Do Youth Employment Programs Work? Evidence from the New Deal

Anna Aizer
Shari Eli
Guido Imbens
Keyoung Lee
Adriana Lleras-Muney*

Abstract: *We study the lifetime effects of the largest American youth employment and training program – the Civilian Conservation Corps (CCC), 1933-1942. We match newly digitized enrollee records to Census, WWII enlistment, Social Security, and death records. We find that longer duration in the CCC led to improvements in height, health status, longevity, and lifetime earnings but did not improve short-term labor market (employment and wage) outcomes. We address potential selection into CCC duration and correct our estimates of long-term CCC impacts using two newly developed control function approaches that leverage unbiased estimates of the short-term effects of Job Corps (the modern version of the CCC) obtained from an RCT. Our findings suggest that short- and medium-term evaluations of employment programs underestimate their overall impacts because they do not observe total lifetime outcomes and ignore health and longevity benefits.*

Keywords: **Training program, long-term evaluation, lifetime outcomes.**

I. Introduction

Unemployment rates are highest among the young, particularly those from poor backgrounds. Recessions exacerbate this difference: at the height of the Great Recession, unemployment rates for those over age 25 peaked at 8.4% in 2010 but were as high as 19.6% for those aged 16 to 24

---

*Anna Aizer, Brown University (aizer@brown.edu), Shari Eli, University of Toronto (shari.eli@utoronto.ca), Guido Imbens, Stanford University (imbens@stanford.edu), Keyoung Lee, Federal Reserve Bank of Philadelphia (Keyoung.Lee@phil.frb.org), and Adriana Lleras-Muney, UCLA (allerasmuney@gmail.com). We are indebted to Joe Price and the BYU Record Linking Lab for helping us collect the data for this project. We are very grateful to many research assistants that worked on this project, especially Ryan Boone, Taehoon Kang and Kyle Sherman. We have benefitted from comments from participants in the various conferences. We are particularly indebted to Rodrigo Pinto for many valuable contributions, to Nathan Hendren for his careful look at the MVPF computations and to Carlos and Alfonso Flores for their extensive help with Job Corps material. This research was funded by the Social Science and Humanities Research Council of Canada and by the Social Security Administration Grant #NB17-16. This research was also supported by the U.S. Social Security Administration through grant #5-RRC08098400-10 to the National Bureau of Economic Research as part of the SSA Retirement Research Consortium. This project was additionally supported by the California Center for Population Research at UCLA (CCPR), which receives core support (P2C-HD041022) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). Finally, this material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1650604. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation, the SSA, any agency of the Federal Government, the NBER, Federal Reserve Bank of Philadelphia, or the Federal Reserve System. All errors are our own.

(US Bureau of Labor Statistics 2018). To address high rates of youth unemployment, government-run employment training programs target young adults. However, the short-run labor market effects of these programs have been shown to be modest at best. Moreover, although there are many randomized control trials evaluating the labor market impacts of youth training programs, there is limited evidence of their effectiveness over the long-run or their effects on non-labor market outcomes (Card, Kluve, and Weber 2018, Barnow and Smith 2015, Crepon and van den Berg 2016).

We evaluate the *short*- and *long-run* effects of means-tested youth employment and training programs by studying the impact of the Civilian Conservation Corps (CCC), the first and largest employment program in U.S. history.[1] We use new econometric techniques that combine lifetime observational data and results from a modern RCT to obtain causal estimates of the lifetime effects of the program. The CCC, created in 1933 to address high youth unemployment during the Great Depression, employed young men aged approximately 17 to 25 in unskilled, manual labor. Under the Army's supervision, enrollees were sent to work in camps in rural areas where they were also fed, housed, and given access to medical care. In addition to work experience, the CCC provided academic and vocational courses as well as cash transfers to participants' families. Between 1933 and 1942, the CCC trained three million enrollees across 2,600 camps. Several programs in existence today, such as Job Corps, Youth Conservation Corps, and JobsFirstNYC, are modeled after the CCC (Levine, 2010).

We collect a new, large individual-level data set of CCC participants and their lifetime outcomes. We digitize administrative records for roughly 25,000 men from the CCC program in Colorado and New Mexico covering the population served by the program between 1938 and 1943. Our data include information on their demographic characteristics, compensation, enlistment duration and reasons for leaving the program. We match these enrollee records to 1940 Census records, WWII enlistment records, Social Security Administration records, and individual death records. These data allow us to investigate the effects of the CCC on important short- and long-run outcomes, including employment, earnings, and longevity, as well as potential mediators, such as education, health, and geographic mobility.

[1] Salmond (1967) reports that in 1932, 25 percent of youths were unemployed, and another 29 percent were only employed part-time. Rawick (1957) estimates that about 20% of youths were unemployed and another 30% were working part-time.

To estimate the effect of the program, we exploit variation in service duration of the enrollees. Treatment duration varied from a few days to more than two years, with the average enrollee participating for nine months. We estimate two sets of models. Our baseline model attempts to address possible endogeneity by examining the determinants of duration and assessing the extent of omitted variable bias. We show that the determinants of duration are complex and that those who trained for long periods were not necessarily from higher or lower SES backgrounds. Moreover, many enrollees with both short and long service durations ended their training for arbitrary reasons. Controlling for extensive individual and camp-level covariates has little impact on our estimates. We assess the sensitivity of our results to adding various covariates, informally and formally, as suggested by Oster (2017). We also use rich data from Colorado's CCC records to perform placebo tests. We find that neither pre-CCC employment nor health (height and weight) predict duration, though we do find a modest relationship between duration and pre-program education.

In our second model, we apply recently developed econometric methods, which combine RCT results on short-run outcomes and observational data on lifetime outcomes, to address any remaining selection and estimate the lifetime causal effects of the program. We use the experimental data from the Job Corps (JC) randomized controlled trial (RCT) – in the spirit of seminal work by Lalonde (1986) who used experimental data on job training – to shed light on the internal validity of research using observational data. Although the JC data pertains to youth training that took place in the 1990s, the program was modeled after the CCC and retained many similar features. Moreover, JC participants are quite similar to CCC participants with regard to most socio-economic characteristics, duration of training, and reasons for quitting. Evaluation of the JC RCT documents modest short-run labor market effects that do not persist in the medium term (Schochet et al. 2008 and 2018).

To leverage the results from short term impact evaluations from RCTs to correct for potential section bias in long term observational estimates, we apply and extend new econometric techniques developed by Athey, Chetty and Imbens (2020). Specifically, we make use of the fact that the RCT for JC allows us to estimate: 1) the causal impact of job training duration on short-run outcomes; and 2) the amount of bias in OLS (observational) estimates of the impact of the duration of job training on short-run outcomes.

We use this information to address remaining omitted variable bias in the CCC long-run estimates following two control function approaches. In the first approach, we assume that short-run treatment effects are similar across the JC and CCC. In the second approach, we assume that short-run omitted variable bias is similar in the JC and CCC. Our estimates change very little when we include the controls for selection based on the JC RCT, suggesting little bias in our observational CCC estimates. Furthermore, in both cases we explicitly bound the remaining bias generated from violations of each assumption.

We find no short- or medium-run *labor market* benefits associated with job training as part of the CCC, consistent with most of the existing work on contemporary job training programs for youth. The literature on contemporary job training programs, however, cannot trace lifetime effects. Moreover, a growing body of evidence suggesting that the short and long run impacts of programs targeting health and human capital are often very different, with some programs showing fade-out of initial gains and others showing that benefits increase over time (Almond et al. 2018).

When we examine long-term outcomes, we find significant long-term benefits associated with longer training. Those who spend one year in the CCC have higher lifetime earnings by 5.2%, live 1.3 years longer, claim benefits (disability or pensions) at older ages and have 10% lower rates of SSDI claims. These gains are consistent with and likely mediated by the improved health of the participants (measured by height and weight in young adulthood) as well as their increased geographic mobility towards healthier and richer areas.

We conclude that our long-run estimates of job training based on the CCC likely represent causal estimates. Therefore, job training evaluations that focus only on the labor market impact of the program and/or consider only short- and medium-term effects, may underestimate the overall benefits. Our findings also suggest that there are positive returns to investing in young adults, contrary to the commonly stated findings that returns on human capital investment are low after age 18. Our conclusion differs from that of the evaluators of the JC experiment who write, “Because overall earnings gains do not persist, the benefits to society of Job Corps are smaller than the substantial program costs” (Schochet et al, 2018). They also differ somewhat from those of Hendren and Sprung-Keyser (2020), who report low values of the MVPF for JC. The difference is due to the fact that we are able to incorporate large increases in longevity and changes in SSDI claims, as well as increases in lifetime earnings that were not previously known.

This paper makes three contributions to the existing literature. We provide the first estimates of the lifetime effects of a youth training program across many dimensions. There is an extensive literature investigating the effects of youth training programs. Card, Kluve, and Weber (2018) collect estimates from more than 200 program evaluations and conclude that the short-term effects of these programs on labor market outcomes is modest, though there is some important heterogeneity. They also observe that treatment effects appear to increase overtime but the evaluations they consider only track outcomes up to three years after the conclusion of the program. In their comprehensive literature reviews, Barnow and Smith (2015) and Crepon and van den Berg (2016) come to similar conclusions and highlight the need for longer term evaluations. In addition, except for a few studies that look at criminal behavior (eg, Heller, 2014), very few evaluations of job training programs examine non-labor market outcomes – they typically focus on employment and wages.

Second, we demonstrate how to combine observational data on lifetime outcomes with RCTs to make progress on estimating unbiased causal lifetime effects. Starting with Lalonde (1986), a large methodological literature has investigated whether observational approaches (e.g. OLS, propensity scores, matching) can be used to obtain unbiased estimates of the effects of youth training programs on labor market outcomes by comparing the results from these approaches to those obtained from randomized trials (See Heckmann et al., 1999, for a review). We depart from this literature by using the experimental results from JC to learn about the bias in the short term estimated impacts of the program and use that to adjust our estimates of the long-term effects of the program derived from observational data.

Last, this paper contributes to the broader evaluation of the New Deal programs developed during the Great Depression. The Great Recession of 2008 and the recent pandemic-induced recession have renewed interest in understanding whether and for whom government programs deployed during large economic crises can be effective. Fishback (2017) provides a comprehensive survey of the literature on the short-run effects of New Deal programs, and reports that New Deal programs increased internal migration, lowered crime, and reduced mortality in the short run (see also Fishback, Haines and Kantor, 2007, and Vellore 2014.) To our knowledge, there are no empirical studies of the causal effects of the CCC program or of any other New Deal program on individual lifetime outcomes. Overall, the results are consistent with the hypothesis that the program provided important in-kind goods and services to disadvantaged populations in a time of need, improving their long-term health, productivity, and longevity.

II. Background: The CCC Program

Program Overview. The CCC was created to provide "relief of unemployment through the performance of useful public work and for other purposes." The CCC had two objectives: 1) to provide relief to unemployed youth; and 2) to preserve and enhance natural resources. The prevailing view at the time was that the provision of work ("relief through work,") would be more beneficial to the unemployed than the receipt of cash transfers ("direct relief"). Moreover, work would reduce the probability that youth would commit crimes and cause social disturbances (Brock 2005).

The untapped work capacity of idle youth would be used to create national parks, preserve forests, irrigate land, and address the damage of the Dust Bowl. Most camps had 200 enrollees at a time and were located close to work sites. The CCC program was popular, and many communities welcomed the camps and the resources they brought (Parham, 1981). Moreover, the enrollees did not directly compete in terms of labor with private sector activities. A nation-wide poll in 1936 showed that more than 80 percent supported the continuation of the program (Paige, 1985). As the program evolved, it added education components in 1934, which became mandatory in 1937, and in 1941 military training was added to the program. [2] The program ended in 1942 due to the onset of World War II.

Eligibility. The CCC program was only open to men who were unmarried, unemployed, primarily between the ages of 17 and 25, and U.S. citizens.[3] Preference was given to those in greater need and CCC enrollees were often selected from families already enrolled in relief programs.[4] Government reports at the time confirm that enrollees were also poorly educated,

[2] Although perhaps unintended, because the military was in charge of running the camps, another perceived benefit of the program was "enrollees made splendid soldier material" (McEntee 1942).
[3] There were some changes to these initial criteria, importantly age eligibility of juniors was modified twice. See Data Appendix Figure 1.
[4] In 1935, it became a requirement that enrollees be drawn from relief rolls, though in practice this was not always the case. In 1937, this requirement was eliminated.

with little work experience, as well as undernourished (McEntee 1942).[5] Enrollees had to present themselves in good physical condition upon their enlistment examination and have no prior history of criminal activity.[6] Finally, they had to be willing to send a substantial portion of their wages to an assigned family member and be willing to move to their designated camp location for the duration of the enrollment period. After the enrollee signed the contract, there was a two-week conditioning period, after which enrollees were sent to their assigned camps.

Compensation and program cost. Enrollees were required to work 40 hours per week and were paid $30 per month, of which generally $25 was sent home to a designated family member.[7] The government paid for the transportation to and from the camp, provided housing, uniforms, food, dental and medical care, and workers' compensation insurance, costing an additional $36 per month. The estimated total annual cost per enrollee was $1,004.[8]

Duration of enrollment. Individuals initially enrolled for a six-month period, and were allowed to re-enroll, for a maximum of two years (4 terms). Although the *average* enrollee in our sample worked for 9.8 months, there is large variation. CCC contracts could be terminated unilaterally by the government, based on governmental needs, at any point. Individuals also deserted, resigned or were expelled prior to completing their contract. Enrollees could leave early if they had secured employment, were enrolled in a formal schooling program or for "urgent and proper call" reasons, for instance the death of a parent or some other personal emergency. Enrollee turnover was costly, and efforts were made to keep it low.

The CCC in Colorado and New Mexico. CO and NM were relatively poor states in this period. National Income Accounts for 1930 suggest that per capita annual personal income was $571 in CO, and $329 in NM, while the nationwide average was $618.[9] Due to the large number of parks and forests in these states as well as the severe impact of the Dust Bowl, Colorado and New Mexico had disproportionate participation in the CCC program. In CO, a total of 57,944 men served, of whom 35,000 came from CO. In NM, a total of 54,500 served, of whom 32,300

---

[5] E.g. in 1939 and 1940, about 52% had 8 years of schooling or less (Annual Report 1940).
[6] Enrollees were vaccinated against typhoid, paratyphoid and smallpox at enlistment.
[7] Later in the program, a portion was retained as savings and given to enrollees upon dismissal.
[8] See BLS (1941). Levine (2010) reports this program was considerably more expensive than Works Progress Administration as it was estimated to cost approximately $800 per enrollee. Critics of the program pointed out that direct relief would have cost an estimated $250 per year instead (McEntee 1942). The value of the training and of the work achieved in terms of conservation is not considered in this estimate.
[9] Bureau of Economic Analysis NIPA 1929-today. SA1-3

came from NM (Cohen, 1980). Enrollees in Colorado and New Mexico were disproportionately Hispanic.[10]

III. Estimation Strategy and Identification Issues

We estimate the effect of the program on lifetime outcomes by comparing outcomes for those who served longer and shorter periods. This strategy is similar to what Flores et al. (2012) do to estimate the returns to the number of courses taken in JC and to Lechner et al. (2011), who evaluate impacts of short and long training programs in Germany. The intuition behind this approach is the following: if training increases skills through some standard production function, then more training should result in greater skills. We use the following specification,

$$Y_{ibj} = c + b * \left(duration\ of\ CCC\ service_{ibj}\right) + X_{ibj}B + e_{ibj} \qquad (1)$$

where Y_{ibj} is an outcome, such as employment or log of age at death for individual *i* born in year *b* training in CCC camp *j*, and X_{ibj} includes individual-level and camp-level covariates. The independent variable of interest is $duration\ of\ CCC\ service_{ibj}$, the duration of training in years. We estimate equation (1) clustering the standard errors at the application county and enrollment year-quarter level, though the results are not sensitive to this choice.[11]

The coefficient *b* identifies the causal effect of duration on a given outcome only if duration is uncorrelated with other determinants of the outcome, conditional on the observables. There are several threats to identification. First, duration is measured with error because dates are often incomplete or missing, possibly causing downward bias in the estimates. Second, there is a possible omitted variable bias on the individual-level. It may be that individuals with higher abilities trained longer because they benefitted more from the program, were less financially constrained and were able to better adapt to military life in camps (positive selection). Alternatively, poorer individuals may have had stronger incentives to train in the CCC because they were more in need of the CCC monthly payment (negative selection). Third, camp characteristics that may affect both duration and outcomes are omitted. For example, individuals

[10] New Mexico also had a large share of Native Americans. Native Americans had their own CCC programs which operated separately within Indian reservations and were administered by the Bureau of Indian Affairs. See Parman (1971) for details. We have no data from this program.
[11] We also experimented with alternative approaches and estimate results clustering at the application county, enrollment year level. Overall, we found these alternatives do not materially impact our conclusions. The evidence suggests that there is little correlation across individuals.

might have stayed longer in camps with good weather, and good weather could improve long-term health (positive selection). Alternatively, demand for work might have been greater in places hardest hit by the dust bowl, leading enrollees to stay longer in unhealthy locations (negative selection). In either case, the coefficient on duration would be biased.

To address these concerns, we take multiple approaches. First, we investigate the determinants of duration to determine the extent of possible selection issues. We also make use of the reasons why individuals dropped out to understand who leaves early and why. Then, we explore how the inclusion of individual- and camp-level covariates affect the estimates of the effect of duration, allowing us to estimate bias due to selection on observables. We estimate bounds using the method proposed by Oster (2017). For a subset of the data, we also conduct placebo tests to see if duration predicts pre-CCC enrollment outcomes (education, labor market experience, height and weight).

Finally, we use the data from the JC experiment to investigate and address potential selection in our estimates of the impact of the CCC on long run outcomes. We rely on and extend methods developed by Athey, Chetty, and Imbens (2020) that use experimental data on the impact of a given treatment on short-run outcomes to generate causal estimates of the impact of the treatment on long run outcomes found in observational data (but not the experimental data). Broadly, these methods use control functions generated using short-run estimates from the experimental JC data to adjust the estimates of the long-run impact of the program generated from the observational CCC data. We describe these methods in greater detail in section VIII.

IV. Data and descriptive statistics

A. Data collection

Colorado (CO) Enrollees. We digitized the entirety of CCC records contained at the State Archives of Colorado. These records include original applications of all individuals who applied. The collection, which includes 18,644 individuals, accounts for the population of individuals who trained between 1937 and 1942.[12] The applications contain: name, address, date of birth, place-of-birth, height, weight, race, and social security number (SSN), marital status, whether the father or mother is living, number of brothers, number of sisters, number of family members in

[12] We established based on published reports from the CCC that the records account for the complete population of records starting in 1937 (see Data Appendix Figure 4).

household, rural status, farm ownership, occupation of main wage earner in household, educational details, employment status and history, name of designated allottee and whether the individual was rejected. With the exception of information on height, weight and race, which were collected upon medical examination, the rest was self-reported. We observe the discharge information detailing the company and camp the individual attended, reason for dismissal, the date of dismissal, and whether the dismissal was honorable.

New Mexico (NM) Enrollees. New Mexico CCC records include information on 9,699 individuals, covering the population of individuals that trained in state from 1938 to 1942.[13] For each individual, the records contain the following: name, date of birth, address, family information (head of family, address of family, and relationship to enrollee), allottee information, enrollment date, assigned camp, date and reason for dismissal and whether the dismissal was honorable. NM records contain substantially less information on participants than CO records because only discharge forms are available.

Camp-level Data. We collected information on the exact location of camps, allowing us to link camps to historical weather patterns (temperature and precipitation), and the (Euclidian) distance of the camp to the closest towns and to each enrollee's hometown. Using the camp name, we can construct indicators for the agency (and thus the type of work) that created the camp. We also construct average characteristics of enrollees (such as the fraction under age 18) in each camp and point in time. Finally, we match camps to census county-level characteristics, such as unemployment rates.

Death Records. The administrative data from CO and NM was matched to death records (including the Social Security Death Master File and state-level death records) to identify the date of death and social security number of each enrollee. This match was done manually by trained genealogists at the BYU Record Linking Lab who found CCC enrollees in the collection of records kept by Ancestry.com and FamilySearch.org. A summary of this process is available in Online Data Appendix 1D. We find death dates for about 82% of recipients (Table 1), (88% of CO recipients and 75% of NM recipients) representing much higher match rates than typically

[13] We established based on published reports from the CCC that the records account for the complete population of records starting in 1938 (see Data Appendix Figure 4).

found in the literature.[14] We use these data to compute the age at death.[15] We match the data using automated methods as a robustness check.

Social Security Records. Using social security numbers, we match our data to the Master Beneficiary Record File (MBR), maintained by the Social Security Administration, which contains information on individual lifetime earnings, disability, and retirement.[16] (More details are available in Data Appendix 1E.) We are able to match 52% of our records to the MBR records. Only those that apply for benefits (social security pensions or disability) appear in the MBR. We have information on 80% of individuals who survived to age 65, indicating a high match rate for the targeted population. For all individuals we observe the age at retirement and whether or not they claimed Social Security Disability Insurance (SSDI) benefits. For individuals retiring in 1979 and later, we can observe the Average Indexed Monthly Earnings (AIME) which is computed as the average of the highest 35 years of earnings after adjusting for inflation.

1940 and WWII records. We match our records to the Federal Census of 1940 and to WWII Enlistment Records. These matches are made using the Abramitzky, Mill, and Perez (2018) algorithm. Details of the procedure are available in Data Appendix 6D and 6E. The 1940 census includes location, demographics (race and ethnicity, marital status, place of birth, household information), and labor market information (employment occupation and wages). We successfully match 44% of individuals to the census, and 29% to WWII enlistment records. This

[14] Our match rates are higher than those typically found in the literature (which range from 20 to 50%) for two reasons (Bailey et al. 2017, Abramitzky et al 2019). First, administrative records contain information not just on individuals but also on their family members. This greatly improves our ability to find individuals by using information from family trees and various vital registration records. Second, the death records come from various sources. Most commonly these come from the Death Master File (DMF) which includes the universe of death certificates in the US starting in the mid 1970s. But the collection also includes records from other sources, including state vital registration sources, deaths during WWII, and gravestones. A few individuals are observed as dying during CCC training.

[15] Mortality information is missing for some individuals for several reasons. First, some individuals died prior to 1975, which is the first year of complete death records in the Social Security Death Master File (For more information about coverage of the DMF, refer to Hill and Rosenwaike (2001). In this case, we might find a death record for them if one exists in state vital records. Second, some individuals might still be alive, so the age at death is censored. Based on SSA life tables we compute that about 1.1% of individuals born in 1920 (our median birth year) would be expected to be alive by 2017. Lastly, we might not have found individuals who died in the 1975-2017 interval due to measurement error and matching errors.

[16] We only observe SSN if they person reported it in the application in CO, or if it is available in the death certificate. However, SSNs are not available for anyone who died after 2008 (these are masked for privacy reasons) or for those who died young and never applied for a SS card.

lower match rate to WWII records is expected: not all individuals enlisted or served in WWII, even when they were eligible and not all records of those who served survived.[17]

B. Sample Selection

For our analysis, we restrict attention to individuals for whom we can observe duration of training, camp, and the outcome of interest. This results in a sample of 23,722 men out of 26,290 (Appendix Table 1 row "Final analytic sample" out of row "All").

For the mortality analysis, we include only the 17,639 men with age of death information. This estimation sample generally is representative of the initial data (Table 1). For the lifetime outcomes from the SSA, our sample includes 12,455 individuals, 52.5% of the original analytic sample. Again, this sample is fairly representative of the initial full sample in many dimensions (duration, YOB, age, height, weight, education, father alive, mother alive, household size, farm) with some notable exceptions (Table 1). By construction, the age at death in this sample is higher because only those who survive to at least 62 are eligible to apply for pensions, unless claiming for disability. We also see fewer Hispanics, more people who lied about their age, and more people who sent money to their mothers. But these differences are small. Moreover, later we investigate the extent of sample selection and the effects of missing data and use imputations in alternative specifications.[18]

C. Summary Statistics: CCC Training and Lifetime Outcomes

Pre-CCC Characteristics. More detailed data for CO suggest that the enrollees were relatively disadvantaged (Table 1). On average, enrollees were 18.7 years old, had completed 8.7 years of schooling and came from a household with 5 individuals. One in four came from a farm, 20% had a deceased father and 15% had a deceased mother. Despite height and weight examinations to exclude the unhealthy, 7% were underweight. Imputing the ethnic origin of the participants, we estimate that about 43% were Hispanic (see online Data Appendix 7). CCC enrollees came from poorer counties than the average males of the same age in CO and NM in the 1930 and 1940 census, consistent with their being recruited from relief rolls. Consistent with

[17] Several cards were lost to fire or were unreadable. See https://aad.archives.gov/aad/series-description.jsp?s=3360&cat=all&bc=sl
[18] We provide the full set of summary statistics in Appendix Tables 2a and 2b

the fact that CO and NM were very poor states, CO and NM enrollees were even more disadvantaged than the average CCC enrollee in the nation—they are substantially younger, shorter, weigh less, have more dependents, and more of them have fewer than 4 years of schooling (Data Appendix Figure 6).[19] Data on the camps suggest that they were typically rural in nature, located relatively far from the enrollees' hometowns (150 miles on average).

Post CCC outcomes. Consistent with CCC enrollees being more disadvantaged at entry, they also have worse long-term outcomes than average for their cohorts: they died younger and earned less. The average enrollee eventually lived to age 70 (one year less than male cohorts born in 1920 who survived to age 17).[20] They also earned $405 in annual ages in 1940, compared $593 for men aged 18-32 in 1940 Census.

V. Determinants of Training Duration

Average training duration was 9.8 months, but there is large variation in the distribution.[21] Duration in months spikes exactly at 6, 12, 18 and 24 months, corresponding to 1, 2, 3 and 4 terms (Figure 1, Panel A.) However, most individuals (62%) dropped out in the middle of their assignment. Among those who left before completing their term, 21% deserted, around 14.5% were dismissed "for the convenience of the government" (e.g., the camp closed), 12% left for a job, and another 12% left because of an "urgent and proper call," for example family sickness or death (Figure 2 Panel A).

To investigate the determinants of duration we estimate simple OLS regressions of the duration of training as a function of individual, family, and camp characteristics. We include year-of-birth fixed effects (YOB) because different cohorts were eligible to train for different amounts of time (Data Appendix Figure 5). We include county-of-enlistment by quarter-of-enlistment (CQE) fixed effects for two reasons. This addresses the fact that the number and types of camps that were opened varied over time and space, affecting where individuals ended up serving and potentially the duration of training. It also addresses differential selection based on

---

[19] We compare the means in our estimation sample to the published national means. These were published in Appendix H of *Annual Report of the Director of the Civilian Conservation Corps: Fiscal Year Ended June 30, 1937* "Appendix H: Census of Civilian Conservation Corps Enrollees."

[20] This information comes from SSA cohort life tables.

[21] Aggregate data on the national CCC program from a 1937 CCC Census shows that the distribution of duration in our states (using CO) is skewed slightly towards shorter durations than the national distribution (Data Appendix Figure 6).

location and time over the program years: the type of individuals who apply for training (and other government benefits) varies substantially with local economic conditions (Méndez and Sepúlveda. 2012).

No clear relationship between personal characteristics and duration emerges in the data (Figure 3 and Appendix Table 3). Individuals who were older trained for longer durations. Those who lied about their age, trained for shorter. Those who were farther away from home also trained for shorter. Surprisingly, individuals with a higher weight, who were presumably healthier individuals, trained for *shorter* durations. Height, which is a marker of improved nutrition and health during the growing years, does not predict training duration. Those with more education trained for longer but so did those who came from larger households or whose parents were deceased.[22]

This evidence is not consistent with a single narrative of selection. There appear to be three groups of enrollees. First, those who served for longer because they were positively selected, such as those with more education or older. A second group seems to be negatively selected, coming from farms, and larger, Hispanic households. Third, some appear to have more or less random reasons to drop out due to luck, such as a job appearing, a camp closing or having an emergency at home.

The evidence also suggests that conditional on individual characteristics and place and time of enrollment, camp conditions mattered (Appendix Table 3). For instance, in places with less rain and milder weather, individuals trained for longer, as did those assigned to camps farther from cities. Peer characteristics also mattered. Durations were longer in camps with larger Hispanic shares of the population or with more men under 18, but shorter in camps with many men who misrepresented their age or sent smaller amounts to their families.

In sum, the primary evidence shows that desirable traits in an enrollee or in a camp did not necessarily lead to longer durations, and there is no single narrative of selection.

VI. The Long-Term Effect of CCC Training on Mortality, Lifetime Earnings and Disability

[22] These results are qualitatively similar if we estimate regressions separately for CO and NM (see Appendix Table 3) but some coefficients are only significant in one state. There are no cases in which the coefficients are statistically significant and of opposite signs.

We now investigate the effect of enrollment duration on lifetime outcomes: mortality, earnings, age at retirement and disability claiming.

A. Impact of CCC Duration on Mortality

For this analysis, we restrict attention to individuals who we linked to a death certificate and who died after age 45 (to avoid WWII related deaths). The results are not sensitive to these restrictions. The longer an enrollee trained, the longer he lived (Figure 4). The relationship is positive and linear.

Next, we estimate an accelerated failure time model of the age at death on duration in which we add controls for the characteristics of the enrollees and the camps to examine whether and how our estimates change in response. The first column of Table 2 Panel A with no controls shows a very precise coefficient on duration of 0.013. Controlling for cohort fixed-effects and county-of-enrollment-quarter-of-the-year (CQE) fixed-effects (column 2) does not change the coefficient estimate. Including family and individual characteristics in column 3 (ever rejected from the CCC, disabled, non-junior member, age, dollars per month allotted, gap in service, distance from the camp to home, whether Hispanic, and for those in CO only, highest grade completed, household size, life on a farm, height and weight at enlistment, whether mother or father deceased, tenure in the county prior to CCC enrollment and reason for discharge) lowers the coefficient to 0.011.[23] Adding camp characteristics in column 4 (mean precipitation in the camp, min and max temperature, type of camp, distance to closest city), peer characteristics in column 5 (average age, share Hispanic, average allottee amount and gaps in service), or camp fixed effects in column 6 changes the coefficient very little. The magnitudes imply that one more year of training increased the age at death by one year (roughly 1.3 percent of 73.6 years of life). Given that the average duration was 9.84 months, the program increased age at death by 0.8 years for the average enrollee. When we limit our sample to CO where the records contain more important baseline information, such as education, height, etc., the results are again similar (column 7).[24]

[23] Full regression with coefficients for controls in Appendix Table 4.

[24] For NM and for CO records with missing data we impute using the mean and include a series for dummies to indicate when the covariate is missing.

The fact that the coefficient is essentially unchanged from columns 1-7 as we add more detailed controls suggests that selection bias may be small. However, to more formally assess the magnitude of the omitted variable bias, we re-estimate these coefficients under various assumptions about the unobservables following Oster (2017). If delta (the proportionality value) is assumed to be 1 (i.e., unobservables as important as observables) then our coefficient would be 0.0136. Alternatively, if delta is assumed to be -1, we would estimate 0.0127. Thus, one more year of training would increase the age at death between 0.96 and 1.02 years.[25]

Finally, to examine possible non-linearities, Figure 5 shows the results of the regression of probability of survival to age *x* on duration for every age between 45 and 90. The coefficients are small and statistically insignificant at younger ages, when the survival is very high. They become positive and statistically significant starting at age 56, continue to increase and peak between ages 68 and 78, and then decline thereafter. As a function of the baseline survival rate, which is declining throughout, the effects rise until age 67, and then decline.

Sample attrition. About 20% of the original sample is missing information on age at death. We assess whether missing age at death is systematically related to training duration (with or without conditioning on covariates). Table 3 Panel A shows that, without controls, the missing rates are not a function of training duration. But conditional on camp, family and individual characteristics, age at death is about 9% (1.7/18) *less* likely to be missing for those who trained for an additional year. This suggests that differential attrition could bias our OLS estimates. To address this issue, we estimate survival models where we make various assumptions about the missing data. The results in Appendix Table 5 show that our findings are robust to various imputation approaches.

Quality of the longevity data. Our main results use the information found by trained genealogists from multiple sources to determine the age at death. To assess the quality of the data and whether the hand matching procedure introduces unknown biases, we replicate the results using machine matches only. To do this we use the EM algorithm to match our records to the Death Master File. The results in Appendix Table 6 show that we still obtain a positive and

[25] We also examined a specification including an indicator variable "completed term" which equals 1 when participants complete increments of 6-month terms. We did not see large differences in the duration coefficient. However, when splitting the sample to those who completed 0 terms and those who completed at least 1 term, we see no effect of duration for those with 0 terms and a similar effect for those who completed at least 1 term.

statistically significant coefficient of duration on age at death, similar in magnitude to our main estimates.

B. Impact of CCC duration on lifetime income

We estimate the impact of program duration on lifetime income as proxied by the AIME (Average Indexed Monthly Earnings), which is the average of an individuals' best 35 years of real earnings as reported to the SSA. This amount is used to calculate pensions. The AIME is only available for those claiming after 1979 but a specification check suggests that the results would be similar if we were to extend to those claiming prior to 1979.[26]

The mean AIME as a function of duration for the sample claiming after 1979 shows a flat or slightly negative relationship between duration and AIME. This relationship reverses once we control for year of birth (Figure 6). This is because more recently born cohorts had shorter durations (due to the end of the program) but larger incomes, generating a spurious negative correlation. In OLS regressions without covariates there is indeed no relationship between duration and the AIME (Table 2 Panel B, column 1). Consistent with the figure the coefficient becomes large, positive and statistically significant when we add controls for birth cohort and for quarter and county of enlistment (column 2). The estimates imply an increase of $67 additional monthly earnings for those who participated for one year, representing an increase of 7% relative to average earnings. As we add more controls across the columns, the estimated coefficient declines to $50, or 5.2% of average monthly earnings (Table 2 Panel B, column 6).

These results do not appear to be driven by sample selection or attrition in the SSA data. There is no effect of duration on whether we match an enrollee to MBR. Nor does the effect of duration on longevity change when we limit to the sample matched on the MBR (Table 3 Panels B and C).

We can compare our estimated returns to a year in the CCC to the returns from a year of schooling. OLS estimates of the returns to schooling from other sources range from 5% (Goldin and Katz 2000) to 8% (Clay et al 2012). Thus, the returns to one year of CCC training (5.2%) are on the lower end of the returns of a year of schooling.

[26] We do have another, slightly noisier, proxy of earnings (the PIA which is based on the AIME) that is available prior to 1979 as well. We find similar results for the PIA pre- and post-1979. See Appendix Table 7 for results.

C. Impact of CCC Duration on Age at Benefit Claiming and SSDI Claiming Rate

We are also able to estimate the impact of CCC duration on age at which individuals first claim benefits from the Social Security Administration (either disability or pensions) and on whether individuals become disabled, measured by Social Security Disability Insurance (SSDI) claiming. We find that one year of CCC enrollment increases the age at claiming benefits by half a year, relative to mean age at claiming of 60 years (Table 2 Panel C), suggesting CCC men were in better health, retired later and lived longer. This is consistent with existing work showing that early retirement is associated with death at younger ages (Waldron, 2001, Fitzpatrick and Moore 2018).

We can examine this health channel directly by looking at how duration affects SSDI claiming, a measure of health. Twenty one percent of the sample claims SSDI benefits. When the full set of controls is included, we find that one year in the CCC reduces claiming by 2.2 percentage points, or 10 percent (Table 2 Panel D). Overall, we find that CCC participation improves health in the long run as measured by delayed benefit claiming, reduced SSDI claiming and greater longevity.

D. Treatment Effect Heterogeneity

Recent reviews of training programs (Card, Kluve, and Weber, 2018, Barnow and Smith 2015, Crépon and van den Berg 2016) suggest substantial heterogeneity in the effects of training programs. We explore heterogeneity in Appendix Table 8. We find that the poorest and most disadvantaged benefitted more. The effects were also larger for those that came from counties with higher unemployment rates. These findings are consistent with Card et al. (2018)'s finding of larger effects of job training in recessions and among the more disadvantaged. Our results differ in two dimensions: we find larger gains for the young, and significant benefits for Hispanics.[27]

VII. Short-Term Outcomes: Evidence from the 1940 Census & WWII Records

[27] We suspect these differences are due to several factors: 1-we compute Hispanic ancestry and do not rely on self-reports, b-our enrollees are from only 2 states with large number of Hispanics in the population; 3-the country of origin among our enrollees differs substantially from today.

We estimate the short-run effect of CCC enrollment in an effort to both compare our estimates with existing work on the short run impact of more recent job training programs and to understand the mechanisms behind our long run impacts of the CCC. We first investigate the effects of CCC duration on employment and wages, the standard outcomes that are typically assessed in job training programs. Next, we investigate other mechanisms such as formal education, health improvements, and geographic mobility, all of which have been associated with improved longevity and labor market outcomes in previous work.

A. Labor market outcomes: Evidence from the 1940 census

For this analysis, we constrain our sample to 9,518 men who participated in CCC before January 1st, 1940, of whom we find 43% in the 1940 census. Duration is unrelated to whether we locate an enrollee's census record. [28]

CCC training duration appears to have little effect on the short-run labor market outcomes of CCC participants (Table 4).[29] Most men (91%) are in the labor force, and longer CCC training had at best a very small effect on this outcome: a 2.1% increase relative to the mean. We observe no effect on employment (conditional on labor force participation) during the week prior to the Census. There is a small, negative and imprecise effect of duration on earnings.[30] Overall, our results are consistent with the conclusions reached in recent reviews that the labor market effects of job training are more positive in the long run than in the short run.

B. Health and Military Service: Evidence from WWII Enlistment Records

We estimate the short run impact of the CCC on health as measured by height and weight using WWII enlistment records. Unlike the 1940 census records, duration does predict their presence in the WWII records: an additional year of training leads to a robust and significant 3.8 percentage point increase in the probability we find the individual in the WWII enlistment records, a 12% increase relative to the mean (Table 4 Column 6). This result is not surprising:

---

[28] Duration does not predict whether we find an enrollee in the 1940 census once we include birth cohort and county-quarter fixed effects (Table 4 top panel).

[29] In Table 4, we only present specification with camp fixed effects corresponding to Column (6) in Tables 2 and 3. Appendix Table 9 presents results on all specifications for Census outcomes and Appendix Table 10 for WWII outcomes.

[30] For example, the largest coefficient for weeks worked is -0.937 which corresponds to 3.4% change relative to the mean of 28 weeks worked. Similarly, we observe a negative but statistically insignificant effect on earnings, corresponding to about a 3% decrease in wages at the mean.

the army organized and administered life in the camps, and CCC men who trained for a long time were well acquainted with military life. Two percent of men in our sample ended their CCC engagement to enlist in the military directly. Given that we have not found differential matching rates in any of our other data, we do not believe differential matching explains this result. Rather, we conclude that the program increased the likelihood of serving in the military. This could reflect greater familiarity with the military after serving in the CCC or it could reflect the acquisition of additional non-cognitive skills that increased the likelihood of success in the military.

For the 5,500 observations we match to WWII records which contain height, weight and schooling, we find that one more year of training translated into roughly 1 more inch of height. This effect is large by historical standards: for example, it took British men 100 years for their average height to increase by 6 inches (Fogel 1994). This result holds conditional on height at CCC enlistment, indicating additional growth after CCC enrollment rather than initial differences in height.

There are multiple reasons why the program increased heights among CCC men despite the fact their average age was 19. First, these results are consistent with existing work showing that undernourished populations grow more slowly and achieve their final adult height at older ages (Steckel 1986). Individuals in the CCC were poor and they received food and medical care, including vaccinations, as part of their participation in the program, likely improving their nutritional status. Second, national reports of the CCC program show that the average height gain in the CCC was half an inch (McEntee 1942). Our estimates are likely larger because our population is more disadvantaged than the average CCC enrollee. Finally, 9% of our sample of CCC enrollees were likely younger than they reported.

Consistent with this, we also observe a 5-6 percent increase in BMI, a common indicator of short-term nutrition. The final CCC report documents an average weight gain of enrollees during the program of 11 pounds (McEntee 1942), and our results suggests that 40-60% of these gains persisted.[31] These results are also consistent with our finding that CCC service lowered SSDI claiming and increased longevity. We conclude that the CCC improved overall health among participants.

[31] For an average enrollee in our sample, adding 11 pounds would translate to a gain of 8%.

C. Effects on Education, and Geographic Mobility

We also show results for formal years of schooling and geographic mobility, which are observed in both the Census of 1940 and WWII Enlistment Records. For these outcomes, we combine information from the two sources to maximize sample size.[32] We control for the time since discharge (or equivalently the year of observation) to account for the fact the outcomes are not observed at the same time for all enrollees.

We find a positive and statistically significant effect of duration on years of schooling of about 0.17 years, relative to a mean of 9.2 years of schooling, controlling for education at baseline (Table 4 Column 9).[33] This represents one tenth of the standard deviation in schooling in the WWII records, and is larger than the effect of many education policies, such as child labor laws, on educational attainment during the early 20th century.[34] This effect likely represents a combination of additional schooling completed as part of the CCC and schooling obtained after the CCC. CCC reports indicate that 8% of men obtained additional schooling during the program.[35] Assuming 8% obtained one more year of school, this would result in a gain in years of schooling of 0.08. Given that 3.5% of enrollees in our data cited education as the reason for leaving the program, post-CCC education gains likely accounts for the rest.

Finally, we examine the relationship between duration and short- and long-term geographic mobility by comparing the county of individuals in their original CCC application with the county of residence indicated in the 1940 Census records, the WWII records and in the death certificates. Thirty five percent of participants moved in the short-term (by 1940 Census or WWII enrollment). Training in the CCC substantially increased the likelihood of moving to another county: one more year of training increased the probability of moving by 5.7 percentage

---

[32] Because the WWII records contain the latest information, we take information from WWII if the enrollee can be found in WWII record and 1940 Census if cannot be found in WWII record and discharged before 1940. For education and marriage, we take the value at WWII, which is later than 1940, if observed in WWII and the value at 1940 Census if only observed in the Census. For moving, we code someone as moved if they moved counties in either 1940 or in WWII.

[33] When we restrict our analysis to those with non-missing baseline education, the estimate declines to 0.12 and remains significant at the 1% level (Appendix Table 10 Column 7).

[34] For example, see Lleras-Muney (2002) or Goldin and Katz (2008). One more year of compulsory schooling led to about 0.05 years of schooling.

[35] The final report states that over one hundred thousand enrollees (3%) were taught how to read and write in the CCC program, 4% of men received primary school degrees (8th grade), 0.6% got their high school diplomas and a handful (270 out of more than 3 million) obtained college degrees. Thus, about 7-8% obtained some schooling.

points, or 17% relative to the mean (Table 5 Column 2).[36] This is substantial particularly during this period, which was characterized by historically low migration nationwide.[37] Moreover, when CCC men moved, they moved to locations with higher paying weekly or annual wages (as of 1940) and lower mortality, measured by the average county level mortality from 1950 to 1968 (Table 5 Columns 3-4).[38] Over the long run, however, most individuals moved and the effect of duration on mobility fades (Table 5 Columns 5-7).

In sum, enrollees who served longer had better health, more schooling and were more likely to move to healthier, richer places, but training longer had no detectable effects in the short-run on labor market outcomes.

Before moving on, we conduct one last check of the validity of our OLS approach. For CO enrollees, we have baseline measures for several outcomes: height, weight, education and prior labor market experience. In our main results we control for these. However, this allows us to test if duration predicts these pre-intervention outcomes. Appendix Table 12 shows that duration does not predict these pre-CCC outcomes, except for education. These results suggest that by in large our approach produces unbiased estimates of the effects of the program, but some bias may remain. In the next section we estimate the long run and lifetime effects of the CCC, incorporating evidence from the JC RCT to address any remaining selection.

VIII. Incorporating Evidence from the Job Corp RCT to Address Selection Bias in the CCC

In estimating the long-run and lifetime effects of the CCC, we build on the seminal work of Lalonde (1986) who used experimental evidence on the impact of job training to assess results based on observational data. In particular, we apply and extend methods developed by Athey, Chetty and Imbens (2020) that enable researchers to exploit data from randomized controlled trials to address potential bias in results based on observational data in a control function framework. In particular, we exploit the availability of data for a 1994 *randomized evaluation* of

[36] As in Table 4, we only present results on specification with camp fixed effects corresponding to Column (6) in Tables 2 and 3. Appendix Table 11 presents full table of all specifications.
[37] In the 1940 census 12% of people report living in a different county than in 1935. https://www.census.gov/dataviz/visualizations/010/
[38] We use this measure instead of the county mortality from 1940 onwards because of the disruptions that occurred during the WWII.

the modern equivalent of the CCC: the Federal Job Corp Program (JC).[39] We address potential endogeneity of our CCC estimates using two approaches that make use of the JC RCT.

Both approaches require the JC to be an externally valid experiment for the CCC, either for the treatment effect directly or for the degree of selection bias. The two programs have considerable similarities in both the observable characteristics of program participants and the short-run impacts, suggesting that the JC may be a comparable experiment. After presenting comparisons of the CCC and JC in the next section, we proceed to describe our two approaches to correct for bias in the CCC estimates in greater detail, and present our bias-corrected estimates of the impacts of the CCC on long term outcomes. Finally, we quantify how relaxing the assumptions changes our estimates.

A. Comparing CCC and JC Enrollees

Comparing participant characteristics. If we restrict attention to men in JC, second column of Table 6 shows that overall, JC and CCC participants are similar.[40] Both are young (19 years old on average) and have relatively few years of schooling. JC participants have completed 10 years of schooling, compared with 8.5 for the CCC enrollees, and 19% have graduated from high school compared with 12% of the CCC enrollees. The CCC sample has considerably more Hispanics, due to the fact we concentrate on CO and NM, whereas the JC data is national.

Participants are also similar in terms of duration of enrollment and reasons for leaving. Mean duration is 9.8 months (s.d. 7.47) for CCC and 5.8 months (s.d. 6.6) for JC. The main reason for the lower duration of the JC participants is that 20% never trained (Figure 2). Conditional on training, the duration among the treated group in JC is 7.8 months. Reasons for leaving are also similar across the two programs.[41] Finally, and perhaps most importantly, when we try to predict

[39] The current website (https://www.doleta.gov/job_corps/) states that "The program helps eligible young people ages 16 through 24 complete their high school education, trains them for meaningful careers, and assists them with obtaining employment." "Students can earn a high school diploma or the equivalent, and college credits. Job Corps also offers tuition-free housing, meals, basic health care, a living allowance, and career transition assistance."

[40] JC participants differ from CCC participants in two key respects: JC includes women and married individuals, whereas the CCC excluded both.

[41] About 30% of JC enrollees complete the program, compared with 38% of the CCC. And of those who leave before completing, 30% in the JC and 22% in the CCC "deserted" while 12% and 4%, respectively, left because of employment opportunities.

duration in the JC, we also find evidence of both positive and negative selection into duration, as well as evidence that duration is random for some, just as we found in the CCC data (Figure 3).[42]

Comparing short term effects. We reproduce the short-run JC randomized evaluation results in Schochet et al. (2008) using only the sample of males (Table 7).[43] In the first column we present estimates that compare the outcomes of those assigned to treatment to those of the control. In the second column we present the implied effects of training duration by estimating the 2SLS effect of duration using the randomized treatment status as an instrument. These estimates represent the causal effect of duration under a certain set of assumptions.[44] The third and fourth columns show OLS estimates for duration for JC and for CCC, respectively. Overall, short term labor market outcomes in JC were more positive than in CCC, but in both the CCC and the JC data, education and mobility increased. The table also shows that the JC increased self-reported health, consistent with our findings that CCC improved health, measured by height and weight.[45]

Comparing longer-term effects. The estimated long-run effects on income and disability are also similar to estimates obtained from the RCT of JC. Given that enrollees participated for 0.83 years, the effect of CCC on lifetime earnings are about 4.6%. The latest evaluation of JC, which tracks individual tax records 20 years after the program (Schochet 2018), finds that participation in JC had a statistically insignificant increase in wages of 2%, with our effects (4.6) well within their 95% confidence interval [-4%; 8%]. It also reports a 40% reduction in SSDI benefits among older JC participants, though not in the overall sample.[46]

In sum, participants in JC are similar in many dimensions and they experienced qualitatively similar long-run improvements in income and health. They also experienced similar improvements in their education, health and mobility in the short run, but they differed in the

---

[42] We find that education, Hispanic ethnicity, non-native speakers trained longer and individuals with a criminal history or those with shorter work histories trained for shorter periods of time (Figure 3 Panel B). As in the CCC, participants that found employment and those that deserted, were rejected or had urgent and proper calls also served shorter durations compared to those that completed their term.

[43] The results in the first column are almost identical to those in Schochet et al. (2008) except that we are restricting the sample to males, and we constructed a few new outcomes (years of education, mobility and marriage).

[44] The key assumption is that there is no direct effect of the assignment to treatment on the outcome beyond its effect on the duration of the training. Appendix Table 13 presents balance tests of baseline characteristics of the JC applicants.

[45] Results are similar if we use the entire scale or only look at whether the respondents are in excellent health alone.

[46] The JC evaluation only uses 15 years of labor market outcomes, whereas we use 35 years. The shorter length of the evaluation may lower the estimated returns.

short-term labor market outcomes. We now formally describe our method of using short-run, experimental estimates from JC to adjust long-run, observational estimates from the CCC.

B. Combining Experimental and Observational Samples: Set-up

We pursue two approaches that exploit experimental data from the JC to address potential selection in the estimates of long-run impact from the CCC. The first approach assumes comparability of short-run treatment effect of the JC and the CCC. The second assumes comparability of selection into longer duration in the JC and the CCC.

The setting for both approaches is as follows. We assume the short term (denoted ST) outcome is a linear function of the treatment and observed and unobserved covariates:

$$y_{iS}^{ST} = \tau_S^{ST} W_{iS} + X_{iS}\gamma_S^{ST} + \alpha_{iS}^{ST}$$

y_{iS}^{ST} is the short-term outcome for sample $S \in \{CCC, JC\}$; W_{iS} is duration of training in program (either CCC or JC); τ_S^{ST} is the short-term treatment effect; X_{iS} includes other controls; and α_{iS}^{ST} is the unobserved component (residual), which is possibly correlated with W_{is} in the CCC sample and is the source of endogeneity of W_{iCCC}. In JC only, we observe a binary treatment status dummy T_i, uncorrelated with α_{iS}^{ST} given the experimental nature of the data, but correlated with training duration W_{iJC}, which allows us to correct for the endogeneity using the random assignment as an instrument.

Similarly for the long-term (LT) outcomes, we have:

$$y_{iS}^{LT} = \tau_S^{LT} W_{iS} + X_{iS}\gamma_S^{LT} + \alpha_{iS}^{LT}$$

Going forward we make the following two assumptions:

Assumption 1: $\alpha_{iJC}^{ST} \perp W_{iJC} | X_i, T_i$ in JC (duration is random given treatment status and X)

Assumption 2: LT and ST residuals are linearly related as:

$$\alpha_{iCCC}^{LT} = \delta\alpha_{iCCC}^{ST} + \varepsilon_{iCCC}^{LT} \textit{ and } \varepsilon_{iCCC}^{LT} \perp W_{iCCC} | X_i, \alpha_{iCCC}^{ST}$$

This is a key assumption: the sources of the endogeneity for the short-term and long-term outcomes share a common component. Therefore, if we are able to control for α_{iCCC}^{ST} in our regression of CCC duration on long-term outcomes, we generate an unbiased estimate of the long-run treatment effect.

C. First Approach: Assuming Identical Short-run Treatment Effects for JC and CCC

For our first approach, we follow Athey, Chetty, and Imbens (2020) in assuming that the experimental sample has external validity and therefore $\tau^{ST}_{CCC} = \tau^{ST}_{JC}$. This implies that the estimate of the ST treatment effect obtained from the RCT of JC is an unbiased estimate of the ST effect in the CCC. We extend their procedure to allow for endogeneity of duration in the JC sample that we address using instrumental variables.[47] Appendix A details the procedure.

After obtaining unbiased estimate of the JC's short-term effects ($\hat{\tau}^{ST}_{JC}$), which is also an unbiased estimate of CCC's short-term effects, we can calculate the short-term CCC residuals, $\hat{\alpha}^{ST}_{iCCC} = Y^{ST}_{iCCC} - \hat{\gamma}X_{iCCC} - \hat{\tau}^{ST}_{JC}W_{iCCC}$. We then include the residuals as control functions in the long-term regressions. If 1) there is equality of the short-term causal effects in JC and CCC, and 2) the residual unobserved component in the long-term regression after controlling for the short-term residual in CCC is uncorrelated with duration, then the coefficient on duration estimated including the control functions gives us the LT causal effect of interest. For robustness, we also try varying short-run residuals we include that we can calculate from multiple short-run outcomes.

The results from this exercise are in Table 8 for our long-run outcomes of longevity, AIME, retirement age, and SSDI claiming with control function generated using education as the short-run outcome. For all samples, we compare the OLS estimate without including any control function (Panel A) on a consistent sample with the estimate including control functions generated from short-run regressions in this approach (Panel B).

The LT estimates of the effect of duration on longevity are unaffected by the inclusion of control functions for all samples and specifications, suggesting the bias in the OLS estimates is small (Table 8 Panels A and B). This can occur because a) the LT unobservables are uncorrelated with duration, indicating little endogeneity; b) the LT unobservables are uncorrelated with LT outcomes, or c) the ST and LT unobservables are different and the controls do not capture the endogeneity. Note that this is not a result of the small treatment effects estimated in the JC trial – the bias correction is always based on the difference between JC RCT

[47] We extend the methods in that paper to account for the differences in settings. In the original Athey et al. (2020) paper the treatment is binary, whereas in our case it is continuous. Secondly the experiment does not directly yield estimates of the effect of the continuous duration variable, so we employ 2SLS methods instead to obtain causal estimates. As a result, we need to make additional assumptions. See Appendix A for details.

and the CCC, not on the treatment effect in JC alone.[48] The results are similar if we use other ST outcomes to compute the control functions or include different set of controls.[49]

While encouraging, these results rely on an assumption that ST treatment effects in JC would be identical in the CCC, after adjusting for some covariates. This assumption seems reasonable for education or geographic mobility but perhaps unreasonable for labor market outcomes. The returns to training, like the returns to other human capital investments, are also likely to depend on post-investment market conditions. There is evidence that the returns to schooling are stochastic and vary considerably over time (Goldin and Katz, 2008 and Rosenzweig and Udry, 2020). The post-WWII economy was better for low-skilled labor than the economy of the early 2000s, which had stagnant wages for low-income groups (Piketty, Saez and Zucman 2018) and low-skill laborers (Autor, Katz and Kearney 2008). This suggests that the assumption of constant treatment effects over time and place might be too restrictive.

D. Second Approach: Assuming Equivalent Selection into JC and CCC

For our second control function approach, we allow for the possibility that the short-term treatment effects are *not* the same in JC and CCC, but instead we assume that the selection bias arising from endogeneity of the duration of the training is the same in JC and CCC. For this approach, we estimate the ST treatment effect in the JC using both OLS and IV techniques, using random assignment to treatment as the instrument for duration. The difference in these two estimates, $\hat{\mu} = \hat{\tau}^{ST}_{JC,OLS} - \hat{\tau}^{ST}_{JC,2SLS}$, represents the omitted variable or selection bias in observational estimates of job training on short-run outcomes. For example, the 2SLS estimate of the effect of duration on education in JC is 0.393, whereas the OLS estimate of duration in the treated arm of JC is 0.360 (Table 7). Any difference between the two estimates is indicative of selection bias. Under the assumption that the selection bias is similar in CCC, the difference between the OLS and the 2SLS estimates in JC can be used to construct an estimate of the

[48] The bias correction is based additionally on the relationship of two different sets of outcomes (ST and LT) in CCC, not on the treatment effect in JC alone. As long as ST and LT selection are related within CCC, the process should correct any bias of our estimate. If the ST TE is 0 in the RCT then the control function is the ST outcome, which potentially still generates a change in the estimated coefficients.

[49] In Appendix Table 14 we show results with control functions calculated using different short-run outcomes, including mobility, short-run labor market outcomes, and including all control functions simultaneously. The result that including control functions do not change the OLS result *keeping consistent the sample of enrollees* does not change. The sample size varies significantly across the inclusion of residuals due to observability of short-run outcomes varying.

unbiased short-run treatment effects, which in turn is used to construct a control function for the LT regressions.

Using the JC estimates of selection bias ($\hat{\mu}$), we adjust our OLS estimate of CCC short-term outcomes, $\hat{\tau}^{ST}_{CCC} = \hat{\tau}^{ST}_{CCC,OLS} - \hat{\mu}$. We then use this adjusted estimate of the treatment effect to calculate the control functions as before. That is, we use this adjusted estimate of the short run treatment effect of the CCC to calculate a residual in regressions of short run outcomes. We use this residual to construct the control function. See Appendix A for the detailed procedure.

The results from this exercise are presented in Table 8 Panel C, using the control function calculated from our short-run education outcome.[50] Again, we find that regardless of the control functions that we include the estimates are stable and not different from the original OLS estimates, suggesting little bias in our LT estimates of the effects of training.

To assess whether the assumption of equivalent selection on unobservables is warranted, we compare selection into duration on observables in the CCC and the JC. Examining individual observable characteristics common across both settings, we conclude that there are important similarities across the two. Specifically, we find that: 1) Individuals who have never worked are more likely to serve for longer in the JC and the CCC, 2) Hispanic enrollees also serve for longer in both the JC and the CCC and 3) individuals with more schooling are more likely to serve for longer in both programs (See Figure 3).

E. Quantifying Violations of Assumptions in Both Approaches

Finally, we quantify how violations of the different assumptions in each approach changes the estimated coefficient of the long-run effect. Essentially, both our approaches rely on obtaining an unbiased estimate of the short-run treatment effect in the CCC. In both cases, define $\phi = (\hat{\tau}^{ST}_{CCC} - \tau^{ST}_{CCC})$, the difference between our "recovered" short-run estimate and the true treatment effect. Any $\phi \neq 0$ leads to a bias of $-\delta * \phi$ in the estimate of the long-run treatment effect with control functions generated from our approaches. In case of the first approach, ϕ is the difference between JC and CCC short-run treatment effect. In the second approach, ϕ stems from the difference in the estimate of the selection bias in the short-run regression for JC and CCC. In

[50] Appendix Table 15 presents results using other control functions, as in Appendix Table 14. Again, results change very little by including the control functions while keeping sample consistent, though the main effects are different and more imprecisely estimated in small samples.

Appendix A, we fully quantify 1) in the first approach, how the bias relates to percentage difference in the treatment effects between JC and CCC, or 2) in the second approach, how it relates to percentage difference in the correlation between duration and the unobserved component in the short-run regression. This allows us to examine how much our estimate of the long-run effect will differ if we were to assume various degrees of violation in our key assumptions.

We present the “bias” terms as the last row of each panel as symmetric bounds around the estimate. These terms can be multiplied by a desired percentage differences in either 1) the JC and CCC short-run treatment effect for the first approach, or 2) the correlation between duration and unobserved component in the second approach. We see that because the bias terms are extremely small, only very large differences in either the short-run treatment effects between JC and CCC or selection bias between JC and CCC would make a meaningful difference in the estimates. We take this as additional evidence that our long-run estimates are robust to different kinds of adjustments for omitted variable bias.

Implications for current JC participants. The only way to infer long term benefits for modern job training programs such as JC is to rely on historical data. However, historical data, by definition, will have been collected in a different context, raising concerns about comparability. There are many reasons to believe the lifetime estimates from the CCC apply to JC participants. The first is the fact that both the participants and the programs share important similarities. The second is that both show positive and statistically significant effects on education, income, geographic mobility, and health. If we assume that mortality going forward is determined by the same factors which have historically determined mortality (if we assume the production function for longevity has remained stable over time), then we should observe increases in longevity for current JC participants. While the short run labor market effects differ for the JC and the CCC, that is likely because the labor markets in the 1940s and the 1990s were very different. In contrast, if one were to look over a participant’s lifetime, differences in labor markets would decline (with both CCC and JC participants experiencing strong and weak labor markets over their lifetimes). We conclude that the common effects on education, income, health and mobility (some of the most important predictors of lifetime income and health) across the two programs suggest that the lifetime benefits estimated in the CCC are likely to generally hold for JC participants as well.[51]

IX. Discussion

Was the program worth it? To answer this question, we calculate the Marginal Value of Public Funds (MVPF) following the approach by Hendren and Sprung-Keyser (2019). The CCC program costs include: 1) upfront cost of the program and 2) increases in social security payouts from enrollees both living longer and having increased PIA. These costs are mitigated by 1) tax increases from earnings benefits of the program, 2) decrease in SSDI claiming, 3) decrease in either SSDI length or retirement length from increase in claiming age. The program benefits include: 1) willingness to pay (WTP) for life extensions, 2) increase in after-tax earnings, 3) $30 per month wage paid (most of which went to families), and 4) the value of other services received by enrollees during the program, such as room and board. These benefits are decreased by the loss of SSDI income from decrease in claiming rate. The MVPF is estimated to be 6.0 including the WTP for life increases, disability reductions, and increases in claiming ages, but only 2.5 if we only count the earnings effects, with reductions in mortality accounting for most of the difference.[52] Thus our conclusions differ from those in Hendren and Sprung-Keyser (2019) who find that the MVFP is below one for JC participants. [53] The key difference is due to our ability to look at *lifetime* effects on *multiple* outcomes, and particularly health and longevity. Moreover, our MVPF likely misses additional benefits. The program likely benefitted not just enrollees, but also their families, and the communities and the landscape where the CCC operated. Future work should attempt to measure these effects to conduct a more extensive cost-benefit evaluation.

---

[51] The literature on the determinants of mortality shows that health, education, income and residential location are important determinants of mortality. Height and normal BMI are both associated with longevity (Fogel 1994). Education (Cutler et al. 2006) and greater lifetime earnings are also associated with lower mortality (Chetty et al. 2016). Finkelstein et al. (2019) and Deryugina and Molitor (2019) show that individuals who move to low-mortality locations experience subsequent lower mortality themselves.

[52] Assumptions made and details of calculation are presented in the Online Data Appendix. Some of the increases in life expectancy could lead to greater government spending through Medicare, potentially lowering the marginal value of public funds (Hendren and Sprung-Keyser 2019)—we ignore these. Families received transfers, which could have benefitted them but also potentially distorted their behaviors. We do not have estimates of these effects. We also do not assess the general equilibrium effects of job training programs. Recent research suggests these effects could be substantial and possibly offset the benefits to individuals (Crepon et al. 2013).

[53] They compute an MVFP of 0.18 for JC. This computation does not incorporate the lower SSDI claims or the potential life extensions we compute here.

While we believe that improved health in early adulthood is a likely mediator, historical accounts suggest that the program may have positively affected other "soft skills," improved mental health, and enlarged social networks. For example, enrollees reported making many life-long friendships and experiencing improvements in their state of mind. Additionally, the Army ran the CCC camps and imposed rules of behavior that were likely unusual for most individuals and may have been beneficial. Criminality is an important outcome which may have been affected as well. Though we do not observe these outcomes directly, we do observe that the CCC increased the probability that young men served in the Army, consistent with a change in either discipline or attitudes towards national service. Future work should assess these claims.

Our results have important implications for evaluations of job training programs. The majority of evaluations focus on labor market outcomes in the short- to medium-term and find small and/or insignificant effects. We confirm these findings in our data. But we observe large changes in lifetime incomes and in outcomes that are not usually studied, namely health, military service, and geographic mobility. These findings suggest that it is essential to evaluate multiple mechanisms and indicators of well-being when assessing the impacts of various interventions in the short and the long term.

References

Abramitzky, Ran, and Mill, Roy and Perez, Santiago, "Linking individuals across Historical sources: A Fully Automated Approach," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 53 (2020), 94-111.

Abramitzky, Ran, Leah Platt Boustan, Katherine Eriksson, James J. Feigenbaum, and Santiago Pérez, "Automated linking of historical data" NBER Working Paper No. w25825, 2019.

Almond, D., Currie, J., & Duque, V, "Childhood circumstances and adult outcomes: Act II," *Journal of Economic Literature, 56* (2018), 1360-1446.

Andrews, Isaiah and James Stock, "Robust Inference with Weak Instruments," NBER Econometrics minicourse, 2018.

Attanasio, Orazio, Arlen Guarín, Carlos Medina, and Costas Meghir, "Vocational Training for Disadvantaged Youth in Colombia: A Long-Term Follow-Up," *American Economic Journal: Applied Economics*, 9 (2017), 131-43.

Autor, David H. Lawrence F. Katz, and Melissa S. Kearney, "Trends in U.S. Wage Inequality: Revising the Revisionists," *The Review of Economics and Statistics*, 90(2008), 300-323.

Athey, Susan, Raj Chetty and Guido Imbens, "Combining Experimental and Observational Data to Estimate Treatment Effects on Long Term Outcomes," Working Paper, 2020.

Bailey, Martha, et al. "How Well Do Automated Linking Methods Perform? Lessons from US Historical Data," NBER Working Paper No. w24019, 2017.

Barnow, Burt S. and Jeffrey Smith, "Employment and Training Programs" NBER Working Paper No. 21659, 2015.

Bell, Felicite C. and Miller, Michael L, “Life Tables for the United States Social Security area 1900-2100,” SSA Pub. No. 11-11536, 2005.
Britton, James Ensign, “The education program of the Civilian Conservation Corps,” University of Richmond Master’s theses Paper No. 135, 1958.
Brock, Julia K, “Creating Consumers: The Civilian Conservation Corps in Rocky Mountain National Park,” Florida State University Treatises and Dissertations, Paper 3012, 2005.
Bureau of Labor Statistics, “Eight Years of CCC Operations, 1933 to 1941,” *Monthly Labor Review*, 52 (1941), 1405-1413.
Bureau of Labor Statistics, "Great Recession, Great Recovery? Trends from the Current Population Survey," *Monthly Labor Review,* April 2018.
David Card, "Origins of the Unemployment Rate: The Lasting Legacy of Measurement without Theory," *American Economic Review*, 101 (2011), 552-557.
Card, David & Jochen Kluve & Andrea Weber, “What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations,” *Journal of the European Economic Association*, 16(2018), 894-931.
Card, David, & Alan B. Krueger, “Does school quality matter? Returns to education and the characteristics of public schools in the United States,” *Journal of Political Economy*, 100 (1992), 1-40.
Chetty, Raj, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler, "The association between income and life expectancy in the United States, 2001-2014." *JAMA* 315 (2016), 1750-1766.
Clay, Karen, Jeff Lingwall, and Melvin Stephens Jr, “Do schooling laws matter? evidence from the introduction of compulsory attendance laws in the united states,” NBER Working Paper No. w18477, 2012.
Cohen, Stan, *The tree Army. A pictorial history if the Civilian Conservations Corps, 1933-1942* (Missoula: Pictorial Histories, 1980).
Cook, Thomas D., William R. Shadish and Vivian C. Wong, “Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons,” Journal of Policy Analysis and Management: *The Journal of the Association for Public Policy Analysis and Management*, *27*(2008), 724-750.
Crépon, Bruno and Gerard J. van den Berg, “Active Labor Market Policies,” *Annual Review of Economics*, 8 (2016), 521–546.
Crépon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot, and Philippe Zamora, “Do labor market policies have displacement effects? Evidence from a clustered randomized experiment,” *The Quarterly Journal of Economics*, *128*(2013), 531-580.
Cunha, Flavio and James J. Heckman, and Susanne M. Schennach, 2010. “Estimating the technology of Cognitive and Noncognitive Skill Formation,” *Econometri*ca, 78 (2010) 883-931.
Cutler, David, Angus Deaton, and Adriana Lleras-Muney, “The Determinants of Mortality,” *Journal of Economic Perspectives*, 20 (2006), 97-120.
Dahl, Gordon B., Andreas R. Kostol, and Magne Mogstad, Family Welfare Cultures,” The *Quarterly Journal of Economics*, 129 (2014), 1711-1752.
Davis, Jonathan M. V. and Sara B. Heller, “Rethinking the benefits of youth employment programs: The heterogeneous effects of summer jobs,” NBER Working Paper No. w23443, 2017.

Deryugina, Tatyana, and David Molitor, “Does When You Die Depend on Where You Live? Evidence from Hurricane Katrina,” NBER Working Paper No. w24822, 2019.

Dobbie, Will, Hans Gronqvist, Susan Niknami, Marten Palme, and Mikael Priks, “The Intergenerational Effects of Parental Incarceration,” NBER Working Paper No. w24186, 2018.

Fechner, Robert, “The Educational contribution of the Civilian Conservation Corps,” *The Phi Delta Kappan*,19 (1937), 305-307, 309.

Federal Security Agency, *Annual Report of the Director of the Civilian Conservation Corps: Fiscal Year Ended June 30 1937*, (Washington: United States Government Printing Office, 1937).

Federal Security Agency, *Annual Report of the Director of the Civilian Conservation Corps: Fiscal Year Ended June 30 1938*, (Washington: United States Government Printing Office, 1938).

Federal Security Agency, *Annual Report of the Director of the Civilian Conservation Corps: Fiscal Year Ended June 30 1939*, (Washington: United States Government Printing Office, 1939).

Federal Security Agency, *Annual Report of the Director of the Civilian Conservation Corps: Fiscal Year Ended June 30 1940*, (Washington: United States Government Printing Office, 1940)..

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams, “Place-Based Drivers of Mortality: Evidence from Migration,” NBER Working Paper No. w25975, 2019.

Fishback, Price, "How Successful Was the New Deal? The Microeconomic Impact of New Deal Spending and Lending Policies in the 1930s," *Journal of Economic Literature*, vol 55(2017), pages 1435-1485.

Fishback, Price, Michael Haines, and Shawn Kantor, “Births, Deaths, and New Deal Relief During the Great Depression,” *Review of Economics and Statistics* 89 (2007), 1-14.

Fitzpatrick, Maria D., and Timothy J. Moore, “The Mortality Effects of Retirement: Evidence from Social Security Eligibility at age 62,” *Journal of Public Economics*, 157 (2018), 121-137.

Flores, Carlos A., Alfonso A. Flores-Lagunes, Arturo Gonzalez and Todd C. Neumann, Estimating the effects of length of exposure to instruction in a training program: the Case of Job Corps. *Review of Economics and Statistics*, *94*(2012), 153-171.

Fogel, Robert W., “Economic Growth, Population Theory, and Physiology: The Bearing of Long-Term Processes on the Making of Economic Policy,” *American Economic Review*, 84 (1994), 369-395.

Gelber, Alexander, Adam Isen, and Judd B. Kessler, “The Effects of Youth Employment: Evidence from New York City Lotteries,” *The Quarterly Journal of Economics*, 131(2016), 423-460.

Goldin, Claudia and Lawrence Katz, "Education And Income In The Early Twentieth Century: Evidence From The Prairies," *Journal of Economic History*, 60 (2000), 782-818.

Goldin, Claudia, & Lawrence F. Katz, “Mass secondary schooling and the state: the role of state compulsion in the high school movement,” In Understanding long-run economic growth: Geography, institutions, and the knowledge economy, University of Chicago Press (2008), 275-310.

Harper, Charles Price, The Administration of the Civilian Conservation Corps, (Clarksburg WV: Clarksburg Publishing Co, 1939).

Heckman, James, Robert LaLonde, and Jeffrey A. Smith, “The economics and econometrics of active labor market programs,” In Handbook of labor economics, Elsevier: 3,(1999), 1865-2097.

Heller, Sara, “Summer jobs reduce violence among disadvantaged youth,” *Science*, 2014, 346 (6214), 1219–1223.

Hendren, Nathaniel and Ben Sprung-Keyser. “A Unified Welfare Analysis of Government Policies NBER Working Paper, No. w26144, 2019.

Hill, Mark E. and Ira Rosenwaike, “The Social Security Administration’s Death Master File: The Completeness of Death Reporting at Older Ages,” *Social Security Bulletin*, 64 (2001), 44-51.

Kugler, Adriana, Maurice Kugler, Juan Saavedra and Luis Omar Herrera Prada, “Long-Term Direct and Spillover Effects of Job Training: Experimental Evidence from Colombia,” NBER Working Paper No. w21607, 2015.

LaLonde, Robert J, “Evaluating the econometric evaluations of training programs with experimental data,” *The American Economic Review*, 76(1986), 604-620.

Lechner Michael, Ruth Miquel and Conny Wunsch, “Long Run Effects of Public Sector Sponsored Training in West Germany,” *Journal of the European Economic Association*, 9 (2011), 742–784.

Levine, Linda, “Job Creation Programs of the Great Depression: the WPA and the CCC” Congressional Research Service, 2010, 7-5700.

Lleras-Muney, Adriana, “Were compulsory attendance and child labor laws effective? An analysis from 1915 to 1939,” *The Journal of Law and Economics*, 45 (2002), 401-435.

McEntee, JJ, *Final Report of the Director of the Civilian Conservation Corps, fiscal year ended*, (Washington, DC: United States Government Printing Office, 1940).

McEntee. JJ, “Final Report of the Director of the Civilian Conservation Corps, April, 1933 through June 30, 1942,” Federal Security Agency M-2125, 1942.

Melzer, Richard Anthony, *Coming of Age in the Great Depression: The Civilian Conservation Corps in New Mexico*, (Las Cruces: Yucca Tree Press, 2000).

Méndez, Fabio, and Facundo Sepúlveda, "The Cyclicality of Skill Acquisition: Evidence from Panel Data," *American Economic Journal: Macroeconomics*, 4 (2012), 128-52.

Montoya, Maria, “The roots of Economic and Ethnic Divisions in Northern New Mexico: The case of the Civilian Conservation Corps” *Western Historical Quarterly*, 26 (1995), 14-34.

Oster, Emily, “Unobservable Selection and Coefficient Stability: Theory and Evidence,” *Journal of Business & Economic Statistics*, 37 (2017), 187-204

Paige, John C, “The Civilian Conservations Corps and the National Park Service: An administrative History,” National Park Service, Department of the Interior, Washington DC, Report number NPS-D-189, 1985.

Parham, Robert Bruce, “The Civilian Conservation Corps in Colorado, 1933-1942,” University of Colorado Master Thesis, 1981.

Parman, Donald L, “The Indian and the Civilian Conservation Corps,” *Pacific Historical Review*, 40 (1971), 39-56.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman, “Distributional national accounts: methods and estimates for the United State,” *The Quarterly Journal of Economics*, 133 (2018), 553-609.

Rawick, George Philip, “The New Deal and Youth: The Civilian Conservation Corps, the National Youth Administration and the American Youth Congress,” Doctoral thesis, History Department, University of Wisconsin, 1957.
Ripani, Laura, Pablo Ibarraran, Jochan Kluve and David Rosas-Schady, “Experimental Evidence on the Long Term Impacts of a Youth Training Program,” *Industrial and Labor Relations Review*, 20 (2018), 1-38.
Rosenzweig, Mark R. and Christopher Udry, “External validity in a stochastic world: Evidence from Low-Income Countries,” *The Review of Economic Studies*, 87 (2020), 343-381.
Salmond, John A, “The Civilian Conservation Corps, 1933-1942,” Durham, North Carolina: Duke University Press, 1967.
Steckel, Richard, “A Peculiar Population: The Nutrition, Health, and Mortality of American Slaves from Childhood to Maturity,” *The Journal of Economic History*, 46 (1986), 721-741.
Schochet, Peter Z, John Burghardt and Sheena McConnell, “Does Job Corps Work? Impact Findings from the National Job Corps Study,” *The American Economic Review*, 98 (2008), 1864-1886.
Schochet, Peter Z, “National Job Corps Study: 20-Year Follow-Up Study Using Tax Data”. Mathematica Policy Research Report, 2018.
Wickens, James F, Colorado in the Great Depression, New York: Garland publishing, 1979.
U.S. Department of Labor, “Handbook for Agencies selecting men for emergency conservation work” Emergency Conservation Work, Bulletin No. 3, Washington GPO May 1, 1933.
Vellore, Arthi, “The Dust Was Long in Settling: Human Capital and the Lasting Impact of the American Dust Bowl,” *The Journal of Economic History*, 78 (2018): 196-230.
Wolfenbarger, Deon, “New Deal Resources on Colorado’s Eastern plains,” National Park Service. United States Department of the Interior, 1992.

Table 1: Summary Statistics

	Analytic Sample			Mortality Sample			Analytic Sample (matched to MBR)		
	N	mean	sd	N	mean	sd	N	mean	sd
Characteristics in Enrollment Application									
Birth year	23,722	1,920	3.712	17,639	1,920	3.649	12,455	1920	3.546
Age at enrollment	23,488	18.75	2.122	17,449	18.73	2.170	12,330	18.74	2.242
Enrollment year	23,722	1,939	1.902	17,639	1,939	1.894	12,455	1939	1.889
Reported age younger than DMF*	23,722	0.0888	0.284	17,639	0.113	0.317	12,455	0.130	0.336
Reported age older than DMF*	23,722	0.167	0.373	17,639	0.219	0.413	12,455	0.253	0.435
Allottee is father	23,722	0.334	0.472	17,639	0.332	0.471	12,455	0.330	0.470
Allottee is mother	23,722	0.466	0.499	17,639	0.475	0.499	12,455	0.475	0.499
Hispanic (imputed using hispanic index)	23,722	0.484	0.500	17,639	0.451	0.498	12,455	0.432	0.495
Additional information in CO records									
Highest grade completed	14,507	8.592	2.109	11,235	8.674	2.081	8,225	8.700	2.055
Household size excluding applicant	7,870	4.745	2.600	6,283	4.763	2.591	4,730	4.725	2.575
Live on farm?	8,101	0.248	0.432	6,460	0.253	0.435	4,846	0.252	0.434
Height (Inches)	8,141	67.80	3.089	6,475	67.88	3.083	4,860	67.92	3.053
Weight (100 pounds)	8,234	1.385	0.171	6,561	1.390	0.172	4,922	1.391	0.171
Body Mass Index	8,115	21.21	2.178	6,461	21.23	2.174	4,849	21.23	2.190
Underweight	8,115	0.0694	0.254	6,461	0.0689	0.253	4,849	0.0685	0.253
Father Living	7,943	0.799	0.401	6,339	0.803	0.398	4,765	0.806	0.396
Mother Living	8,006	0.850	0.357	6,391	0.855	0.352	4,808	0.855	0.352
Service Characteristics									
First allottee amount (dollars per month)	22,970	21.63	3.772	17,088	21.67	3.721	12,097	21.70	3.683
Duration of service (yrs)	23,722	0.821	0.706	17,639	0.826	0.708	12,455	0.816	0.701
Ever Rejected?	23,722	0.0194	0.138	17,639	0.0201	0.140	12,455	0.0199	0.140
Camp Characteristics									
Distance from home to camp in miles (derived)	22,405	154.8	207.1	16,645	157.2	208.0	11,740	159.5	209.1
1st closest city distance form camp (miles)	23,480	26.68	22.50	17,454	26.57	22.26	12,322	26.40	22.06
Mean precipitation in camp 1933-1942	23,202	33.43	9.281	17,253	33.52	9.321	12,174	33.66	9.382
Mean min temp in camp 1933-1942	23,202	1.459	3.474	17,253	1.382	3.457	12,174	1.265	3.450
Mean max temp in camp 1933-1942	23,202	17.51	4.114	17,253	17.39	4.108	12,174	17.24	4.106
Death Certificate Data									
Age at death	19,377	69.82	16.84	17,639	73.62	12.03	12,348	74.76	9.25
=1 if missing age at death	23,722	0.183	0.387	17,639	0	0	12,455	0.009	0.092
Survive at 70	19,377	0.587	0.492	17,639	0.644	0.479	12,348	0.706	0.456
P(70), imputed to 0 if missing	23,722	0.479	0.500	17,639	0.644	0.479	12,455	0.700	0.458
Imputed Prob of Survival at 70 Using Age at Discharge	23,718	0.589	0.446	17,636	0.644	0.479	12,455	0.705	0.454
1940 Census Data									
Matched to 1940 Census	23,722	0.449	0.497	17,639	0.479	0.500	12,455	0.487	0.500
For those that served before 1940									
Year of birth	4,217	1,918	3.836	3,410	1,918	3.803	2,451	1918	3.559
Age at last birthday (in years)	4,217	21.77	3.836	3,410	21.75	3.803	2,451	21.74	3.559
Hispanic	4,217	0.279	0.449	3,410	0.258	0.438	2,451	0.245	0.430
White	4,217	0.991	0.0933	3,410	0.992	0.0903	2,451	0.991	0.092
In labor force	4,217	0.909	0.288	3,410	0.912	0.283	2,451	0.909	0.288
Working, conditional on labor force	3,833	0.711	0.453	3,110	0.718	0.450	2,228	0.711	0.453
Wage, conditional on working	2,983	405.3	361.0	2,424	401.8	337.4	1,764	410.8	360.7
Years of educ	4,159	8.770	2.477	3,363	8.842	2.445	2,415	8.873	2.420
Moved Residence Counties	4,215	0.299	0.458	3,408	0.291	0.454	2,450	0.296	0.457
WWII Records									
Matched to WWII records	23,722	0.306	0.461	17,639	0.338	0.473	12,455	0.347	0.476
Birth year	7,263	1,920	2.810	5,954	1,920	2.831	4,321	1920	2.815
Enrollment year	7,262	1,942	1.424	5,954	1,942	1.439	4,321	1942	1.45
Years of education	7,263	9.395	1.787	5,954	9.404	1.785	4,321	9.399	1.766
Height in inches**	5,971	67.52	6.089	4,876	67.70	6.098	3,510	67.73	6.164
Weight in lbs***	5,641	138.6	26.19	4,595	138.7	25.70	3,327	139.4	27.17
BMI	5,466	21.55	4.500	4,451	21.50	4.101	3,214	21.55	4.399
Ever Married	7,256	0.215	0.411	5,947	0.221	0.415	4,316	0.224	0.417
Moved Residence Counties	7,215	0.303	0.460	5,914	0.296	0.457	4,290	0.303	0.46
Birthplace Rest of US	7,215	0.230	0.421	5,913	0.237	0.425	4,295	0.244	0.429

Notes: Basic sample includes records with duration (begin and end date of enrollment), camp id and enrollment county. The analytical sample for the mortality analysis only includes those not missing death age and death age more than 45. When multiple records were found for a single individual we use the information in the first enrollment record. * Reported age being younger (older) than DMF OR than the oldest (youngest) reported if the individual has multiple enrollment spells. ** Dropped values below 40. *** Dropped values below 90 and over 350

Table 2: Effect of Service Duration on Longevity and Lifetime Earnings

Depdent variable	(1) No Controls	(2) Add Birth, County-qtr Dummies	(3) Add Indiv Controls	(4) Add Camp Chars	(5) Add Peer Chars	(6) Add Camp FE	(7) CO Only
Panel A: Longevity for the full sample							
Duration of service (yrs)	0.013***	0.013***	0.011***	0.011***	0.013***	0.013***	0.013***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Observations	17,086	17,086	17,086	17,086	17,086	17,086	10,944
R-squared	0.003	0.117	0.126	0.127	0.128	0.138	0.149
Mean Dep	73.62	73.62	73.62	73.62	73.62	73.62	73.30
Panel B: AIME (MBR sample claimed 1979 and later)							
Duration of service (yrs)	-0.083	67.178***	62.791***	62.450***	56.717***	50.134***	48.707***
	(9.563)	(12.565)	(12.840)	(12.889)	(14.378)	(15.555)	(18.481)
Observations	10,241	10,241	10,241	10,241	10,241	10,241	6,525
R-squared	0.000	0.188	0.204	0.205	0.206	0.222	0.236
Mean Dep	963.62	963.62	963.62	963.62	963.62	963.62	1010.70
Panel C: Retirement age							
Duration of service (yrs)	0.506***	0.507***	0.452***	0.462***	0.427***	0.401***	0.554***
	(0.069)	(0.092)	(0.094)	(0.095)	(0.105)	(0.114)	(0.127)
Observations	11,712	11,712	11,712	11,712	11,712	11,712	7,768
R-squared	0.005	0.157	0.167	0.168	0.169	0.184	0.192
Mean Dep	60.27	60.27	60.27	60.27	60.27	60.27	60.43
Panel D: SSDI (excluding unknowns)							
Duration of service (yrs)	-0.016**	-0.022***	-0.020**	-0.021**	-0.017*	-0.021**	-0.031***
	(0.006)	(0.008)	(0.009)	(0.009)	(0.010)	(0.010)	(0.012)
Observations	10145	10145	10145	10145	10145	10145	6480
R-squared	0.001	0.154	0.161	0.163	0.164	0.181	0.205
Mean Dep	0.21	0.21	0.21	0.21	0.21	0.21	0.20

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Sample is restricted only to those that died after age >= 45. Column (1) includes only duration of service as regressor. Column (2) adds Birth and County-Year-Quarter of Enrollment fixed effects. Column (3) adds individual controls. Column (4) adds camp characteristics, such as distance from nearest city and average temperature. Column (5) adds peer characteristics, where peers are defined as other enrollees serving in the same camp at the same time. Column (6) adds camp fixed effects and removes camp characteristics. Column (7) runs the regression specification in Column (6) for only enrollees from our Colorado Records. For complete list of controls, refer to text or Appendix Table 3.

Table 3: Effect of Service Duration on Missing Data and Sample Selection

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	No Controls	Add Birth, County-qtr Dummies	Add Indiv Controls	Add Camp Chars	Add Peer Chars	Add Camp FE	CO Only
Panel A: Does duration predict whether longevity is missing?							
Duration of service (yrs)	0.001	-0.017***	-0.020***	-0.020***	-0.017***	-0.015***	-0.008
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Observations	22,964	22,964	22,964	22,964	22,964	22,964	14,116
R-squared	0.000	0.111	0.196	0.197	0.198	0.206	0.200
Mean Dep	0.18	0.18	0.18	0.18	0.18	0.18	0.15
Panel B: Does duration predict being in the MBR sample?							
Duration of service (yrs)	-0.006	0.004***	0.010*	0.011*	0.009	0.005	0.002
	(0.005)	(0.001)	(0.006)	(0.006)	(0.007)	(0.007)	(0.009)
Observations	22,980	22,980	22,980	22,980	22,980	22,980	14,116
R-squared	0.000	0.102	0.205	0.206	0.206	0.212	0.187
Mean Dep	0.53	0.53	0.53	0.53	0.53	0.53	0.57
Panel C: Is the effect of duration on longevity for the MBR sample the same as in the full sample?							
Duration of service (yrs)	0.013***	0.010***	0.009***	0.009***	0.012***	0.011***	0.014***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Observations	11,953	11,953	11,953	11,953	11,953	11,953	7,913
R-squared	0.005	0.157	0.169	0.169	0.170	0.185	0.190
Mean Dep	74.81	74.81	74.81	74.81	74.81	74.81	74.78

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. See Notes on Table 2 for specifications in each column. Panel A explores the outcome of = 1 if death age is missing, = 0 otherwise. Panel B explores the outcome of = 1 if in the MBR sample, = 0 otherwise. Panel C explores the outcome log death age (same as Table 2 Panel A), but only for the sample of individuals found in the MBR sample.

Table 4: Effect of Service Duration on Labor Market Outcomes Observed in the 1940 Census

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Data Source:	1940 census outcomes (individuals enrolling pre 1940)					WWII enlistment records			combined
Outcome	Found in Census Records	In Labor Force	Weeks Worked in 1939^	Total Annual Wage in 1939^	Ln Total Annual Wage | Working^	Found in WWII Records: served	Height (inches)	BMI	Education (yrs)
Duration of service (yrs)	0.012	0.019*	0.265	-14.497	-0.014	0.038***	1.143***	1.018***	0.169***
	(0.012)	(0.010)	(1.199)	(26.389)	(0.062)	(0.007)	(0.221)	(0.204)	(0.040)
Observations	9,518	4,052	2,360	2,148	1,749	22,963	5,770	5,287	9,586
Mean Dep	0.43	0.91	27.88	383.71	471.25	0.31	67.55	21.53	9.23

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. This table only displays specification in Column (6) of Table 2 on different outcomes observed in the 1940 Census. Sample are enrollees who serve before 1940 and can be matched to a 1940 Census Record. For results on all specifications, refer to Appendix Tables 9 and 10.

Table 5: Effect of Service Duration on Geographic Mobility Over the Lifetime

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Short term mobility (place in 1940 census or WWII enslitment differs from place of residence at enrollment in CCC)				Long term mobility (place of death differs from place of enrollment in CCC)		
	Moved to a Different State	Moved to a Different County	New County Has Higher Yearly Wage Than Sending County	New County Has Above Median Mortality Rate (1950-1968)	Died in a Different State	Died in a Different County	New County Has Above Median Mortality Rate (1950-1968)
Duration of service (yrs)	0.026***	0.057***	0.077**	-0.065***	-0.029*	0.005	0.006
	(0.007)	(0.011)	(0.034)	(0.024)	(0.015)	(0.011)	(0.015)
Observations	9,568	9,568	1,452	3,003	7,235	7,231	5,313
Mean Dep	0.09	0.33	0.59	0.38	0.5	0.8	0.25

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. This table only displays specification in Column (6) of Table 2 on different outcomes observed in the 1940 Census. Sample are enrollees who serve before 1940 and can be matched to a 1940 Census Record. For results on all specifications, refer to Appendix Table 10.

Table 6: Characteristics of Eligible Job Corps Applicants and Comparison to CCC

	Job Corps Data		CCC
Characteristic	All Applicants	Males only	Males Only
Baseline Characteristics			
Duration (in years, only positive durations)	0.67	0.652	0.819
Male	0.6	1	1
Age at application	18.8	18.728	18.75
White, non-Hispanic	0.3	0.304	NA
Black, non-Hispanic	0.5	0.451	NA
Hispanic	0.2	0.169	0.484
Other	0.1	0.076	NA
Years of education	10.2	10.042	8.581
High school diploma or more (including GED)	0.2	0.19	0.12
Ever arrested	0.3	0.332	NA
Had a job in the past year	0.6	0.662	NA
Ever had job	0.8	0.808	0.375
Average earnings in the past year (dollars)	2974.9	3255.739	NA
Mean for outcomes			
Duration for treated (years, duration > 0)	0.67	0.652	0.826
Duration for treated (years)	0.483	0.487	0.819
Years of school	11.145	11.07	9.403
Employment (in week of the survey)^	0.606	0.631	0.71
Weeks worked in previous year	30.62	32.17	27.88
Total ann. earnings in prev. yr	10538.31	11947.78	382.43
Total ann. earnings in prev. yr (weeks worked > 0)	12990.85	14471.77	466.69
Moved^^	0.198	0.207	0.34
Self-reported health status in 12 months^^^	1.786	1.733	NA
Self-reported health status in 48 months^^^	1.809	1.757	NA
Self-reported health excellent or good (12-month)*	0.838	0.855	NA
Self-reported health excellent or good (48-month)*	0.828	0.842	NA
Reason ended: End of term	0.31	0.302	0.378
Reason ended: Employment	0.042	0.038	0.116
Reason ended: Convenience of the government	0.001	0	0.145
Reason ended: Urgent and Proper Call	0.09	0.056	0.116
Reason ended: Deserted	0.331	0.373	0.223
Reason ended: Rejected upon examination	0	0	0.0101
Reason ended: No Record	0.228	0.232	0.0127
Observations: Baseline	14327	8646	NA
Observations: Outcomes	11313	6528	NA

Source: Jobs Corps Baseline data. ^employment is not conditional on labor force participation. ^^for Job Corps it is defined as living more than 20 miles away from baseline residence. For CCC it is defined as living in a different county than the county of residence at the time of enrollment. For Job Corps, employment is defined as having a job during the 208th week after the baseline survey (four years). ^^^Self-reported health status with 1 = excellent health, 2 = good, 3 = fair, and 4 = poor health. *Constructed variable that is equal to 1 if self-reported health status is 1 or 2 (excellent health or good health).

Table 7: Comparison to Job Corps

	Jobs Corps Data			CCC
	RCT		OLS	OLS
	Coefficient on Treatment Dummy (ITT)	2SLS Instrument Duration with Treatment	Coefficient on Duration (years)+	Coefficient on Duration (years)
Years of school	0.184***	0.393	0.360***	0.169***
	(0.039)	(0.084)	(0.041)	(0.040)
Observations	6,280	6,280	3,407	9,620
Employment (in week of the survey)^	0.026**	0.056	0.060***	-0.015
	(0.013)	(0.027)	(0.015)	(0.022)
Observations	6,022	6,022	3,285	3,684
Weeks worked in previous year	1.615***	3.443	2.629***	0.265
	(0.536)	(1.142)	(0.610)	(1.199)
Observations	6,235	6,235	3,382	2,360
Total Annual Earnings in previous year	969.765***	2,083.466	1,055.435***	-14.497
	(280.804)	(603.598)	(336.311)	(26.389)
Observations	6,081	6,081	3,317	2,148
ln(Earnings)	weeks worked>0	0.038	0.080	0.078**	-0.014
	(0.027)	(0.057)	(0.031)	(0.062)
Observations	5,009	5,009	2,753	1,749
Moved^^	0.018*	0.038	0.060***	0.057***
	(0.011)	(0.023)	(0.014)	(0.011)
Observations	6,301	6,301	3,419	9,568
Self-reported health excellent or good (12-month)^^^	0.035***	0.073	0.020*	
	(0.009)	(0.020)	(0.010)	
Observations	5,920	5,920	3,234	
Self-reported health excellent or good (48-month)^^^	0.016*	0.034	0.013	
	(0.010)	(0.020)	(0.011)	
Observations	6,279	6,279	3,407	
Duration of training in months			5.829	
Individual controls?	No	No	Yes	Yes

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Sample is Jobs Corps data males only. +Sample includes all treated, including those with zero duration. Controls include year and quarter of baseline, year and quarter of 48-mo followup survey, whether individual was enrolled in non-residential program and baseline characteristics such as whether individual had child, was ever arrested, had ever used drugs, had a job, had a job in the previous year, ever had a job, race, native language, on welfare as a child, education, baseline marital status and others. ^ Employment is not conditional on labor force participation. ^^ For Job Corps it is defined as living more than 20 miles away from baseline residence. For CCC it is defined as living in a different county than the county of residence at the time of enrollment. For Job Corps, employment is defined as having a job during the 208th week after the baseline survey (four years). Earnings conditional on employment only includes the earnings of individuals employed during the 208th week after the baseline survey. ^^^ Constructed variable that is equal to 1 if self-reported health status is 1 or 2 (excellent health or good health).

Table 8: Long term estimates using education control functions for identification

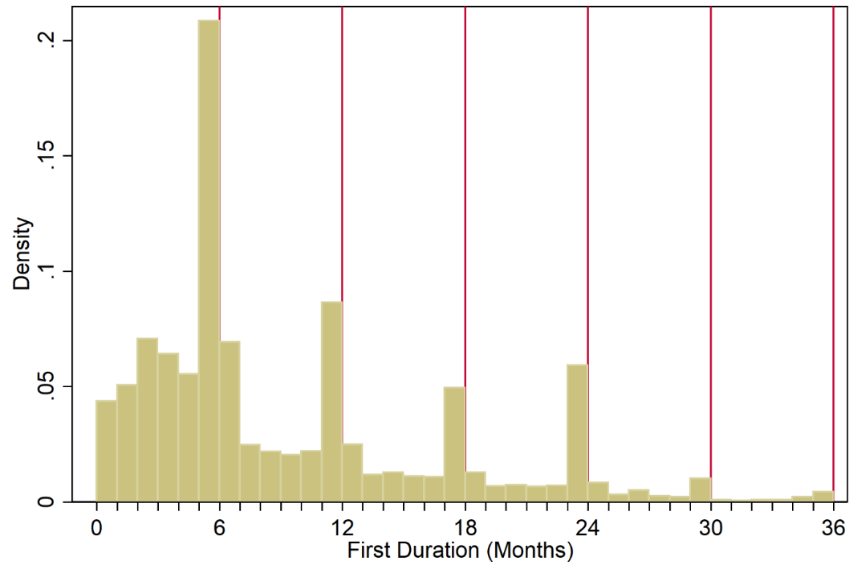
	(1)	(2)	(3)	(4)
Dependent Variable:	*Log Death Age*	*AIME*	*Retirement Age*	*SSDI*
Panel A: OLS Without Control Functions				
Duration of service (yrs)	0.013***	47.882**	0.509***	-0.02
	(0.004)	(21.770)	(0.189)	(0.014)
Panel B: Control Function Approach 1 (Athey et al 2020)				
Duration of service (yrs)	0.013***	52.363**	0.536***	-0.022
	(0.005)	(21.843)	(0.190)	(0.014)
Bounds to account for assumption violations*	±7.22E-5	±1.152	±6.93E-3	±5.78E-4
Panel C: Control Function Approach 2 (This Paper)				
Duration of service (yrs)	0.013***	46.809**	0.502***	-0.019
	(0.004)	(21.763)	(0.189)	(0.014)
Bounds to account for assumption violations*	±4.59E-06	±0.079	±5.42E-04	±4.19E-05
Observations	7,722	4,613	4,575	4,575

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. This table reports the coeffcients on duration in a regression of log age at death, AIME, retirement age, and SSDI claiming. The sample is enrollees for which the control function using education can be computed using only common covariates between JC and CCC (enrollment age, age less than 18 indicator, highest grade level, hispanic status, whether helpd a previous job, whether graduated high school, household size, from rural hosehold, whether father is living, whether mother is living). See text for a description about how the control functions in Panels B and C are constructed.

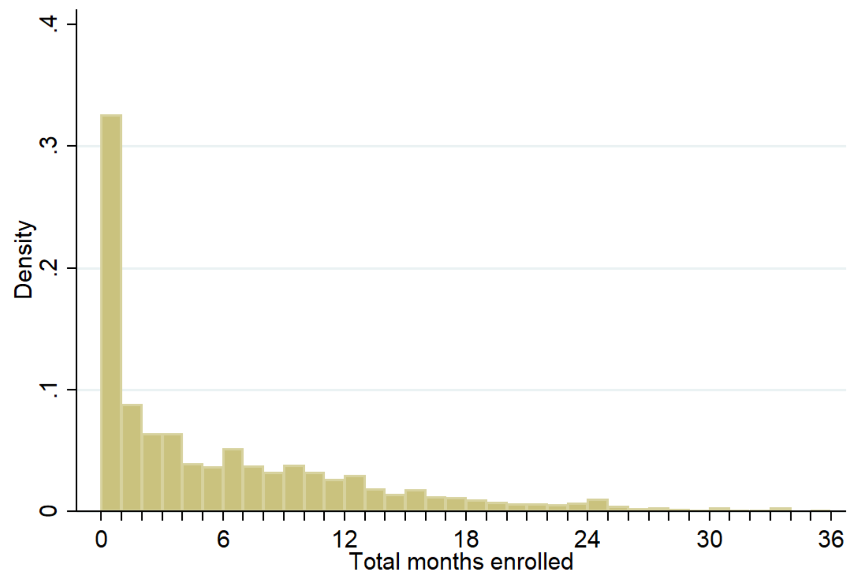
* This term can be multiplied by desired percentage difference in treatment effect between JC and CCC (Panel B) or omitted variable bias between JC and CCC (Panel C) to calculate the final bounds.

Figure 1: Distribution of Service Duration in the CCC Records and Jobs Corps

Panel A: CCC



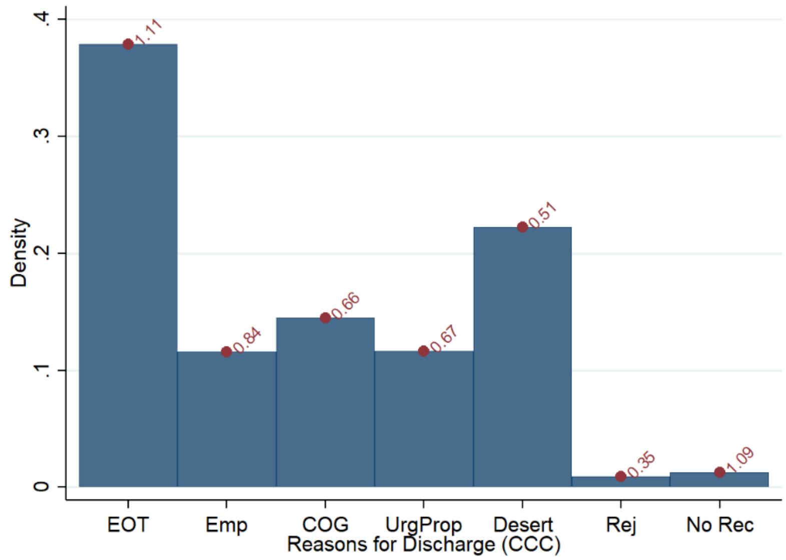
Panel B: Jobs Corps



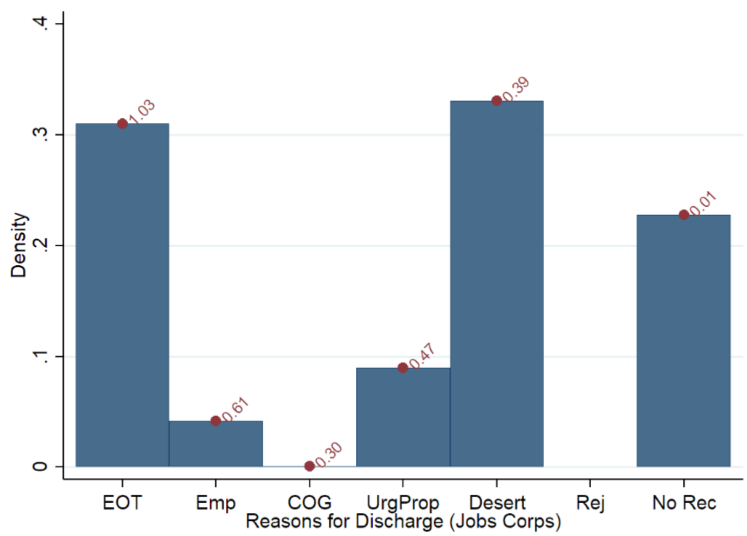
Notes: We exclude durations greater than 3 years (less than 1% of the observations) in this figure. Mean duration is 9.44 months (s.d. 7.47) for CCC and 5.8 months (s.d. 6.6) for Jobs Corps.

Figure 2: Distribution of Reason for Discharge

Panel A: CCC



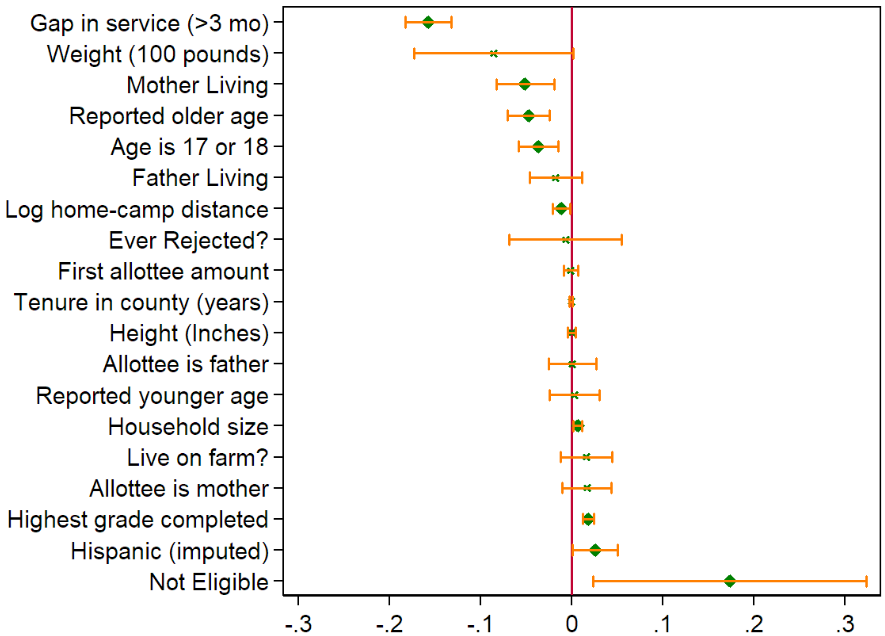
Panel B: Jobs Corps



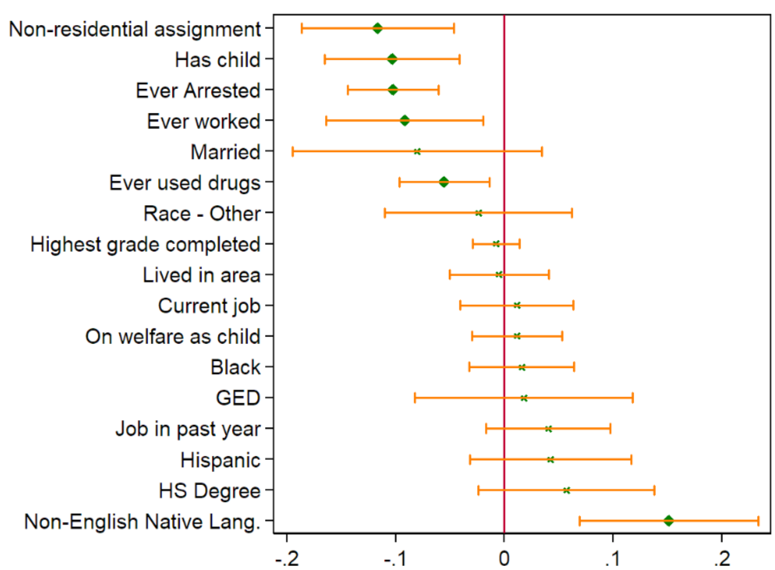
Note: Values on top of the bar graph are mean duration (in years) for each category: EOT (End of Term), Emp (employment outside the program), COG (Convenience of the Government), UrgProp (Urgent and Proper Call), Desert, Rej (Rejected), No Rec (No record). Reasons for Jobs Corps was harmonized to match with CCC's reasons for discharge.

Figure 3: Determinants of Duration

Panel A: CCC

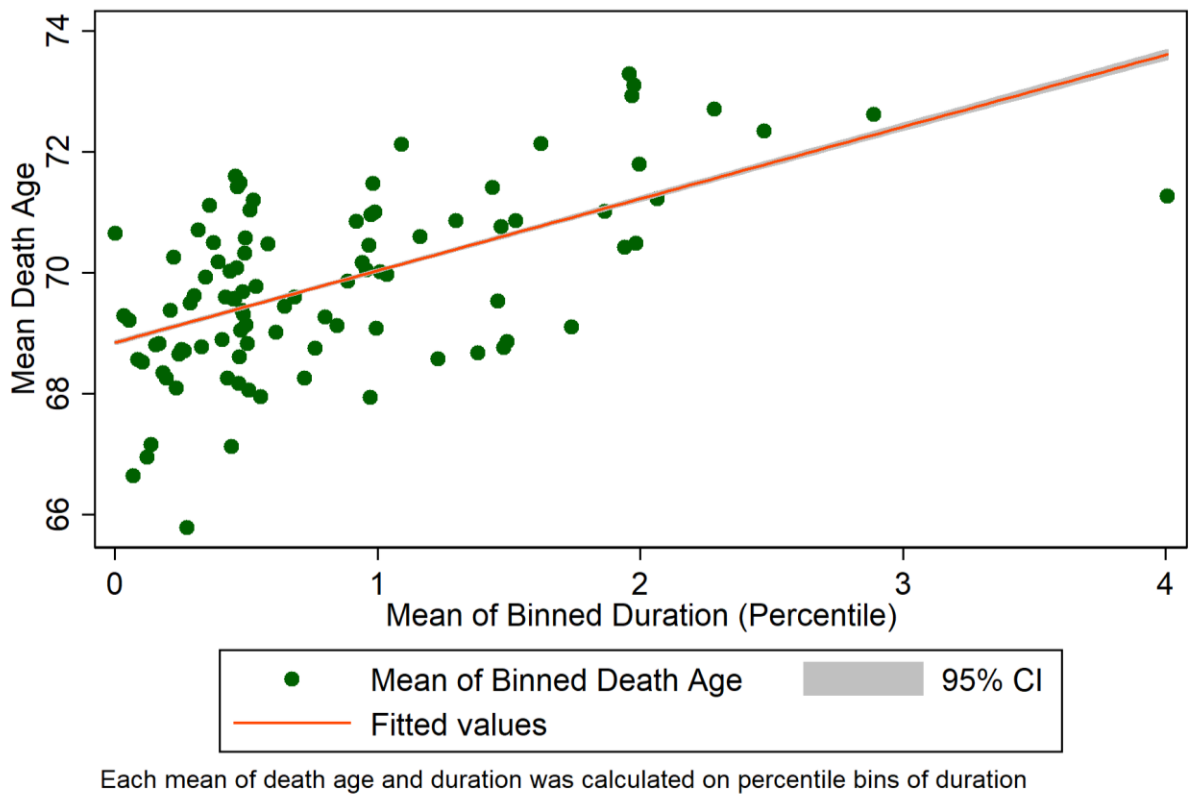


Panel B: Jobs Corps



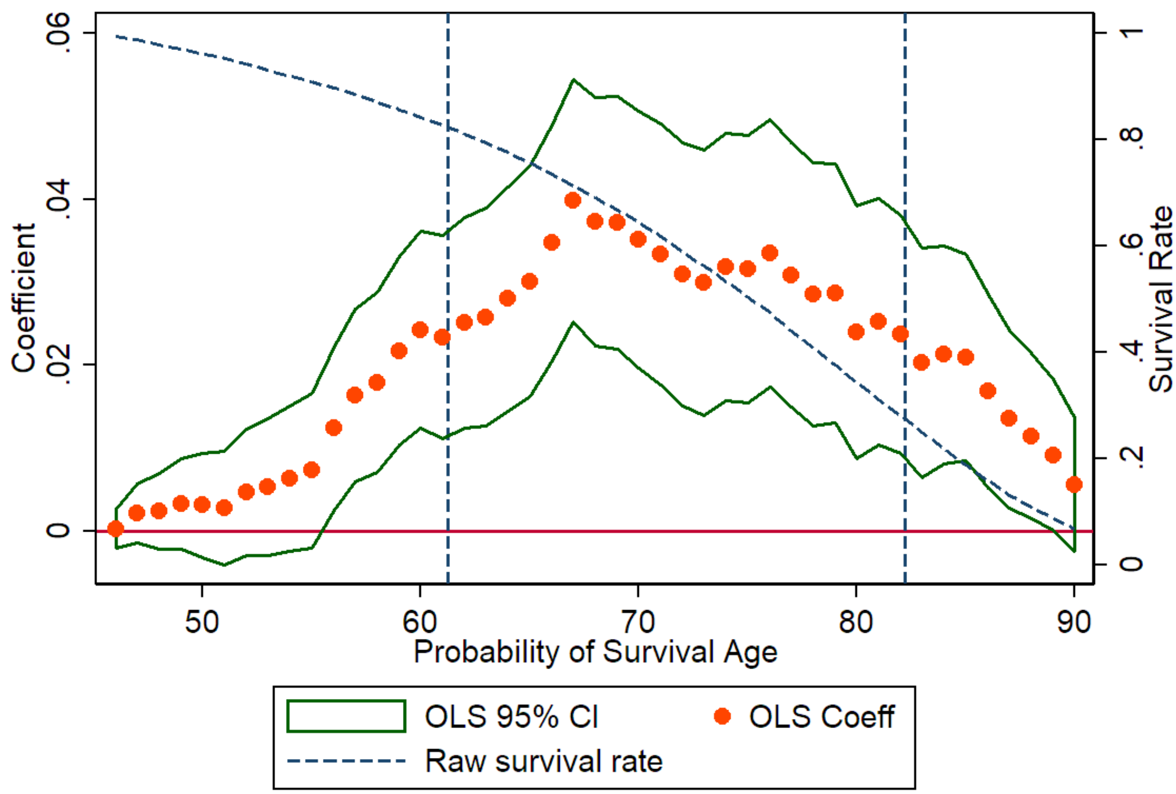
Note: Estimates and 95% confidence intervals plotted for coefficient estimates on selected variables from regressing duration on various individual, camp, and peer characteristics. Coefficients in diamond are statistically significant at the 95% level. Mean duration for the estimation sample is 0.84 years for CCC and 0.49 years for Jobs Corps. Full results of the regression estimates are shown in Appendix Table 2.

Figure 4: Longevity Increases with CCC Service Duration



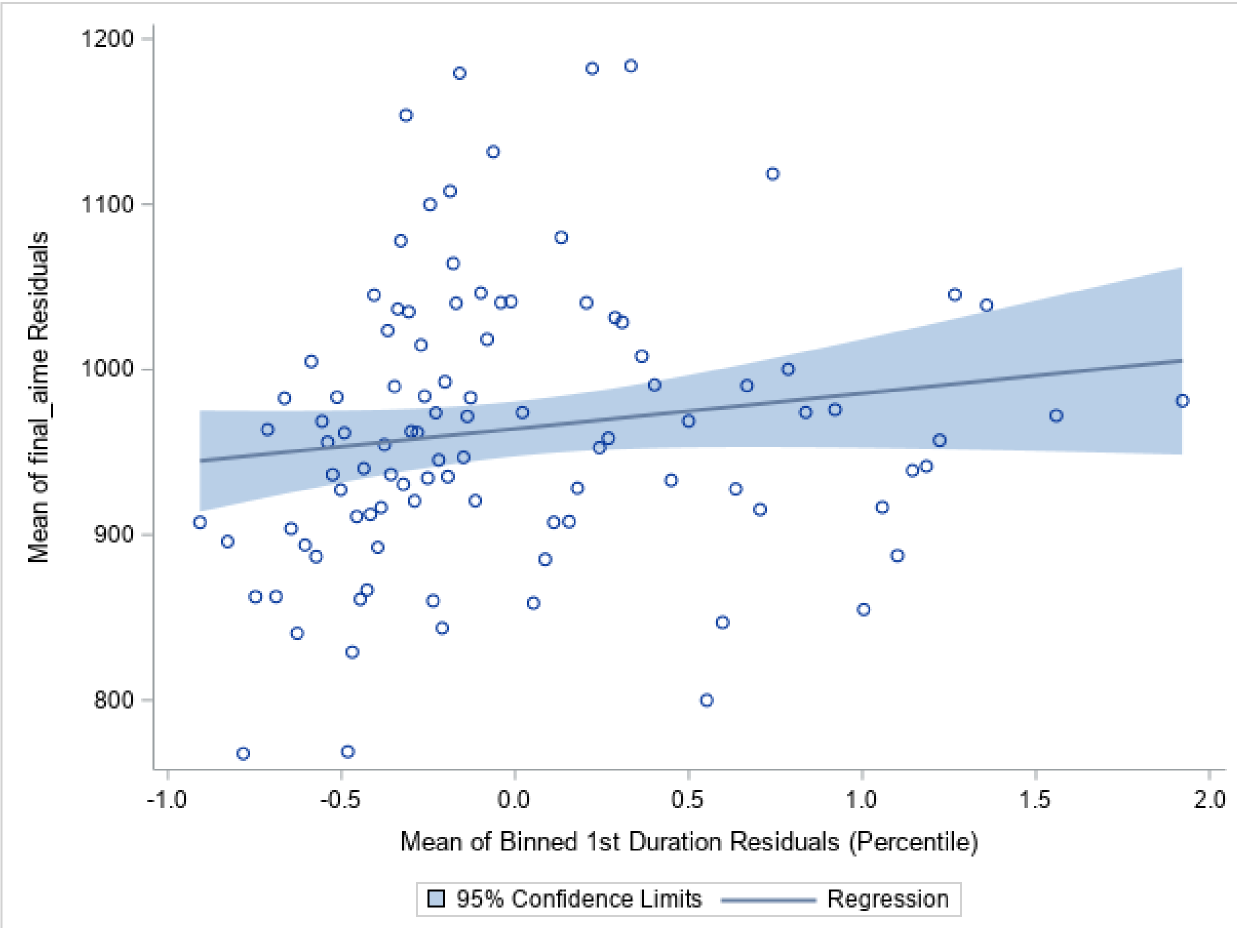
Notes: figure plots the linear fit of mean death age within each percentile bin of duration. Data: Administrative records matched to death certificates. See text for more details.

Figure 5: Effect of Service Duration on the Probability of Survival to Different Ages



Notes: On the left y-axis, this figure reports the coefficients (and standard errors) from running linear regressions of the probability that the person survived to a given age a on duration, where age ranges from age 45 to age 90. The regressions use the administrative data we collected and control for all observables at baseline (see Table 2 for details). On the right y-axis we plot the survival rate.

Figure 6: CCC Duration and AIME



Notes: Authors computation based on administrative program data matched to the Master Beneficiary Records, for those claiming 1979 or later. This restricts the sample to enrollees serving less than 3 years. It plots residuals from regressing AIME on birth year (y-axis), and regressing duration of service on birth year (x-axis).

Appendix A Control Function Approach

In this section we explore the control function approach in detail, beginning with the original approach in Athey, Chetty, and Imbens (2020) then discussing our extension.

A.1 Athey Chetty Imbens (2020)

In Athey Chetty Imbens (2020) (henceforth ACI) the set-up is an experimental sample with only the secondary (short-term) outcome and observational sample with both the secondary and primary (long-term) outcomes. The question they address is how the experimental sample can be used to obtain the treatment effect on the long-term outcome that is observed only in the observational sample.

ACI has four assumptions that allows us to recover τ_O^P, reproduced here:

Assumption 1. (External Validity of the Observational Study) *The observational sample is a random sample of the population of interest.*

This assumption exists to set the baseline of the analysis to the observational sample, and is essentially definitional.

Assumption 2. (Internal Validity of the Experimental Sample) *For $w = 0, 1$,*

$$W_i \perp\!\!\!\perp \left(Y_i^P(w), Y_i^S(w)\right) | X_i, G_i = E \tag{A1}$$

This assumption allows us to estimate treatment effects in the experimental sample without bias. **Assumption 3.** (Conditional External Validity) *The experimental study has conditional external validity if*

$$G_i \perp\!\!\!\perp \left(Y_i^P(0), Y_i^P(1), Y_i^S(0), Y_i^S(1)\right) | X_i \tag{A2}$$

Assumption 3 implies that the conditional average treatment effect in both samples is the same as $E[Y_i^S(1) - Y_i^S(0)|X_i, G_i = O] = E[Y_i^S(1) - Y_i^S(0)|X_i, G_i = E]$. Assumption 3 also implies that $\tau_O^S = \tau_E^S$ and $\sigma_O^S = \sigma_E^S$.

Finally, the last assumption relates the secondary (short-term) outcomes to primary (long-term) outcomes:

Assumption 5. (Latent Unconfoundedness) *For w = 0,1,*

$$W_i \perp\!\!\!\perp Y_i^P(w)|X_i, Y_i^S(w), G_i = O \tag{A3}$$

This allows ACI to identify τ_O^P by inferring the bias in the observational sample from the estimated treatment effects on the secondary outcome in the two samples, and transfer that to the primary (long term) outcome.

A.2 ACI Linear Setting and Our Approach

In ACI linear setting, the short-term outcomes have the following formulation,

$$\begin{aligned} Y_i(0) &= X_i^T\gamma + \alpha_i \\ Y_i(1) &= Y_i(0) + \tau_g \\ Y_i &= \tau_g W_i + X_i^T\gamma + \alpha_i \end{aligned}$$

Furthermore, they assume a stronger version of A5,
Assumption 5' LINEAR LATENT UNCONFOUNDEDNESS

$$\begin{aligned} &\alpha_i^P = \delta\alpha_i^S + \varepsilon_i^P \\ &W_i \perp\!\!\!\perp \varepsilon_i^P | X_i, \alpha_i^S, G_i = O \end{aligned}$$

In our approach, we differ with ACI's linear setting in two ways. First, we use continuous treatment, which makes Assumption 5 into a stronger one. Second, instead of the ACI Assumption 3 that the experimental sample is externally valid for the observational sample, we consider two different approaches: first, we assume that the short-term treatment effect are the same between the two samples, and second, we assume that the short-term bias is the same and utilize the instrument in the JC sample. In the most favorable case both lead to the same results because the observational study has internal validity from the outset.

Our first approach takes the assumption that the short-term treatment effect between CCC and JC samples are the same, or in our notation, $\tau_E^S = \tau_O^S$. Using the IV approach in the JC sample, we can obtain an unbiased estimate of τ_E^S, which in turn gives us an unbiased estimate of τ_O^S. Finally, we can construct the control function as in ACI

$$\hat{\alpha}_i^S = Y_i^S - W_i\hat{\tau}_O^S - X_i^T\hat{\gamma}^S \tag{A4}$$

and include the control function in the long-term regression of the observational sample.

Our second approach assumes that the (linear) selection bias is the same between CCC and JC. In this approach, we exploit the fact that we have an instrument for duration in the JC sample. Therefore, we can think of the difference between the IV estimate and the OLS estimate gives us an estimate of the bias in the JC sample,

$$\hat{\sigma} = \hat{\tau}_{E,OLS}^S - \hat{\tau}_{E,2SLS}^S \tag{A5}$$

Then, adjusting the OLS estimate of the short-term treatment effect from the CCC sample $\hat{\tau}_O^S - \hat{\mu}$ gives us an unbiased estimate of the short-term treatment effect of the CCC sample. Finally, we construct the control function as before and include in the long-term regression of the observational sample.

We present a complete step-by-step description here. To make notations easier to interpret in the description of the approaches, we replace the experimental sample subscript E by JC for Jobs Corps and observational sample subscript O by CCC for CCC. Additionally, we replace secondary outcome sample superscript S by ST for short-term and the primary outcome sample superscript P by LT for long-term.

Approach 1: Assuming treatment effect is the same

1. Using the (experimental) JC data we estimate the short-term treatment effects for outcomes available in both the CCC and JC data. These include schooling, employment, earnings and geographic mobility. Using the JC sample, we instrument for the duration W using the random assignment T. This procedure gives us an unbiased estimate of the short-run treatment effect in the JC, as well as in the CCC (by assumption).

2. Estimate the residual in the CCC data using the estimated ST treatment effect from the JC RCT ($\hat{\tau}_{JC}^{ST}$)
$$\hat{\alpha}_{iCCC}^{ST} = Y_{iCCC}^{ST} - \hat{\gamma}X_{iCCC} - \hat{\tau}_{JC}^{ST}W_{iCCC} \tag{A6}$$

3. Include the ST residuals calculated in step 2 ($\hat{\alpha}_{iCCC}^{ST}$) as controls in the LT CCC regressions:
$$Y_{iCCC}^{LT} = X_{iCCC}\gamma_{CCC}^{LT} + \tau_{CCC}^{LT}W_{iCCC} + \delta\alpha_{iCCC}^{ST} + \varepsilon_{iCCC}^{ST} \tag{A7}$$

Approach 2: Assuming selection bias is the same

1. Estimate ST treatment effect from RCT using both OLS and 2SLS. We use random assignment to treatment as instrument for duration to construct the 2SLS estimates. We construct the OLS estimates using the treated arm of the experiment only.

2. Estimate the selection or omitted variable bias term ($\hat{\mu}$) by subtracting JC's 2SLS estimate from JC's OLS estimate of ST treatment
$$\hat{\mu} = \hat{\tau}_{JC,OLS}^{ST} - \hat{\tau}_{JC,2SLS}^{ST} \tag{A8}$$

3. Estimate ST treatment effect in the CCC sample and adjust it by the estimated selection or OVB term ($\hat{\mu}$)
$$\hat{\tau}_{CCC}^{ST} = \hat{\tau}_{CCC,OLS}^{ST} - \hat{\mu} \tag{A9}$$

4. Estimate the residual of the ST treatment effect using our adjusted estimate of the short-term treatment effect ($\hat{\tau}^{ST}_{CCC}$)

$$\hat{\alpha}^{ST}_{iCCC} = Y^{ST}_{iCCC} - \hat{\gamma} X_{iCCC} - \hat{\tau}^{ST}_{CCC} W_{iCCC} \tag{A10}$$

5. Include estimated residual in LT treatment effect regression in order to generate an unbiased estimate of the long-term impact of the CCC on outcomes.

$$Y^{LT}_{iCCC} = X_{iCCC}\gamma^{LT}_{CCC} + \tau^{LT}_{CCC} W_{iCCC} + \delta\alpha^{ST}_{iCCC} + \varepsilon^{ST}_{iCCC} \tag{A11}$$

A.3 Quantifying the Effect of Violations of Assumptions

In each of our two approaches, we make an assumption that allows us to recover τ^S_O without bias in large samples. In the first approach, we assume that $\tau^S_{CCC} = \tau^S_{JC}$, and in the second approach, we assume that $\sigma^S_{CCC} = \sigma^S_{JC}$.

In practice it is plausible that neither assumption holds exactly. So, let us suppose that both these assumptions are violated, and we estimate short-term TE in the observational sample with bias. let the bias be denoted by so $\phi = \hat{\tau}^S_O - \tau^S_O$. In our first approach, ϕ is the difference between JC and CCC short-term treatment effects, In our second approach, ϕ is the difference in the short-term bias between JC and CCC. We can characterize the biases for the two approaches. In general if the short term effects are similar, even if not identical, the first approach is preferable, whereas if the biases are similar, but not identical, the second approach is preferable.

Then,

$$\begin{aligned}
\hat{\alpha}^S_i &= Y^S_i - W_i\hat{\tau}^S_O - X^T_i\hat{\gamma}^S \\
&= \alpha^S_i - W_i * \phi \\
\hat{\alpha}^P_i &= \alpha^P_i - (\delta * \phi)W_i
\end{aligned}$$

and so regressing primary outcomes on duration, X, and control function will be mis-specified

$$Y^P_i = (\tau_P - \delta * \phi)W_i + X^T_i\gamma + \delta\alpha^S_i + \varepsilon^P_i \tag{A12}$$

which yields a final bias of $-\delta * \phi$.

In our first approach, where we assume that short-term treatment effects are identical, ϕ term is the difference between JC and CCC short-term treatment effects, so $bias = -\delta * (\tau^S_{CCC} - \tau^S_{JC})$. Expressing this in terms of percentage difference in short-term treatment effects,

$$bias_1 = -\delta * \tau^S_{JC} * \%\Delta\tau^S$$

where $\%\Delta\tau^S = \frac{\tau^S_{CCC} - \tau^S_{JC}}{\tau^S_{JC}}$.

In our second approach, ϕ is the difference in the short-term bias between JC and CCC, so the bias is $bias = -\delta * (\Delta(\text{short-term bias}))$ or

$$bias_2 = -\delta * \left(\Delta\frac{1}{sd(W_i)^S} * \beta^S_2 corr(W_i, U_i)^S_{JC} + \frac{1}{sd(W_i)^S_{CCC}} * \Delta[\beta^S_2 corr(W_i, U_i)^S]\right)$$

where $\beta^S_2 corr(W_i, U_i)^S_{JC}$ is a component of the omitted variable bias in short-run regression of JC and $\Delta[\beta^S_2 corr(W_i, U_i)^S]$ is the difference in the components between CCC and JC.[1] Everything except $\Delta[\beta^S_2 corr(W_i, U_i)^S]$ is observed.

So after the estimate of the long-term treatment effect is first adjusted by, $-\delta * \Delta\frac{1}{sd(W_i)^S} * \beta^S_2 corr(W_i, U_i)^S_{JC}$, the remaining bias for the long term effect, expressed in terms of percentage difference of the short-term bias term is,

$$-\delta * \frac{\beta^S_2 corr(W_i, U_i)^S_{JC}}{sd(W_i)^S_{CCC}} * \%\Delta[\beta^S_2 corr(W_i, U_i)^S]$$

[1] In a regression setting $Y_i = \beta_0 + \beta_1 W_i + \beta_2 U_i + \eta_i$, the omitted variable bias when U_i is omitted can be expressed as $\beta_2 corr(W_i, U_i)\frac{sd(U_i)}{sd(W_i)}$

Appendix Table 1: Sample Selection

Sample Restriction	Itself	Sequential
All	26290	26290
Camp Exist	25165	25165
Enrollment Exist	24832	23943
Duration Exist	26050	23722
Final analytic sample	**23722**	**23722**
Death Age Exist	21457	19377
Death Age Restrict	24386	17639
Final analytic sample for mortality	**17639**	**17639**

The rows show many observations survive after dropping for each restriction. Itself column shows how many observations survive if we drop for just the restriction in the row. Sequential column shows the final observations that survive when we drop for each reason sequentially. Our working sample is 23,722, where we additionally lose observations to Death Age Exist for death age analysis, resulting in a sample of 17,639.

Appendix Table 2a: Summary Statistics From Enrollment Records

	Analytic Sample			Mortality Sample			Analytic Sample (matched to MBR)		
	N	mean	sd	N	mean	sd	N	mean	sd
<u>Characteristics in Enrollment Application</u>									
Birth year	23,722	1,920	3.712	17,639	1,920	3.649	12,455	1920	3.546
Age at enrollment	23,488	18.75	2.122	17,449	18.73	2.170	12,330	18.74	2.242
Enrollment year	23,722	1,939	1.902	17,639	1,939	1.894	12,455	1939	1.889
Reported age younger than DMF*	23,722	0.0888	0.284	17,639	0.113	0.317	12,455	0.130	0.336
Reported age older than DMF*	23,722	0.167	0.373	17,639	0.219	0.413	12,455	0.253	0.435
Age is 17 or 18	23,488	0.564	0.496	17,449	0.535	0.499	12,330	0.513	0.500
Not Eligible	23,722	0.0151	0.122	17,639	0.0143	0.119	12,455	0.0139	0.117
Allottee is father	23,722	0.334	0.472	17,639	0.332	0.471	12,455	0.330	0.470
Allottee is mother	23,722	0.466	0.499	17,639	0.475	0.499	12,455	0.475	0.499
Non-junior	23,722	0.00628	0.0790	17,639	0.00675	0.0819	12,455	0.0067	0.0818
Hispanic (imputed using hispanic index)	23,722	0.484	0.500	17,639	0.451	0.498	12,455	0.432	0.495
Additional information in CO records									
Highest grade completed	14,507	8.592	2.109	11,235	8.674	2.081	8,225	8.700	2.055
Household size excluding applicant	7,870	4.745	2.600	6,283	4.763	2.591	4,730	4.725	2.575
Live on farm?	8,101	0.248	0.432	6,460	0.253	0.435	4,846	0.252	0.434
Height (Inches)	8,141	67.80	3.089	6,475	67.88	3.083	4,860	67.92	3.053
Weight (100 pounds)	8,234	1.385	0.171	6,561	1.390	0.172	4,922	1.391	0.171
Body Mass Index	8,115	21.21	2.178	6,461	21.23	2.174	4,849	21.23	2.190
Underweight	8,115	0.0694	0.254	6,461	0.0689	0.253	4,849	0.0685	0.253
Overweight	8,115	0.0450	0.207	6,461	0.0461	0.210	4,849	0.0462	0.210
Father Living	7,943	0.799	0.401	6,339	0.803	0.398	4,765	0.806	0.396
Mother Living	8,006	0.850	0.357	6,391	0.855	0.352	4,808	0.855	0.352
Tenure in county (years)	5,432	12.66	6.483	4,326	12.68	6.504	3,353	12.59	6.522
Ever had a paid regular job?	8,841	0.375	0.484	7,022	0.386	0.487	5,256	0.394	0.489
Male White Unemployed / Male White Pop 1937	23,709	0.0885	0.0397	17,629	0.0864	0.0388	12,450	0.085	0.0378
Male White Unemployed / Male White Pop 1940	23,709	0.0710	0.0308	17,629	0.0696	0.0299	12,450	0.0688	0.0291
<u>Service Characteristics</u>									
First allottee amount (dollars per month)	22,970	21.63	3.772	17,088	21.67	3.721	12,097	21.70	3.683
Duration of service (yrs)	23,722	0.821	0.706	17,639	0.826	0.708	12,455	0.816	0.701
Ever Rejected?	23,722	0.0194	0.138	17,639	0.0201	0.140	12,455	0.0199	0.1397
=1 if disabled	23,722	0.00847	0.0917	17,639	0.00686	0.0825	12,455	0.0069	0.0828
Gap in service (more than 3 months)	23,722	0.160	0.366	17,639	0.173	0.378	12,455	0.180	0.384
Reason ended: End of term	23,722	0.379	0.485	17,639	0.379	0.485	12,455	0.372	0.483
Reason ended: Employment	23,722	0.116	0.320	17,639	0.124	0.329	12,455	0.125	0.331
Reason ended: Convenience of the government	23,722	0.145	0.352	17,639	0.151	0.358	12,455	0.154	0.361
Reason ended: Urgent and Proper Call	23,722	0.117	0.321	17,639	0.122	0.327	12,455	0.125	0.330
Reason ended: Deserted	23,722	0.222	0.416	17,639	0.206	0.404	12,455	0.205	0.404
Reason ended: Rejected upon examination	23,722	0.00915	0.0952	17,639	0.00754	0.0865	12,455	0.0069	0.0828
Reason ended: No Record	23,722	0.0128	0.112	17,639	0.0120	0.109	12,455	0.012	0.109
Honorable Discharge	23,722	0.767	0.423	17,639	0.785	0.411	12,455	0.786	0.410
<u>Camp Characteristics</u>									
Distance from home to camp in miles (derived)	22,405	154.8	207.1	16,645	157.2	208.0	11,740	159.5	209.1
1st closest city distance form camp (miles)	23,480	26.68	22.50	17,454	26.57	22.26	12,322	26.40	22.06
2nd closest city distance form camp (miles)	23,480	49.86	22.49	17,454	49.33	22.32	12,322	48.71	22.17
Mean precipitation in camp 1933-1942	23,202	33.43	9.281	17,253	33.52	9.321	12,174	33.66	9.382
Mean min temp in camp 1933-1942	23,202	1.459	3.474	17,253	1.382	3.457	12,174	1.265	3.450
Mean max temp in camp 1933-1942	23,202	17.51	4.114	17,253	17.39	4.108	12,174	17.24	4.106
Camp Mean Hispanic (imputed using hispanic index)	23,722	0.482	0.313	17,639	0.462	0.312	12,455	0.430	0.329
Camp Type: Department of Grazing	23,671	0.135	0.341	17,593	0.132	0.339	12,455	0.131	0.337
Camp Type: Federal Reclamation Project	23,671	0.0553	0.229	17,593	0.0566	0.231	12,455	0.056	0.230
Camp Type: Fish and Wildlife Service	23,671	0.0118	0.108	17,593	0.0111	0.105	12,455	0.0106	0.102
Camp Type: National Forest	23,671	0.295	0.456	17,593	0.290	0.454	12,455	0.292	0.454
Camp Type: National Monument	23,671	0.0191	0.137	17,593	0.0184	0.134	12,455	0.0188	0.136
Camp Type: National Park	23,671	0.105	0.307	17,593	0.108	0.310	12,455	0.108	0.310
Camp Type: Soil Conservation	23,671	0.307	0.461	17,593	0.311	0.463	12,455	0.306	0.461
Camp Type: State Park	23,671	0.0524	0.223	17,593	0.0527	0.223	12,455	0.054	0.226
Camp Type: Other	23,671	0.0202	0.141	17,593	0.0206	0.142	12,455	0.0214	0.145

Notes: Basic sample includes records with duration (begin and end date of enrollment), camp id and enrollment county. The analytical sample for the mortality analysis only includes those not missing death age and death age more than 45. When multiple records were found for a single individual we use the information in the first enrollment record. *Reported age being younger (older) than DMF OR than the oldest (youngest) reported if the individual has multiple enrollment spells.

Appendix Table 2b: Summary Statistics From Death Certificate, 1940 and WWII Records

	Analytic Sample			Analytic Sample for mortality Analysis			Analytic Sample (MBR matched)		
	N	mean	sd	N	mean	sd	N	mean	sd
<u>Death Certificate Data</u>									
Age at death	19,377	69.82	16.84	17,639	73.62	12.03	12,348	74.76	9.25
=1 if missing age at death	23,722	0.183	0.387	17,639	0	0	12,455	0.009	0.092
Survive at 70	19,377	0.587	0.492	17,639	0.644	0.479	12,348	0.706	0.456
P(70), imputed to 0 if missing	23,722	0.479	0.500	17,639	0.644	0.479	12,455	0.700	0.458
Imputed Prob of Survival at 70 Using Age at Discharge	23,718	0.589	0.446	17,636	0.644	0.479	12,455	0.705	0.454
<u>1940 Census Data</u>									
Matched to 1940 Census	23,722	0.449	0.497	17,639	0.479	0.500	12,455	0.487	0.500
Panel a: those that served before 1940									
Year of birth	4,217	1,918	3.836	3,410	1,918	3.803	2,451	1918	3.559
Age at last birthday (in years)	4,217	21.77	3.836	3,410	21.75	3.803	2,451	21.74	3.559
Hispanic	4,217	0.279	0.449	3,410	0.258	0.438	2,451	0.245	0.430
White	4,217	0.991	0.0933	3,410	0.992	0.0903	2,451	0.991	0.092
In labor force	4,217	0.909	0.288	3,410	0.912	0.283	2,451	0.909	0.288
Working, conditional on labor force	3,833	0.711	0.453	3,110	0.718	0.450	2,228	0.711	0.453
Wage, conditional on working	2,983	405.3	361.0	2,424	401.8	337.4	1,764	410.8	360.7
Lives in CO	4,217	0.776	0.417	3,410	0.787	0.409	2,451	0.790	0.407
Lives in NM	4,217	0.166	0.372	3,410	0.152	0.360	2,451	0.144	0.351
Years of educ	4,159	8.770	2.477	3,363	8.842	2.445	2,415	8.873	2.420
Moved Residence Counties	4,215	0.299	0.458	3,408	0.291	0.454	2,450	0.296	0.457
Panel b: those that served after 1940									
Year of birth	636	1,920	3.486	532	1,920	3.493	418	1921	2.621
Age at last birthday (in years)	636	19.66	3.486	532	19.62	3.493	418	19.45	2.621
Hispanic	636	0.365	0.482	532	0.340	0.474	418	0.330	0.471
White	636	0.994	0.0791	532	0.992	0.0865	418	0.995	0.069
In labor force	636	0.879	0.326	532	0.883	0.321	418	0.880	0.325
Working, conditional on labor force	559	0.719	0.450	470	0.711	0.454	368	0.712	0.453
Wage, conditional on working	440	253.8	167.2	366	258.6	172.1	282	252.9	149.6
Lives in CO	636	0.855	0.352	532	0.868	0.338	418	0.864	0.344
Lives in NM	636	0.134	0.341	532	0.122	0.328	418	0.129	0.336
Years of educ	629	8.347	2.135	526	8.390	2.114	413	8.370	2.097
Moved Residence Counties	636	0.145	0.352	532	0.139	0.346	418	0.136	0.344
<u>WWII Records</u>									
Matched to WWII records	23,722	0.306	0.461	17,639	0.338	0.473	12,455	0.347	0.476
Birth year	7,263	1,920	2.810	5,954	1,920	2.831	4,321	1920	2.815
Enrollment year	7,262	1,942	1.424	5,954	1,942	1.439	4,321	1942	1.45
Years of education	7,263	9.395	1.787	5,954	9.404	1.785	4,321	9.399	1.766
Height in inches*	5,971	67.52	6.089	4,876	67.70	6.098	3,510	67.73	6.164
Weight in lbs**	5,641	138.6	26.19	4,595	138.7	25.70	3,327	139.4	27.17
BMI	5,466	21.55	4.500	4,451	21.50	4.101	3,214	21.55	4.399
Ever Married	7,256	0.215	0.411	5,947	0.221	0.415	4,316	0.224	0.417
Home State CO	7,232	0.591	0.492	5,928	0.605	0.489	4,300	0.617	0.486
Moved Residence Counties	7,215	0.303	0.460	5,914	0.296	0.457	4,290	0.303	0.46
Home State NM	7,232	0.319	0.466	5,928	0.305	0.460	4,300	0.289	0.453
Birthplace CO	7,215	0.444	0.497	5,913	0.451	0.498	4,295	0.462	0.499
Birthplace NM	7,215	0.322	0.467	5,913	0.309	0.462	4,295	0.292	0.455
Birthplace Rest of US	7,215	0.230	0.421	5,913	0.237	0.425	4,295	0.244	0.429

Notes: Basic sample includes records with duration (begin and end date of enrollment), camp id and enrollment county. The analytical sample for the mortality analysis only includes those not missing death age and death age more than 45. When multiple records were found for a single individual we use the information in the first enrollment record. * Dropped values below 40. ** Dropped values below 90 and over 350

Appendix Table 3: Determinants of CCC Service Duration

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Indiv Controls	Camp Controls	Indiv+Camp	Add County-Quarter FE	CO Only	CO Non-missing Only
Individual characteristics						
Ever Rejected?	-0.201***		-0.020	-0.007	-0.009	0.060
	(0.033)		(0.034)	(0.031)	(0.034)	(0.038)
=1 if disabled	-0.446***		-0.464***	-0.328***	-0.363***	-0.237*
	(0.055)		(0.055)	(0.050)	(0.061)	(0.127)
Non-junior	0.834***		0.840***	0.509***	0.574***	0.005
	(0.122)		(0.119)	(0.097)	(0.127)	(0.235)
Reported Age Younger than DMF^	0.033*		0.026	0.003	0.003	-0.005
	(0.019)		(0.019)	(0.014)	(0.020)	(0.024)
Reported Age Older than DMF	0.081***		0.089***	-0.047***	-0.029*	-0.033
	(0.015)		(0.015)	(0.012)	(0.016)	(0.025)
Not Eligible	0.300**		0.265*	0.174**	0.186*	0.662***
	(0.139)		(0.141)	(0.077)	(0.106)	(0.134)
Age is 17 or 18	0.100***		0.103***	-0.037***	-0.045***	-0.020
	(0.014)		(0.014)	(0.011)	(0.014)	(0.021)
Allottee amount	0.058***		0.060***	-0.001	0.009	0.026***
	(0.004)		(0.005)	(0.004)	(0.006)	(0.009)
Allottee is father	0.045***		0.045***	0.001	0.001	-0.003
	(0.017)		(0.017)	(0.013)	(0.019)	(0.027)
Allottee is mother	0.045***		0.045***	0.017	0.030	0.012
	(0.017)		(0.016)	(0.014)	(0.019)	(0.027)
Gap in service	-0.201***		-0.156***	-0.158***	-0.126***	-0.113***
	(0.016)		(0.015)	(0.013)	(0.016)	(0.020)
Log distance from home to camp (miles)	-0.016***		-0.013**	-0.011**	-0.015***	-0.021**
	(0.005)		(0.005)	(0.005)	(0.006)	(0.008)
Hispanic (imputed using hispanic index)	0.078***		0.058***	0.026**	-0.014	0.007
	(0.014)		(0.014)	(0.013)	(0.017)	(0.019)
Highest grade completed (CO only)	0.024***		0.021***	0.019***	0.016***	0.007*
	(0.003)		(0.003)	(0.003)	(0.003)	(0.004)
Household size excluding applicant (CO only)	0.012***		0.013***	0.007***	0.008***	0.007***
	(0.003)		(0.003)	(0.002)	(0.002)	(0.003)
Live on farm? (CO only)	0.053***		0.053***	0.016	0.012	0.017
	(0.016)		(0.017)	(0.014)	(0.015)	(0.017)
Height (Inches) (CO only)	0.002		0.001	0.000	-0.000	-0.001
	(0.003)		(0.003)	(0.002)	(0.002)	(0.002)
Weight (100 pounds) (CO only)	-0.189***		-0.154***	-0.085*	-0.113**	-0.019
	(0.054)		(0.052)	(0.045)	(0.047)	(0.045)
Father Living (CO only)	-0.054***		-0.055***	-0.018	-0.015	-0.006
	(0.019)		(0.019)	(0.015)	(0.015)	(0.017)
Mother Living (CO only)	-0.088***		-0.095***	-0.051***	-0.056***	-0.032
	(0.021)		(0.021)	(0.016)	(0.017)	(0.024)
Tenure in county (years) (CO only)	-0.001		-0.001	-0.001	-0.001	-0.001
	(0.001)		(0.001)	(0.001)	(0.001)	(0.001)

Camp characteristics						
=1 if camp is in enrollment state		-0.094***	0.053	0.154***	0.165***	-0.027
		(0.034)	(0.051)	(0.058)	(0.059)	(0.066)
Mean precipitation in camp 1933-1942		-0.001	-0.001	-0.004***	0.001	0.001
		(0.001)	(0.001)	(0.001)	(0.002)	(0.003)
Mean min temp in camp 1933-1942		0.010	0.014**	0.030***	0.027***	0.012
		(0.006)	(0.006)	(0.008)	(0.008)	(0.010)
Mean max temp in camp 1933-1942		-0.018***	-0.021***	-0.034***	-0.022**	-0.006
		(0.006)	(0.006)	(0.007)	(0.009)	(0.011)
Camp Type: Department of Grazing		0.131***	0.123***	-0.075	0.117	-0.052
		(0.044)	(0.041)	(0.063)	(0.087)	(0.116)
Camp Type: Federal Reclamation Project		0.118**	0.099**	-0.055	0.147	0.031
		(0.047)	(0.045)	(0.070)	(0.096)	(0.120)
Camp Type: Fish and Wildlife Service		0.106**	0.024	-0.383***		
		(0.051)	(0.048)	(0.131)		
Camp Type: National Forest		0.008	-0.006	-0.106*	0.024	-0.091
		(0.043)	(0.041)	(0.060)	(0.078)	(0.109)
Camp Type: National Monument		0.145*	0.121	-0.303***	-0.265*	-0.166
		(0.088)	(0.084)	(0.090)	(0.147)	(0.179)
Camp Type: National Park		0.069	0.060	-0.117*	-0.012	-0.165
		(0.044)	(0.042)	(0.063)	(0.079)	(0.101)
Camp Type: Soil Conservation		0.121***	0.100***	-0.075	0.092	-0.070
		(0.040)	(0.038)	(0.059)	(0.080)	(0.108)
Camp Type: State Park		-0.031	-0.041	-0.119*	-0.078	-0.176
		(0.054)	(0.050)	(0.069)	(0.090)	(0.147)
Log distance to closest city (miles)		-0.007*	-0.007**	0.011**	0.000	0.022**
		(0.004)	(0.004)	(0.005)	(0.007)	(0.008)
Log distance to 2nd closest city (miles)		0.028	0.035*	-0.017	-0.044*	0.012
		(0.019)	(0.019)	(0.022)	(0.025)	(0.037)
Peer Char: Hispanic at enrollment		0.386***	0.239***	0.249***	0.015	0.051
		(0.044)	(0.047)	(0.070)	(0.071)	(0.098)
Peer Char: Age at enrollment		-0.200***	-0.235***	-0.319***	-0.313***	0.052
		(0.021)	(0.023)	(0.034)	(0.035)	(0.041)
Peer Char: Reported Age Younger than DMF		0.483***	0.381**	-0.607***	-0.579**	0.478*
		(0.170)	(0.169)	(0.211)	(0.254)	(0.262)
Peer Char: Reported Age Older than DMF		-0.276**	-0.452***	-1.025***	-0.814***	0.397
		(0.127)	(0.137)	(0.200)	(0.236)	(0.318)
Peer Char: Not Eligible (First enrollment)		1.861***	1.587***	1.349***	-0.295	1.949*
		(0.256)	(0.273)	(0.389)	(0.452)	(1.041)
Peer Char: Allottee amount		0.083***	0.030***	-0.255***	-0.360***	-0.305***
		(0.005)	(0.007)	(0.017)	(0.024)	(0.018)
Peer Char: Allottee: Father		-0.083	-0.120	0.019	-0.040	0.088
		(0.126)	(0.122)	(0.149)	(0.177)	(0.198)
Peer Char: Allottee: Mother		-0.163	-0.117	-0.032	-0.078	-0.221
		(0.126)	(0.128)	(0.133)	(0.147)	(0.202)
Peer Char: Gap in service		-0.931***	-0.692***	-0.652***	-0.156	-1.462***
		(0.098)	(0.099)	(0.133)	(0.140)	(0.191)
Constant	-1.457***	3.342***	2.800***	12.992***	14.686***	6.747***
	(0.458)	(0.518)	(0.569)	(0.868)	(0.991)	(0.807)
Observations	17,639	17,086	17,086	17,086	10,944	3,013
R-squared	0.181	0.160	0.222	0.574	0.482	0.465
Mean Dep	0.83	0.84	0.84	0.84	0.76	0.67
FE	BD	BD	BD	BD,CYQ	BD,CYQ	BD,CYQ
Sample	All	All	All	All	CO	CO
Reason	N	N	N	N	N	N
Number of County-Quarter Groups				1,789	1,231	477

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1. Only Duration <= 3 years, death age >= 45 are included in regression. Variables imputed if missing and missing dummies included. County Unemployment is from ICPSR compilation of County statistics from 1937 Census of Unemployment and 1940 Decenniel Census. Those values are given to enrollment years 1937, 1938 for 1937 Census and 1939-1942 for 1940 Census. ^ =1 if reported age in CCC documents is smaller than in the DMF, or maximum of all reported age for enrollee.

Appendix Table 4: Full Regressions of Log Death Age on Duration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	No Controls	Add Birth, County-qtr Dummies	Add Indiv Controls	Add Camp Chars	Add Peer Chars	Add Camp FE	CO only
Duration of service (yrs)	0.013***	0.013***	0.011***	0.011***	0.013***	0.013***	0.009**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)
Ever Rejected?			-0.031***	-0.031***	-0.031***	-0.030***	-0.031***
			(0.011)	(0.011)	(0.011)	(0.011)	(0.012)
=1 if disabled			-0.006	-0.006	-0.006	-0.004	0.008
			(0.016)	(0.016)	(0.016)	(0.016)	(0.023)
Non-junior			0.002	0.004	0.003	-0.000	-0.034
			(0.018)	(0.019)	(0.019)	(0.019)	(0.025)
Reported age younger than DMF^			-0.019***	-0.019***	-0.019***	-0.019***	-0.010*
			(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Reported age older than DMF			-0.022***	-0.022***	-0.022***	-0.022***	-0.017***
			(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
Not Eligible			0.010	0.011	0.010	0.011	0.014
			(0.017)	(0.017)	(0.017)	(0.017)	(0.022)
Age is 17 or 18			0.007*	0.007*	0.007*	0.007*	0.004
			(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
First allottee amount (dollars per month)			0.000	0.000	0.000	0.000	0.000
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Allottee is father			0.008*	0.008*	0.008*	0.008	0.003
			(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Allottee is mother			0.001	0.001	0.001	0.001	-0.001
			(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Gap in service (more than 3 months)			0.001	0.001	0.001	0.001	-0.006
			(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
Log distance from home to camp			0.001	0.002	0.002	0.002	0.002
			(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Hispanic (imputed using hispanic index)			0.018***	0.018***	0.018***	0.019***	0.018***
			(0.004)	(0.004)	(0.004)	(0.004)	(0.006)
Highest grade completed (CO only)			0.004***	0.004***	0.004***	0.005***	0.005***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Household size excluding applicant (CO only)			0.003***	0.003***	0.003***	0.003***	0.003***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Live on farm? (CO only)			0.011*	0.011*	0.011*	0.011*	0.011**
			(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Height (Inches) (CO only)			0.001	0.001	0.001	0.001	0.001
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Weight (100 pounds) (CO only)			-0.042**	-0.041**	-0.041**	-0.041**	-0.041**
			(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Father Living (CO only)			0.000	0.001	0.000	-0.000	-0.000
			(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Mother Living (CO only)			0.008	0.008	0.008	0.008	0.007
			(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Tenure in county (years) (CO only)			-0.001	-0.001	-0.001	-0.001	-0.001
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
=1 if camp is in enrollment state				-0.015	-0.017		
				(0.012)	(0.012)		
Mean precipitation in camp 1933-1942				0.000	0.000		
				(0.000)	(0.000)		
Mean min temp in camp 1933-1942				-0.002	-0.002		
				(0.001)	(0.002)		
Mean max temp in camp 1933-1942				0.000	0.001		
				(0.001)	(0.001)		

Camp Type: Department of Grazing				-0.020	-0.019		
				(0.024)	(0.024)		
Camp Type: Federal Reclamation Project				-0.017	-0.019		
				(0.025)	(0.026)		
Camp Type: Fish and Wildlife Service				-0.012	-0.013		
				(0.032)	(0.033)		
Camp Type: National Forest				-0.015	-0.013		
				(0.024)	(0.025)		
Camp Type: National Monument				-0.006	-0.001		
				(0.028)	(0.028)		
Camp Type: National Park				-0.021	-0.017		
				(0.024)	(0.025)		
Camp Type: Soil Conservation				-0.010	-0.007		
				(0.024)	(0.024)		
Camp Type: State Park				-0.013	-0.012		
				(0.024)	(0.025)		
Log distance to closest city				-0.002**	-0.002**		
				(0.001)	(0.001)		
Log distance to 2nd closest city				0.003	0.005		
				(0.006)	(0.006)		
Peer Char: Hispanic at enrollment					0.002	-0.024	-0.010
					(0.014)	(0.021)	(0.023)
Peer Char: Age at enrollment					0.011**	0.014**	0.011
					(0.005)	(0.006)	(0.007)
Peer Char: Reported Age Younger than DMF					0.006	-0.031	-0.058
					(0.043)	(0.057)	(0.066)
Peer Char: Reported Age Older than DMF					-0.017	-0.007	-0.054
					(0.029)	(0.037)	(0.040)
Peer Char: Not Eligible (First enrollment)					-0.029	-0.070	-0.188*
					(0.051)	(0.077)	(0.098)
Peer Char: Allottee amount					0.002	-0.000	0.004
					(0.002)	(0.003)	(0.004)
Peer Char: Allottee: Father					-0.050*	-0.079**	-0.081*
					(0.030)	(0.038)	(0.044)
Peer Char: Allottee: Mother					-0.004	0.003	0.019
					(0.025)	(0.031)	(0.036)
Peer Char: Gap in service					-0.025	-0.026	0.008
					(0.026)	(0.033)	(0.033)
Reason for discharge (code) = 2, Emp							-0.006
							(0.006)
Reason for discharge (code) = 3, COG							-0.007
							(0.006)
Reason for discharge (code) = 4, UrgProp							-0.009
							(0.006)
Reason for discharge (code) = 5, Desert							-0.017
							(0.019)
Reason for discharge (code) = 6, Rej							-0.016
							(0.031)
Reason for discharge (code) = 7, No Rec							0.001
							(0.019)
Honorable Discharge							0.007
							(0.019)
Constant	4.274***	4.391***	4.308***	4.294***	4.063***	4.363***	4.309***
	(0.002)	(0.137)	(0.159)	(0.168)	(0.206)	(0.162)	(0.183)
Observations	17,086	17,086	17,086	17,086	17,086	17,086	10,944
R-squared	0.003	0.117	0.126	0.127	0.128	0.138	0.149
Mean Dep	73.62	73.62	73.62	73.62	73.62	73.62	73.30
FE	None	BD,CYQ	BD,CYQ	BD,CYQ	BD,CYQ	3D,CYQ,Camp	3D,CYQ,Camp
Sample	All	All	All	All	All	All	CO
Number of County-Quarter Groups		1,789	1,789	1,789	1,789	1,789	1,231

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Sample only includes duration <= 3 years and death age >= 45.

Appendix Table 5: Effect of Service Duration on Survival Rates by Age

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Survival to age 70	Mean Dep	0.65				
Duration of service (yrs)	0.030***	0.032***	0.028***	0.035***	0.030***	0.030***
	(0.005)	(0.006)	(0.006)	(0.007)	(0.008)	(0.008)
Observations	17,086					
Panel B: Survival to age 70 missing imputed	Mean Dep	0.64				
Duration of service (yrs)	0.022***	0.026***	0.023***	0.028***	0.023***	0.016**
	(0.004)	(0.005)	(0.005)	(0.006)	(0.006)	(0.007)
Observations	21,269					
Panel C: Survival to age 70 missing imputed to 0	Mean Dep	0.52				
Duration of service (yrs)	0.024***	0.037***	0.037***	0.040***	0.034***	0.020***
	(0.005)	(0.006)	(0.006)	(0.007)	(0.007)	(0.008)
Observations	21,269					
County-Quarter FE	N	Y	Y	Y	Y	Y
Controls	N	N	Y	Y	Y	Y
Peer + Camp Controls	N	N	N	Y	Y	Y
Camp FE	N	N	N	N	Y	Y
Type of Dismissal	N	N	N	N	N	Y

Standard errors (clustered at the application county and enrollment year-quarter level) in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Sample only includes death ages >= 45. Panel B imputes survival probability using the age at discharge, birth year, and life tables from SSA. Panel C imputes 0 for missing survival probability.

Appendix Table 6: The Effect of Service Duration for Machine-Matched Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	No Controls	Add Birth, County-qtr Dummies	Add Indiv Controls	Add Camp Chars	Add Peer Chars	Add Camp FE	CO Only
Panel A: Longevity from CCC for the machined-matched sample							
Duration of service (yrs)	0.013***	0.011***	0.010***	0.010***	0.011***	0.010**	0.017***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)
Observations	8,833	8,833	8,833	8,833	8,833	8,833	5,904
R-squared	0.003	0.186	0.192	0.194	0.195	0.212	0.220
Mean Dep	72.64	72.64	72.64	72.64	72.64	72.64	72.41
Panel B: Longevity from DMF for the matchine-matched sample							
Duration of service (yrs)	0.013***	0.011***	0.010***	0.010***	0.012***	0.012***	0.019***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)
Observations	9,175	9,175	9,175	9,175	9,175	9,175	6,071
R-squared	0.003	0.181	0.186	0.188	0.189	0.205	0.214
Mean Dep	72.65	72.65	72.65	72.65	72.65	72.65	72.42
Panel C: Does duration predict whether they are machine-matched to DMF?							
Duration of service (yrs)	0.015***	0.024***	0.026***	0.026***	0.024***	0.022***	0.020**
	(0.005)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.009)
Observations	22,964	22,964	22,964	22,964	22,964	22,964	14,116
R-squared	0.000	0.110	0.153	0.153	0.154	0.161	0.165
Mean Dep	0.41	0.41	0.41	0.41	0.41	0.41	0.44

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. In Panel A, we use death age calculated from CCC birth year and death age from hand-matched sources. In Panel B we use death age calculated from DMF birth date and death date from the machine match. Sample is restricted only to those that died after age >= 45 for Panels A and B.

Appendix Table 7: Effect of Service Duration on Longevity and Lifetime Earnings

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	No Controls	Add Birth, County-qtr Dummies	Add Indiv Controls	Add Camp Chars	Add Peer Chars	Add Camp FE	CO Only
Panel A: What is the effect of duration on PIA in the MBR sample? (Claimed 1979 and later)							
Duration of service (yrs)	-1.675	21.706***	19.893***	19.717***	18.979***	17.083***	15.459***
	(2.869)	(3.743)	(3.827)	(3.841)	(4.284)	(4.636)	(5.414)
Observations	10,241	10,241	10,241	10,241	10,241	10,241	6,525
R-squared	0.000	0.200	0.215	0.216	0.218	0.233	0.254
Mean Dep	437.70	437.70	437.70	437.70	437.70	437.70	449.34
Mean Implied AIME	904.62	904.62	904.62	904.62	904.62	904.62	940.99
Implied AIME Increase	-5.23	67.83	62.17	61.62	59.31	53.38	48.31
Panel B: What is the effect of duration on PIA in the MBR sample? (Claimed earlier than 1979)							
Duration of service (yrs)	13.075***	12.552**	12.692**	10.713*	8.819	8.792	8.088
	(3.857)	(6.107)	(6.313)	(6.481)	(7.394)	(10.585)	(11.020)
Observations	1,562	1,562	1,562	1,562	1,562	1,562	1,284
R-squared	0.007	0.456	0.503	0.507	0.511	0.557	0.526
Mean Dep	314.02	314.02	314.02	314.02	314.02	314.02	317.41
Panel C: What is the effect of duration on SSDI claiming in the MBR sample? (excluding unknowns)							
Duration of service (yrs)	-0.016**	-0.022***	-0.020**	-0.021**	-0.017*	-0.021**	-0.031***
	(0.006)	(0.008)	(0.009)	(0.009)	(0.010)	(0.010)	(0.012)
Observations	10145	10145	10145	10145	10145	10145	6480
R-squared	0.001	0.154	0.161	0.163	0.164	0.181	0.205
Mean Dep	0.21	0.21	0.21	0.21	0.21	0.21	0.20
Panel D: What is the effect of duration on SSDI claiming in the MBR sample? (unknowns grouped with those who claimed)							
Duration of service (yrs)	-0.019***	-0.022***	-0.022**	-0.023***	-0.020**	-0.022**	-0.030**
	(0.006)	(0.008)	(0.009)	(0.009)	(0.010)	(0.011)	(0.012)
Observations	10373	10373	10373	10373	10373	10373	6613
R-squared	0.001	0.154	0.161	0.163	0.164	0.179	0.201
Mean Dep	0.22	0.22	0.22	0.22	0.22	0.22	0.22
Panel E: What is the effect of duration on SSDI claiming in the MBR sample? (unknowns grouped with those who did NOT claim)							
Duration of service (yrs)	-0.015**	-0.020**	-0.018**	-0.019**	-0.014	-0.019*	-0.030**
	(0.006)	(0.008)	(0.008)	(0.008)	(0.009)	(0.010)	(0.012)
Observations	10373	10373	10373	10373	10373	10373	6613
R-squared	0.001	0.151	0.157	0.160	0.161	0.178	0.202
Mean Dep	0.20	0.20	0.20	0.20	0.20	0.20	0.20

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Sample is restricted only to those that died after age >= 45.

Appendix Table 8: Heterogeneity in OLS effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Sample	CO	NM	Age <= 18	Age > 18	Allottee Mother	Allottee Father	Allottee Other	Urate above median	Urate below median
Panel A: Log Death Age									
Duration of service (yrs)	0.013***	0.014**	0.014***	0.013***	0.018***	0.009	0.011	0.017***	0.013
	(0.003)	(0.005)	(0.005)	(0.004)	(0.004)	(0.006)	(0.009)	(0.004)	(0.009)
Observations	11,148	6,243	8,042	9,349	8,253	5,801	3,337	8,238	2,742
Panel B: PIA									
Duration of service (yrs)	15.677***	19.203**	25.859***	11.645	12.425*	15.848*	31.316*	29.252***	15.337*
	(5.375)	(9.202)	(6.878)	(7.209)	(6.956)	(9.045)	(17.003)	(9.434)	(7.893)
Observations	6,641	3,779	5,680	4,740	5,077	3,536	1,807	3,415	3,520
NEW Panel B: AIME									
Duration of service (yrs)	49.685***	46.739	84.050***	22.809	33.324	48.270	78.381	79.175**	55.935**
	(18.333)	(30.033)	(23.494)	(23.595)	(23.447)	(30.341)	(56.854)	(32.064)	(26.157)
Observations	6,641	3,779	5,680	4,740	5,077	3,536	1,807	3,415	3,520

	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Sample	Hispanic	Not Hispanic	BMI < 18.5 (CO)	BMI 18.5-25 (CO)	BMI >= 25 (CO)	Phase 2 (1935-1937)	Phase 3 (1937-1940)	Phase 4 (1940-1942)	Random-ized
Panel A: Log Death Age									
Duration of service (yrs)	0.018***	0.009**	0.008	0.013**	0.098	0.022***	0.022***	0.015	0.020***
	(0.005)	(0.004)	(0.071)	(0.007)	(0.137)	(0.006)	(0.005)	(0.009)	(0.005)
Observations	7,864	9,527	433	5,627	290	3,852	7,256	6,049	5,170
Panel B: PIA									
Duration of service (yrs)	19.052**	19.827***	-17.382	21.666**	-724.479	6.741	23.239***	40.099***	17.739**
	(7.672)	(6.285)	(127.386)	(9.369)	-	(9.789)	(7.662)	(14.172)	(8.346)
Observations	4,720	5,700	309	3,943	204	1,739	4,597	4,049	3,097
NEW Panel B: AIME									
Duration of service (yrs)	47.532*	65.536***	-255.798	70.562**	-2400.048	21.357	68.853***	113.696**	55.689**
	(24.967)	(21.708)	(451.014)	(32.194)	-	(32.579)	(25.771)	(47.589)	(27.912)
Observations	4,720	5,700	309	3,943	204	1,739	4,597	4,049	3,097

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Sample is restricted only to those that died after age >= 45 and restrictions described by column headings. The specification uses the most restrictive specification with Camp FE, which was the specification used in Table 2, Column 6.

Appendix Table 9: Effect of Service Duration on Labor Market Outcomes Observed in the 1940 Census

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Regression of Outcome on Duration	No Controls	Add Birth, County-qtr Dummies	Add Indiv Controls	Add Camp Chars	Add Peer Chars	Add Camp FE	CO Only
Census							
Found in Census Records	Mean Dep	0.43					
Duration of service (yrs)	-0.015**	0.009	0.007	0.009	0.006	0.012	0.011
	(0.007)	(0.010)	(0.010)	(0.010)	(0.011)	(0.012)	(0.013)
Observations	9,518	9,518	9,518	9,518	9,518	9,518	7,553
R-squared	0.001	0.137	0.152	0.154	0.155	0.166	0.154
In Labor Force	Mean Dep	0.91					
Duration of service (yrs)	0.014**	0.013*	0.013*	0.015**	0.016*	0.019*	0.018*
	(0.006)	(0.007)	(0.007)	(0.007)	(0.009)	(0.010)	(0.011)
Observations	4,052	4,052	4,052	4,052	4,052	4,052	3,374
R-squared	0.001	0.272	0.279	0.280	0.280	0.305	0.286
Working in Census Week | Labor Force	Mean Dep	0.71					
Duration of service (yrs)	0.006	-0.004	-0.005	-0.004	-0.010	-0.015	-0.011
	(0.011)	(0.014)	(0.014)	(0.014)	(0.019)	(0.022)	(0.023)
Observations	3,684	3,684	3,684	3,684	3,684	3,684	3,067
R-squared	0.000	0.265	0.279	0.283	0.286	0.310	0.295
Weeks Worked in 1939^	Mean Dep	27.88					
Duration of service (yrs)	0.669	-0.691	-0.911	-0.937	-0.896	0.265	0.227
	(0.732)	(1.044)	(1.049)	(1.029)	(1.082)	(1.199)	(1.213)
Observations	2,360	2,360	2,360	2,360	2,360	2,360	2,208
R-squared	0.000	0.314	0.345	0.351	0.354	0.383	0.361
Total Annual Wage in 1939^	Mean Dep	383.71					
Duration of service (yrs)	16.773	-12.266	-18.948	-20.038	-21.185	-14.497	-14.633
	(16.061)	(23.145)	(23.911)	(23.533)	(25.577)	(26.389)	(26.652)
Observations	2,148	2,148	2,148	2,148	2,148	2,148	2,011
R-squared	0.001	0.318	0.352	0.357	0.359	0.391	0.376
Ln Total Annual Wage | Working^	Mean Dep	471.25					
Duration of service (yrs)	0.047	-0.035	-0.047	-0.042	-0.051	-0.014	-0.012
	(0.039)	(0.052)	(0.051)	(0.052)	(0.058)	(0.062)	(0.062)
Observations	1,749	1,749	1,749	1,749	1,749	1,749	1,649
R-squared	0.001	0.396	0.447	0.452	0.454	0.487	0.456

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Sample are those whose first term in CCC is before 1940 and are not enrolled in 1940. The 1940 Census was taken on April 1, 1940. ^ Sample are those whose first term in CCC is before 1939 and are not enrolled in 1939. Census asks labor force and work status on the week before the Census enumeration, while wage information and weeks worked is asked for the year before the Census 1939.

Appendix Table 10: Effect of Service Duration on WWII Service, Health and Education Observed in WWII

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Regression of Outcome on Duration	No Controls	Add Birth, County-qtr Dummies	Add Indiv Controls	Add Camp Chars	Add Peer Chars	Add Camp FE	CO Only
WW2							
Found in WWII Records	Mean Dep	0.31					
Duration of service (yrs)	0.018***	0.036***	0.034***	0.035***	0.038***	0.038***	0.042***
	(0.005)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.009)
Observations	22,963	22,963	22,963	22,963	22,963	22,963	14,116
Enlistment Year	Mean Dep	1942.24					
Duration of service (yrs)	-0.181***	0.976***	0.975***	0.976***	0.966***	0.962***	0.964***
	(0.025)	(0.008)	(0.008)	(0.008)	(0.009)	(0.010)	(0.011)
Observations	7,018	7,018	7,018	7,018	7,018	7,018	4,785
Height	Mean Dep	67.55					
Duration of service (yrs)	-0.022	1.098***	1.098***	1.097***	1.161***	1.143***	1.207***
	(0.103)	(0.190)	(0.191)	(0.190)	(0.209)	(0.221)	(0.276)
Observations	5,770	5,770	5,770	5,770	5,770	5,770	3,816
BMI	Mean Dep	21.53					
Duration of service (yrs)	-0.134**	0.789***	0.829***	0.822***	0.874***	1.018***	1.157***
	(0.064)	(0.191)	(0.191)	(0.190)	(0.195)	(0.204)	(0.265)
Observations	5,287	5,287	5,287	5,287	5,287	5,287	3,454
Combined WW2 Census							
Education	Mean Dep	9.23					
Duration of service (yrs)	-0.072**	0.299***	0.185***	0.186***	0.188***	0.169***	0.115***
	(0.035)	(0.041)	(0.035)	(0.036)	(0.038)	(0.040)	(0.043)
Observations	9,586	9,586	9,586	9,586	9,586	9,586	6,907

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Sample are those found in WWII records. WWII: additionally includes the age at enlistment dummies. Combined: additionally includes age at observation dummies, where if observed in Census, the age is 1940 - birth year.

Appendix Table 11: Effect of Service Duration on Geographic Mobility Over the Lifetime

Regression of Outcome on Duration	(1) No Controls	(2) Add Birth, County-qtr Dummies	(3) Add Indiv Controls	(4) Add Camp Chars	(5) Add Peer Chars	(6) Add Camp FE	(7) CO Only
Panel a: Short-term geographic mobility (Combined WW2 and Census)							
Moved to a Different State	Mean Dep	0.09					
Duration of service (yrs)	-0.014***	0.023***	0.024***	0.023***	0.028***	0.026***	0.033***
	(0.004)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.009)
Observations	9,568	9,568	9,568	9,568	9,568	9,568	6,891
Moved to a Different County	Mean Dep	0.33					
Duration of service (yrs)	0.006	0.053***	0.054***	0.053***	0.062***	0.057***	0.067***
	(0.007)	(0.009)	(0.009)	(0.009)	(0.011)	(0.011)	(0.012)
Observations	9,568	9,568	9,568	9,568	9,568	9,568	6,891
New County Has Higher Yearly Wage Than Sending County				Mean Dep	0.59		
Duration of service (yrs)	-0.005	0.046**	0.049**	0.047**	0.062**	0.077**	0.079**
	(0.020)	(0.020)	(0.020)	(0.021)	(0.026)	(0.034)	(0.035)
Observations	1,452	1,452	1,452	1,452	1,452	1,452	1,209
New County Has Above Median Mortality Rate (1950-1968)				Mean Dep	0.38		
Duration of service (yrs)	-0.061***	-0.071***	-0.068***	-0.072***	-0.068***	-0.065***	-0.064**
	(0.012)	(0.019)	(0.019)	(0.020)	(0.021)	(0.024)	(0.026)
Observations	3,003	3,003	3,003	3,003	3,003	3,003	2,403
Panel b: Long-term geographic mobility							
Died in a Different State	Mean Dep	0.5					
Duration of service (yrs)	-0.016*	-0.020*	-0.025**	-0.025**	-0.026**	-0.029*	-0.027*
	(0.008)	(0.012)	(0.012)	(0.012)	(0.013)	(0.015)	(0.015)
Observations	7,235	7,235	7,235	7,235	7,235	7,235	4,784
Died in a Different County	Mean Dep	0.8					
Duration of service (yrs)	0.003	-0.002	-0.004	-0.003	0.003	0.005	0.002
	(0.007)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.012)
Observations	7,231	7,231	7,231	7,231	7,231	7,231	4,781
New County Has Above Median Mortality Rate (1950-1968)				Mean Dep	0.25		
Duration of service (yrs)	-0.030***	0.004	0.006	0.005	0.003	0.006	0.009
	(0.008)	(0.012)	(0.012)	(0.012)	(0.013)	(0.015)	(0.016)
Observations	5,313	5,313	5,313	5,313	5,313	5,313	3,678

We assume that the person lived in the county of application when definining wheter a person moved. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Sample are those found in WWII records. WWII: additionally includes the age at enlistment dummies. Combined: additionally includes age at observation dummies, where if observed in Census, the age is 1940 - birth year.

Appendix Table 12: Placebo Tests for CO Only

Regression of Outcome on Duration	(1) No Controls	(2) Add Birth, County-qtr Dummies	(3) Add Indiv Controls	(4) Add Camp Chars	(5) Add Peer Chars	(6) Add Camp FE	(7) CO Only
Education	Mean Dep	8.72					
Duration of service (yrs)	0.223***	0.225**	0.261***	0.257***	0.216**	0.212*	0.212*
	(0.080)	(0.095)	(0.091)	(0.090)	(0.107)	(0.118)	(0.118)
N	2,987	2,987	2,987	2,987	2,987	2,987	2,987
Height	Mean Dep	67.94					
Duration of service (yrs)	-0.035	-0.218	-0.062	-0.054	-0.162	-0.209	-0.209
	(0.125)	(0.170)	(0.146)	(0.149)	(0.179)	(0.186)	(0.186)
N	2,334	2,334	2,334	2,334	2,334	2,334	2,334
Weight (100 pounds)	Mean Dep	1.40					
Duration of service (yrs)	-0.012*	-0.016	-0.008	-0.008	-0.005	-0.002	-0.002
	(0.007)	(0.011)	(0.008)	(0.008)	(0.010)	(0.010)	(0.010)
N	2,067	2,067	2,067	2,067	2,067	2,067	2,067
Ever Had a Paid Job	Mean Dep	0.45					
Duration	-0.007	-0.018	-0.048	-0.061	-0.065	-0.048	-0.048
	(0.032)	(0.051)	(0.048)	(0.047)	(0.049)	(0.059)	(0.059)
Observations	1,104	1,104	1,104	1,104	1,104	1,104	1,104

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Dependent variables are pre-program characteristics of individuals. Each column's specification corresponds to column specifications in Table 5. Regressions do not include imputed values.

Appendix Table 13: Balance Test of Baseline Characteristics for Job Corps Applicants

	Full sample				Males only			
Characteristic	Treatment	Control	Difference		Treatment	Control	Difference	
Male	0.591	0.599	-0.008	(0.009)				
Age	18.861	18.826	0.035	(0.038)	18.735	18.717	0.018	(0.047)
White - Non-Hispanic	0.274	0.265	0.009	(0.008)	0.309	0.295	0.014	(0.01)
Black - Non-Hispanic	0.476	0.478	-0.002	(0.009)	0.45	0.452	-0.002	(0.011)
Hispanic	0.174	0.181	-0.007	(0.007)	0.163	0.178	-0.015*	(0.008)
Non-English Native Language	0.141	0.143	-0.001	(0.006)	0.14	0.144	-0.004	(0.008)
Has Child	0.181	0.179	0.002	(0.007)	0.106	0.108	-0.002	(0.007)
Childhood Household Head - Mother	0.483	0.49	-0.007	(0.009)	0.45	0.467	-0.016	(0.011)
Highest Grade Completed - Mother	11.516	11.539	-0.022	(0.051)	11.678	11.658	0.02	(0.062)
Highest Grade Completed - Father	11.471	11.578	-0.107	(0.064)	11.605	11.608	-0.003	(0.079)
Never on Welfare During Childhood	0.47	0.459	0.012	(0.009)	0.489	0.485	0.004	(0.012)
Highest Grade Completed	10.069	10.081	-0.012	(0.027)	9.953	9.969	-0.016	(0.032)
High School Degree	0.178	0.182	-0.004	(0.007)	0.139	0.142	-0.003	(0.008)
GED	0.047	0.055	-0.008*	(0.004)	0.05	0.052	-0.001	(0.005)
Ever Worked	0.8	0.788	0.011	(0.007)	0.812	0.801	0.011	(0.009)
Worked in Past Year	0.649	0.64	0.009	(0.008)	0.666	0.655	0.012	(0.01)
Currently has Job	0.215	0.208	0.007	(0.007)	0.221	0.204	0.017*	(0.009)
Months Worked in Past Year	6.055	6.127	-0.072	(0.092)	6.028	6.067	-0.039	(0.113)
Earnings in Past Year (if employed during past year	3019.377	2903.822	115.556	(103.731)	3319.099	3156.064	163.035	(137.756)
Typical Hours Worked (if employed during past yea	35.635	35.344	0.291	(0.348)	36.922	36.73	0.192	(0.44)
Typical Wage (if employed during past year)	5.062	5.078	-0.017	(0.033)	5.167	5.194	-0.027	(0.042)
Received AFDC	0.316	0.316	-0.001	(0.009)	0.244	0.242	0.002	(0.01)
Received Food Stamps	0.437	0.446	-0.009	(0.009)	0.37	0.378	-0.008	(0.011)
Received Any Welfare	0.578	0.585	-0.007	(0.009)	0.511	0.518	-0.007	(0.012)
Ever Used Drugs	0.386	0.376	0.01	(0.009)	0.43	0.423	0.007	(0.011)
Ever Arrested	0.264	0.266	-0.001	(0.008)	0.337	0.326	0.011	(0.01)
Non-residential Job Corps Participant	0.137	0.141	-0.004	(0.006)	0.067	0.072	-0.005	(0.005)
Obs	8813	5514	14327		5036	3610	8646	

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Data source is baseline data for Job Corps program from Schochet et al. (2008). If employed during past year is measured as the individual worked for at least 2 weeks in the previous year.

Appendix Table 14: Control Function Approach 1 (Assuming Treatment Effect is Same)

	(1)	(2)	(3)	(4)	(5)
Dependent Variable:	*Log Death Age*	*PIA*	*AIME*	*Age*	*SSDI*
Panel A: Using Education Only					
OLS	0.013***	13.068**	47.882**	0.509***	-0.02
	(0.004)	(5.853)	(21.770)	(0.189)	(0.014)
Ctrl Common	0.013***	14.314**	52.363**	0.536***	-0.022
	(0.005)	(5.873)	(21.843)	(0.190)	(0.014)
Ctrl All	0.013***	14.244**	52.109**	0.535***	-0.022
	(0.005)	(5.871)	(21.836)	(0.189)	(0.014)
N	7,722	4,613	4,613	4,575	4,575
Panel B: Using Moved Only					
OLS	0.013***	13.102**	48.286**	0.507***	-0.018
	(0.004)	(5.857)	(21.781)	(0.189)	(0.014)
Ctrl Common	0.013***	13.142**	48.421**	0.507***	-0.018
	(0.004)	(5.859)	(21.786)	(0.189)	(0.014)
Ctrl All	0.013***	13.107**	48.303**	0.507***	-0.018
	(0.004)	(5.858)	(21.783)	0.189	(0.014)
N	7,703	4,600	4,600	4,562	4,562
Panel C: Using Others Only					
OLS	0.024	-9.955	-17.604	1.057	-0.112*
	(0.015)	(18.872)	(78.572)	(0.646)	(0.058)
Ctrl Common	0.025*	-7.344	-8.394	1.133*	-0.112*
	(0.015)	(19.084)	(79.492)	-0.649	(0.059)
Ctrl All	0.025	-7.244	-8.041	1.127*	-0.111*
	(0.015)	(19.091)	(79.524)	(0.650)	(0.059)
N	1,382	621	621	621	621
Panel D: All Control Functions					
OLS	0.025	-10.07	-18.237	1.072	-0.112*
	(0.015)	(18.767)	(78.103)	(0.650)	(0.059)
Ctrl Common	0.025*	-9.707	-18.255	1.087*	-0.109*
	(0.015)	(18.857)	(78.404)	(0.651)	(0.059)
Ctrl All	0.025*	-9.516	-17.625	1.090*	-0.108*
	(0.015)	(18.857)	(78.403)	(0.651)	(0.059)
N	1,362	611	611	611	611

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Panel A-Panel B uses different includes different sets of control functions. In Panel A, we only include control function generated using education as the short-run outcome. In Panel B, the short-run outcome is short-run mobility from the 1940 Census and WWII rolls. Panel C uses whether working, weeks worked, log wage as short-run outcomes. Panel D includes all control functions in Panel A to Panel C simultaenously. Three results for each of these samples are presented. OLS row presents the OLS estimate without including the control functions on the sample of observations where each control function can be calculated. Ctrl Common row presents the results with control functions using only common covariates between JC and CCC (enrollment age, age less than 18 indicator, highest grade level, hispanic status, whether helpd a previous job, whether graduated high school, household size, from rural hosehold, whether father is living, whether mother is living). Ctrl All row presents the results with control functions using common covariates as well as other variables included in the full specifications corresponding to Table 2 Column 6.

Appendix Table 15: Control Function Approach 2 (Assuming Selection Bias is Same)

	(1)	(2)	(3)	(4)	(5)
Dependent Variable:	*Log Death Age*	*PIA*	*AIME*	*Retirement Age*	*SSDI*
Panel A: Using Education Only					
OLS	0.013***	13.068**	47.882**	0.509***	-0.02
	(0.004)	(5.853)	(21.770)	(0.189)	(0.014)
Ctrl Common	0.013***	12.764**	46.809**	0.502***	-0.019
	(0.004)	(5.851)	(21.763)	(0.189)	(0.014)
Ctrl All	0.013***	13.180**	48.291**	0.512***	-0.02
	(0.004)	(5.851)	(21.760)	(0.189)	(0.014)
N	7,722	4,613	4,613	4,575	4,575
Panel B: Using Moved Only					
OLS	0.013***	13.102**	48.286**	0.507***	-0.018
	(0.004)	(5.857)	(21.781)	(0.189)	(0.014)
Ctrl Common	0.013***	13.171**	48.517**	0.508***	-0.018
	(0.004)	(5.861)	(21.793)	(0.189)	(0.014)
Ctrl All	0.013***	13.155**	48.421**	0.510***	-0.018
	(0.004)	(5.862)	(21.799)	(0.189)	(0.014)
N	7,703	4,600	4,600	4,562	4,562
Panel C: Using Others Only					
OLS	0.024	-9.955	-17.604	1.057	-0.112*
	(0.015)	(18.872)	(78.572)	(0.646)	(0.058)
Ctrl Common	0.024	-7.683	-9.744	1.070*	-0.107*
	(0.015)	(19.041)	(79.311)	(0.648)	(0.059)
Ctrl All	0.023	-9.207	-15.239	1.018	-0.108*
	(0.015)	(18.977)	(79.018)	(0.647)	(0.059)
N	1,382	621	621	621	621
Panel D: All Control Functions					
OLS	0.025	-10.07	-18.237	1.072	-0.112*
	(0.015)	(18.767)	(78.103)	(0.650)	(0.059)
Ctrl Common	0.025*	-5.819	-0.582	1.153*	-0.106*
	(0.015)	(18.850)	(78.348)	(0.651)	(0.059)
Ctrl All	0.024	-9.443	-15.682	1.038	-0.107*
	(0.015)	(18.769)	(77.989)	(0.647)	(0.059)
N	1,362	611	611	611	611

Notes: Robust standard errors in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Panel A-Panel B uses different includes different sets of control functions. In Panel A, we only include control function generated using education as the short-run outcome. In Panel B, the short-run outcome is short-run mobility from the 1940 Census and WWII rolls. Panel C uses whether working, weeks worked, log wage as short-run outcomes. Panel D includes all control functions in Panel A to Panel C simultaenously. Three results for each of these samples are presented. OLS row presents the OLS estimate without including the control functions on the sample of observations where each control function can be calculated. Ctrl Common row presents the results with control functions using only common covariates between JC and CCC (enrollment age, age less than 18 indicator, highest grade level, hispanic status, whether helpd a previous job, whether graduated high school, household size, from rural hosehold, whether father is living, whether mother is living). Ctrl All row presents the results with control functions using common covariates as well as other variables

Data Appendix

CCC Data

Contents

1. Overview of data sources

Data used for CCC is assembled from various sources. The major sources of data are:

1) **Archival documents** that include application and discharge forms newly digitized by us and various information about CCC camps of primarily New Mexico and Colorado.
2) **FamilySearch / Ancestry.com** data that links the individuals found in the archival files to various historical sources available online from familysearch.com and ancestry.com, assembled by the BYU Record Linking Lab
3) **Social Security Administration Death Master File** data where we use the SSN, death date, and birth dates found in (2) to link people to correct identifiers

These sources are combined to create the final record-level data. Because some records in the archive belong to the same individual, the record-level data contain more observations than the number of individuals. We tag records so that records belonging to the same individual are assigned the same *PersonID*. We detail the procedure in Section 2.

We use the person-level data and add in additional sources of data to complete the Analysis Sample. The records we link are:

1) **1940 Census** that we machine-match for demographic and family characteristic variables
2) **WWII enlistment records** that we machine-match for demographic variables

The individuals in the Analysis Sample are uniquely identified by variables *state* (of enrollment) and *PersonID*. This is the final dataset used for analysis.

More details on each section:

A. CCC Archives

Colorado (CO)
The Colorado data is from transcriptions of following records: (i) *Certificate of Selection for the Civilian Conservation Corps*, (ii) *Application for the Enrollment*, (iii) *Discharge Form* (Unofficial name). The records are found in the Colorado State Archive under the title "Civilian Conservation Corps Enrollments (Statewide) 1936-1942."

New Mexico (NM)
The New Mexico data is from transcriptions of *Civilian Conservation Corps, New Mexico District records.* (Citation number: collection 1959-030)

B. CCC Camps

The opening and closing dates of CO camps come from Robert W. Audretsch, who supplied us with a list of camps, their associated companies, and the beginning and start dates of the company numbers within the camps.

The CO camp location comes from various historical records that we hand-coded.

The camp type code information comes from http://www.ccclegacy.org/CCC_Camp_Lists.html.

C. Colorado Name Index

Colorado Name Index contains information on a subset of enrollees and their camp assignment that was retrieved from searching through mentions of enrollees' names in contemporary local newspaper articles. Local newspapers often announced young men in their area who enrolled in the CCC and contained basic information about their enrollment. We have used this information to impute camp numbers in cases we were missing them. The procedure is detailed in Section 4.

The *Colorado Name Index* is from the following book:

A Colorado Civilian Conservation Corps Enrollee Name Index
by Robert W. Audretsch
Publisher: CreateSpace Independent Publishing Platform; 1 edition (April 5, 2017)
ISBN-10: 1545102910
ISBN-13: 978-1545102916
Amazon link: https://www.amazon.com/Colorado-Civilian-Conservation-Corps-Enrollee/dp/1545102910

D. FamilySearch (BYU Record Linking Lab)

After the records from the state archives were transcribed and cleaned, individuals in the data were sent to the BYU Record Linking Lab to be found in various historical genealogy websites including Ancestry.com and FamilySearch.org. Their date of death and social security numbers were collected. The individuals' names, date of birth, place of birth, allottee (usually a family member) names were used to find these individuals. The match is performed by trained historians, using records from multiple data sources and information from CCC.

The BYU Record Linking Lab found two major variables:

i. SSN

Social security numbers were mostly found on Ancestry.com. The sources of the SSNs on Ancestry are:

1) Ancestry.com. *U.S., Social Security Death Index, 1935-2014* [database on-line]. Provo, UT, USA: Ancestry.com Operations Inc, 2011. Original data: Social Security Administration. *Social Security Death Index, Master File*. Social Security Administration.

2) Ancestry.com. *U.S., Social Security Applications and Claims Index, 1936-2007* [database on-line]. Provo, UT, USA: Ancestry.com Operations, Inc., 2015. Original data: Social Security Applications and Claims, 1936-2007.

Note: SSN is only available for those who have been dead for 10 years. Therefore, we cannot find SSN for those who died before 2005/2006.

For reference, see:
SSDI: http://search.ancestry.com/search/db.aspx?dbid=3693
SSACI: http://search.ancestry.com/search/db.aspx?dbid=60901

ii. Death Dates

Death dates were found using various sources including the aforementioned social security administration data, Find A Grave Index, and other sources.

1) Ancestry.com. *U.S., Social Security Death Index, 1935-2014* [database on-line]. Provo, UT, USA: Ancestry.com Operations Inc, 2011. Original data: Social Security Administration. *Social Security Death Index, Master File*. Social Security Administration.

2) Ancestry.com. *U.S., Find A Grave Index, 1600s-Current* [database on-line]. Provo, UT, USA: Ancestry.com Operations, Inc., 2012. Original data: *Find A Grave*. Find A Grave. http://www.findagrave.com/cgi-bin/fg.cgi.

E. Social Security Administration

Finally, we get information on individual's Average Indexed Monthly Earnings (AIME), retirement age, and SSDI claiming behavior by matching our individuals to Social Security Administration's (SSA) Master Beneficiary Record File (MBR).

In order to find our individuals in SSA's MBR, we need the individuals' SSN, first and last names. As described above, for some of our individuals, we have SSN information directly found by BYU Record Linking Lab from various historical sources. For others whose SSNs were not found by the Lab, we use the combination of date of death, date of birth, place of death, first and last names to locate them on the Social Security Death Master File to retrieve their SSNs. The combination of the SSNs and first and last names were used to match these individuals to the SSA's MBR.

2. Assignment of individual ids for multiple records

Individuals can generate multiple records in the CCC record-keeping system. For example, a person who enrolled twice could generate two records: one enrollment form for each time he enrolled. Because our raw data consists of records of enrollment and discharge, our raw data is in the record-level, not in the individual-level. We convert the record-level raw data into an individual-level data by using the information in the records to assign records to unique individuals.

We use the following information in each record to determine whether records belong to the same individual: enrollee's first and last names, birth dates, CCC serial number, social security number (if available in the original records for CO), allottee's first and last names, and allottee's relation to the participant. All of these fields in each record are subject to transcription and record-keeping errors. In addition, SSN data is only sparsely available for CO enrollees. Therefore, we first use a "fuzzy" matching algorithm for each record to group records with similar field values. Then, we verify the matches manually. Additional information from the BYU Record Linking Lab allowed them to tag more records as coming from the same individuals.

Records vs Individuals Statistics

	CO	NM
Number of Records	21,538	10,713
Number of Individuals	18,644	9,699
Number of Individuals with…		
- 1 record	16,082	8,746
- 2 records	2,263	894
- 3 records	269	57
- 4 records	27	2
- 5 records	3	0

3. CCC Regulations

The rules and regulations regarding the operation of CCC camps as well as allotment of funds to CCC employees changed from the program's inception in 1933 to its closure in 1942. Below is a compilation of CCC regulations that are pertinent to our research.

2. Each employee of the CCC was given a serial number which was composed in the following:
   a. Serial numbers started with the letters "CC" to denote the Civilian Conservation Corps as opposed to other emergency relief programs. The letters "CC" were followed by the number of the area corps number. In the case of Colorado and New Mexico, the area number was 8. See the map below (source: National Parks Service. https://www.nps.gov/parkhistory/online_books/ccc/ccc/chap2.htm).

   b. Serial numbers then contain information on the company number. "In order that the numerical designation of the company may indicate its origin by corps area, blocks of numbers are assigned in accordance with the following system: 100-199 to First Corps Area, 201-299 to Second Corps Area, 901-999 to Ninth Corps Area.. When this series becomes exhausted, 1,000 will be added to each block of numbers; e.g. 1101-1199 to First Corps Area, 1201-1299 to Second Corps Area, and so on" (quote found here: https://babel.hathitrust.org/cgi/pt?id=mdp.39015020215433;view=1up;seq=17).

3. Allocation of funds received by CCC employees:
    a. Our data show that there was variation in the amounts received by CCC employees. This is consistent with the regulations found here: https://babel.hathitrust.org/cgi/pt?id=mdp.39015020215433;view=1up;seq=25 In particular, enrollees without special status (such as leaders or assistant leaders) were paid $30 per month. Of the $30 received, enrollees were required to pay at least $22 to their families.
4. Enrollment over time:
    a. CCC enrollment: Our data mostly contain information for those who enrolled after 1937. The most likely reason for this is that the CCC changed from being a program that was part of the Emergency Conservation Work program to its own entity known as the Civilian Conservation Corps in 1937. See quote here: "There are hereby transferred to the Corps all enrolled personnel, records, papers, property, funds, and obligations of the Emergency Conservation Work established under the Act of March 31, 1933 (48 Stat. 22), as amended; and the Corps shall take over the institution of the camp exchange heretofore established and maintained, under supervision of the War Department, in connection with and aiding in administration of Civilian Conservation Corps work camps conducted under the authority of said Act as amended: Provided, That such camp exchange shall not sell to persons not connected with the operation of the Civilian Conservation Corps" (source here: https://www.nps.gov/parkhistory/online_books/ccc/cccaa.htm)

4. Imputing camp numbers for CO data

We have used various sources to impute camp numbers for individuals that do not have camp information in the CO data.

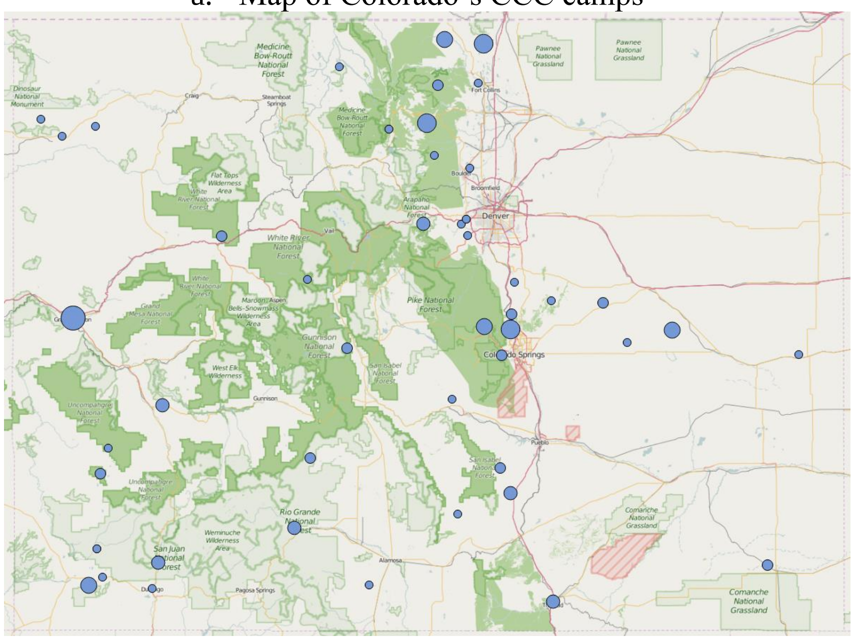
1) **Company Numbers**: For some enrollees, we have company numbers but not camp numbers. The correspondence between company and camp numbers were obtained from Robert W. Audretsch, who documented the company number assigned to specific camps over time.
2) **CCC Serial Numbers**: Each enrollee was assigned a serial number when they first enrolled. The serial number contains the area of enrollment (as described in Section 2) and the company number they were assigned to. The company numbers were then used to impute the camp of assignment.
3) **Colorado Name Index:** For enrollees with enrollment date information but no camp information (either directly from the records or that could be imputed from the serial numbers), we supplemented the camp information through the *Colorado Name Index*. As described in Section 1, the Index contains information from local newspapers on

enrollees and their camps at a point in time (when the article was published). We used enrollees' first and last names, place of birth or place of enrollment application, and their enrollment and discharge date to manually match the enrollee to a newspaper record in the Index. Then, we assigned the camp information from the Index as the enrollee's first camp of assignment.

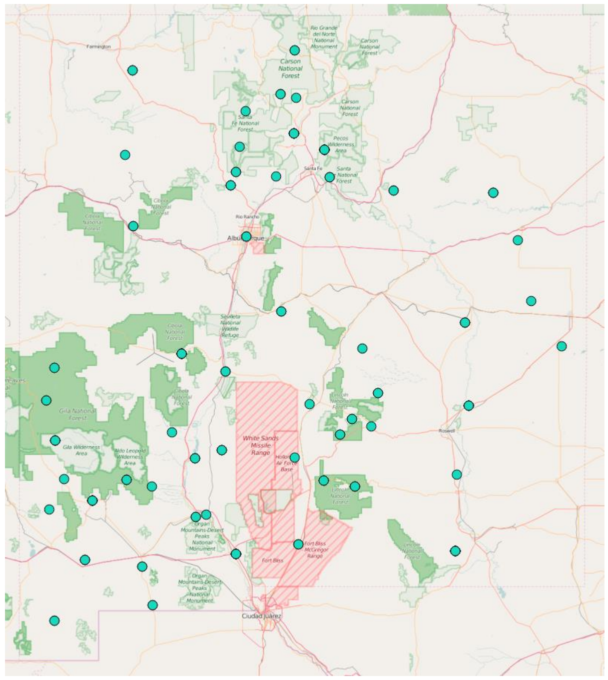
5. Construction of camp location and characteristics from historical records (incomplete)

Camp ID in administrative records merged with camp-information from multiple sources. Dates of operation of camp were obtained from Robert W. Audretsch. Camp location was approximated by location descriptions in historical documents.

a. Map of Colorado's CCC camps



b. Map of New Mexico's CCC camps



Distance to closest town was computed taking the list of Colorado and New Mexico towns and their latitude and longitude from United States Geological Survey's Geographic Names Information System (USGS GNIS). Pairwise distances from each camp to each city was calculated, then for each camp, the town with the smallest distance value was selected as the distance to closest town.

Camp weather information was obtained from historical weather data at the PRISM Climate Group at Oregon State University. The data contains minimum and maximum temperature and precipitation at the monthly level and covers the entire United States from 1985-1980 at the spatial scale of 4km x 4km. It is a climatologically aided interpolation and takes as first guess the long-term averages in the area. For more information, visit the PRISM website at http://prism.oregonstate.edu/historical/. We obtain the historical monthly weather data for each camp from the GIS raster files using camp location (longitude and latitude).

Camp peer characteristics are computed using information of individuals at each point in time in our dataset. The peer characteristics for enrollee i is the weighted average of demographic characteristics of other enrollees in our data who served in the same camp overlapped in service duration with i, where the days of overlap are used as the weights. Thus, enrollees that overlapped for a longer period of time get higher weights in the peer characteristics calculation.

In other words, the peer characteristics PX_i of enrollee i is calculated by,

$$PX_i = \sum_{j \in K_i} \frac{d_{ij}}{\sum_{j \in K_i} d_{ij}} x_j$$

where K_i is the set of enrollees that overlap with i, d_{ij} is the days of overlap between i and j, x_j is the demographic characteristic of j.

6. Matching Individuals to 1940 Census and WWII Enlistment Records

This appendix overviews the matching approach used to match CCC participants to Census and WW2 Army enlistment records. We rely on the Expectation Maximization approach to match records. Overall, the match rates are consistent with standard literature and the matches seem consistent. There seems to be some selection in terms of who is matched.

A. Introduction to Matching Approach

The matching approach follows "Linking Individuals Across Historical Sources: a Fully Automated Approach" by Ran Abramitzky, Roy Mill, and Santiago Perez (2018).[1] Any matching approach has to balance three competing goals:

1. Minimize false negatives (Type II errors)
2. Minimize false positives (Type I errors)
3. Create a representative sample

Ideally records would be identified by a unique administrative identifier that is stable across datasets (e.g., social security number). In most historical cases, we are forced to rely on a combination of less definitive information, such as year of birth, name, place of birth, and place of residence to match records. Therefore, choosing how to match on these characteristics is a major decision. There are three major sources of variation in variables across records for a given individual. First, the respondent introduces variation. They could state the wrong age or change their name (e.g., ``Nick" instead of ``Nicholas"). This issue is especially prevalent in historical Censuses due to lower literacy and education levels. Secondly, the interviewer can make transcription errors (e.g., write the name as ``Brian" instead of ``Ryan"). Finally, additional errors are introduced during the digitization of physical Census rolls.

We choose to rely on the Expectation Maximization (EM) approach outlined in Abramitzky et al. (2018). Individuals are matched to 1940 Census and WWII enlistment records primarily using automated methods. One alternative approach would be to rely on exact matches but relying solely on exact matches would significantly lower match rates and increase Type II errors. There are significant transcription errors in these records and the EM approach allows some flexibility when dealing with errors.

The EM approach falls under the umbrella of automated methods. The advantages of automated methods include the fact that they are reproducible, rule-based, can compare all records, and are cheaper. The disadvantages are that they do not have the same contextual information that humans do (e.g., "Bill" is short for "William") and humans are better able to incorporate additional information in a flexible manner.

Bailey et al. (2018) raised substantive concerns about using automated methods as opposed to linking by hand. They find that automated linking algorithms produce high rates of incorrect matches ranging from 13 to 69 percent when assuming hand-linked sample is the ground truth.

[1] Please see this article for a more detailed description of the approach

Match rates are especially poor when automated methods are combined with phonetic name cleaning. They tested three automated methods, Ferrie (1996), iterative method of Abramitzky et al. (2012 and 2014), and the regression prediction approach of Feigenbaum (2016), though not the EM approach. These results are an issue because poor matches can significantly attenuate estimates.

Abramitzky et al. (2018) find much better results for more modern automated methods, such as the EM approach, than the approaches tested in Bailey et al. (2018). Additionally, they find that automated methods perform similarly to hand-linking methods when the same information is used. Conservative EM methods tested by Abramitzky et al. had <10% false match rate, which was lower than hand-linking methods with the same information, though hand-linking methods also made significantly more matches. Moreover, when both methods made a match then there was greater than 90% agreement.

In order to address concerns of false matches we rely on conservative matching criteria and do not conduct phonetic cleaning or significant name standardization. Finally, we validate a subset of matches against hand matches provided by Family Search (FS).

B. Overview of Matching Procedure

There are several decisions to make before beginning any estimation. The first decision is which variables to match on. The standard approach is to match on pre-determined characteristics. Typically, this means birth year, place of birth, first name and last name.

The second decision is which variables to block on. The approach will only compute distance between individuals who are exact matches on certain characteristics. Fundamentally, blocking is used to reduce computational complexity by avoiding computing distances between every potential pair of individuals. For example, it is common to block on the first letter of the first name.

The third decision is how to measure string distance. Some approaches effectively use an indicator for whether names are an exact match or they combine this approach with a phonetic cleaning algorithm, such as the NYSIIS. Phonetic cleaning is especially useful if most errors are due to translating a heard name to a written one. Continuous string distance measures can also be used and are most useful when errors are due to transcription mistakes during digitization.

Now, we present the basic concept behind the Expectation Maximization algorithm (Dempster, Laird, and Rubin 1977; Winkler 1989). For any observation, there are many match candidate pairs, i. For each candidate pair we observe distances γ_i. Assume each of these candidates are drawn from one of two distributions. Each candidate pair has two associated probabilities: one for true matches, $P(\gamma_i|Match_i)$, and one for false matches, $P(\gamma_i|NotMatch_i)$. Using Bayes Rule, the probability that our candidate pair i, with distance γ_i, is a true match is given by:

$$P(Match|\gamma_i) = \frac{P(\gamma_i|Match_i)}{P(\gamma_i|Match_i)p_{m+}\,P(\gamma_i|NotMatch_i)(1-p_m)}$$

Using these expressions, we can take the following approach to estimate match probabilities:

1. Define distribution families for each of the distance variables to get $P(\gamma_i|Match_i)$ and $P(\gamma_i|NotMatch_i)$. Assume distances for each variable are independently distributed conditional on match status
2. Guess initial parameter values $\theta_m^{(t)}, \theta_{nm}^{(t)}$ for each distribution and the probability of a true match, $p_m^{(t)}$
3. Loop over the following two steps until convergence:
    A. Calculate for each pair the probability of a match, $w_i^{(t)} = P(Match|\gamma_i)$ for a given $(\theta_m^{(t)}, \theta_{nm}^{(t)}, p_m^{(t)})$
    B. Get updated parameter estimates $(\hat{\theta}_m^{(t+1)}, \hat{\theta}_{nm}^{(t+1)}, \hat{p}_m^{(t+1)})$ by maximizing:

$$\log L(\gamma, \theta, p_m) = \sum_{i=1}^{n} w_i^{(t)} \log p_m P(\gamma_i|\theta_m) + \left(1 - w_i^{(t)}\right) \log(1 - p_m) P(\gamma_i|\theta_{nm})$$

Once we have the converged estimates then we can compute $P(\gamma_i|Match_i)$ for any candidate pair. The final major choice is choosing what qualifies as a match. There are two components to this decision:

1. The minimum threshold in order to qualify as a match
2. The maximum threshold for the second closest match

(1) means that if there are no ``good" matches then it is better not to declare any a match. (2) means that if there are at least two ``good" candidates then there is a high Type II error rate when selecting one over the other. For the primary analysis, we take a conservative approach, setting a high threshold for (1) and (2).[2]

C. Implementation

One significant issue is that New Mexico CCC records do not contain data on the birthplace of participants. When matching to the 1940 Census and WW2 records we rely on a two step procedure to create matches:

- **First stage:** Colorado and New Mexico CCC participants are matched to 1940 Census and WW2 enlistment records
    - *Blocking variables:* State of residence, first letter of first name and first letter of last name
    - *Matching variables:* Year of birth and name distances
- **Second stage:** Next, we remove matched individuals and for unmatched individuals in the Colorado CCC we conduct a second round of matching
    - *Blocking variables:* Place of birth, first letter of first name and last name
    - *Matching variables:* Year of birth and name distances

In the first stage we look only within the current state of residence (e.g., only look at residents of Colorado in the 1940 Census for CO CCC participants). In the second stage, we use the additional information on place of birth for CO CCC participants to search across the United States.

[2] The threshold for (1) is 0.8 and the minimum distance for the second best match (2) is 0.3

The primary concern with using the state of residence is that we will miss migrants. There are two reasons that this should not be a major issue in our case. First, the 1940 Census, most CCC enlistment, and most WW2 enlistment take place in a relatively short time frame. Secondly, we can check the number of migrants in the Family Search hand-links. For both CO (91.4%) and NM (96.8%) most of the CCC participants are still in the same state during the 1940 Census. For New Mexico it seems very reasonable to only look within the state. The percentage is somewhat lower for Colorado, which is why we conduct the second stage and also match on place of birth.

Next, we decide to use the Jaro-Winkler string distance (Jaro 1989, Winkler 2006). The Jaro-Winkler string distance calculates the number of transpositions required to match two strings, weighting errors in the early part of the string more heavily. The distance is measured from 0 (no matching characters) to 1 (exact match). We invert this scale so that 0 is exact match and 1 is no matching characters so our measure is increasing in distance. In our case the largest concern is transcription errors during digitization so it makes sense to use a string distance measure.

The next choice is the creation of distributions for distance variables. We follow Abramitizky et al. (2018) and specify multinomial distributions for year of birth and name distances. Year of birth distances are segmented into groupings of 0, 1, or 2 years distance.[3] Name distances are segmented based on Jaro-Winkler scores into groupings: [0,0.067], (0.067,0.120],(0.120,0.250],(0.250,1]. These groups run from closest to farthest distance.

We also add in the hand-matches from Family Search. If the Family Search matches conflict with the automated methods then we use the Family Search match. Finally, we also conduct a tie-breaking procedure using additional information in cases where the best match clears the minimum threshold but the second best match is too close. If the first best match passes the tiebreak criteria and second best match fails then we count it as a match. Middle initial is used as a tiebreaker in both stages, while place of birth is used as a tiebreaker in the first stage for Colorado. For example, if the CCC record has middle initial "F", the first best match also has the middle initial "F" but the second best match has the middle initial "M" then it is counted as a match.

D. Census Matching Results

Matching Appendix Table 7-1: Match rates between CCC records and 1940 Census

[3] Matches with larger distances are not considered

Census Match Rates by Type	CO	NM	Overall
EM and FS	0.08	0.06	0.07
Only EM	0.34	0.22	0.30
Only FS	0.05	0.09	0.07
No match	0.53	0.63	0.56
Observations	18644	9699	28343

Note: Values represent match rates as percentages of column totals. Match rates are for CCC participants to 1940 Census. EM stands for Expectation Maximization approach, FS stands for hand matches by FamilySearch team

False Negatives (Type II error): Matching Appendix Table 7-1 shows that 44% of CCC participants have been matched to 1940 Census records. 30% of participants have been matched through EM only, 7% through FS only, and 7% through both methodologies. This match rate for the EM approach is in line with the literature. Additionally, there is an upper bound on potential matches. In order to find this upper bound for matches to the 1940 Census, Abramitzky et al. (2018) linked a copy of the 1940 Census digitized by Family Search and one digitized by Ancestry.com. Even in this case they can only link up to 67% of the Census due to individuals with similar attributes and ``brutally bad transcriptions" due to difficulties reading cursive.

Matching Appendix Table 7-2: Match consistency between EM and FS for CCC-1940 Census matches

Census Match Consistency	CO	NM	Overall
% of participants matched by both EM and FS	0.08	0.06	0.07
% of overlap matched to same individual	0.95	0.91	0.94
Observations	18644	9699	28343

Note: Values represent percentages of column totals. Match rates are for CCC participants to 1940 Census. EM stands for Expectation Maximization approach, FS stands for hand matches by Family-Search team. Consistent FS and EM match measures whether EM and FS approaches matched CCC participant to the same Census individual in cases when both approaches make a match

False Positives (Type I error): While we do not have an absolute ``ground truth" sample, one way of examining Type I errors is to see if the EM and FS approaches match the same individual when they overlap. As seen in Matching Appendix Table 7-2, there is a high degree of consistency when both methods made a match - 94% of the time they matched the same CCC participant to the same Census record. We can go a step further and examine the discrepancies to understand if there is a reason to prefer the EM approach or FS hand matches. We use additional information (e.g., county of residence) and classify the discrepancies. In about 1/3 of cases the EM match is preferred, in 1/3 of cases the FS match is preferred, and the remaining cases are indeterminate. Therefore, there does not seem to be a clear reason to prefer either method.

Representativeness: Finally, we check which individuals are matched by regressing an indicator of whether matched on CCC participant characteristics at the time of their first enrollment. If matches are at random then there should be no clear pattern.

Matching Appendix Table 7-3: Predictors of CCC-1940 Census matches by type of match for CO

	EM match	FS match	EM and FS	FS or EM match
Age at enrollment	0.00	-0.00**	-0.00	-0.00
	(1.00)	(0.01)	(0.19)	(0.29)
Age of death	0.00	-0.00*	-0.00**	0.00
	(0.15)	(0.04)	(0.01)	(0.13)
Enroll year	-0.00	-0.00	-0.00	-0.00
	(0.80)	(0.27)	(0.46)	(0.54)
Dist. to camp (mi)	-0.00***	-0.00***	-0.00***	-0.00***
	(0.00)	(0.00)	(0.00)	(0.00)
Born in CO	-0.01	0.00	-0.01	-0.00
	(0.14)	(1.00)	(0.09)	(0.60)
Height (in)	0.02*	0.01	0.01**	0.01*
	(0.01)	(0.08)	(0.01)	(0.03)
Weight (lb)	-0.00	-0.00	-0.00	-0.00
	(0.49)	(0.41)	(0.10)	(0.74)
BMI	0.01	0.01	0.01*	0.00
	(0.35)	(0.30)	(0.04)	(0.61)
Missing parent	-0.06***	-0.07***	-0.04***	-0.08***
	(0.00)	(0.00)	(0.00)	(0.00)
Farm	0.01	0.05***	0.03***	0.03
	(0.45)	(0.00)	(0.00)	(0.07)
Urban	-0.00	-0.02	-0.01	-0.01
	(0.76)	(0.18)	(0.15)	(0.65)
Years educ	0.01***	0.00*	0.00***	0.01***
	(0.00)	(0.02)	(0.00)	(0.00)
Unemployed	-0.02	-0.01	-0.02	-0.02
	(0.29)	(0.25)	(0.08)	(0.38)
Constant	1.17	5.37	2.47	4.07
	(0.87)	(0.29)	(0.54)	(0.58)
Observations	18644	18644	18644	18644
R^2	0.044	0.018	0.018	0.049

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: p-values in parentheses. Match rates are for CO CCC participants to 1940 Census. EM stands for Expectation Maximization approach, FS stands for hand matches by FamilySearch team

Matching Appendix Table 7-3 shows the results for Colorado CCC participants broken out by match type. In general, CCC participants who were matched seem slightly better off. For example, matched individuals have higher education levels, less likely to be missing parents, and are taller on average. These differences do not seem to be large in absolute magnitude though, so it seems as though the matches are reasonably well representative.

Matching Appendix Table 7-4: Predictors of CCC-1940 Census matches by type of match for NM

	EM match	FS match	EM and FS	FS or EM match
Age at enrollment	-0.00	-0.00**	-0.00*	-0.00**
	(0.07)	(0.01)	(0.05)	(0.01)
Age of death	0.00	-0.00	-0.00	0.00
	(0.74)	(0.15)	(0.05)	(0.86)
Enroll year	-0.00	0.00	-0.00	-0.00
	(0.42)	(0.92)	(0.66)	(0.65)
Dist. to camp (mi)	-0.00	0.00	0.00	-0.00
	(0.20)	(0.35)	(0.48)	(0.40)
Constant	4.98	-0.24	1.47	3.27
	(0.39)	(0.96)	(0.63)	(0.60)
Observations	9699	9699	9699	9699
R^2	0.019	0.003	0.003	0.016

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: p-values in parentheses. Match rates are for NM CCC participants to 1940 Census. EM stands for Expectation Maximization approach, FS stands for hand matches by FamilySearch team

Matching Appendix Table 7-4 shows the results for New Mexico matches. For New Mexico, we have significantly fewer indicators of participant characteristics; however, there again seems to not be large differences in terms of the type of individual matched.

E. WW2 Matching Results

Matching Appendix Table 7-5: Match rates between CCC records and WWII Enlistment records

WW2 Match Rates	CO	NM	Overall
EM match	0.31	0.24	0.29
EM match (Adj)	0.78	0.59	0.72
Observations	18644	9699	28343

Note: Values represent match rates percentages of column totals. Adjusted values are scaled by state-age cohort enlistment percentages

False Negatives (Type II error): Matching Appendix Table 7-5 shows that 29% of CCC participants are matched to WW2 army enlistment records. There are two primary reasons that this match rate is lower than the 1940 Census. First, there is no supplementary source of matches to augment the EM approach with (FS matches). Secondly, the Census has universal coverage while only a subset of men will be in the WW2 army enlistment records. We can compute an adjusted match rate by estimating the percentage of men in each state-year of birth cell that are in the records.[4] This procedure assumes CCC participants are no more likely to enlist than other of the same age in the same state. Based on these calculations we would expect 40% of the Colorado CCC participants and 41% of the New Mexico CCC participants to have be in the WW2 army enlistment records. The adjusted match rates (match percentage of those we expect to find) and is 78% for Colorado and 59% for New Mexico. Note that these adjusted match rates seem high but cannot account for whether CCC individuals were more likely to serve in the

[4] Using state-year of birth-years of education cells does not substantively alter the results

Army. For example, CCC camps typically involved significant Army administration which could increase the likelihood to serve due to familiarity with the military.

False Positives (Type I error): Without another source of matches for the WWII data it is difficult to conduct any sort of consistency analysis. Therefore, we rely on the findings of high consistency in the CCC to 1940 Census matches in order to support the EM approach in this case.

Representativeness: We repeat the regression of match status on characteristics, but the interpretation is slightly complicated in this case. There are two forms of selection: first, selection into who is drafted (and meets minimum standards) or enrolled in the Army, and secondly there is selection through who is matched.

Matching Appendix Table 7-6: Predictors of CCC-WWII enlistment matches by type of match for CO

	EM match
Age at enrollment	-0.01***
	(0.00)
Age of death	0.00
	(0.94)
Enroll year	0.01***
	(0.00)
Dist. to camp (mi)	-0.00*
	(0.01)
Born in CO	0.02**
	(0.00)
Height (in)	-0.01
	(0.15)
Weight (lb)	0.00
	(0.14)
BMI	-0.01
	(0.12)
Missing parent	-0.01
	(0.63)
Farm	0.02
	(0.22)
Urban	-0.02
	(0.19)
Years educ	0.01***
	(0.00)
Unemployed	0.01
	(0.68)
Constant	-26.86***
	(0.00)
Observations	18644
R^2	0.041

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: p-values in parentheses. Match rates are for CO CCC participants to WW2 enlistment records. EM stands for Expectation Maximization approach

Matching Appendix Table 7-6 shows the results for Colorado CCC participants. Matched individuals are again better educated, but most indicators are not statistically significant. Matching Appendix Table 7-7 shows the results for New Mexico CCC participants.

Matching Appendix Table 7-7: Predictors of CCC-WWII enlistment matches by type of match for CO

	EM match
Age at enrollment	0.00**
	(0.00)
Age of death	-0.00
	(0.79)
Enroll year	-0.02***
	(0.00)
Dist. to camp (mi)	-0.00
	(0.06)
Constant	40.74***
	(0.00)
Observations	9699
R^2	0.035

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: p-values in parentheses. Match rates are for NM CCC participants to WW2 enlistment records. EM stands for Expectation Maximization approach

7. Imputing Hispanic Origin

We follow the approach of Fryer and Levitt (2004) to construct a Hispanic name index for any first or last name using the 1940 Census. The name index is constructed using the Hispanic indicator variable in the 1940 Census. Each first and last name is given a value (0-1) based on:

$$HispIndex_i = \frac{\#\ of\ indiviuals\ with\ name\ who\ are\ Hispanic}{\#\ of\ individuals\ with\ name}$$

Individuals were not directly asked whether they are Hispanic during the Census until 1980 so an algorithm was used to classify individuals in prior Censuses retroactively. Eight rules were used, but at their most basic they are:

1. Individual or their parents/grandparents were born in a Hispanic area
2. Individual has a Spanish surname and was born in the US
3. Individual is a relative or spouse of someone who qualifies by (1) or (2)

Once the indexes are created, they are matched to CCC participants. There is an index for first name, last name, and a combined index created by combining them. Individuals above certain thresholds are classified as likely Hispanic.

8. Imputing Probability of Survival

Probability of survival of individuals are imputed in two ways. First, we can impute the probability as 0 for those with missing age of death (presuming that they are dead). Second, we can take a more sophisticated approach using the fact that the person was at least alive at the time of discharge and using the conditional probably of survival after having survived to age a_d at the time of discharge. This probability of survival uses information of survival probabilities from age a_d to a desired age threshold, e.g. $\bar{a}$ = 70. These rates can be obtained from the corresponding cohort life tables put out by the SSA (Bell and Miller 2005) for each enrollee's birth cohort, b.

We estimate survival models where we make various assumptions about the missing data. We concentrate on survival to age 70, which is slightly below the median age at death (73). Because the number 70 is a round multiple of ten, it avoids issues of age heaping. Appendix Table 4 shows the results. We start by estimating survival models using only the sample without missing data for reference (Panel A). Panel A shows the same basic patterns we found in Table 2: those who trained longer were more likely to survive and the estimates are very stable. In the last specification, the results imply that one more year of training increased the probability of survival to age 70 by about 4.6% relative to the mean. Panel B shows the results when we impute the probability of survival using life tables and information on the age at the time of training. Here, we find that the effect of training duration (once we add all controls) is somewhat lower (2.3. instead of 3 percentage points) but still statistically significant.

In Panel C, we impute all missing as zero (we assume that all the men for whom survival is missing died before age 70). The rationale for doing this is that the DMF and other sources of death tend to be complete starting in the 1970s (Hill and Rosenwaike, 2001). If most of the missing data is missing because of death certificates are not available to researchers (rather than due to errors in matching) then all the missing deaths occurred between the CCC training and 1970, much before our CCC men turned 70 (recall most of the men were born around 1920).

When we do this, we find that one more year of training is associated with about a 5% increase in survival relative to the mean.

9. Calculation of Marginal Value of Public Funds

We first calculate the cost of the program. The cost measure of MVPF incorporates both the direct cost to the government and various mitigation of cost. In particular, the CCC cost measure includes the following:

1. Upfront cost of the program
2. Increased social security payout from both the increase in pension amount and increase in longevity of enrollees
3. Cost mitigation from increased tax revenue from increased earnings of the enrollees
4. Cost mitigation from decrease in social security disability (SSDI) payout from decrease in claiming rate
5. Cost mitigation from decrease in social security payout from increase in retirement age
6. Cost mitigation from decrease in SSDI payout from increase in claiming age
7. Goods produced during the program, namely conservation work

We get information on (1) from Levine 2010, who estimated the annual cost per enrollee to be $1,004. Assuming the figure is in 1939 dollars, using Consumer Price Index All Urban Consumers (CPI-U) January-to-January growth, that amounts to $14,384.81 in 2017 dollars for our average enrollee who served around 0.8 years (9.6 months).

For (2), we use the mortality profile from our regression results illustrated in Figure 6. We assume that enrollees survive to age 45 with probability 1. For each age $x > 45$, we take the average survival rate to age x of our regression sample to be the baseline survival rate, and the estimate of the coefficient on duration to be the increase in survival rate for an enrollee that served one year. Multiplying the estimate by the average duration gives us the increase in the rate of survival for our average enrollee to age x, for each age x from 46-90. We assume after age 90, the survival rate declines to 0 evenly until age 95.

The average person in our sample receives the average PIA amount of $437.70 per month, assumed to be in 1982 dollars, as 1982 is the year on which our average enrollee turns 62 when SSA starts calculating PIA using AIME. Converting that to an annual benefit amount in 2017 dollars gives us an annual benefit of $13,525.85. We assume that 65 is the claiming age for social security benefit. Multiplying i) the PIA with ii) the probability of survival to age x for each $x \geq 65$, iii) by the discount factor, and finally iv) summing the yearly amounts gives us the present value of the baseline social security benefit.

The average enrollee receives an extra $14.11 of PIA (Appendix Table 7 Panel A, Column 6 multiplied by average duration), which is an annual increase of $436.05 in 2017 dollars. Taking into account the increased survival rates to age x, multiplying the baseline benefit by the total increased survival rate and by the discount rate gives the increase in the PV of benefits from increased rate of survival. Multiplying the total increased survival rate by the discount rate and the additional PIA amount gives the increase in the PV of benefits from increased PIA amount. Summing these two and subtracting it from the baseline PV of social security benefit gives us the final cost increase from increase in social security benefit over the lifetime. The final measure amounts to $2,514.17.

Calculating (3) is similar to the above, but instead of multiplying the PIA amount for ages above 65, we multiply AIME for ages 30-65. The average enrollee in our sample has an AIME of $963.62, and an increase of $44.10 (Table 2 Panel B Column 6 multiplied by average duration), both assumed to be in 1982 dollars. We impose an additional assumption that the earnings increase does not kick in until age 30—this is to incorporate our null result of service duration on short-term labor market outcomes as well as uniformly applying the earnings increase over the last 35 years of earnings, to mimic SSA's rule of using 35 years of highest earnings. This gives us the total PV of earnings and PV of earnings increase. We calculate the tax portion of this by assuming a tax rate of 33.6%, which is the CBO estimated average tax rate for FPL 100-149% provided in Appendix G of Hendren and Spruce-Keyser (2019). The final measure comes out to be $6,965.46 in 2017 dollars.

We calculate (4) by first calculating the baseline cost of SSDI. We assume that the average claiming age is 50 and the average SSDI amount is $1,171.80 in 2017 dollars, which is the average benefit in current payment status at the end of June 2017 from SSA's Selected Data From Social Security's Disability Program (https://www.ssa.gov/oact/STATS/dib-g3.html). Like how we calculated (2), we multiply i) this amount, ii) the average claiming rate of our sample (0.21), iii) the probability of survival at each age, iv) the discount factor, then v) sum the amounts over all years. This gives us the baseline value of SSDI claiming. We compare this with the change in SSDI amounts by taking into account the decrease of SSDI claiming probability for our average enrollee of about -0.017 (Table 2 Panel D Column 6 multiplied by average duration) as well as the increase in the probability of survival. Here we assume that the decrease in SSDI claiming probability applies uniformly across all ages. This gives us the final value of $910.61.

For (5) and (6), we use our average enrollee's increase in claiming age by 0.33 (Table 2 Panel C Column 6 multiplied by average duration). We assume that this age increase applies to retirement with probability (1 - 0.21) and to disability with probability 0.21, which is the average rate of disability claiming in our sample. For retirement, we multiply 0.33 by (1 - 0.21) and the amount the average enrollee would receive at age 65 calculated in (2), giving us $732.51. For disability, we multiply 0.33 by 0.21 and the amount average enrollee would receive at age 51 calculated in (4), which gives us $72.61.

We abstract from (7), as we have no good estimate of the total value of conservation work provided by the program. Thus, our estimate could be thought of as an upper bound of the cost.

Now, on to the WTP (or value) of the program. CCC provided the following short- and long-term benefits to enrollees:

1. Willingness to pay (WTP) for increase in longevity
2. Increase in earnings
3. Monthly real wage of $66.25 while enrolled, which includes the benefits enrollees received during the program (BLS 1941)
4. Decrease in benefit from loss of SSDI income as the enrollee claims at lower rate

Calculating (1) is again similar to the above cost calculation on increased social security payment. Instead of multiplying the PIA amount, we multiply the statistical value of life, assumed to be $150,000 in 2017 dollars (Lee et al 2009), for ages 45 to 95. We obtain an estimate of $25,456.40. For (2), since we already obtained the PV of earnings increase and the subsequent tax increase in calculating the cost, it is simply the after-tax portion of the PV of earnings calculated there. Therefore, we have $13,642.41 for the post-tax earnings benefit. (3) is straightforward, where we take the average amount enrollees received ($66.25 multiplied by average duration), which is $11,390.36 in 2017 dollars. For (4), this value is identical to what is calculated in (4) in the cost side, $910.61.

The final measures of cost and benefit are $8,217.78 and $49,578.56, respectively. Finally, MVPF is equal to the ratio of WTP to Cost, which is estimated to be 6.03. Without the WTP for increase in longevity, the MVPF comes out to be 2.52.

10. Special Acknowledgements

A. Dirk Van Hart provided us with dates for which New Mexico camps were open for enrollment.
B. Robert W. Audretsch provided us with dates for which Colorado camps were open for enrollment.

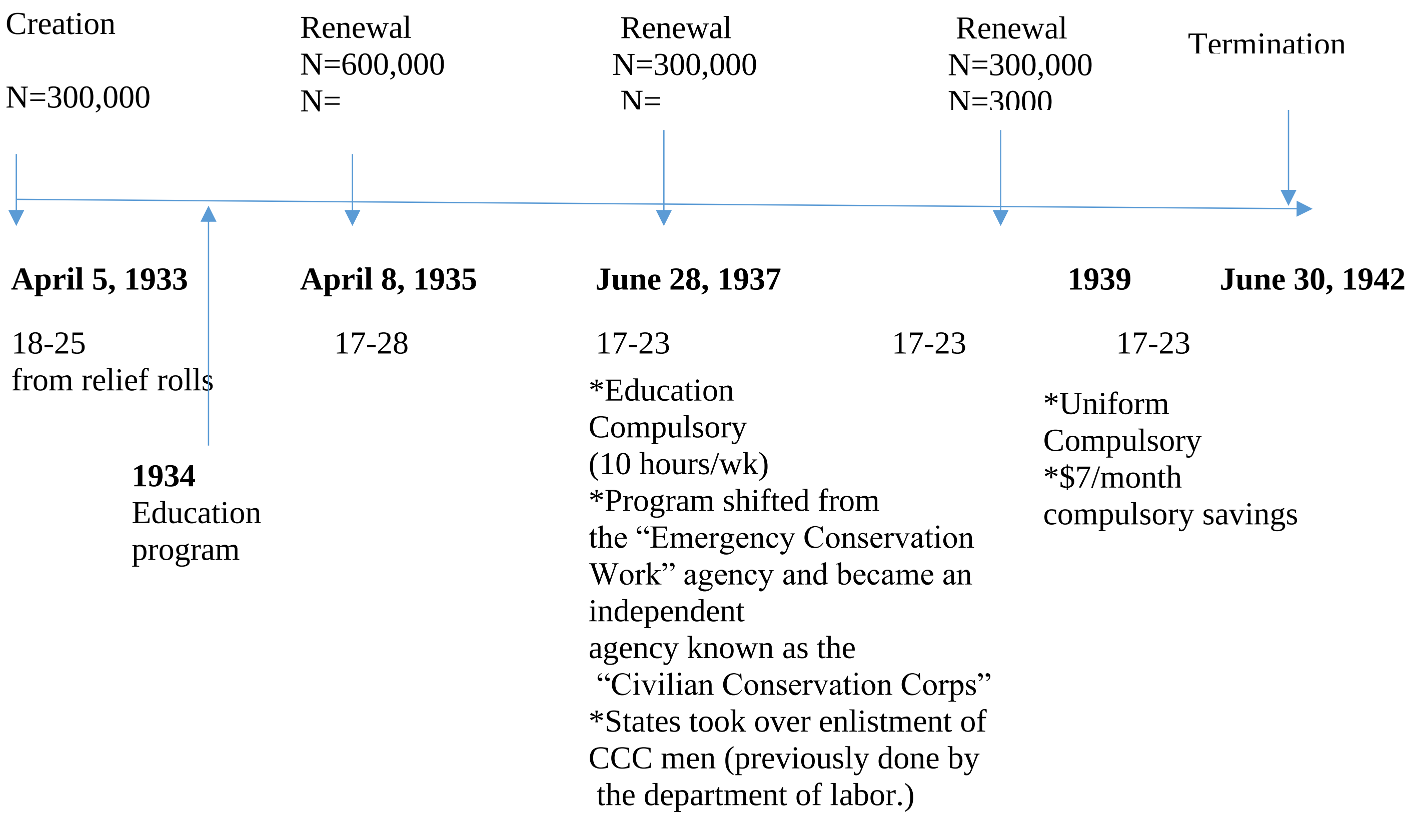
11. Additional Tables and Figures

Data Appendix Table 1: Comparison of Counties of Enrollees vs Whole State

Year	1930				1940			
State	CO		NM		CO		NM	
Geography	State	CCC	State	CCC	State	CCC	State	CCC
Variables								
Share Urban	0.50	0.40	0.25	0.22	0.53	0.42	0.33	0.28
Share in Farm	0.27	0.33	0.37	0.38	0.22	0.28	0.32	0.35
Share Owns Home	0.50	0.50	0.59	0.64	0.47	0.48	0.61	0.65
Mean Rent	38.88	37.60	26.39	23.09	102.99	95.43	219.27	271.40
Mean Age	29.57	28.35	25.26	25.24	31.40	30.12	26.14	25.84
Share Male	0.51	0.52	0.52	0.52	0.51	0.51	0.51	0.51
Share White	0.98	0.99	0.92	0.95	0.99	0.99	0.93	0.96
Share Mexican	0.06	0.07	0.14	0.11	0.07	0.13	0.34	0.44
Share Ever Married	0.51	0.49	0.45	0.44	0.54	0.52	0.47	0.45
Share Students	0.24	0.25	0.25	0.25	0.21	0.23	0.25	0.26
Share Foreign-born	0.10	0.09	0.06	0.05	0.07	0.06	0.03	0.02
Mean Occscore	21.78	20.59	19.05	18.34	22.54	21.38	20.10	19.19
Share Employed	0.90	0.90	0.93	0.92	0.90	0.89	0.88	0.85
Mean Income					392.11	332.25	326.73	277.49
Mean Educ Years					7.75	7.25	5.86	5.45
Share Hisp Origin					0.08	0.13	0.34	0.44

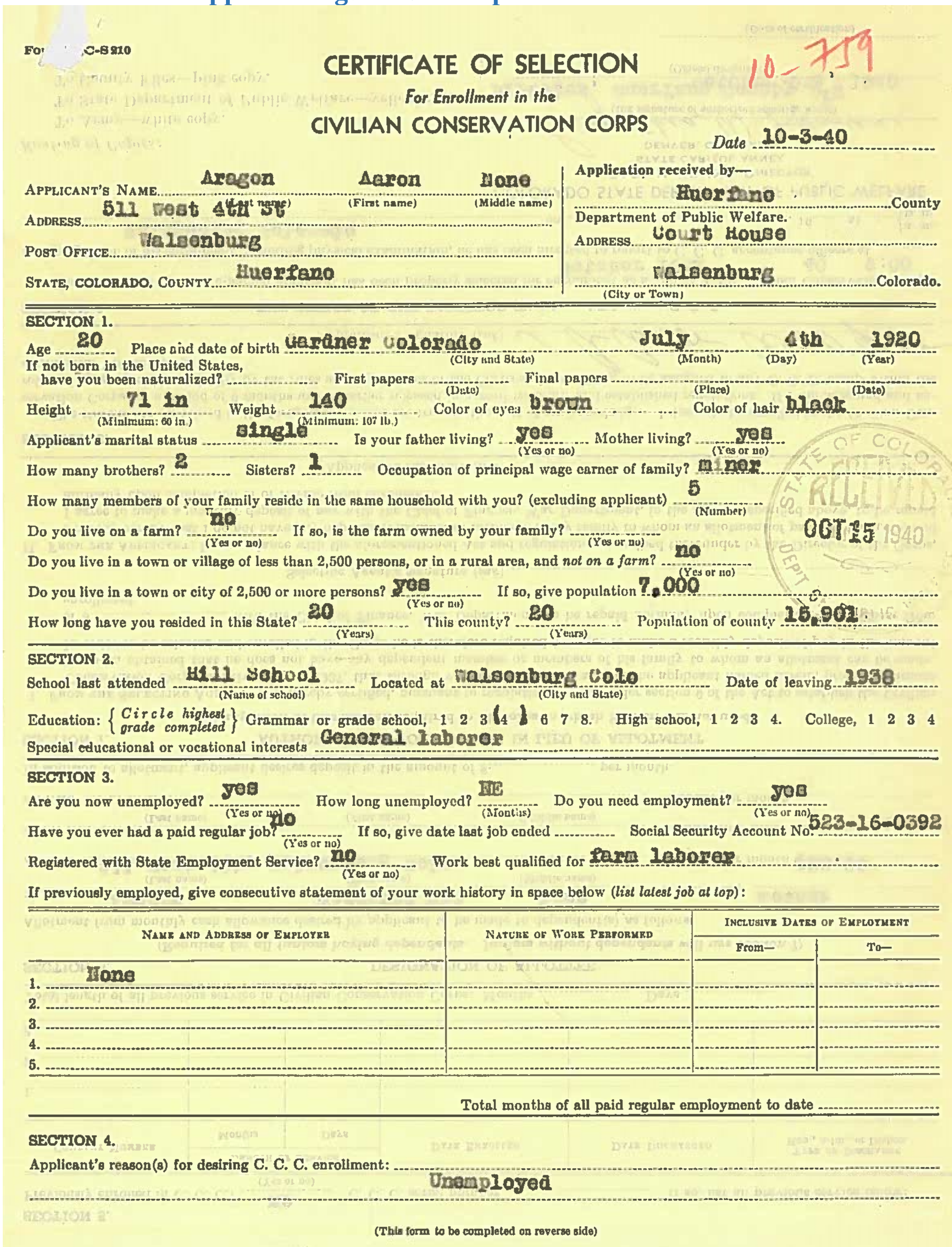
Note: Columns “State” are the state average of variables in each row. Columns CCC is the weighted average of county characteristics, where the weights are the share of CCC enrollees in our data enrolling from each county.

Data Appendix Figure 1: History of CCC program[5]



[5] Information on the history of the CCC program used in Appendix Figure 1 come from the following sources: https://babel.hathitrust.org/cgi/pt?id=mdp.39015004052794;view=1up;seq=13 On June 28, 1937, the CCC was once again renewed with funding for three additional years according to Public Law No. 163 (effective on July 1, 1937) see here: https://www.nps.gov/parkhistory/online_books/ccc/cccaa.htm

Data Appendix Figure 2: Example Colorado enrollment record



Form C-S210

10-759

CERTIFICATE OF SELECTION

For Enrollment in the

CIVILIAN CONSERVATION CORPS

Date 10-3-40

APPLICANT'S NAME Aragon (Last name) Aaron (First name) None (Middle name)

ADDRESS 511 West 4th St

POST OFFICE Walsenburg

STATE, COLORADO, COUNTY Huerfano

Application received by— Huerfano County Department of Public Welfare.

ADDRESS Court House

Walsenburg (City or Town) Colorado.

SECTION 1.

Age 20 Place and date of birth Gardner Colorado (City and State) July (Month) 4th (Day) 1920 (Year)

If not born in the United States, have you been naturalized? .... First papers .... (Date) Final papers .... (Place) .... (Date)

Height 71 in (Minimum: 60 in.) Weight 140 (Minimum: 107 lb.) Color of eyes brown Color of hair black

Applicant's marital status single Is your father living? yes (Yes or no) Mother living? yes (Yes or no)

How many brothers? 2 Sisters? 1 Occupation of principal wage earner of family? miner

How many members of your family reside in the same household with you? (excluding applicant) 5 (Number)

Do you live on a farm? no (Yes or no) If so, is the farm owned by your family? .... (Yes or no)

Do you live in a town or village of less than 2,500 persons, or in a rural area, and *not on a farm*? no (Yes or no)

Do you live in a town or city of 2,500 or more persons? yes (Yes or no) If so, give population 7,000

How long have you resided in this State? 20 (Years) This county? 20 (Years) Population of county 15,901

STATE OF COLORADO ... RECEIVED OCT 15 1940 ... DEPT

SECTION 2.

School last attended Hill School (Name of school) Located at Walsenburg Colo (City and State) Date of leaving 1938

Education: { *Circle highest grade completed* } Grammar or grade school, 1 2 3 4 5 6 7 8. High school, 1 2 3 4. College, 1 2 3 4

Special educational or vocational interests General laborer

SECTION 3.

Are you now unemployed? yes (Yes or no) How long unemployed? NE (Months) Do you need employment? yes (Yes or no)

Have you ever had a paid regular job? no (Yes or no) If so, give date last job ended .... Social Security Account No. 523-16-0392

Registered with State Employment Service? no (Yes or no) Work best qualified for farm laborer

If previously employed, give consecutive statement of your work history in space below (*list latest job at top*):

	Name and Address of Employer	Nature of Work Performed	Inclusive Dates of Employment From—	To—
1.	None			
2.				
3.				
4.				
5.				

Total months of all paid regular employment to date ....

SECTION 4.

Applicant's reason(s) for desiring C. C. C. enrollment: Unemployed

(This form to be completed on reverse side)

SECTION 5.

Previously enrolled in C. C. C.? no (Yes or no) C. C. C. serial number ______ If so, list all previous service below:

Company Number	Length of Service		Date Enrolled	Date Discharged	Type of Discharge Hon., Adm., or Dishon.
	Months	Days			
1.					
2.					
3.					

Total length of all previous service in Civilian Conservation Corps: Months ______ Days ______

SECTION 6. **DESIGNATION OF ALLOTTEE**

(Required for all juniors having dependents. Juniors without dependents will use Section 7)

Allotment from monthly cash allowance desired by applicant to be made to dependent(s) as follows:

Name Aragon (Last name) Margaret Mrs (First name) None (Middle name) Relationship Mother

Address 511 West 4th, Walsenburg Colo Amount per month $22.00

Name ______ (Last name) (First name) (Middle name) Relationship ______

Address ______ Amount per month ______

In addition to allotment, applicant desires deposit in the amount of $______ per month.

SECTION 7. **AUTHORIZATION FOR DEPOSIT IN LIEU OF ALLOTMENT**

(Completion of this Section required in all cases in which Section 6 is not used)

I. From the Selecting Agency: It is hereby certified, pursuant to regulations issued under section 9 of the Act to establish the Civilian Conservation Corps effective July 1, 1937, that through verification of the status of the applicant named herein, proper assurance has been obtained that he does not have any dependent member or members of his family to whom an allotment can be made. In order to be selected and enrolled in the Corps he is therefore required to agree to make a monthly deposit of pay in the amount of $______ with the Chief of Finance, War Department, to be repaid normally upon completion of or release from enrollment.

Selecting Agent's signature (*ink*) ______

II. From the Applicant: In accordance with the aforementioned Act and regulations prescribed thereunder by the Director of the Corps, I hereby certify that I do not have any dependent member or members of my family to whom an allotment of pay can be made, and I agree to make a monthly deposit of pay with the Chief of Finance, War Department, in the amount specified above, to be repaid normally upon completion of or release from enrollment.

Applicant's signature (*ink*) ______

SECTION 8.

The statements contained in the foregoing Sections are true, to the best of my knowledge. I desire to be enrolled in the Civilian Conservation Corps for a period of 6 months unless earlier released in accord with law and established regulations. If I am accepted and enrolled, I agree to abide faithfully by the rules and regulations of the Corps and am willing to be assigned to any C. C. C. camp within the continental United States.

Applicant's signature (*ink*) J. Aaron Aragon

SECTION 9. **THE OFFICE OF THE DIRECTOR (Division of Selection) C. C. C.**

Certifies that the above-named applicant has been properly selected for enrollment as a Junior in the Civilian Conservation Corps. For completion of his enrollment, including physical examination, he has been directed to report to C. C. C. acceptance officers at Walsenburg Colorado on October 10th, 1940 at 9:00 a. m./p. m.

COLORADO STATE DEPARTMENT OF PUBLIC WELFARE
EARL M. KOUNS, DIRECTOR
STATE CAPITOL ANNEX
DENVER, COLORADO

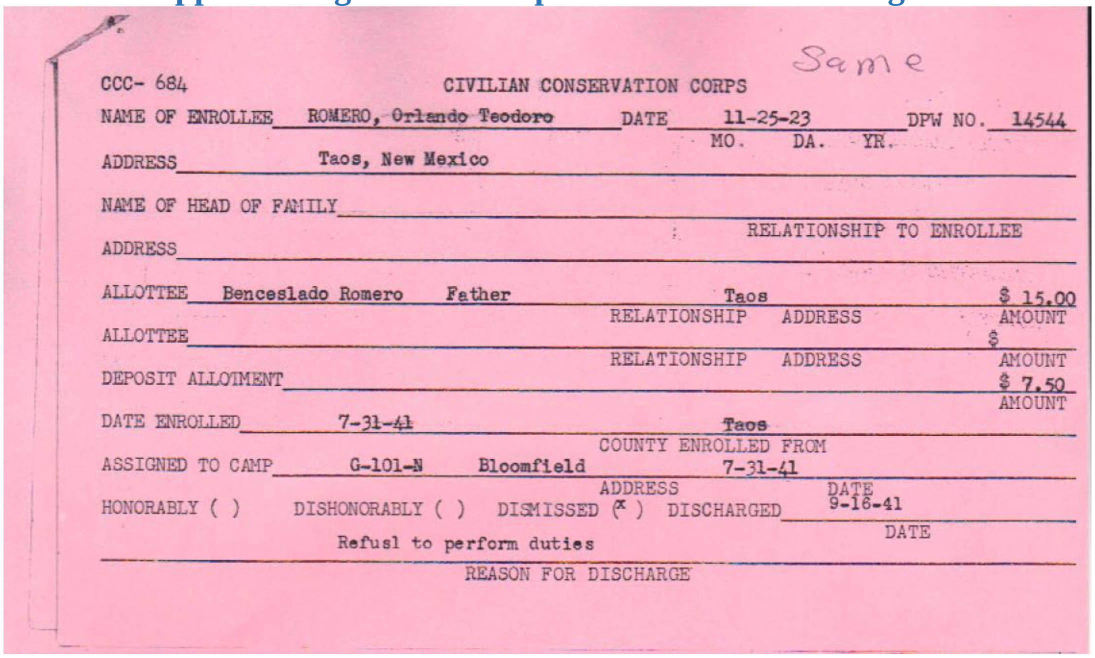
Routing of Copies:

To Army—white copy.
To State Department of Public Welfare—yellow copy.
To County Files—pink copy.

By Cynthia W. Hanes (Ink signature of authorized selecting agent)

Director, Huerfano County DPW October 9th, 1940 (Official designation)

Data Appendix Figure 3: Example New Mexico discharge record



Same

CCC- 684 CIVILIAN CONSERVATION CORPS

NAME OF ENROLLEE ROMERO, Orlando Teodoro DATE 11-25-23 DPW NO. 14544

MO. DA. YR.

ADDRESS Taos, New Mexico

NAME OF HEAD OF FAMILY

RELATIONSHIP TO ENROLLEE

ADDRESS

ALLOTTEE Benceslado Romero Father Taos $ 15.00

RELATIONSHIP ADDRESS AMOUNT

ALLOTTEE $

RELATIONSHIP ADDRESS AMOUNT

DEPOSIT ALLOTMENT $ 7.50

AMOUNT

DATE ENROLLED 7-31-41 Taos

COUNTY ENROLLED FROM

ASSIGNED TO CAMP G-101-N Bloomfield 7-31-41

ADDRESS DATE

HONORABLY ( ) DISHONORABLY ( ) DISMISSED (x) DISCHARGED 9-16-41

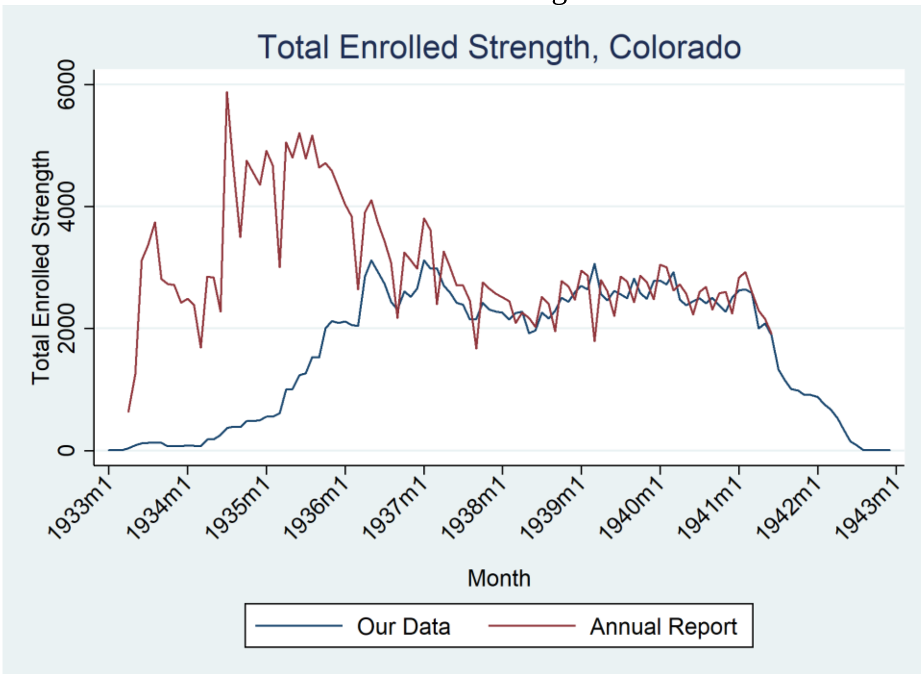
DATE

Refusl to perform duties

REASON FOR DISCHARGE

Data Appendix Figure 4: CO and NM data completeness

a. Archival data coverage in Colorado



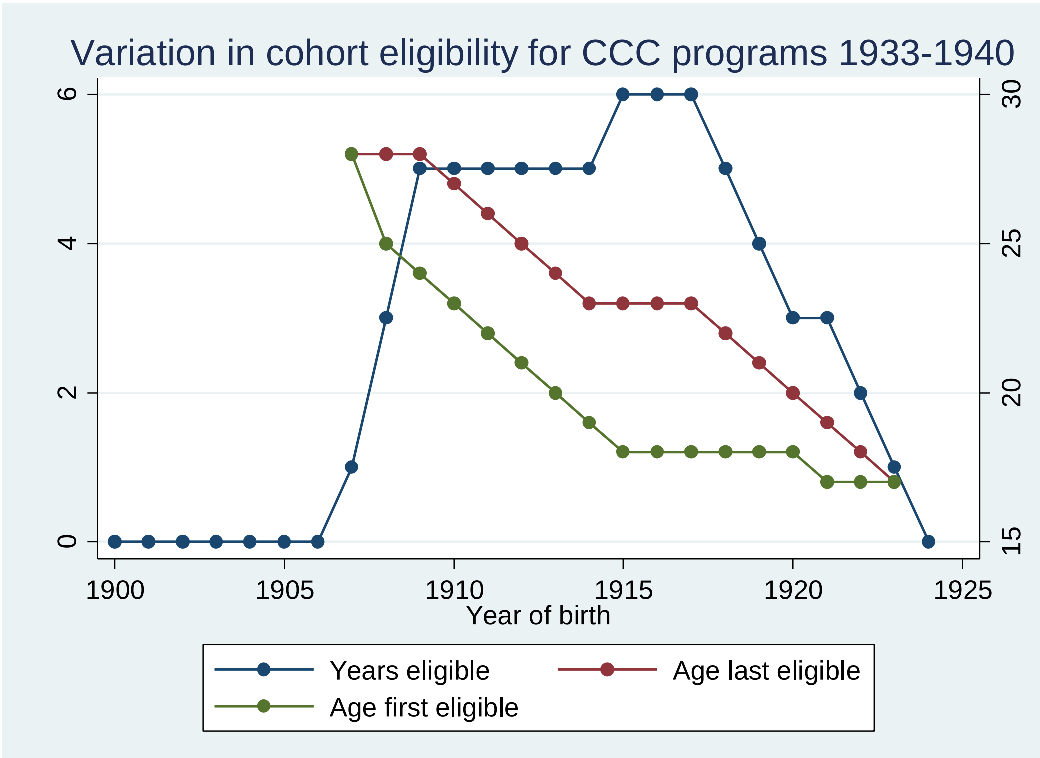
a. Archival data coverage in New Mexico

Note: Total enrolled strength is the number of enrollees at each month. Data from the Annual Report come from the following sources:

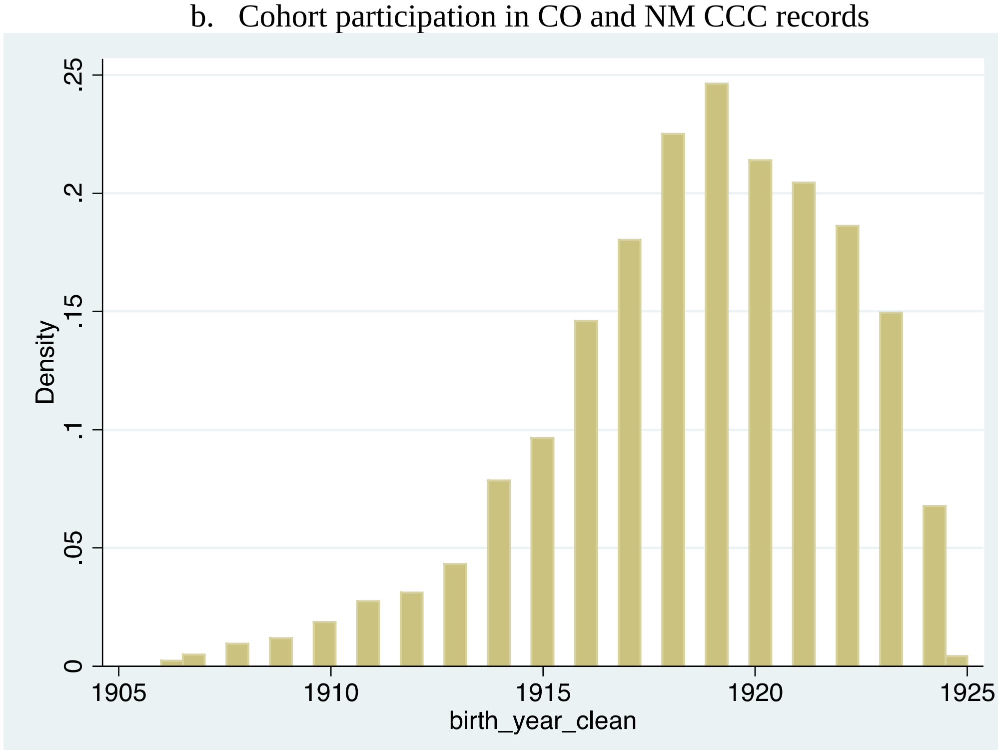
- Summary Report of the Director of Emergency Conservation Work on the Operations of Emergency Conservation Work: For the period extending from April 1933 to June 30, 1935, Appendix E
- Annual Report of the Director of Emergency Conservation Work: Fiscal Year Ending June 30, 1936, Appendix E
- Annual Report of the Director of Emergency Conservation Work: Fiscal Year Ending June 30 1937, Appendix D
- Annual Report of the Director of the Civilian Conservation Corps: Fiscal Year Ended June 30 1938, Appendix E
- Annual Report of the Director of the Civilian Conservation Corps: Fiscal Year Ended June 30 1939, Appendix I
- Annual Report of the Director of the Civilian Conservation Corps: Fiscal Year Ended June 30 1940, Appendix E
- Annual Report of the Director of the Civilian Conservation Corps: Fiscal Year Ended June 30 1941, Appendix E

Data Appendix Figure 5: Cohort eligibility and participation in CCC

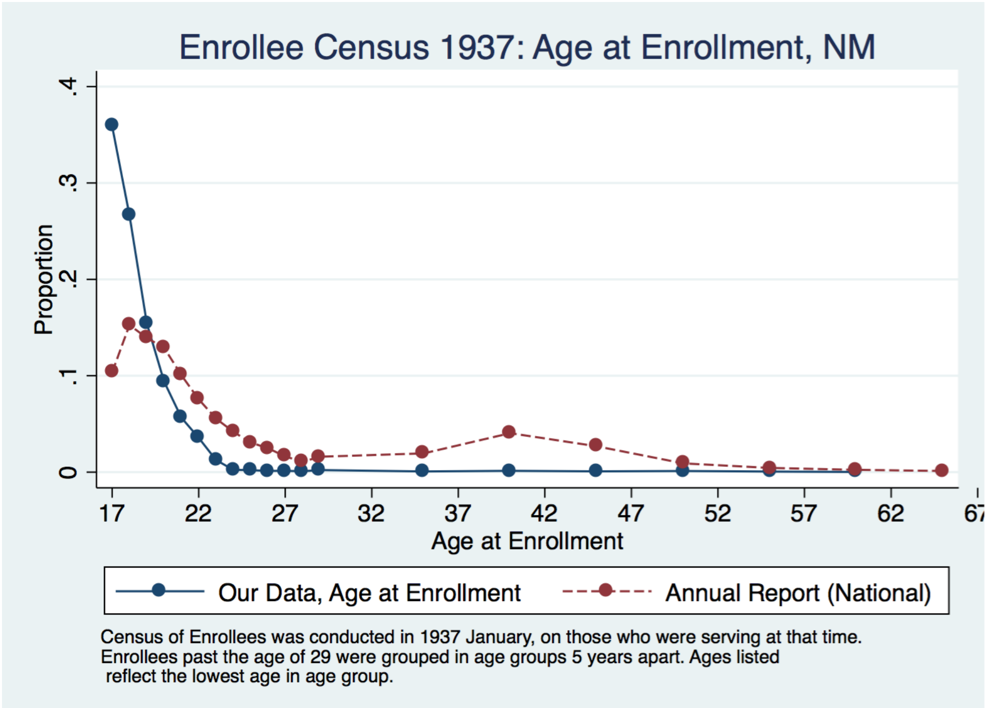
a. Variation in Cohort eligibility during the years CCC operated

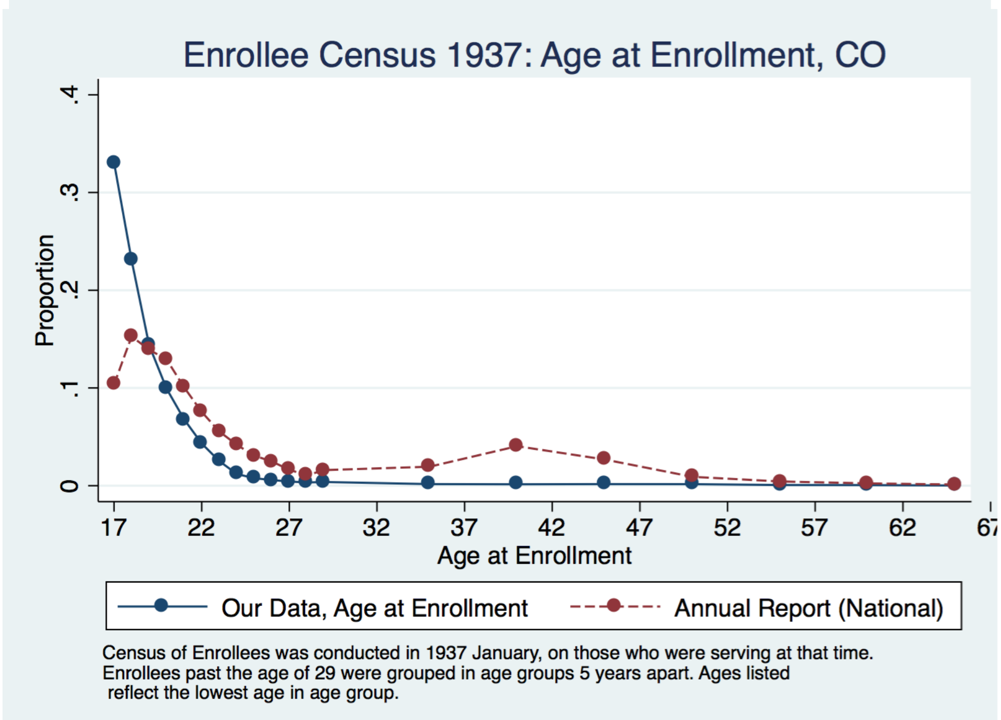


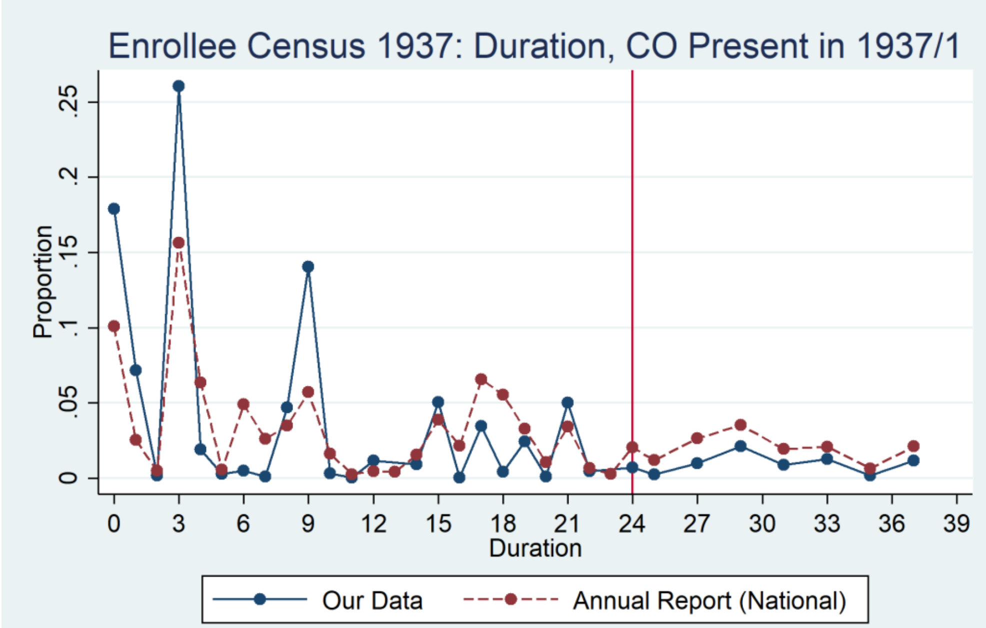
b. Cohort participation in CO and NM CCC records



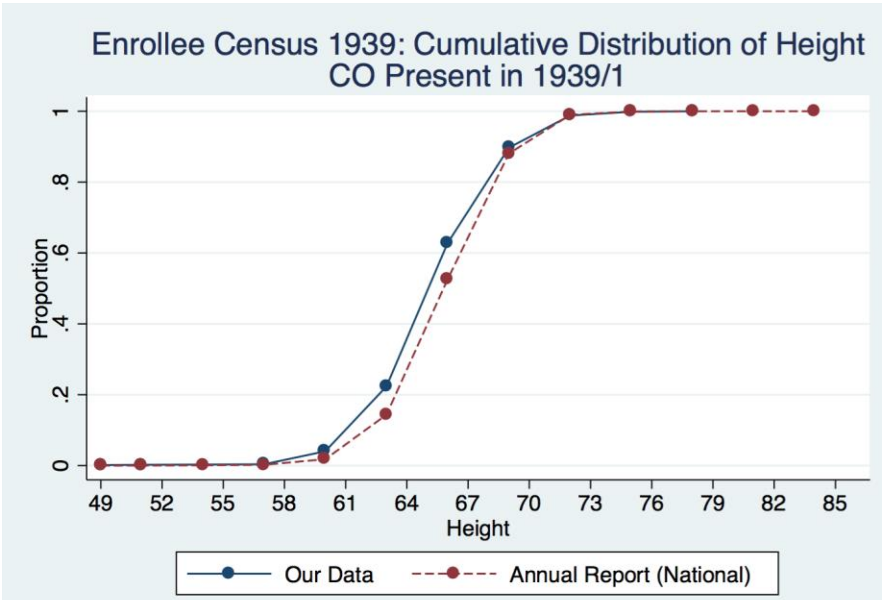
Data Appendix Figure 6: CCC enrollees in CO and MN are more disadvantaged than enrollees nationwide



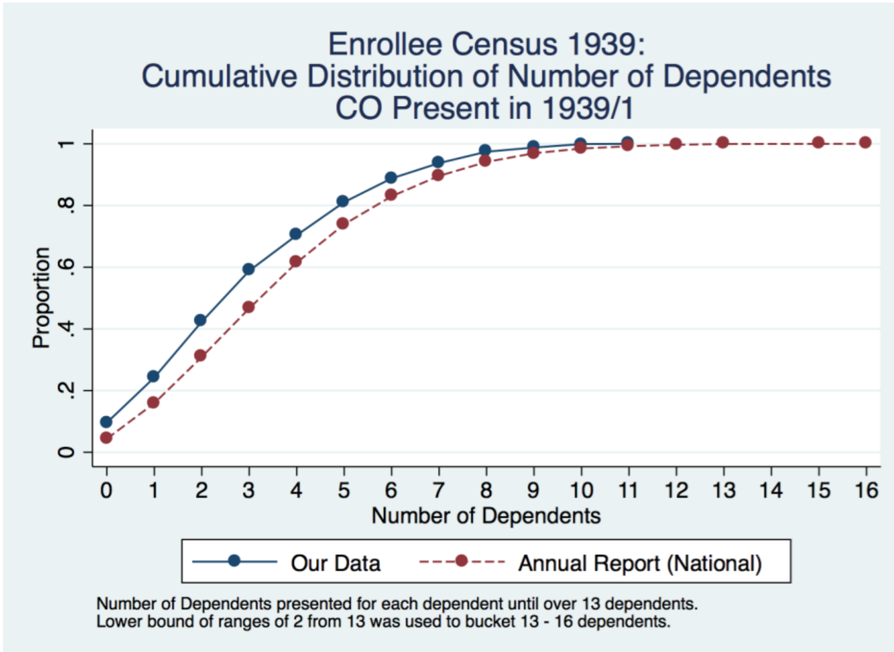


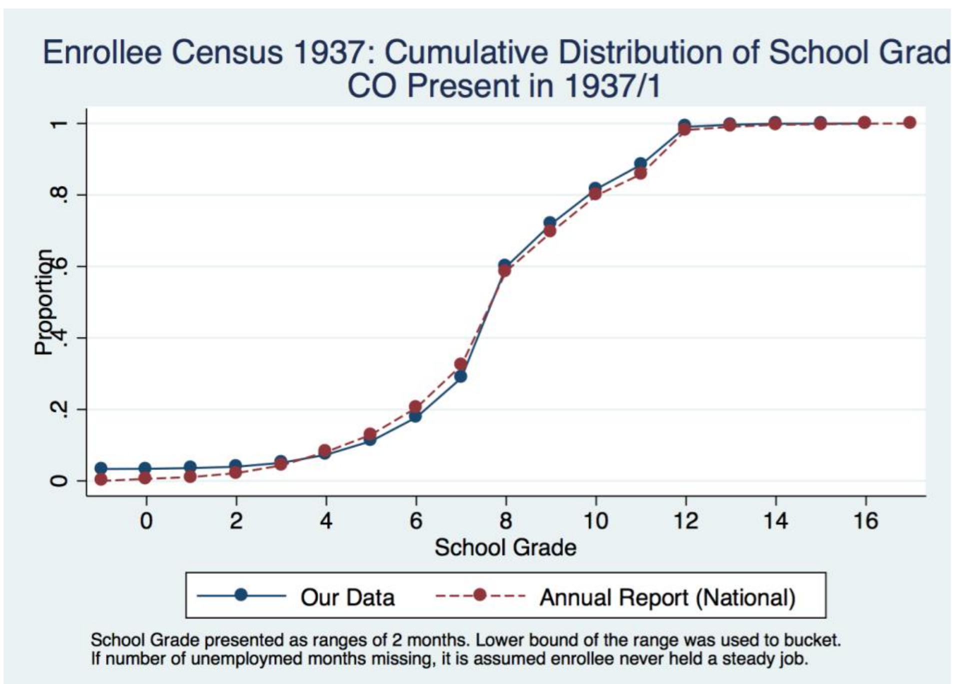


Census of Enrollees was conducted in 1937 January, so here we restrict the sample to be enrolled in 1937 Jan, and duration to be calculated analagously to the Report (duration until 1937 Jan



Height reported in ranges of 10. Lower bound of range is used to bucket height.


Enrollee Census 1939:
Cumulative Distribution of Number of Dependents
CO Present in 1939/1
Proportion
0
.2
.4
.6
.8
1
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Number of Dependents
Our Data
Annual Report (National)
Number of Dependents presented for each dependent until over 13 dependents.
Lower bound of ranges of 2 from 13 was used to bucket 13 - 16 dependents.


Enrollee Census 1937: Cumulative Distribution of School Grad
CO Present in 1937/1
Proportion
0
.2
.4
.6
.8
1
0 2 4 6 8 10 12 14 16
School Grade
Our Data
Annual Report (National)
School Grade presented as ranges of 2 months. Lower bound of the range was used to bucket.
If number of unemploymed months missing, it is assumed enrollee never held a steady job.