inverse probability weighting

	Std diff Raw	Std diff Weighted	Var ratio Raw	Var ratio Weighted
Female	.127	-.003	.967	1.001
Black female	.115	-.006	1.720	.973
Hispanic female	.048	.000	1.309	1.003
Asian female	.123	-.004	1.493	.987
Other race female	.068	-.005	1.591	.970
Black	.107	-.008	1.496	.972
Hispanic	-.005	.002	.979	1.010
Asian	.149	-.004	1.372	.991
Other race	.026	-.003	1.144	.982
In state	.314	-.011	.757	1.014
Intent STEM	.408	-.009	.867	1.007
HS GPA	.201	-.010	.945	1.004
International student	-.073	.001	.666	1.011
Pell grant	.282	-.013	1.477	.987
Pell grant * female	.271	-.008	1.827	.986
ACT composite score	.061	.002	1.140	.981
ACT composite score sq.	.073	.004	1.155	1.036
AP CALAB	.104	-.000	1.104	.999
AP CALBC	.114	-.002	1.299	.993
AP BY	.188	.003	1.377	1.004
AP CH	.201	-.010	1.509	.982
AP CSA	.043	-.000	1.470	.997
AP CSAB	-.013	.000	.790	1.012
AP PHYSB	.045	.000	1.223	1.002
AP PHYSE	.043	-.005	1.374	.966
AP PHYSM	.110	-.002	1.525	.991
AP STAT	-.024	.001	.928	1.004

Notes: Standardized differences between the treatment and comparison groups.

Table 15: Standardized differences for research jobs as treatment and STEM graduation as outcome, inverse probability weighting

	Std diff Raw	Std diff Weighted	Var ratio Raw	Var ratio Weighted
Female	.066	-.001	.986	1.000
Black female	.025	-.003	1.138	.981
Hispanic female	.017	.000	1.108	1.002
Asian female	.134	-.001	1.538	.995
Other female	.058	.004	1.488	.970
Black	.007	-.005	1.033	.977
Hispanic	-0.25	.002	.896	1.011
Asian	.173	-.003	1.431	.994
Other race	.033	-.004	1.187	.980
In state	.328	-.010	.740	1.014
Intent STEM	.553	-.007	.763	1.008
HS GPA	.258	-.007	.913	.999
International student	-.055	.000	.738	.998
Pell grant	.180	-.006	1.305	.992
Pell grant * female	.163	-.003	1.471	.993
ACT composite score	.164	0.001	1.041	.980
ACT composite score sq.	.176	.002	1.083	1.035
AP CALAB	.166	-.000	1.159	.999
AP CALBC	.166	-.002	1.441	.995
AP BY	.258	.004	1.509	1.005
AP CH	.266	-.010	1.671	.984
AP CSA	.049	-.001	1.546	.989
AP CSAB	-.007	.003	.887	1.065
AP PHYSB	.086	.000	1.443	1.002
AP PHYSE	.060	-.006	1.535	.959
AP PHYSM	.147	-.005	1.722	.984
AP STAT	.002	-.000	1.009	.997

Notes: Standardized differences between the treatment and comparison groups.

Table 16: Sensitivity analysis using Mantel-Haenszel bounds for grant employment and graduation

Subsample	Γ'	Γ''
Female	1.25	1.60
Male	1.20	1.65
White	1.15	1.45
Black	1.35	2.40
Hispanic	1.25	3.05
Asian	1.10	1.70
Pell grant	1.35	1.80

Notes: Results shown only for the Mantel-Haenszel upper bounds (assumption: overestimation of treatment effect).
Γ': lowest value of $\Gamma = e^{\gamma}$ for which the test statistic p^{+}_{MH} is not significant at a 5 percent significance level.
Γ'': highest value of $\Gamma = e^{\gamma}$ for which the test statistic p^{+}_{MH} is not significant at a 5 percent significance level.
p^{+}_{MH}: significance level (assumption: overestimation of treatment effect)

Table 17: Sensitivity analysis using Mantel-Haenszel bounds for research jobs and graduation

Subsample	Γ'	Γ''
Female	1.15	1.55
Male	1.15	1.60
White	1.20	1.60
Black	1.00	1.45
Hispanic	1.25	3.55
Asian	1.00	1.60
Pell grant	1.30	1.90

Notes: Results shown only for the Mantel-Haenszel upper bounds (assumption: overestimation of treatment effect).
Γ': lowest value of $\Gamma = e^{\gamma}$ for which the test statistic p^+_{MH} is not significant at a 5 percent significance level.
Γ'': highest value of $\Gamma = e^{\gamma}$ for which the test statistic p^+_{MH} is not significant at a 5 percent significance level.
p^+_{MH}: significance level (assumption: overestimation of treatment effect)

Table 18: Sensitivity analysis using Mantel-Haenszel bounds for grant employment and STEM graduation

Subsample	Γ'	Γ''
Female	1.30	1.65
Male	1.70	2.20
White	1.55	1.90
Black	1.00	2.15
Hispanic	1.50	3.95
Asian	1.50	2.25
Pell grant	1.25	1.75

Notes: Results shown only for the Mantel-Haenszel upper bounds (assumption: overestimation of treatment effect).
Γ': lowest value of $\Gamma = e^{\gamma}$ for which the test statistic p^+_{MH} is not significant at a 5 percent significance level.
Γ'': highest value of $\Gamma = e^{\gamma}$ for which the test statistic p^+_{MH} is not significant at a 5 percent significance level.
p^+_{MH}: significance level (assumption: overestimation of treatment effect)

Table 19: Sensitivity analysis using Mantel-Haenszel bounds for research job and STEM graduation

Subsample	Γ'	Γ''
Female	1.60	2.05
Male	1.80	2.40
White	1.70	2.10
Black	1.10	3.10
Hispanic	1.00	2.60
Asian	1.20	1.85
Pell grant	1.15	1.70

Notes: Results shown only for the Mantel-Haenszel upper bounds (assumption: overestimation of treatment effect).
Γ': lowest value of $\Gamma = e^{\gamma}$ for which the test statistic p^{+}_{MH} is not significant at a 5 percent significance level.
Γ'': highest value of $\Gamma = e^{\gamma}$ for which the test statistic p^{+}_{MH} is not significant at a 5 percent significance level.
p^{+}_{MH}: significance level (assumption: overestimation of treatment effect)