capsee

CENTER FOR ANALYSIS OF
POSTSECONDARY EDUCATION
AND EMPLOYMENT

Estimating Returns to College Attainment: Comparing Survey and State Administrative Data Based Estimates

Appendices A, B, C, D, and E

A CAPSEE Working Paper

Judith Scott-Clayton
Community College Research Center
Teachers College, Columbia University
and NBER

Qiao Wen
Community College Research Center
Teachers College, Columbia University

January 2017

List of Appendices

Appendix A
Definitions and Additional Notes for Key Variables

A1. Educational Attainment

We determine respondents' educational attainment using both self-reported survey data and administrative data from the postsecondary transcripts study, which was administered by the NLSY97 research team during 2010–2011 (see Appendix C for a more detailed introduction to the study). For respondents whose transcripts data are available, we identify their "highest degrees obtained" as the highest degrees listed on transcripts from all the postsecondary institutions. If such information is unavailable, we use their self-reported survey data.[1] More specifically, we use their self-reported "highest degree obtained" as of 2009 (in round 14), the year prior to the year of earnings considered. For individuals who were not interviewed in that year, we use their educational attainment as of their most recent round of interview, if they were not enrolled in school in 2009 and afterwards. We provide crosschecking results regarding postsecondary enrollment and attainment between the self-reports and transcripts data in Appendix D.

We further classify educational attainment into six categories: high school diploma or GED only (i.e., without any postsecondary enrollment), attended some two-year college but obtained no degree, attended some four-year college but obtained no degree, undergraduate certificate, associate degree, and bachelor's degree and above.[2]

A2. Job Earnings

Our analysis uses total earnings from all jobs in 2010, which includes either self-employed jobs or employee-type jobs. Job earnings include wages, salary, commissions, and tips.[3] Apart from job earnings, NLSY97 also documents income from business or farm work, investment revenues, rental property, dividend and interest, worker compensation, child support and all other earnings. As an alternative, we could use total earnings, but it might not be a better measurement because: first, for other earnings categories, the survey questions ask for income earned by the respondents together with his or her spouse or partner, instead of the respondents alone as is the case for job income; and second, some other categories of earnings, such as investment and business income, include possibly negative values which would complicate the

---

[1] 34.6 percent of the respondents who have reported any postsecondary enrollment have no transcript data available for us to validate their degree receipt information.

[2] There are 459 respondents (approximately 5 percent) who have obtained a graduate degree (including master's degrees, Ph.D.s, and professional degrees). Due to this small size, we combine them with respondents who have bachelor's degrees and create the "bachelor's degree and above" group.

[3] For respondents who were not interviewed in 2011 and thus do not have income information, we use job earnings in 2009 (inflated to 2010 dollars) if the individuals were not enrolled in school since 2009. This affects 4.44 percent (approximately 300 individuals) of the total analysis sample. We create a variable to indicate which income year is used for each individual.

notion of earnings when adding up all those categories of income. Thus, we decide to use earnings from jobs only.

A3. Other Labor Market Outcomes

Apart from earnings, we also look at other labor market outcomes in order to capture variations in employment and labor supply characteristics by degree level. These outcomes include the probability of having at least a job, the probability of having positive earnings from jobs, the probability of having any source of positive earnings, total weeks and hours worked during the calendar year and probability of working full-time and year-round.

More specifically for the last indicator, we use respondents' labor supply characteristics to proxy for their full-time and year-round working status. According to U.S. Department of Labor, the Fair Labor Standards Act (FLSA) does not define full-time employment or part-time employment. This is a matter generally to be determined by the employer.[4] However, for statistical purposes, both the U.S. Bureau of Labor Statistics (BLS) and U.S. Census Bureau define "full-time workers" as persons who usually work 35 hours or more per week, and "year-round workers" as those who work 50 to 52 weeks during a calendar year.[5] For this study, we use two methods to proxy for "full-time and year-round" working status: first, following the BLS and Census approach, we use respondents' total hours worked in 2010, and classify respondents who worked 1,750 hours (i.e., 35 hours per week multiplied by 50 weeks) or more as "full-time, year-round" workers. However, we are afraid that information on total hours worked during a calendar year might not be available in most datasets, especially for administrative data. Therefore, we use an alternative approach that determines respondents' working status, by using their job income. More specifically, we compute a minimum annual job income by multiplying the federal minimum wage rate in a given year by 1,750 hours, and classify respondents whose job income is equal to or above this threshold as "full-time, year-round" workers. We further use these two proxies for sensitivity analysis to compare the estimates of returns to degree level when different samples are selected.

A4. Interstate Mobility

We derive migration information from the restricted-use geocode data, and construct two variables to capture respondents' interstate migration behaviors. We create two dummies to indicate whether the respondents have ever moved across states between two years or survey rounds since he or she turned 18 years old, and whether the respondent's state of residence in

---

[4] https://www.dol.gov/general/topic/workhours/full-time
[5] https://www.census.gov/hhes/www/laborfor/faq.html#Q7 and
http://www.bls.gov/opub/ted/2014/ted_20141223.htm

2010 is different from his or her state of residence at age 17.[6] We intend to use state of residence in 2010 to proxy for where the respondents work in the most recent year the survey data are available, and use state of residence at age 17 to proxy for respondents' home state before college entrance. For our sensitivity analysis, we use the second dummy to identify potential bias in returns to education if only single-state administrative earnings data are available for use.

Ideally, in order to mimic single-state administrative databases, we would like to use the states where respondents attended college, rather than their home states, to see whether they have migrated out of the states where they attend colleges to work. But, the NLSY97 interviews ask respondents' state of residence by asking the address where they live in a given year, and a substantial proportion of respondents reported their permanent home address, even if they attended college out of their own states. Moreover, only 4 percent out of our analysis sample migrated across states from age 17 to attend (first) college,[7] suggesting that home states are an appropriate proxy for constructing our dummy for whether respondents moved across states between college and work.

A5. Key Control Variables

NLSY97 contains very detailed information on respondents' family background and pre-college performance. We incorporate in our regression models these variables that might affect both respondents' college experience and labor market outcomes. Our model specifications include four sets of controls. Basic controls include gender, age, and race/ethnicity, and we classify respondents' race/ethnicity into four categories: White non-Hispanic, Black non-Hispanic, Hispanic, and other races/ethnicities.

The second set of controls includes the level and squared forms of working experience, and we measure working experience as respondents' cumulative hours worked from the year he or she turned 18 through 2010. If respondents miss hours of work information for certain year(s), we impute the hours as zero for that year.

The third set of controls further includes geographic and family background controls. The publicly available geographic information captures individuals' residence in the first round when they were 13–17 years old. More specifically, they include regions of residence (i.e., Northeast, South, West, and North Central) and characteristics of the residence (i.e., urban/rural/unknown area and MSA/non-MSA area). Family background controls include parental educational attainment, household net worth in 1997 (round 1), and household size. The parental educational

---

[6] The "ever moved across states since 18 years old" dummy equals to 1 if individuals have changed their state of residence at least once starting from when they were 18 years old to the year 2011; this dummy equals to 0 if "state of residence" is the same for every year when the respondent was 18 years old and older, or "state of residence" is the same for all available years since 18 years old and "state of residence" information for the remaining years is missing. This dummy equals to *missing* if information for "state of residence" is missing for every single year or it is only available for one year since the respondent was 18 years old.

[7] Again, respondents might report their home address even if they attend college out of their home states.

attainment variable comes from the parent's survey that registered at round 1.[8] Household net worth is calculated by subtracting total debts from total assets. For the majority, this variable is also derived from the parent's survey. For a small number of youths who were considered "independent" in 1997, household net worth was reported by the youths.[9]

The last set of controls further include pre-college performance, including type of high school attended, overall high school GPA and test score on the Armed Services Vocational Aptitude Battery (ASVAB). High school type includes public, private or parochial, and other types of high school. Overall high school GPA indicates grade point averages across all high school courses on a 5-point grading scale. It is weighted by course credits. High school GPA is derived from the supplemented survey of high school transcript study. We also include the test score on the Armed Services Vocational Aptitude Battery (ASVAB) as a measurement of individual ability. ASVAB is a military enlistment test. We use the adjusted ASVAB score constructed by the NLSY97 research team, which indicates respondents' percentile ranking on the four math and verbal subtests that take into consideration the respondents' age. Moreover, the high school GPA variable is derived from the supplemented survey of high school transcript study, and only 69 percent of the overall NLSY97 sample has valid high school transcript data. Likewise, only 79 percent of the NLSY97 sample completed the ASVAB test during 1997–1998. A more detailed discussion about our approach for missing imputations is presented in Appendix B.

---

[8] We use respondents' bio-mother's highest degree obtained to construct for parental educational attainment. We replace it with respondents' bio-father's highest degree obtained if his or her bio-mother's education information is missing. We further replace it with his or her stepparents' education if both his or her bio-parents' education information is missing.

[9] NLSY97 uses the following criteria for "independent": NLSY97 youths were considered independent if they have had a child, were enrolled in a four-year college, were no longer enrolled in school, were not living with any parents or parent-figures, or had ever been married or were in a marriage-like relationship (defined in rounds 1–8 as a sexual relationship in which partners of the opposite sex live together) at the time of the survey. Reaching the age of 18 was another criterion for independence, but the reference date for that age varied between surveys and questionnaire sections.

Appendix B
Additional Details of Data Cleaning

B1. Top Coding for Income and Assets Values

In order to protect the confidentiality of respondents, NLSY97 "top codes" the highest income and asset values. More specifically, for income variables in our analysis, the top 2 percent of reported values are top coded and replaced with the mean of the high values. For the family net worth variable, respondents with family net worth above $600,000 in 1997 dollars are top coded to the value of $600,000.

B2. Imputations for Missing Values

Missing values for variables in the NLSY97 can occur for five reasons: (1) the respondent refused to answer the question; (2) the respondent did not know the answer to the question, after the interviewer giving hints and clarifications to the question; (3) the respondent was not interviewed for the entire round of the survey and thus had missing values for all questions in that round; (4) for computer-based questions, if the respondent skipped the question when he or she should have answered that question (i.e., the question applies to him/her), a missing value was assigned to that question; and (5) the question did not apply to the respondents. For our analysis, we recode missing values due to the last reason based on the logic and answers to previous questions. For example, a follow-up question asking "how many jobs did you have" applies only to respondents who answered "yes" to the previous question that asks whether a respondent held any jobs in a certain year and codes respondents who answered "no" as missing. For our analysis, we recode above-mentioned missing values to zero because we know they are equivalent to "zero" jobs in a certain year. For the other four categories of missing values, we do not distinguish them from each other and recode them all to a single category of missing.

After this initial recoding, we further impute missing values for some variables to avoid substantial sample size reduction. For example, for control variables that have a high percentage of missing values, such as respondent's high school GPA and family net worth, we impute missing values to zero and thus include them in regression models. For labor market outcomes in 2010, if variables have missing values due to respondents missing the entire round of the interview, we impute those missing values by using respondents' information from the previous round (i.e., labor market outcomes in 2009) and inflate variables that measured in dollar values to 2010 dollars.[10] We create flags to indicate all imputed values.

---

[10] We replace missing values with labor market outcomes in 2009 if the respondents were not enrolled in school in 2009 and afterwards.

Appendix C
Comparing Self-Reported and Administrative College Enrollment and Attainment Data

C1. Introduction to the Postsecondary Transcript Study

The Postsecondary Transcript Study collects undergraduate transcripts for the NLSY97 respondents who have reported any postsecondary enrollments in any survey round from 1997 (round 1) to 2011 (round 15). This study was carried out separately in addition to the NLSY97 surveys, and data are now available in the public-use NLSY97 dataset.[11] The transcripts data contain variables at institution, term, and course levels, providing detailed information about respondents' postsecondary enrollment and course-taking patterns, college performance, and degree receipt.

For our analysis, we compare data from the postsecondary transcript study and self-reported survey data to see whether there are any discrepancies regarding information on postsecondary enrollment and degree attainment between these two sources and to detect patterns of misreporting for the self-reports.

C2. Previous Studies on Crosschecking Multiple Sources of Educational Data

The NLSY97 research team has not yet conducted any internal consistency check between the two sources of college experience data that are available in the NLSY97. However, the NLSY97 has conducted a similar transcript study about high school experience, and Datta and Krishnamurty (2008) compare that with self-reported high school information from the survey by crosschecking key variables, such as number of high schools attended, high school GPA and performance on math courses, and receipt of high school credentials. They found that students with better academic results are more likely to have matched information between the two sources. Due to some administrative errors in the high school transcript data, they cannot claim which source of the data is more reliable.

Apart from NLSY97, several studies use other data sources to address the issues of misreporting of educational information. "Estimating Returns to Schooling When Schooling Is Misreported" by Kane, Rouse, and Staiger (1999) compares educational attainment data from self-reported National Longitudinal Study of High School Class of 1972 (NLS72) and the Postsecondary Education Transcript Studies (PETS) to investigate degrees of disagreement between the two sources and how the discrepancies may affect estimates of returns to schooling.[12] Assuming that transcript data represent true information, they find that people are

---

[11] The Postsecondary Transcript Study is conducted by researchers from the University of Texas-Austin and University of Wisconsin-Madison. Data collection and cleaning are carried out by NORC at the University of Chicago.

[12] For more information about major transcript studies in the U.S., see: https://nces.ed.gov/surveys/pets/about.asp

more likely to over-report than under-report, and that the misreporting rate is lower when educational attainment is measured by degree attainment, rather than years of schooling. They find that people misreport their educational information because they lie, they do not know whether the schooling they have counts as credentials, or they do not remember.

"Measurement of Higher Education in the Census and Current Population Survey" by Black, Sanders, and Taylor (2003) compares information regarding educational attainment from three sources of self-reported datasets: the 1990 Decennial Census, the post-1991 Current Population Survey (CPS) and the 1993 National Survey of College Graduates (NSCG). The Census uses a mail system and relies on a single question to measure educational attainment; the CPS uses face-to-face interviews, and interviewers can seek clarifications if they think the respondents' answers are inconsistent with their previous answers; the NSCG asks respondents to write down the name and address of each school from which they obtained the credentials, time of degree receipt, and type and major of the degree. The authors assume that the NSCG is the most reliable source and then compare degree distributions in Census and CPS to that of the NSCG. They find a higher disagreement rate between the Census and NSCG than between CPS and NSCG. They also conduct a sub-group analysis and find that there is a higher degree of discrepancies among minorities. They attribute it to minorities' language barriers and unfamiliarity with the U.S. higher education system. This study has two main caveats: first and foremost, this study does not have a source to validate true degree receipt status (for example, by using an administrative dataset). Even the NSCG depends on self-reported data, and people can misreport on that as well. Moreover, they use respondents' occupations to "best guess" whether the respondents have "professional degrees," and categorize people who work as nurses or hairdressers but report that they have professional degrees as misreporting their degrees (because the researchers believe that they actually have vocational certificates rather than professional degrees like an MBA). This approach might involve substantial measurement errors, which can render their estimates inaccurate and inefficient; lastly, their samples of the Census, CPS, and NSCG include non-identical respondents. However, the authors assume that all samples have the same distribution in terms of educational levels, but their estimates would be biased if the true distributions differ among the three samples.[13]

To sum up, previous studies on comparing self-reports and administrative datasets show that discrepancies do exist, and they tend to believe that administrative datasets are more likely to reflect true information. In the following section, we compare self-reports and transcripts data for the NLSY97 regarding information on postsecondary enrollment and degree attainment.

---

[13] More specifically, they construct comparable samples from the full samples.

Appendix D
Results of Crosschecking Self-Reported Survey Data and Postsecondary Transcripts Data for the NLSY97

Our crosschecking method is as follows: for degree attainment information from the self-reported NLSY97 survey data, we use "highest degree obtained" as of 2011 (round 15). We then compare them with degrees listed in the transcripts from the postsecondary transcript study and assume that the latter represent true degree attainment information.

According to the NLSY97 survey data, 68.5 percent of the full NLSY97 sample has attended postsecondary schools as of 2011 (round 15). For the Post-Secondary Transcript Study (PSTRAN), permissions to collect transcripts were received from 52.4 percent of those who claimed to have attended college, and at least one transcript was received for 42.5 percent of them. Table D1 below shows transcript status for the full NLSY97 sample.

As Table D1 shows that, out of the 4,709 respondents who gave permission to obtain all their postsecondary transcripts, 207 respondents (2.3 percent) have confirmed to be "never enrolled in a degree program at the named institution." Most of them had been registered but never attended, or enrolled in non-degree coursework only. Table D2 below lists the "highest degree obtained" reported in the survey for these 207 respondents who are confirmed to have over-reported their postsecondary enrollment information.

Table D1: Transcript Status for the Full NLSY97 Sample

Transcript Status			N	%	Implication
No postsecondary enrollment reported in survey data			2,830	31.5	Cannot check enrollment or degree receipt information; might under-report college experience
Post-secondary enrollment reported in survey data (6,154)	Did not participate in the Transcript Study		1,445	16.1	Cannot check enrollment or degree receipt information; might over- or under- report
	Participated in the Transcript Study (4,709)	Enrollment in at least one postsecondary institution confirmed; at least one transcript received	3,818	42.5	Confirmed to have enrolled in post-secondary institution(s); can check degree receipt information associated with received transcript(s)
		Enrollment in at least one postsecondary institution conformed; no transcript received	231	2.6	Confirmed to have enrolled in post-secondary institution(s); cannot check degree receipt information; might over- or under- report
		None of the reported enrollment in postsecondary institution(s) was valid	207	2.3	Over-reported postsecondary enrollment; degrees associated with such enrollments are also over-reports
		No confirmed enrollments; at least one reported postsecondary institution cannot be located	453	5.0	Cannot check enrollment or degree receipt; most likely to be over-reports
Total			8,984	100	

Table D2: Self-Reported Education Attainment for Respondents Who Over-Reported Post-Secondary Enrollment but Actually Never Enrolled in College

Self-Reported Highest Degree Obtained	N	%
None	21	10.1
GED	56	27.1
High school diploma	109	52.7
Associate degree	10	4.8
Bachelor's degree	6	2.9
Professional degree	1	0.5
Missing information	4	1.9
Total	207	100.0

Apart from these 207 people, we also crosscheck the 3,818 respondents from whom at least one transcript has been received. As Table D3 shows, of these people, 20.6 percent over-reported the number of undergraduate enrollments they had attended.[14] For the other 79.4 percent of respondents, all of their self-reported "undergraduate degree programs attended" are confirmed to be valid.

Table D3: Crosschecking Number of Undergraduate Institutions Attended Between Self-Reports and Transcript Data

Number of Undergraduate Institutions Attended	N	%
Match	3,030	79.4
Do not match	788	20.6
Total	3,818	100.0

Of these 3,030 people, transcripts associated with *all* the institutions reported were received for 2,428 respondents (63.6 percent). At least one transcript has *not* been received for the other 1,390 respondents (36.4 percent) (see Table D4. below).

Table D4: Transcript Status for Respondents Whose Self-Reported Number of Postsecondary Institutions Attended Matches With Transcript Data

Dummy: Received all transcript(s) for all self-reported postsecondary institutions	N	%
Yes	2,428	63.6
No	1,390	36.4
Total	3,818	100.0

We first focus on the 2,428 respondents whose transcript(s) for all their self-reported postsecondary institutions attended were *all* received. More specifically, we compare degree receipts information derived from the transcripts study with their self-reported "highest degree obtained" from the survey data. Out of these 2,428 respondents, 5 respondents have missing values regarding their highest degree obtained in the survey data, and degree receipts information listed in transcripts cannot be coded for 7 respondents. Out of the remaining 2,416 respondents, nearly 48 percent have not received any postsecondary credentials according to their transcripts (see Table D5 below).

---

[14] It is possible that these respondents enrolled in institutions that did not issue transcripts at the time the postsecondary transcripts study was conducted.

Table D5: Degree Receipt Status for Respondents Whose Transcripts Are All Received

Dummy: Received any post-secondary degrees according to transcript data	N	%
Yes	1,266	52.1
No	1,162	47.9
Total	2,428	100.0

For those who have at least one degree listed in their transcripts, we crosscheck types of degrees received between transcripts and self-reports. Note that the two data sources use a slightly different categorization of degree types. The postsecondary transcript study categorizes degrees into four types: undergraduate certificate, associate degree, bachelor's degree, and graduate degree (grad).[15]

As Table D6 shows, for these 2,416 respondents whose transcripts are all received, 2,040 respondents (84.4 percent, cells shaded green in Table D6) have matched degree receipt information between their transcript(s) and self-reports. Sixty-six respondents or 2.7 percent (cells shaded red in Table D6) over-report their degree attainment in the survey, and 95 respondents or 3.9 percent (cells shaded yellow in Table D6) under-report their degree attainment. Truthfulness of self-reports for the remaining 215 respondents or 8.9 percent (cells shaded grey in Table D6) are undetermined, either because they have a certificate listed on their transcript(s), which does not have an equivalent category in the survey data, or because they report a graduate degree, which is out of scope of the transcript study. Setting aside these 215 respondents and also taking into consideration the over-reporting of degrees presented in Table D2, the matching rate in terms of postsecondary degree receipts is approximately 92 percent for those whose transcripts for all self-reported postsecondary institutions attended were all received. The corresponding over- and under-report rates are approximately 3.7 percent and 4.3 percent respectively.

---

[15] The transcript study only asks for undergraduate transcripts, but for people who had undergraduate and graduate coursework within the same institution that issues one single transcript, the degree type listed on such transcripts may show "graduate degree."

Table D6: Comparing "Degree(s) Received" in Transcript Data and Self-Reported "Highest Degree Received" in Survey Data for Respondents Whose Transcripts Are All Received

Transcript: Degree Received	Survey: Highest Degree Obtained				
	No Post-secondary Degree	Associate Degree	Bachelor's Degree	Graduate	Total
No degree	1,096	36	25	2*	1,159
Undergraduate certificate only**	129*	22*	1	0	152
Associate degree only	31	164	4	0	199
Bachelor's degree only	34	0	600	35*	669
Certificate + associate degree	5	24	0	1*	30
Certificate + bachelor's degree	0	0	20	0	20
Associate degree+ bachelor's degree	1	6	95	2*	104
Certificate + associate degree+ Bachelor's degree	1	0	6	0	7
Grad. only*	6*	16*	2*	0	24
Certificate + grad.	2	2	0	0	4
Associate degree + grad.	0	2	0	0	2
Bachelor's degree + grad.	0	0	8	30	38
Certificate + associate degree + grad.	1	0	0	0	1
Certificate + bachelor's degree + grad.	0	0	1	3	4
Associate degree+ bachelor's degree + Grad.	0	0	1	2	3
Total	1,306	272	763	75	2,416

*The postsecondary transcripts study focuses on undergraduate study. Those who only have transcripts at the graduate level are excluded from the study. The 24 people who have "graduate degree only" are confirmed to have no undergraduate degree(s). (Maybe these people attended undergraduate institutions abroad, or failed their college course but had a graduate degree. Also, maybe there are administrative errors associated with their "graduate" transcripts.)

**"Undergraduate certificate" refers to degrees obtained from technical or occupational programs. The NLSY97 survey does not have this category of credential, thus people were likely to regard their certificates either as associate degrees or as "no postsecondary degrees."

Further including respondents from whom at least one but not all transcripts were received yields higher rates of misreporting. As Table D7 shows, the matching (green), over-report (red), and under-report (yellow) rates for such respondents are 87.9, 8.2, and 3.9 percent respectively. However, note that the "misreporting behaviors" might be due to the fact that their self-reported highest degrees obtained are associated with the un-received transcript(s).

Table D7: Comparing "Degree(s) Received" in Transcript Data and Self-Reported "Highest Degree Received" in Survey Data for Respondents Whose Transcript(s) Are Not All Received

Transcript: Degree Received	Survey: Highest Degree Obtained				
	No Post-secondary Degree	Associate Degree	Bachelor's Degree	Graduate	Total
No degree	1,551	114	122	34*	1,821
Undergraduate certificate only**	173**	29**	7	1*	210
Associate degree only	44	218	24	7*	293
Bachelor's degree only	40	3	873	227*	1,143
Certificate + associate degree	6	29	0	1*	36
Certificate + bachelor's degree	1	0	28	6*	35
Associate degree + bachelor's degree	3	8	130	13*	154
Certificate + associate degree + bachelor's degree	1	1	8	0	10
Grad. only*	7*	23*	4*	0	34
Certificate + grad.	3	3	0	0	6
Associate degree + grad.	0	2	0	0	2
Bachelor's degree + grad.	0	0	11	36	47
Certificate + associate degree + grad.	1	1	0	0	2
Certificate + bachelor's degree + grad.	0	0	1	3	4
Associate degree + bachelor's degree + grad.	0	0	1	5	6
Total	1,830	431	1,209	333	3,803

*The postsecondary transcripts study focuses on undergraduate study. Those who only have transcripts at the graduate level are excluded from the study. The 34 people who have "graduate degree only" are confirmed to have no undergraduate degree(s). (Maybe these people attended undergraduate institutions abroad, or failed their college course but had a graduate degree. Also, maybe there are administrative errors associated with their "graduate" transcripts.)

**"Undergraduate certificate" refers to degrees obtained from technical or occupational programs. The NLSY97 survey does not have this category of credential, thus people were likely to regard their certificates either as associate degrees or as "no postsecondary degrees."

Table D8 below summarizes the crosschecking results. Over-reporting includes over-reporting enrollment or degree attainment. If we assume that information listed on the transcripts represent true postsecondary enrollment and degree attainment information, there are 3.7 to 8.2 percent of respondents who over-reported their postsecondary degree attainment. There are 4.4 to 20.6 percent of respondents who over-reported their postsecondary enrollment, either by claiming to have enrolled in college but never enrolled, or enrolled in at least one but claimed larger number of institutions attended than the true number of institutions that were confirmed as valid enrollment. Approximately 3.9 to 4.3 percent of respondents under-reported their degree attainment information in the survey dataset.

Table D8: Summary of Misreporting Behaviors in the NLSY97 Survey Dataset

Over-Reports and Under-Reports		%
Over-reporting enrollment:	Reported postsecondary enrollment, but never enrolled	4.4
	Enrolled; number of institutions enrolled reported > confirmed	20.6
Over-reporting degree attainment:		3.7–8.2
Under-reporting degree attainment:		3.9–4.3

There are several limitations of our approach of crosschecking, and we note that results presented in Table D8 might either over- or under- estimate true misreporting behaviors in the survey data for the following reasons. First, transcripts were requested in 2010 and 2011 (round 14 and round 15). Individuals not reporting college enrollment prior to 2010 and not completing the round 14 interview are not systematically asked to provide a "waiver" that permits the researchers to obtain transcript(s) from all the institutions they reported to have enrolled in. Respondents who completed the round 14 or round 15 interviews, or who reported postsecondary enrollment prior to round 14, have all been asked for the waiver to release their transcripts. The dates of degree receipts listed on transcripts span from 1998 to 2013. Therefore, it is possible that respondents received degrees after their most recent round of interview when they reported their most recent degree information. As a result, "under-reporting" of degree attainment as shown in Table D8 may be due to different timings of the survey interviews and transcript release dates.[16] Different timings can affect over-reports as well: if transcripts were obtained in round 14, and people enrolled in school after that and obtained additional degrees and thus reported higher degree attainment in round 15, then the over-reporting shown in Table D8 might actually reflect truthful reports.

However, as noted before, the number of respondents who obtained their highest credentials in round 15 is negligible. Therefore, we believe that results presented in Table D8 are an accurate estimate of the misreporting behaviors in the NLSY97 survey.[17]

Table D9 below summarizes our decisions on how to recode self-reported postsecondary educational data according to the above crosschecking results. Briefly, we use data from the postsecondary transcript study as the main source of postsecondary educational enrollment and attainment. We make flags to indicate respondents for whom we find discrepancies regarding "highest degree obtained" between their self-reports and transcripts. For respondents whose educational information cannot be confirmed by transcripts, we use the self-reported survey data.

---

[16] However, very few respondents (approximately 0.4 percent) obtained their highest degrees in 2012 or 2013, and most people who have at least one valid transcript completed round 15.

[17] The other possible biases of our estimate might come from administrative errors of transcripts. For our analysis, we assume that the transcripts reflect true and correct information, but in reality transcripts can have errors as well.

Table D9: Decisions for Coding Postsecondary Enrollment and Degree Receipt Information for Respondents With Different Transcript Receipt Status

Transcript Status			*N*	%	Implication	Decision
No postsecondary enrollment reported in survey data			2,830	31.5	Cannot check enrollment or degree receipt information; might under-report college experience	Use survey data
Postsecondary enrollment reported in survey data	Did not participate in the Transcript Study		1,445	16.1	Cannot check enrollment or degree receipt information; might over- or under- report	Use survey data
	Participated in the Transcript Study	Enrollment in at least one postsecondary institution confirmed; at least one transcript received	3,818	42.5	Confirmed to have enrolled in post-secondary institution(s); can check degree receipt information associated with received transcript(s)	Use degree information from transcripts if all were received; use degree information from survey if at least one transcript(s) was not received
		Enrollment in at least one postsecondary institution conformed; no transcript received	231	2.6	Confirmed to have enrolled in post-secondary institution(s); cannot check degree receipt information; might over- or under- report	Use degree information from survey data
		None of the reported enrollment in postsecondary institution(s) was valid	207	2.3	Over-reported postsecondary enrollment; degrees associated with such enrollments are also over-reports	Change survey data to "never attended postsecondary institutions" and obtained "no postsecondary degrees"
		No confirmed enrollments; at least one reported postsecondary institution cannot be located	453	5.0	Cannot check enrollment or degree receipt; most likely to be over-reports	Use survey data
Total			8,984	100		

15

Appendix E
Selected Studies on the Labor Market Returns to Postsecondary Education

Table E1: Selected Studies on Estimating Returns to Higher Education

Authors and Year	Study	Data and Sample	Method	Results
Mincer (1974)	"Schooling, experience, and earnings"	1/1000 sample of the 1960 U.S. Census; White and non-farm men who have earnings in 1959 only	OLS	Return to schooling is 10% and return to experience is 8%
Angrist & Krueger (1991)	"Does compulsory school attendance affect schooling and earnings?"	U.S. 1970 and 1980 Census; men born in 1920–29, 1930–39, and 1940–49	IV (using quarter of birth interacted with birth year)	6%–10.1% (corresponding OLS estimate is 5.2%–7%)
Angrist & Krueger (1992)	"Estimating the payoff to schooling using the Vietnam-era draft lottery"	U.S. 1979–1985 CPS; men born in 1944–1953 (thus were exposed to Vietnam War draft)	IV (using draft lottery number)	6.6% (corresponding OLS estimate is 5.9%)
Kane & Rouse (1993)	"Labor market returns to two- and four-year colleges: Is a credit a credit and do degrees matter?"	NLS Class of 1972; women only	IV (using tuition at 2-year and 4-year state college and distance to nearest college); schooling is measured using units of college credit equivalents	IV returns estimated to be 9.1%, and 9.4% if test scores and parental education are added as controls; credits at 2- and 4- year colleges are interchangeable
Butcher & Case (1994)	"The effect of sibling composition on women's education and earnings"	PSID 1985; White women aged 24 or older	IV (using presence of siblings)	18.5% (corresponding OLS estimate is 9.1%)

Card (1995)	"Earnings, schooling, and ability revisited"	NLS young men (1966 cohort); men who had earnings in 1976	Father–son pairs: using father's education as IV	IV returns estimated to be 13.2% when using college proximity as IV and 9.7% when using college proximity interacted with family background
Ashenfelter & Zimmermann (1997)	"Estimates of the returns to schooling from sibling data: Fathers, sons, and brothers"	NLS young men (1966 cohort), NLS older men; constructed father-son pairs and brother pairs from the NLS 1978 and 1981	For brother pairs: using the other brother's education as IV; for father-son pairs: using father's education as IV	8%–10.9% (corresponding OLS estimate is 4.9%–5.2%)
Rouse (1999)	"Further estimates of the economic return to schooling from a new sample of twins"	Twinsburg Twins Survey; twin pairs interviewed in the 1991, 1992, 1993, and 1995 Twinsburg Twins Festival	Identical twins	11% (corresponding OLS estimate is 7.5%)
Heckman, Lochner, & Todd (2008)	"Earnings function and rates of return"	U.S. decennial Censuses and CPS	OLS; adopt a nonparametric approach that take into account tuition cost, income taxes and nonlinearities in the earnings-schooling-experience relationship	For White males, returns to advancing from 12 to 14 years of schooling in the 1960s, 1970s, 1980s, 1990s, and 2000s range from 6–12%, 6–13%, 5–11%, 7–14% and 8–14 respectively; those for Black males range from 5–11%, 7–12%, 8–12%, 15–18% and 15–19% respectively
Brand & Xie (2010)	"Who benefits most from college? Evidence for negative selection in heterogeneous economic returns to higher education"	NLSY-79 & Wisconsin Longitudinal Study (1957 cohort); a random sample of individuals who graduated from Wisconsin high schools in 1957	Propensity score matching	Individuals who are least likely to obtain a college education benefit the most from college

| Angrist & Chen (2011) | "Schooling and the Vietnam-era GI Bill: Evidence from the draft lottery" | 2000 U.S. Census; men born in 1948–1953 | IV (constructed 5 lottery instruments for different birth cohorts) | Military service increased schooling by around 7% (primarily from more years of college); earnings gains are close to 0 when using annual earnings in 1999; also find a large veteran effect on public-sector employment and a moderate decrease in the probability of living in one's state of birth |

Table E2: Selected Studies on Returns to Sub-baccalaureate Awards

Panel A: Returns to Associate Degrees Relative to High School Credentials:							
Authors and Year	**Study**	**Data and Sample**	**Method**	**Results (College Premiums)**			
Hollenbeck (1993)	"Postsecondary education as triage: Returns to academic and technical programs"	NLSY72	OLS			Men: -0.01	Women: 0.12
Grubb (1993, 1995)	"The varied economic returns to postsecondary education: New evidence from the class of 1972"	NLSY72	OLS			Men:	Women:
				Vocational program		0.00	0.10
				Academic program		0.04	0.03
Kane & Rouse (1995a)	"Labor-market returns to two- and four-year college"	NLSY79	OLS			Men: 0.29	Women: 0.36
Kane & Rouse (1995b)	"Comment on W. Norton Grubb: 'The varied economic returns to postsecondary education: New evidence from the class of 1972'"	NLS72	OLS			Men: 0.08	Women: 0.29
Jaeger & Page (1996)	"Degrees matter: New evidence on sheepskin effects in the returns to education"	CPS91	OLS			Men:	Women:
				Vocational program		0.08	0.31
				Academic program		0.20	0.23
Grubb (1997)	"The returns to education in the sub-baccalaureate labor market, 1984–1990"	SIPP	OLS			Men: 0.18	Women: 0.23
Leigh & Gill (1997)	"Labor market returns to community colleges: Evidence for returning adults"	NLSY79	OLS			Men: 0.24	Women: 0.29
Gill & Leigh (2000)	"Community college enrollment, college major, and the gender wage gap"	NLSY79	OLS			Men: 0.13	Women: 0.21

Averett & Dalessandro (2001)	"Racial and gender differences in the returns to 2-year and 4-year degrees"	NLSY79	OLS		Men:	Women:
				White	0.18	0.19
				Black	0.19	0.33
Surette (2001)	"Transfer from two-year to four-year college: An analysis of gender differences"	NLSY79	OLS		Men: 0.07	Women: 0.13
Ishikawa & Ryan (2002)	"Schooling, basic skills, and economic outcomes"	NALS	OLS		Men:	Women:
				White	0.02	0.05
				Black	-0.01	0.00
				Hispanic	0.06	0.03
Gill & Leigh (2003)	"Do the returns to community colleges differ between academic and vocational programs?"	NLSY79	OLS		Men: 0.22	Women: 0.29
Bailey et al. (2004)	"The return to a sub-baccalaureate education: The effects of schooling, credentials, and program of study on economic outcomes"	HS&B	OLS		Men: 0.12	Women: 0.47
Light & Strayer (2004)	"Who receives the college wage premium? Assessing the labor market returns to degrees and college transfer patterns"	NLSY79	OLS		Men: 0.19	Women: 0.19
Marcotte et al. (2005)	"The returns of a community college education: Evidence from the National Education Longitudinal Survey"	NELS	OLS		Men: 0.17	Women: 0.4

Panel B: Returns to Certificates Relative to High School Credentials Only:

Authors and Year	Study	Data and Sample	Method	Results	
Grubb (1997)	"The returns to education in the sub-baccalaureate labor market, 1984-1990"	SIPP	OLS	Men: 0.08	Women: 0.20
Marcotte et al. (2005)	"The returns of a community college education: Evidence from the National Education Longitudinal Survey"	NELS	OLS	Men: 0.07	Women: 0.24

Panel C: Returns to Community College Enrollment (Without Credentials Obtained) Relative to High School Credentials:

Authors and Year	Study	Data and Sample	Method	Results		
Grubb (1993, 1995)	"The varied economic returns to postsecondary education: New evidence from the class of 1972" & "Response to comment"	NLS72	OLS		Men:	Women:
				Vocational program	0.04	0.02
				Academic program	0.02	0.00
Kane & Rouse (1995)	"Comment on W. Norton Grubb: 'The varied economic returns to postsecondary education: New evidence from the class of 1972'"	NLS72	OLS		Men:	Women:
					0.06	0.07
Jaeger & Page (1996)	"Degrees matter: New evidence on sheepskin effects in the returns to education"	CPS91	OLS		Men:	Women:
					0.09	0.09
Grubb (1997)	"The returns to education in the sub-baccalaureate labor market, 1984-1990"	SIPP	OLS		Men:	Women:
					0.07	0.22
Leigh & Gill (1997)	"Labor market returns to community colleges: Evidence for returning adults"	NLSY79	OLS		Men:	Women:
					0.21	0.04
Gill & Leigh (2000)	"Community college enrollment, college major, and the gender wage gap"	NLSY79	OLS		Men:	Women:
					0.15	0.08
Averett & Dalessandro (2001)	"Racial and gender differences in the returns to 2-year and 4-year degrees"	NLSY79	OLS		Men:	Women:
				White	0.06	0.11
				Black	0.20	0.18
Surette (2001)	"Transfer from two-year to four-year college: An analysis of gender differences"	NLSY79	OLS		Men:	Women:
					0.12	0.13

					Men:	Women:
Bailey et al. (2004)	"The return to a sub-baccalaureate education: The effects of schooling, credentials, and program of study on economic outcomes"	HS&B	OLS		0.00	0.14
Marcotte et al. (2005)	"The returns of a community college education: Evidence from the National Education Longitudinal Survey"	NELS	OLS	Years enrolled:	Men:	Women:
				2+ years	0.17	0.25
				1.5 years	0.13	0.17
				1 year	0.08	0.09
				0.5 year	0.00	0.07
				All	0.06	0.09

Note. The following abbreviations stand for: NLS72: National Longitudinal Survey of the High School Class of 1972; NLSY79: National Longitudinal Survey of Youth 1979; CPS: Current Population Survey; SIPP: Survey of Income and Program Participation; NALS: National Adult Literacy Survey; HS&B: High School and Beyond; NELS: National Education Longitudinal Study of 1988; ACS: American Community Survey.

References

Black, D., Sanders, S., & Taylor, L. (2011). Measurement of higher education in the census and current population survey. *Journal of the American Statistical Association*, *98*(463), 545–554.

Datta, A. R., & Krishnamurty, P. (2008, May). *High school experience: Comparing self-report and transcript data from the NLSY97*. Presented at the NLSY97 10th Anniversary Conference, Chicago, IL.

Kane, T. J., Rouse, C. E., & Staiger, D. (1999). *Estimating returns to schooling when schooling is misreported* (NBER Working Paper No. 7235). Cambridge, MA: National Bureau of Economic Research.