The Long-Run Impact of Cash Transfers to Poor Families: New Estimates using Larger Samples and New Methods

By Anna Aizer, Sungwoo Cho, Shari Eli, Joseph Ferrie, and Adriana Lleras-Muney*

November 2025

Abstract: *We re-estimate the long-run impact of cash transfers on children growing up in poor families who participated in state Mothers' Pension programs 1911-1930. By making use of family trees, we collect a new dataset tracking a large set of additional recipients, including women. We use new econometric models to account for potential selection into the program. We find that transfers increased longevity by about 1%, more so for boys than girls, but the estimated effects are noisy. Transfers increased adult incomes by about 7% and employment by 6%, but had no impacts on education. However, all our new estimates are smaller than in the original study. (JEL I12, I14, I18, I32, I38, J16, N32)*

I. Introduction.

Rich countries today devote a substantial share of their GDP to redistributive programs, broadly known as the "safety net." An important motivation for these programs is to help children growing up in poverty and thus provide a more equal set of opportunities to all children regardless of the circumstances of their parents. Yet despite the ubiquity of these programs, many questions remain unanswered regarding their effectiveness and how to best design them. In the U.S., the first anti-poverty programs distributed cash to poor mothers with children with the objective of helping children escape poverty. Over time, these programs have been replaced with in-kind programs partly as a result of the perception that cash transfers were costly and generated large distortions in the behaviors of recipients (Aizer et al. 2022). Surprisingly, however very little credible

* Aizer: Department of Economics, Brown University, 64 Waterman Street, Providence, RI 02912, and NBER (e-mail: aizer@brown.edu); Cho: Department of Economics, University of California, Los Angeles, Los Angeles 90095 (e-mail chosw13@g.ucla.edu); Eli: Department of Economics, University of Toronto, 150 St. George Street, Toronto, ON M5S 3G7, Canada, and NBER (shari.eli@utoronto.ca); Ferrie: Department of Economics, Northwestern University, 2211 Campus Drive, Evanston, IL 60208, and NBER (e-mail: ferrie@northwestern.edu); Lleras-Muney: Department of Economics, University of California, Los Angeles, 9373 Bunche Hall, Los Angeles, CA 90095, and NBER (e-mail: alleras@econ.ucla.edu). We are grateful to all of the RAs that worked on this project, and in particular to Tomas Guanziroli and Ariadna Jou. Adriana Lleras-Muney benefited from facilities and resources provided 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). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health & Human Development or the National Institutes of Health.

evidence exists on the benefits of these early cash-transfer programs for children, particularly in the long run.

Recent events have renewed interest in this question: proposals for universal basic income have gained popularity and many states and countries are currently conducting experiments to determine the likely effects of such programs. In this context, we set out to investigate whether the original state-operated Mothers' Pension programs in the U.S. (the first welfare programs which operated in the pre-New Deal era 1911-1935) improved the lives of the children they served. An advantage of the setting we study is that these programs were in place when no other programs were available to recipients. As a result, we can assess whether the money alone was effective. As families eligible for cash welfare became eligible for numerous in-kind goods after the War in Poverty was launched in the 1960s, this is very difficult to do in modern populations. Another challenge in conducting this sort of evaluation comprehensively is that one needs to track children into adulthood to observe the ultimate success of these programs. Because the program was in place more than a century ago, we can observe lifetime outcomes.

In Aizer et al. (2016), we assessed the effects of the receipt of cash transfers from state Mothers' Pension (MP) programs on the lifetime outcomes of poor boys. The programs provided poor families with cash transfers equivalent to 30% or more of family income with the median family enrolled for 3 years . We collected individual-level records that survived in county archives. After digitizing the data, we tracked the children's lifetime outcomes in subsequent military, census, and death certificate databases. To assess whether the program benefitted children we relied primarily on lifespan as a summary measure, but we also looked at children's adult incomes and final educational attainment. To estimate the causal impact of the transfers, we compared initially eligible families that did and did not receive benefits. Most families that were denied benefits are observed to be less poor potentially explaining why they were denied. Thus, our approach likely underestimates the benefits of the transfers. We found that men who were children in families accepted and provided benefits under the program lived longer, obtained more schooling, and earned higher adult incomes than children in similar families that were ultimately not provided benefits.

In this paper we revisit our original findings using newly collected data and new estimation methods. We now report results from a larger sample that overcomes several of the limitations in the original paper: women weren't previously followed, match rates were low, and the samples

were correspondingly small, particularly for education and income. Our original research used computer based matching algorithms that rely on the name and date of birth of individuals to match them across databases. As a result, we could only properly study the sons of poor mothers and could do so for less than 60% of the sons. We have linked again all the original administrative program records to the family trees available in FamilySearch which obtains information on death from multiple sources. Rather than relying on names and dates of birth alone, our new linking takes advantage of the fact that in the administrative records we observe *families*. We can use this information to better track individuals. Because we link individuals to family trees, we also observe changes in family names occurring at the time of marriage which allow us to track women as well as men. Where we originally studied the impacts of the program on lifespan for about 8,000 boys in poor families, we now have lifespan information for about 13,000 boys and 11,000 girls.

This newly expanded data has other advantages over our original data. First, our data on outcomes is much improved. Lifespan is no longer left and right censored, Individuals who died before 1965 or after 2012 are now included. This is important: previous research has shown that left and right censoring biases estimates of the effects of early treatments on lifespan (Lleras-Muney et al. 2022; Goldstein et al. 2023). Second, our larger samples facilitate the study of mechanisms.. In the original study we could observe the education and earnings of about 1,500-2,000 boys, whereas we now have this information for roughly 9,000 boys and 8,600 girls. Finally, from these new sources of data, we also collect additional information about applicants that was not available in the application records such as the age, education, race and ethnicity of the mother, which allow us to validate our original claims that families receiving transfers were on average poorer leading to underestimated program effects.

This large sample paired with new estimation techniques allows us to investigate for whom cash transfers matter most. An important unanswered question is whether cash transfers and other redistributive programs equally benefit all recipients. A number of recent papers in economics have shown that males do particularly poorly compared to girls when they grow up in poor or disadvantaged households (Chetty et al. 2016, Autor et al. 2019). A key question is whether the converse is also true: do boys benefit more from public programs if they are growing up in disadvantaged households?

Aside from sex differences, these programs could also have different impacts for other groups. Understanding treatment effect heterogeneity would help improve the design and targeting of cash

transfer programs. We leverage our richer and larger data to estimate heterogeneous effects using novel methods. We now estimate the average treatment effect (ATE) of the cash transfer on outcomes using the causal random forests methods recently developed by Wager and Athey (2018) and Athey, Tibshirani, and Wager (2019). This is a matching-like approach that compares "nearby" treated and control units wher proximity is determined by similarity in their propensities to be treated. It estimates individual-level treatment effects which can be used to compute consistent estimates of the ATE under the standard assumption of unconfoundedness. This method to estimate heterogeneous treatment effects leverages recent advances in machine learning to address limitations of other matching approaches which rely on more parametric assumptions to estimate propensity scores. These methods allow for better matching on observables and can leverage the newly available covariates we have collected. They also produce a distribution of treatment effects which can be used not only to compute average treatment effects but also to formally test for heterogeneity and to investigate its sources which is helpful to understand how to best target such programs. These methods also allow us to confront the challenge posed to our estimation of the treatment effect by the disparity in the sizes of our treated and control groups in the presence of treatment effect heterogeneity (Słoczyński 2022). When treatment effects are heterogenous, OLS reports a weighted average of the treatment effects, where the weights are *not* proportional to the size of each group (e.g. Deaton 1994). We report both OLS and ATE estimates.

We find that cash transfers increased the longevity of children in receiving families, but the estimated impacts, while large, are imprecise. In percentage terms our estimates are close to those we reported originally: receiving the transfers increased lifespan by about 1%, and the confidence intervals cannot reject estimates in the range of our original study. We also re-estimate impacts on potential mechanisms affecting lifespan, primarily education and labor market outcomes. We find that the impact of cash transfers on years of schooling is now null, in contrast to our original findings. We do find statistically significant impacts on income. These income effects are also observed when we look at occupation-based income and at employment. Thus, we do find that cash transfers improved the labor market outcomes of poor children. In the sample of all children, we see that receiving cash transfers led to a 7% increase in income and a 6% increase in employment in adulthood.

We find evidence of heterogeneity in treatment effects, particularly in income. However, we find that observables are not particularly useful to predicted gains from transfers and thus it is

difficult to improve targeting in a setting like ours despite the fact the data suggest the gains for the program could be significantly larger if targeting could be done on the basis of gains. We also find larger impacts for boys than girls in both lifespan and income, but we cannot formally reject that the effects are identical by sex in levels. But this heterogeneity is partly explained by the differential treatment of girls in mixed-sex families. In these families, girls saw small treatment effects and boys see large treatment effects. In same sex families, girls have large positive benefits and boys see no benefits. Because most families are mixed, girls benefit less on average.

This paper makes three contributions to the literature.

First, we contribute to the large and growing efforts to understand whether and how public policies can improve the lot of children growing up in poverty. To our knowledge this is the first and only effort that has evaluated the impact of cash transfers on the lifetime economic outcomes and lifespan of recipients. Most existing work evaluates impacts in the short to medium run. The closest work to ours is Bailey et al. (2024) who look at the impact of the roll-out of Food Stamps on poor children's lifetime outcomes over about 50 years. Our estimates on economic outcomes are similar. Our estimates of the impacts on lifespan are difficult to compare since they only observe survival to a certain year. However, their point estimates, like ours, suggest larger impacts for men, though the results are also not statistically significantly different between men and women. Page (2021) conducted a systematic review of the impacts of childhood income on children's outcomes, and she concludes that across studies an increase of $1,000 in childhood leads to less than 1% increase in earnings in adulthood and a 0.01–0.02 increase in completed years of education. Similar to that work, we find significant labor market gains among both men and women in young adulthood. The magnitude of our findings is larger than in these previous studies for income but similarly modest or smaller for education.

A series of recent Randomized Controlled Trials (RCTs) have failed to find effects of cash transfers on children's outcomes, at least in the short run. The largest of these efforts provided poor families in the U.S. an unconditional cash transfer of $1,000 a month for three years and found no observable improvements in the educational outcomes of children (Krause et al. 2025). While there are important differences between this study and ours, our results are in line with theirs: we do not see meaningful impacts in education in our study. However, we do see improvements in incomes and lifespans demonstrating that there can be a large discrepancy between short- and long-term evaluations. This discrepancy may be due to the fact that the benefits

of the cash do not operate through schooling but through other mechanisms such as soft skills (e.g. Heckman and Kautz 2012) or because the effects only manifest themselves over the long run, as is the case with health (Aizer et al. 2024).

Second, we contribute to the understanding of sex differences in health and other life outcomes. Women live longer than men in most countries of the world today, much more so today than in the past. But the reasons for this growing female advantage are poorly understood. Biological factors, in particular genetic factors, disadvantage males. But why this disadvantage has grown is not well understood (Goldin and Lleras-Muney 2019). Researchers have long hypothesized that males are less buffered than females against the environment during growth and development (see the review by Stinson 1985, DeWitte 2017). This appears to be true in utero and in the first year of life. For example, infant mortality rates of males exceed those of females in almost all the settings in which they has been studied (e.g. Drevenstedt et al. 2008). But this pattern is less well established thereafter partly because whatever biological disadvantage boys may have may be offset by differential sex investments by families and by societies more generally, which typically tend to benefit boys.

Recent research in the U.S. however has found evidence that boys may be particularly disadvantaged by poverty. For example, Chetty et al. (2016) and Autor et al. (2019) find that family disadvantage disproportionately inhibits the academic development of boys, which suggests that boys growing up in poverty might benefit more than girls from cash transfers. Some research suggests indeed that boys benefit more than girls from early education programs (see the review in Magnuson et al. 2016, or the more recent work of Gray-Lobe et al. 2023). But the evidence is less well established for cash transfers at least in the U.S. For example, Akee et al. (2010) find that girls' high school graduation rates are more impacted by casino profits, but the results on years of educational attainment are similar for boys and girls. Our findings are the first we are aware of to study the sex differences of cash transfers on lifespan. In line with the literature, we do find suggestive evidence of larger lifespan impacts on boys, but the economic benefits for girls we observe, while smaller in magnitude, are larger in percentage terms than those for boys.

We do however find evidence that the sex composition of siblings affects the benefits of the program suggesting the boys and girls are treated differently by their parents and this partly explains the smaller effects we observe among girls. Many studies find that parents invest more in their sons, though these studies are typically conducted in contexts with strong son preferences

due to cultural, religious, and legal reasons (Jayachandran 2015). Although differential treatment is extremely hard to prove, in the U.S. most studies have not found evidence of parental sex discrimination, even in the 19th century (Logan 2007, Goldin and Katz 2008, Lleras-Muney and Goldin 2019). Our data suggest differential treatment by sex depends on the sex composition of siblings. In particular we find that boys with mixed-sex siblings benefit more from transfers, whereas girls benefit more from transfers in families with all-girl siblings.

Finally, we contribute to recent efforts to replicate results in the social sciences. A few high-profile studies in different disciplines have shown that many well-known and well-published empirical findings cannot be replicated, suggesting that publication bias and researcher discretion lead to an unusually large number of false positive results being published (Open Science Collaboration, 2015; Camerer et al. 2016, 2018; Christensen and Miguel 2018). Yet replications in economics remain rare (Duvendack et al. 2017). Because so few studies can track the effects of a program over the lifetime of the participants, we deemed it worthwhile to replicate our original findings once the possibility of collecting better data arose. Our initial results for lifespan (our main outcome in the original study) "replicate" in the sense that we cannot reject the null that the new estimates are identical to the ones in the original study. However, point estimates are smaller for this and other outcomes. Smaller treatment effects in larger follow-up studies appear to be the most common outcome in replication studies. Our replication efforts confirm these patterns in our setting. We discuss the reasons for these smaller impacts.

II. The Mothers' Pension Program and the debate about cash transfers

a. Historical background.

Following innovations in Europe, almost all states in the US passed Widows' or Mothers' Pension (MP) programs sometime between 1911 and 1935. The idea to provide impoverished women with children with cash became popular because of a greater understanding was emerging that mothers provided essential care for children and that children who grew up in institutions fare poorly.

These programs provided poor families with cash. They initially targeted impoverished widows with children and eventually eligibility was expanded to include abandoned or divorced women, and women whose husbands were imprisoned or living in institutions for the ill. Although states passed the initial legislation and provided general guidelines for how the programs operated, it was up to each county to create, fund and operate its own program, sometimes with partial funds

from the state. As a result, each county had its own program. These programs were initially quite small, as many counties did not implement a program at all, and those who did typically served a rather small portion of the eligible population.[1]

The programs provided cash transfers to families as a function of the number of children and their ages for different durations depending on the county and on whether the families remained eligible. Only poor women who had no spouses were eligible. Neither income nor wealth thresholds were specified as criteria for eligibility, but the county would often investigate to assess the extent of a women's resources. Typically, only children under the age of 14 were eligible for transfers. Women could continue to receive payments so long as they had eligible children, though in some states they had to reapply with some frequency. Women who remarried, obtained incomes from other sources or moved locations would lose benefits as a result. In our records we do not observe the duration of payments for all families, but the median duration in the records that do report it was about 3 years. State laws typically specified the maximum generosity of the transfer, which ranged from 10 to 25 dollars for the first child in the states we study. Some counties systematically gave the maximum but there was considerable discretion over the allotted amounts. Aizer et al. (2016) report that on average transfers amounted to roughly 30-40% of maternal income at the time. Finally, most states required women to stay home as a condition of the transfers, (but some regulated the extent to which women could work) and to provide an adequate home environment.

The programs were popular and grew overtime. But when the Great Depression arrived, they became insolvent: counties ran out of revenues while applications soared. Thus in 1935 as part of the Social Security Act, the federal government revamped the program under the name of Aid to Dependent Children (ADC) and committed to assist states with funding so long as they met basic federal guidelines. With the creation of Social Security pensions for survivors, the composition of recipients quickly moved away from widows towards single parent households of other types in the subsequent years. In 1962 the program was renamed Aid to Families with Dependent Children (AFDC). The program eventually became popularly known as "welfare" and it remained the main source of cash transfers for poor families until the 1996 Welfare Reform.

[1] For example, the Children's Bureau (1926) estimated that counties served at most 1/3 of eligible families. See Eli et al. (2025) for a more detailed analysis on the diffusion and generosity of these programs across counties.

Today the program has mostly disappeared (following the 1996 welfare reform) and has been de facto replaced by the Earned Income Tax Credit (EITC).

Over the years, the welfare program was subject to a substantial amount of criticism, starting in the 1960s, after a period of intense growth in welfare rolls. The main concerns regarding the program related to the potential behavioral responses of recipients and would-be recipients. The program was seen as disincentivizing work and marriage, while incentivizing fertility and mobility towards generous states. As a result, most research regarding the program focused on investigating these responses among the adult recipients. These concerns also led to the eventual demise of the program in the US and its de facto replacement with the provision of in-kind goods and services to the poor (in the form of food stamps, health care insurance, or housing) or with programs that implicitly condition aid on work effort such as the EITC (Aizer et al. 2022).

However much less attention was paid to the question of whether the program achieved its main aim to help poor children escape poverty as adults (Shah & Gennetian 2024). Thus, whether cash transfers are in fact beneficial for poor children and whether they are better or worse than other programs targeted to poor children remains unclear.

b. Previous literature

The question of whether incomes improve child outcomes in the long run is difficult to investigate. Several papers have been written in the last decade investigate how income, obtained from various sources, affects child outcomes. This research has produced surprisingly varied results, as summarized by Page (2024) who conducted an extensive review of the impacts of money on children's short- and long-term outcomes and by Shah & Gennetian (2024) who conducted a review of studies evaluating the impact of policies that transfer cash to poor families.

From the outset it's worth noting that there is very little evidence from randomized control experiments providing cash and tracking child outcomes into adulthood. As Shah & Gennetian describe, there are a number of experimental studies conducted in the 1960s and 1970s which studied the impacts of modifying the existing welfare programs at the time (Aid to Families with Dependent Children or AFDC) which guaranteed a maximum cash transfer benefit, which was reduced with earned income at a benefit-reduction rate. In these experiments the guaranteed benefit was increased, and the benefit reduction rate of the welfare program was lowered among a random subset of poor families for 3 to 5 years. These "negative income tax" experiments were designed

to estimate the negative labor supply effects of increasing the guaranteed income and/or changing the benefit reduction rate. But some data was collected on children, and it showed improvements on their short-term education and health outcomes. Similarly randomized experiments in the 1990s investigated impacts of the welfare to work experiments which conditioned transfers on work and found improvements in children's short term education outcomes. However, Price and Song (2018) study the long-term effects of the same experiments in Seattle-Denver and find no long run benefits on children.[2] Note however that these experiments did not randomize cash transfers.

Several papers look at related policies using natural experiments instead. Most closely related to this study, Hoynes et al. (2016) and Bailey et al. (2024) investigate the long-term effects of the roll out of the Food Stamp program in the 1960s which provided poor families with vouchers that could be redeemed for food, which they deemed "quasi-cash". They find large benefits in the long run, particularly for children treated very early in life. Several papers have also investigated the impact of the EITC, which provides rebates to working families who qualify when filing annual taxes. These studies also tend to find positive effects on the education and health of affected children (see Aizer et al. 2024 or Shah & Gennetian 2024 for longer description of these studies). However, the EITC is conditional on working so its findings are not clearly pure cash transfers effects. For the same reason (it conditions on work), the EITC does not reach the poorest families in the US.

Closest in spirit to the ideal experiment is the study by Cesarini et al. (2016) who investigate the impact of lottery earnings in Sweden and find no substantial benefits for the children of adult winners, even among poor families. A similar study in the US found that children in poor families who won the lottery, even large amounts, were only slightly more likely to attend college as a result (Bulman et al. 2021). Another highly influential set of studies by Akee et al. (2010, 2013, 2018) traces the impacts of large unanticipated increases in income due to the introduction of casinos in reservations among Cherokee children, a group with extremely high poverty rates. It resulted in increased obesity among children (a potentially negative outcome) but otherwise improved education, and health and reduced criminal behavior and drug use. The

[2] Price and Song find no effects on economic or health outcomes of the children when they are adults, though it is worth noting that due to the large tax on earned income and its effect on labor supply, the actual increase in the income of the treated families was rather modest in this experiment.

literature has not fully resolved why the estimates from different studies vary, as discussed by Page (2024).

There have been several recent efforts to re-evaluate the effect of unconditional cash transfers in the US and in other high-income countries.[3] The largest of these efforts provided poor families in the US an unconditional cash transfer of 1000 a month for three years and found no observable improvements in the educational outcomes of children (Krause et al. 2025). Other smaller and less generous RCTs (XXX) have also found similarly disappointing results. However, these efforts have only investigated changes in short term outcomes such as schooling. Previous work evaluating the impact of childhood interventions has demonstrated that short term outcomes are not always reliable indicators of long-term impacts (e.g. Aizer et al. 2022, Heckman et al. XXX). In particular, standard measures of educational performance such as test scores or educational attainment are not always impacted by interventions, such as early education, that ultimately increase incomes and wellbeing. Thus, in the absence of reliable short-term indicators, it is essential to evaluate the impact of cash transfers on long term outcomes.

A crucial issue is whether the additional income changes the behavior of parents in ways that help or harm children. Using the newly constructed data from the Mothers' Pension program, we tracked the behaviors and outcomes of the mothers that received cash transfers using the same design that compares initially eligible accepted and rejected mothers. In Aizer et al. (2023) we report that cash recipients delayed remarriage by about a year or two and were less likely to move out of their county of residence as a result of the cash transfers. But we found that women married at similar rates over the lifetime and married similar men, they did not have more children as a result, nor do they appear to work less after receiving cash transfers. Overall, we found no changes in their household income in the long run, nor do we estimate statistically significant longevity benefits for mothers. This suggests that the behavioral responses of the mother were modest, and aside from mobility, were unlikely to have large effects on children.

In this paper we use the newly collected data to investigate again the impact of the program on children. We pay particular attention to whether the estimates differ by sex, by program characteristics and by maternal behaviors.

[3] For a review of studies summarizing the effects of unconditional and conditional cash transfers in developing countries see XX.

III. Empirical Approaches

We wish to re-evaluate the impact of cash transfers on children using our improved data. To facilitate comparisons, we first use the same approach we employed in our initial study which consists in estimating unconditional and conditional mean differences in outcomes between children in accepted and rejected families using OLS. To address the issue of treatment effect heterogeneity and the bias it generates, we then employ newer approaches from the machine learning literature as we now discuss.

a. OLS approach

We begin by performing a simple OLS analysis that compares the outcomes accepted and rejected applicants, with or without covariates, following the approach of Aizer et al. (2016). Specifically, we estimate the following model

$$y_{ict} = c + \beta * Accepted_i + X_{ict}\gamma + \theta_c + \mu_t + \varepsilon_{ict} \qquad \text{eq. (1)}$$

where y is the outcome of interest (the age at death, education, income) for individual i whose mother applied for cash transfers in county c and year t. "Accepted" is a dummy equal to 1 if the family was given transfers and 0 if they were deemed ineligible after they applied. The coefficient of interest is β which captures the mean difference in outcomes between the two groups. We report unconditional mean differences without any additional controls. We also present regression coefficients with additional controls. We control for county (θ_c) and year fixed effects (μ_t) to account for the fact that the program was essentially a county-level program, only loosely bound by state laws. In addition, we typically control for all the observed individual covariates in the application (marital status of the mother, number and age of children listed in the application), and include state-level time-varying covariates (X_{ict}).[4] We cluster the standard errors at the family-level. We assess the robustness of the results to controlling for newly observed covariates. We discuss our identification assumption below.

[4] The state level time varying covariates include: laws regulating cash transfers in the MP program (max amount for first child, max amount for each additional child, whether work is required, and whether application is required); education and child labor laws (age at entry, minimum age to get a work permit, continuation school required); state level expenditures on total charities, social programs and schools; and measures of income (state level ratio of manufacturing wages to national wages).

b. Generalized Random Forest (GRF) estimates and other estimators

As noted in the introduction, previous research has shown that in the presence of heterogeneous treatment effects, OLS does not estimate the average treatment effect (ATE) of interest. The main challenge in addressing this, consists in properly estimating different treatment effects for each unit. Here we make use the latest advances in machine learning to address this challenge. Specifically, we use the generalized random forests methods recently developed by Wager and Athey (2018) and Athey, Tibshirani, and Wager (2019). These methods aim to estimate heterogenous treatment effects in the most flexible (non-parametric) way by leveraging machine learning methods. In spirit, this matching-like approach provides consistent estimates of the ATE under the standard assumption of unconfoundedness. It assumes that untreated observations with similar propensities to be treated as treated observations provide appropriate counterfactuals. Unlike traditional propensity score methods, this method leverages machine learning, specifically causal forests, to find the "nearest neighbors" and computes treatment effects for each treated unit using these untreated but similar observations. Specifically, we now estimate the following equation:

$$y_{ict} = c + \beta_i * Accepted_i + F(X_{ict}) + \varepsilon_{ict}. \qquad \text{eq. (2)}$$

There are two key differences between this GRF approach, and the OLS approach described in equation (1). First, the matching is done non-parametrically through the function $F(X_{ict})$. It does not impose that the effect of the covariates affects the outcomes though a linear index - it allows for higher order terms and interactions between variables. Second, the treatment effect is now estimated at the individual level (β_i) for all individuals in the sample allowing to investigate heterogeneity in the effects of the cash.

The algorithm first trains a causal forest using a full set of covariates and then estimates conditional average treatment effects (CATE) at each point in the covariate space. An individual tree in a causal forest is trained by drawing a random subsample from the dataset which is then split into several nodes (leaves) to form a tree. Each split is chosen to divide the data based on one covariate and a threshold value, aiming to maximize differences in estimated treatment effects between the resulting child nodes. At each leaf of a tree, the difference in average outcomes

between treated and control observations is used to estimate the local treatment effect. The GRF algorithm measures the goodness of a split using heterogeneity across nodes and maximizes the difference in treatment effects across nodes. The program estimates many such trees (a forest). Then, a prediction is made using a weighted average of each tree's prediction where the weights reflect the similarity based on how often they appear in the same leaves across trees.

To implement this procedure, we use the algorithms provided by the authors. Researchers have some discretion in how these effects are estimated. We make the following decisions to train a causal forest. First, we use 50% of the full dataset to grow each tree. Second, we train 100,000 trees in a causal forest.[5] In training each tree, we consider $\sqrt{x} + 20$ number of variables for each tree where x is the number of variables and set 20 as the minimum number of observations in each leaf. These choices are suggested by the authors as the preferred default except for the number of trees and the minimum number of observations in each leaf.

We estimate the average treatment effects (ATE) using a doubly robust augmented-inverse-propensity weighting estimation method (Robins et al., 1994). This is a more efficient procedure according to Chernozukhov et al. (2022). We also report the overlap-weighted average treatment effect (ATE overlap) as recommended by Li, Morgan, and Zaslavsky (2018) – this addresses the issue of having estimated propensities close to 0 or 1 which results in very large or very small weights for some observations when computing the ATE.

A key choice in this estimation concerns clustering, which affects the standard errors (as expected) and unlike OLS, it also affects the point estimates (Athey and Wager 2019). This occurs because when there are clusters, the program samples clusters rather than individuals and each unit in a cluster gets equal weight. Our main estimates cluster the standard errors at the family level because everyone in a family is treated or not. But we also check the robustness of the results to alternative clustering levels.

c. **Identification assumptions**.

Both our OLS and GRF estimators compare the average outcomes of the children in accepted families and in rejected families, conditional on some set of controls. Thus, we assume that receiving transfers is as good as random conditional on our controls. The data are not entirely

[5] Davis and Heller (2019) use 100,000 trees and Beaman et al. (2021) use 250,000 trees but find no meaningful increase in stability when increasing the number of trees from 100,000.

consistent with this assumption. We investigated this extensively in our previous work. Using information from the application records, Aizer et al. (2016) documented that the most common reason for rejection was that the applicants had too many alternative sources of income. Consistent with this, accepted families were larger and had younger children, making them more likely to be poor. We also used observable characteristics at the time of application to predict family income using the 1915 Iowa census (the only sources of income data prior to 1940) and found that indeed accepted applicants were predicted to be slightly poorer. This finding would be consistent with the program accepting the poorest families as intended by the legislation. If selection on unobservables is similar to selection on observables, so that accepted families are indeed poorer and more disadvantaged than rejected ones, then the benefits of the MP program on children will be underestimated. We investigate this issue again with the new data.

d. Heterogeneity

To investigate heterogeneity, we again make use of the tools in the proposed by Wager and Athey. The machine learning algorithm we employ estimates individual treatment effects. They propose a formal test of heterogeneity. Once the CATEs have been estimated, they can be ranked into percentiles. Then one can plot the difference between the ATE for individuals at a given percentile and the overall ATE, which they label the *Targeting Operator Characteristic (TOC)* curve. If everyone has the same treatment effect, then the curve is flat, and the gap is equal to 0 everywhere. If there are some individuals with very large treatment effects, then the difference between their ATE and the ATE in the full sample will be large and significantly different than 0. The area under this TOC curve describes how different the treatment effects are across all percentiles, and it's called the *Rank-Weighted Average Treatment Effect* (*RATE*). If there is no heterogeneity this area is small and not statistically significantly different from zero. We conduct this test. If one rejects homogeneity, then one way to investigate heterogeneity with the GRF is to project the estimated ATEs onto observable characteristics or to partition the sample by CATEs, which is referred to as the Best Linear Projection (BLP).

IV. Data Creation and Characteristics

We create the data for this study by matching administrative records from the MP program which contain information on applicant families. We then match individuals to family trees and to Census

records to observe their long-term outcomes. We focus on three main outcomes of interest: The age at death or lifespan of the children in families receiving welfare, their years of education and their income as reported in the 1940 census. This section describes how we collected the data and tracked the outcomes of the applicants in this paper.

A. Data Creation

A.1. Administrative records from the MP program

We found and digitized the original administrative records of applicants to the MP program. Because the program was implemented and administered by counties, each county was searched separately. Records in 3 states contain no information on rejected applicants and are not included in this study as a result. Our records pertain to 77 counties in 11 states (Idaho, Illinois, Iowa, Minnesota, Montana, North Dakota, Ohio, Oklahoma, Oregon, Washington and Wisconsin).[6] The records include the universe of records for applicants in these counties in the years during which the program operated. The records vary widely in detail, but all include the name of the mother, and the name and date of birth of all eligible children as well as the determination (accepted or not) and the allotted monthly amount if the application was accepted. In some records we also observe reasons for application and reasons for denials.

A.2. Obtaining data on outcomes from family trees and Census records.

We contracted with the Record Linkage Lab (RLL) at Brigham Young University to create the final data which includes the original administrative records as well as outcomes for all family members. The RLL employs BYU undergraduates trained in genealogical research to search manually for individuals through a wide variety of sources, including family trees generated by amateur genealogists and available through Ancestry.com and the Family History Library of the LDS Church. The lab matched individuals in the MP applications to a tree, where the birth and death information can be found. These vital statistics can come from a variety of sources. The lab also located individuals in the 1940 Census where we can observe educational attainment and labor market outcomes.

Our administrative records included important information about family structure that allows us to track individuals at very high rates when combined with family trees. Even though names

[6] Eli et al. (2025) describe all the county programs that existed. Unfortunately, the records of many programs (e.g. New York) have not survived and others are not publicly available.

and dates of birth are often insufficient to track an individual across data bases, tracking becomes easier if one knows the name of an individual's mother and the names and dates of birth of their siblings which we observe in the application for transfers and in family trees. The database of family trees available from FamilySearch includes more 1.7 billion people – the largest collection of its kind. By making use of the family trees from FamilySearch (FS), the lab is able to identify women's maiden names and all their married names.[7] This key innovation makes it possible to track women at (roughly) the same rates as men. It also makes it easier to track men.

Note then that our new dataset was *not* constructed by taking the original administrative records and matching them to one (or more) databases using an algorithm (which was how we proceeded in the original study). It was hand matched by individuals who are making use of multiple data sources. Our original data relied on linkage from MP records to just two primary sources: the 1940 U.S. Census of Population, and the Social Security DMF. The linkage was done using an algorithm that matched observations primarily using names and dates of birth. The RLL methodology relies on these databases but also incorporates additional sources that reveal links we did not uncover. For example, the age at death can be now ascertained from a death certificate found in the DMF, but it could also be found in state records, tomb records or be reported by family members. Also, because we observe married as well as single names, we can track women regardless of whether they are married or have remarried.

A.3 Comparison with earlier data.

These data sources allow us to overcome major limitations in our prior research.

First, we now cover deaths occurring from the time of the MP application until death. The DMF data only covers deaths occurring from roughly 1965 to 2012, which results in right and left censored age at death. This results in potentially biased estimates of the effects of the program (e.g. see Goldstein et al. 2023), an issue which we investigated extensively in prior work. For example, if an individual whose family appeared in the MP records died before 1965, they were not in the original data. A family tree or a state death record, however, would make this link possible. This difference in linking capacity by itself will result in the RLL strategy capturing more

[7] The lab proceeded in several steps to create the data. First it identified whether any member of a given family was already in a family tree. If not, then a tree was created. Second the RA were instructed to populate the data in the tree if the data was not already available. For example, if a person did not have a date of death in a tree, then the RA would look for that information and attach it and the source if they found it. The Online Appendix of Aizer et al. (2024) contains detailed information regarding this process.

pre-1965 deaths than we captured initially. The tree also captures deaths occurring more recently, up to 2019. Thus, in the new data the main reason why age at death is missing is that the age at death cannot be found, but we are fairly certain that none of the individuals we study are still alive.

Second, RLL's access to family trees and state marriage records also makes it possible for them to trace females, as both sources overcome the "surname stability" problem. The original methods which merged across databases using names and dates of birth would allow us to track very few women (less than 5%), and we never presented results by sex as a result. Our new data tracks men and women at roughly the same rates.

Third, these sources also make it possible to assess the reliability of links that we did make. Making a link between two sources is often easier when there is information from a third source which "bridges" the original sources, even if surnames remain stable. An MP program record may lack an exact date of birth and lead to several candidate links in the DMF which does include date of birth, leading to ambiguities that can be resolved in the presence of a family tree entry or a state birth certificate or church baptismal record. As a result, the data we obtained is of higher quality. For example, we were able to ascertain that several applicant records correspond to the same family which had applied multiple times, sometimes under different names (if e.g. the mother remarried) or with slightly different information on the dates of birth of children. We were also able to identify more records in which the dates of death are unique, obviating the need to choose a match or use other procedures to account for multiple matches.

Finally, family trees also allow us to observe a set of new predetermined covariates that we can use for our analysis. Specifically, we collect additional information about the mother: her age, her education (if she is still alive in the 1940 census), the number of kids she had that were too old to be listed in the welfare application, and the number of kids she had that had died prior to the application. We also know how many siblings the mother had, which welfare offices considered an alternative source of support. These are important new covariates that help us to assess the socio-economic status of the family at the time of application. We can also track the outcomes of the mother after she applied to the program: we are able to assess if she remarried, moved counties or if she had more kids after participating in the program as well as her age at death.

In sum, the new data are constructed using a different approach that uses many more sources of data and that relies on individuals for matching rather than on algorithms. As a result, our match rates are now substantially higher. We now observe age at death for than 85% of men and 73% of

women (compared to 52% and 0% in the original study), the incomes of 54% of men and 50% of women, and the education of more than 55% of men and women. These match rates are also substantially higher than the match rates typically obtained in historical research which range from 10% (e.g. Abramitzky, Boustan, and Eriksson, 2014) to 30% (Bailey et al., 2022). Data created in this manner, are often described as the "gold standard" and have been used as a benchmark for automated linkages (Bailey et al. 2020). Abramitzky et al. (2021) who compare machine and handmade methods of matching across censuses find that the quality of the resulting data is not very different and conclude that the primary advantage of using humans is their ability to use more sources of data, which is the case in this paper. We compare the data and the results with those reported in our previous paper in more detail in the last section of the paper.

A.4 County and State-level characteristics

We collected county and state level time varying characteristics which we merge with the records at the county and year of application level. At the county level, these include demographic characteristics (% rural, sex ratio, % married with kids in KF, % black and population) and state level characteristics including economic conditions (state manufacturing earnings/US manufacturing earnings), law regulating the employment and education of children (child labor and compulsory schooling laws), effective state-level social program generosity and state-level MP program characteristics.

B. Summary statistics for the analytical sample

We restrict attention to cohorts born 1900-1925 living in counties where we observe rejected applicants. **Table 1** shows summary statistics for the sample we tracked.[8] We include 29,417 individuals who were under the age of 14 when their mother applied for transfers around 1921. There come from 12,045 families who lived in 77 different counties (and thus 77 different MP programs) between 1911 and 1930. The average child in the sample was about 8 years old and lived in a family with about 3.5 children under the age of 14 (thus listed in the application). The family trees reveal that in addition to these listed siblings, children had another 1.5 siblings older than 14 (not listed in the application) and 0.2 siblings who had died prior to the application. Thus, on average these families were large – the mothers in these families had more than five children,

[8] Table A1 reports the state characteristics we also collected.

which is well above the average in the US in 1920.[9] The mothers were on average 37 years old, had less than 8 years of schooling (primary school) and at least half were widowed (marital status is missing in some records). Applicants lived in counties that were predominantly (54%) rural, above the national average in 1920. These predetermined characteristics can be used to predict the income of the family using the 1915 Iowa census.[10] The predicted family income is of $787 dollars.

Our main outcomes of interest are the lifespan (age at death), education and income of the children in applicant families. The mean age at death was almost 71, though the standard deviation of this outcome is large, almost 20 years. These children obtained about 10 years of school, with about a third completing high school. Their average (individual) income (from salary and wages) in 1939 was about $462 dollars, with only 55% reporting a non-zero income (in part because many women did not work).

About 12% of the applicants in our records were initially deemed eligible and thus filed an application, but they were eventually rejected. Among those accepted and receiving transfers the mean duration is 3.3 years and the annual transfers about $274 dollars, which is about 35% of the predicted family income.

C. Re-Assessing the identification assumption

Our key identification assumption is that families receiving transfers are similar or poorer than rejected families, in which case our estimates provide a lower bound of the impact of cash transfers on children. We re-assess this with our new data by comparing accepted and rejected families.

Table 1 shows the mean characteristics for each group. We continue to find that rejected applicants appear to be slightly better off than rejected applicants. Rejected children are older, come from smaller families with fewer young children, in which women are less likely to be widows, consistent with our earlier findings. Our new data reveals that they also have mothers that are older. They also had more children over the age of 14 who could help with additional resources. Most of these differences are statistically significant (the p-value of the differences is shown in the last column of **Table 1**).

[9] The total fertility rates in the US in 1920 was 3.64. See https://eh.net/encyclopedia/fertility-and-mortality-in-the-united-states/.

[10] To predict household income, we first estimate a regression using the 1915 Iowa Census, in which total income is regressed on family size, the ages of all siblings, maternal marital status, foreign-born status, race, educational attainment, occupational score and their interactions. Then we used the estimated coefficients to predict income in our sample.

To assess the magnitude of these differences, we look at differences in predicted family income. As **Figure 1** shows, the estimated density of predicted income for accepted families is shifted to the left, with a large share of families in the accepted group having very little income. On average we find that rejected families had incomes that were about 7% larger ($836 v. $780) and this is a statistically significant difference. This is similar to what we originally reported (a 9% greater income among rejected families) and is consistent with the evidence from rejection reasons. It confirms that the program operated as intended, turning away less needy recipients.

Overall rejected families were somewhat less poor, which will lead us to underestimate the benefits of cash transfers. We will assess the robustness of our results to including our newly collected covariates which do change our predicted incomes substantially, despite the fact they do not meaningfully change the gaps between accepted and rejected families.

D. Missing data

Although our match rates are high, individuals with missing data could be selected. Indeed, prior research finds that individuals from higher SES backgrounds are more likely to be represented in family trees (e.g. Black et al. 2025). Most crucial for our study is whether the match rates differ by accepted status. We test this directly by regressing a dummy equal to one if we found the outcome for the individual (and zero otherwise) on all our predetermined characteristics and the treatment dummy (accepted). We test to see if being in an accepted family predicts whether or not the data are observed.

We find that, once we control for the extensive set of controls, (the ones we are able to collect in the new data but not in the older data), being accepted into the program and thus receiving cash transfers does not predict the rate at which we find individuals' education or income. But it does predict whether or not we see lifespan at the 10% level (see **Table A.2**). However, in all cases the magnitudes of the differences are small. In the case of lifespan (the only outcome for which the gap is close to significant), we find 79.5% of rejected applicants' age at death, and accepted applicants at a rate of 82.3%, a difference of 2.8 percentage points or 3.5% of the baseline for rejected applicants (**Table 1**), which falls to 1.4 percentage points (1.8%) once we include covariates which is a small difference.

V. Re-evaluating the effects of cash transfers on children

We now investigate the impact of cash transfers on all children using our newly collected data.

a. **Main results**

We start by presenting some preliminary evidence of the results by looking at the raw distribution of outcomes for accepted and rejected children. **Figure 2** shows the distribution of our primary outcomes for accepted and rejected children. The distribution of the age at death among children is shifted to the right for accepted children relative to rejected children. On average accepted children lived to age about age 71 whereas rejected children lived to age about age 70 – an increase of about a year, which is statistically significant at the 5% level (**Table 1**).

We also plot the distribution of education. Here it is less clear whether the accepted or rejected children are better off but accepted children appear to have been slightly more likely to have graduated from high school and from primary school than rejected children. Indeed, the mean difference in years of education is only 0.13 years of school, a modest increase among accepted children, which is nevertheless statistically significant. Finally, we plot the distribution of income in 1939. These distributions appear very similar, but on average we find that accepted children earn lower incomes of $459 versus $487 among rejected children, a $28 difference that is marginally significant at the 10% level. These results suggest already that in the new data the estimates of the effects of cash transfers might be lower than in the original study which reported increases in lifespan of 1.5 years, in education of about 0.3 years and in income of about 10%.

However, since the accepted children appear to be on average slightly poorer, controlling for observables could matter. The impacts of cash transfers conditional on observables do affect these conclusions somewhat (**Table 2**). The GRF estimates which yield a point estimate of transfers on the log of lifespan of 0.012. While these estimates are economically meaningful – they imply an increase in the age at death of about 0.7 years – the estimate is not statistically significant at conventional levels. The estimated effect is lower if we use lifespan as the outcome, implying an increase of about half a year of life, also economically meaningful but not statistically significant.

Next, we look at education impacts. We report effects transfers on years of schooling and on the probability of graduating from high school. Both effects are statistically insignificant and small in magnitude. The GRF ATE implies that the transfers increased years of school by 0.04 years, less than 1 percent relative to the rejected mean, and less than 2 percent of a standard deviation for the rejected group. Although the figures suggested a larger fraction of accepted children graduated high school, the estimates for high school graduation rates are also very small, one percentage

point or less (relative to the rejected mean of 31%) and statistically insignificant. A potential concern with these results is that children born between 1920 and 1925 might still be attending school in 1940. But if we restrict attention to only those born between 1900 and 1920, the education effects remain small and statistically insignificant (**Table A14**).

We do find however significant effects on labor market outcomes. The program increased income by $31, which corresponds to a roughly 6.4% increase in income relative to the rejected mean of $487.3. This impact is both economically significant and statistically significant at the 5% level. The results are similar if we look at the log of income (which drops individuals with zero incomes): The coefficient implies an increase in income of 7.5%. We also find a positive and statistically significant effect on the probability of having positive incomes in 1939 of 0.018 percentage points, corresponding to an increase of 3.3% relative to the rejected mean. If we drop individuals born after 1920 these conclusions are similar (**Table A14**).

To confirm these results, we then look at whether individuals are working and at their occupation scores, which assign individuals to the 1950 average income in their occupation. Work is an indicator equal to one if the person reported being in the labor force. The point estimate of 0.028 percentage points implies a 6% increase in the fraction working relative to the rejected mean. Similarly, we find a positive and statistically significant effect of the transfers on occupation scores of about 0.48, roughly 3.8% relative to the rejected mean. These results suggest that children who received cash transfers were more likely to be in the labor force and entered better paid occupations compared to those who didn't.

b. Robustness

ATE v ATE overlap. A well-known issue with using propensity score based ATE estimates is that they are sensitive to having estimated propensities that are very close to 0 or 1. This is because the estimated ATE is a weighted average of the estimated individual treatment effects and uses the estimated propensities to construct weights. **Table 2** also presents results using the ATE overlap which addresses this issue. Our estimates are virtually identical when we use this alternative estimator, except for the estimate of the effects on the log of income which turn from positive and significant in the ATE (7.5%) to a much smaller and statistically insignificant effect (1.1%). However, since the estimated impacts on other labor market outcomes are almost identical regardless of the methods we use, we conclude that these effects are robust to how the ATE is computed.

Clustering levels. Since the level of clustering affects the GRF point estimates as well as their standard errors, we re-estimate the GRF ATEs for lifespan, education and income clustering at increasingly large levels: no clustering, family clustering, clustering at the county*year level and clustering at the county level (**Table A.3**). The estimated ATEs are vary somewhat across clustering choices for lifespan but they are similar in magnitude: they are the largest if we do not cluster (0.016) and the smallest if we cluster at the county*year (0.010), as we did in the original study. For education the point estimates are essentially 0 when we cluster at the family level or don't cluster but they increase very substantially to about 0.15 if we cluster at the county level, though they remain insignificant. For income the point estimates are smaller if we cluster at the county-year or county level, but the magnitudes remain economically significant, and all the point estimates are statistically significant in all cases. Thus, both the point estimates and their statistical significance are robust to our choice of cluster.

c. Heterogeneous treatment effects and targeting

We now explore sources of heterogeneity in the effects of cash treatment. **Figure A6** shows the distribution of the estimated treatment effects for our main outcomes. The mean effect for lifespan is 0.012. But there is significant heterogeneity in these point estimates, which range from -0.02 to 0.06. Though the average treatment effect is positive, about 12% of recipients have negative estimated treatment effects. We also see substantial heterogeneity in education effects, though the estimated effects are clearly centered around 0, which a very large fraction (x%) having negative impacts. All treatment effects fall within the -0.2 and 0.2 range however, so regardless of their sign, these effects are modest – recall that the standard deviation of education is 2.4 so the effects are a tenth of a standard deviation at most. Finally for income we see that most treatment effects are positive, though about 10% of the sample has negative effects. Although the mean treatment effect is $27 some individuals gained as much as $50 or $75.

While these figures suggest important heterogeneity, these treatment effects are estimated with error and thus this heterogeneity is not clearly significant. We formally test whether there is evidence of heterogeneity in treatment effects using the Target Operating Curves (TOC). These are shown for the full sample in **Figure 3** for the three main outcomes. All curves show that the treatment effects vary systematically across the treatment effect percentiles, but the standard errors are large, and they confidence intervals overlap with the 0 line except for income. A formal test confirms finds that we can reject the null of homogeneity for income and for education but not for

lifespan (**Table A19**). Interestingly we would not be able to reject homogeneity in our original data (labeled DMF) for any outcome, highlighting the value of our larger new sample.

A first implication of these results is that the average treatment on the treated (ATT) and the average treatment on the untreated (ATU) estimates might differ. This speaks to whether the program was well targeted: one hopes that the treatment effects would not be larger for the untreated. We do not find this to be the case here. For lifespan, education and income the ATT exceeds the ATU (**Table 2**). We do see slightly larger impacts on the untreated for the occupation score. However, the ATU and the ATT are not statistically different from each other across all outcomes so we cannot reject the null that the program was well targeted.

However, our results raise the question of whether the program could have been better targeted. We can obtain bounds on the gains from targeting using the distributions shown in **Figure A6**. About 12 of individuals in our data were rejected. We can ask whether average gains would have been substantially larger if the program had rejected the 12 percent of individuals that has the lowest treatment effects. This basic computation shows that these maximum gains are significant: if we targeted the top 88% of individuals with the larges treatment effects then the ATE for lifespan would be 0.021, about double what we see in the sample; and the income effects would be $42 instead of $30, an increase of more than a third. Thus there is evidence that the program could have resulted in substantially larger gains with a different targeting approach. But it is not clear whether these can be realized since they require targeting on unobserved gains.

We can investigate whether there are *identifiable* groups that may benefit more than others from the program – this knowledge could be used to improve targeting in practice. The first question we explore is whether targeting on income levels is appropriate. The program targeted the poor. Then as now the belief is that the poorest groups benefit the most from cash transfers and some prior work has indeed found as much (XX). **Figure A7** shows a bin scatter plot of the treatment effects (y-axis) plotted against predicted income levels (x-axis). For lifespan education and income treatment effects appear to decrease with baseline family income. However, if we estimate the best linear projection of the estimated CATEs on predicted income, we find that these relationships are not statistically significant. Moreover, we can see in the Figures that within income levels there is substantial heterogeneity in the CATEs. For example, the effects on lifespan range from 0.006 to 0.013 for predicted family incomes around $1,000. This analysis suggests then that targeting on income, among this already impoverished group, would not improve the total

gains associated with the transfers. This finding aligns with the recent work by Haushofer et al (2025) who also find that targeting solely on the basis of poverty does not maximize the effects of cash transfers in Kenya today.

To identify who would benefit more, we use the new methods proposed by Athey and Wager by projecting the estimated treatment effects on various predetermined covariates (**Table A15**). Consistent with the RATE results, with the exception of a child's age, there are no predetermined covariates that predict the treatment effects for lifespan, at least in a statistically significant fashion. Also recall that we cannot reject homogeneity in the treatment effects for lifespan. Interestingly the coefficient on age suggests larger gains for older (not younger) children. However a joint test that family characteristics matter cannot reject the null, nor can we rejected the null that county and state characteristics matter. So for lifespan, there are not observables in the data that can be used to target the program better. For education two covariates stand out: treatment effects are more positive in larger families and in which the oldest child is young. This suggests that the cash prevents families from sending their oldest child to work which is particularly beneficial if the oldest is young. The results also suggest that families face a quality-quantity trade-off in education that cash transfers alleviate. Maternal education also increases the effect of the cash transfers. However, the results for education are always small in magnitude and thus we do not give them much weight. Moreover again here we find that joint test cannot reject the null that individual and region characteristics matter.

The most interesting results pertain to the labor market outcomes – the results so far suggest these effects are large and statistically significant and there is meaningful statistically significant heterogeneity in the treatment effects. We find that effects are again larger for older children, whose mothers are more educated, and in families in which a child had died prior to the MP application, which suggests these families suffered from other difficult-to-overcome vulnerabilities, like poor maternal health or poor health environments. Jointly these family characteristics do predict gains at the 10% level (p-value 0.082) whereas location characteristics do not.

Before moving on, we investigate one last implication of heterogeneity: if treatment effects are heterogeneous then the estimation method matters and OLS will not give unbiased ATE.[11] **Table 2** shows that the estimation method does matter at least for some outcomes. For lifespan the OLS coefficients are smaller than either the ATE or the ATE overlap, but the differences are small and all estimates are statistically insignificant. For education OLS also yields lower (in fact negative) effects, but again these are very small and also insignificant. Thus, our conclusions for lifespan and education are similar using either method. The same is true for work and occupation scores. For all measures of income (total income, log income and having a positive income) however, the OLS coefficient is smaller than the GRF ATE and it is not significant whereas the GRF ATEs are. This is perhaps not surprising – income is the one outcome for which we find significant evidence of treatment effect heterogeneity.

In sum, the results show that cash transfers improved the labor market outcomes of children and potentially increased their lifespan though these estimates are noisy. But it did not meaningfully increase their education. Although there is evidence that cash transfers had heterogeneous effects on children, particularly in terms of income, the scope for improving targeting appears to be small, since there aren't many observables that could be used to better the program. Interestingly poverty status does not appear to be a good targeting mechanism and there is. In fact, there is some suggestive evidence that that gains from the program are larger for families with educated mothers and smaller in families in which children has died suggesting a trade-off between gains and deprivation.

It is helpful to compare these results to those reported in the original study. All the point estimates are lower than those reported in the original study which reported lifespan gains of 1.8 percent (vs 1.2% here), education gains of 0.3 years (compared to 0.04 here) and gains in income of about 10% (vs 7% here). We cannot reject the null that the estimates for all outcomes are the same as those presented in the original paper. However, all the point estimates are lower than in the original study. We investigate why this occurs. First, we investigate if the addition of girls into

[11] Słoczyński (2022) shows that OLS estimates can be particularly misleading when the treated and control groups differ dramatically in size as they do in our setting in which the "untreated" group account for 12% of the total sample. In the case of a binary treatment, the OLS estimate is a weighted sum of the ATT and the ATU, with the weights inversely related to the size of the group. In our analysis, this implies that the estimated OLS may reflect a substantial ATU and a small ratio of untreated to treated observations. Indeed Słoczyński (2022, pp. 507-8) uses our original data and adjusts for the disparity in the size of the treated and untreated groups, finding that doing so reveals ATE and ATT are both indistinguishable from zero, and that our previous results were driven by a large and precisely estimated ATU that exceeded the ATT.

the sample resulted on lower estimates. Then we investigate other reasons why our estimates are lower.

VI. Do girls benefit as much as boys from cash transfers?

The previous analysis did not uncover major differences in treatment effects by sex, at least conditional on other covariates. We revisit sex differences in more detail now and ask if the unconditional treatment effects differ by sex. A large literature has documented that boys and girls might be treated differently by parents, and that this varies depending on the sex composition of children. We investigate this now. This investigation can also potentially explain why our new estimates are lower than in our previous study.

We start by documenting that the outcomes of boys and girls differed substantially in our sample as shown in **Table 3**. Women in our sample live almost 7 longer than men consistent with the large gender gap in life expectancy that emerged in the early 20th century (Goldin and Lleras-Muney 2019). They have higher levels of education (they obtained 0.4 years more years of schooling, and they are 8 percentage points more likely to have graduated from high school) consistent with previous evidence that women graduated from high school at higher rates among cohorts born in the first half of the 20th century (Goldin and Katz 2008). They also have substantially lower incomes ($173 vs. $737 for men) and are less likely to have any positive incomes (30% vs. 79% of men). This is also to be expected since women from these cohorts worked at lower rates than men (25% v. 75%), and worked in lower paid occupations, as can be seen from their lower occupation scores (5.7 v. 19.6) (**Table 3**). All these differences are large and statistically significant. Thus, cash transfers may have had different effects on them.

a. Comparing the data for boys and girls

First, we verify that girls and boys come from similar families. **Table 3** shows this is by in large the case: except for one case, there are no statistically significant differences in the 30 or so pre-determined characteristics we observe across sexes.[12] Interestingly however, girls are more likely to be in rejected families, and they received larger amounts when accepted. This suggest potential discrimination against families with more girls, as the marginal accepted family was likely poorer. Although not significant, we do see that the predicted income for girls in accepted families is lower than the predicted income for boys in accepted families (thought the predicted

[12] Table A4 shows the mean differences by sex in state-level characteristics.

income of the rejected is also higher for boys). This suggests that the bias against finding positive effects for girls might be slightly larger than for boys.

We also check to see if accepted boys and girls are matched at different rates than rejected boys and girls (**Table A.5**). Once we include all the predetermined controls, for boys the match rates are identical for rejected and accepted for all outcomes. The match rates are also similar among accepted and rejected girls for education and income but we are more likely to find death records for girls in accepted families. This effect is small however: with controls the accepted girls are 2.7 percentage points more likely to have lifespan, an increase of 3.7 percent relative to a mean of 73%. This difference among girls likely caused the differences we observed in the overall sample in **Table A.2** and suggests that although we obtained data at very high rates for women the data are less good for women than men. We will consider how this affects our lifespan results.

b. Results by sex

We start by investigating the unconditional differences in outcomes of accepted and rejected children by sex. For both boys and girls the distribution of the age at death is shifted right for accepted children (**Figure A3**). Indeed, both accepted boys and girls lived about 0.9 years longer than rejected ones. For education both figures show a larger share of accepted children graduating high school, but the differences between accepted and rejected are small for both sexes (about 0.15 for boys and 0.09 for girls). For income the distributions appear identical by accepted status for boys and girls, but accepted boys' income is $46 lower than rejected boys income whereas it is $18 higher for accepted girls compared to rejected ones. However these results do not condition on covariates, and since accepted families are poorer, the effects are biased towards finding negative treatment effects.

The main results that control for covariates are shown separately by sex in **Table 4**. The effect of cash transfers on lifespan for girls are smaller than the effects for boys and this is true regardless of whether we use levels of logs, or estimate effects by GRF or OLS. For example, the ATE for boys is 0.16 and it is 0.007 for girls for the log of lifespan, and it is 0.58 for boys and 0.4 in the levels' specification. However, the estimates for boys and girls are both statistically insignificant. Moreover, we cannot reject the null that the coefficient estimates are identical across sexes.

We do not find any effects of the cash on the educational attainment of either girls or boys. In all cases magnitude of the estimates is small and the estimates are not statistically significant. If we use the upper bound of the confidence interval of the most positive estimates (the ATE for girls

and the ATE overlap for boys), we can reject gains larger than 0.25 years for girls and 0.23 for boys, which correspond to about 10% of a standard deviation in years of schooling.

Next, we look at labor market outcomes. The effects on income for girls are positive (about $14-$17) and stable across specifications, but they are statistically insignificant. Although the standard errors are large, the magnitudes are not trivial: the ATE implies a 10% increase relative to the mean of the rejected applicants. For boys we estimate effects that are larger in magnitude, ranging from $14 (OLS) to $43 (GRF), but they are statistically significant only in the GRF specification so again these estimates are somewhat sensitive to how we estimate treatment effects. Relative to the rejected mean these represent increases of 1.8%-5.8%. So, although in levels the effects are smaller for girls, we cannot reject that the point estimates are the same across sexes, and in percentage terms (relative to the rejected baseline) the effects are in fact larger for women.

We also look at the fraction working and at occupation scores. Cash transfers increase the fraction of women working by roughly 2-3 percentage points, a 8-14% increase relative to the rejected mean. The ATE overlap and OLS estimates are statistically significant at the 5% level. For boys the estimates are less stable ranging from a 0.016 statistically insignificant OLS estimate to a 0.033 statistically significant ATE and ATE overlap estimates. These point estimates imply a 2-4.4% increase relative to the rejected mean. So again, in percentage terms the effects are smaller for men, but we cannot reject that the point estimates are identical across sexes.

Finally, we look at occupation scores. These point estimates are positive for both sexes, but these vary more in magnitude and significance depending on how we estimate them (the effects range from 0.13 to 0.58 for girls and from 0.52 to 0.96 for boys). These estimates are statistically significant for boys but not girls, but again we cannot reject that they are statistically the same across sexes and in percentage terms they are larger for women (2.4-11%) than men (2.8-5%).

In sum, we find larger estimates for boys in terms of longevity and labor market outcomes, but we cannot reject these estimates are the same because they are imprecise. In percentage terms the effects for lifespan are also larger for boys than girls, but for labor market outcomes the opposite is true. Despite not controlling for other covariates, these results are consistent with the results from the best linear projection which also show larger effects for boys for both lifespan and income that are both insignificant.

c. Heterogeneity by type of family and other covariates.

The summary statistics suggested that families with mostly girls might have been treated differently from families with mostly boys. As noted earlier in many context families also treat boys and girls differently, particularly in families with children of both genders. We now study this by estimating the effect of transfers for boys in only boy families compared to boys in mixed families and for girls in only-girl families compared to girls mixed families. We do find evidence that family structure affects the pattern of gains associated with the transfers.

The results (**Table A7**) suggest that in families with both boys and girls of eligible ages, the transfers benefitted boys more than girls in terms of lifespan and income (there are no education effects in any type of families). In mixed families the ATE estimate of the impact of transfers on lifespan is 0.014 for boys and -0.001 for girls, the effects on income is about $62.56 for boys but 7 for girls, and the impacts on the occupation scores is 1.1 for boys but -0.07 for girls. Moreover, all the ATEs for income are statistically significant for boys in mixed sex families but not for girls in mixed sex families. A joint test that the effects for boys differ from the effects on girls in mixed families rejects the null for all labor outcomes at the 5% level (p-value 0.046) and rejects the null across all outcomes at the 10% level (p-value 0.056).

Children in families with only boys or only girls fare differently as well. Girls in girls-only families have systematically higher effects on lifespan and labor market outcomes than girls in mixed families: the lifespan effects are 0.015 in girls-only families compared to 0 in mixed families, and the income effects are $44.3 in girl only families compared to 7 in mixed families. Boys in boys-only families however do worse than boys in mixed families. For example, the income effects are -1.6 in boys-only families compared to 62.6 in mixed families. Because the sample sizes are substantially smaller for children in same-sex families, we cannot reject that the treatment effects are the same, but the point estimates are highly suggestive and in line with the results of previous work in developing countries.[13]

d. Are boys in poor families more sensitive to inputs?

[13] Are the effects of cash transfers heterogeneous for boys and girls? The RATE test rejects homogeneity in lifespan effects for boys and in income effects for girls. We investigate if any covariates predict heterogeneity in outcomes by sex. For boys, there are no covariates that individually predict larger treatment effects at statistically conventional levels. But we find that older boys and boys with more educated mothers had larger treatment effects for income as did those those in household where children had died benefitted less. For girls no covariates predict lifespan treatment effects, but many covariates appear significant in explaining education effects, with girls benefitting more in larger families in which the youngest child is older. For income the number of kids that died pre-application predicts smaller treatment effects as they do for boys.

We find evidence of lifespan gains for both sexes, and while the effects are economically meaningful and larger for boys, they are not statistically significant for either sex or different across sexes. For men we estimate and increase in lifespan of roughly 1.2% to 1.6%. The effect on income ranges from 2% to 5%. If we assume that the entire increase in lifespan is driven by the income increases, then the implied elasticity of longevity with respect to income ranges from 0.24 to 0.6. For women the elasticity is smaller ranging from 0.04 to 0.08. Thus, in its totality the evidence is consistent with the hypothesis that the lifespan of boys from poor families is more responsive to good shocks during childhood and later on, but we cannot fully confirm this in the data due to the imprecision of our estimates. However, these gains are larger for boys than girls in families with both boys and girls suggesting differential treatment within families. Thus the evidence is also consistent with differential treatment rather than differential sensitivity. Because most families are mixed, this might explain why boys see larger gains than girls overall.

VII. Understanding the discrepancies across studies for boys.

The previous section suggests one reason why this follow-up study has found lower estimates thatn the original study: we now include women, and they appear to have smaller gains in levels when compared to men. However, this is not the entire reason why our estimates differ. The point estimates for boys in the new data (reported in **Table 4**) are lower than those reported in the original study which reported lifespan gains of 1.8 percent (vs 1.2% here), education gains of 0.3 years (compared to at most 0.05 here) and gains in income of about 10% (vs 7% here). In this part of the paper, we investigate the reasons for these the discrepancies. We start by describing how the samples for boys differ and then investigate other possibilities.

a. Differences in the characteristics of the samples

In our original study (Aizer et al. 2016) we reported effects of cash transfers on boys using administrative records matched to administrative database using algorithms. We obtained mortality data by matching to death records in the Death Mortality files which cover deaths from 1965 to 2012. For this project we recollected the data for all outcomes by hand using genealogical databases and methods as described earlier.

We start by comparing the means for boys in the data used in previous paper, which we label DMF data (since it is based on the Death Mortality Files) and in the new data, which we label FS

data since it is derived from FamilySearch records (**Table 5**). To better understand how the data has changed we report the means for the boys that were in the original study but whose outcomes may have changed (column 3), the means for the boys that we have added to the sample (column 4) and finally for all the boys in the new data (column 2), which we already used for the analysis by sex.

The DMF data contains 7,860 boys matched to unique death records. The new data includes 14,286 boys, 7,387 from the original DMF sample (with the exception of 473 records we discovered were repeated – see earlier data discussion) and 6,899 newly added boys.[14] In principle the "Old sample" and the DMF sample should be identical in terms of predetermined characteristics (except for the few observations we corrected). However, the outcomes will differ because many boys will be matched to different outcomes using the new matching approach. The "New Sample" consists entirely of boys who were not tracked in the original study.

As expected, the age at death in the original sample is higher (73) than in the new sample (67.8). This is because most of the boys we previously did not find died before 1965 when the DMF records start to cover the population of deaths. This is made clearer when we compare the "Old Sample" column and the "New Sample" column. The age at death in the old sample was 72 (close but not identical to the 73 in the original study) but it is roughly 62 among the newly added boys. **Figure 4** shows distributions of age at death for each sample and shows quite clearly that in the newly added sample there are many more individuals dying before age 50. As a result, the standard deviation of lifespan increases quite substantially from 11.6 in the original data to more than 19 in the full FS data for boys. Thus, while the new data contains a substantially larger number of observations for boys, we do not have more power to detect lifespan changes in the new data, but substantially less.[15] This difference alone is likely to change our conclusions regarding the link between MP receipt and lifespan. The newly added boys also appear somewhat less well-off having lower educational attainment and incomes.

There are other noticeable differences. The duration of transfers and the amount of the transfers is lower among the newly added boys. These less generous benefits are likely driven by pre-

[14] There were 473 boys in the original DMF data that we found out to be repeated observations, repeated applicants or had been given the wrong sex in the original study. The columns labeled DMF still assumes these are distinct to match what was reported in the original study.

[15] The required sample size to detect a statistically significant (at the 5% level) mean change of 1 year given a mean of 73 and a standard deviation of 11.5, as in the DMF data is approximately 2,074 observations (statistical power of 80%). If the mean is 67.8 and the standard deviation is 22.68 then we require 8,066 observations.

existing differences: the newly added boys come from families with older mothers, who were less likely to be widows, and had more older children, suggesting their families are somewhat richer. Our predicted income measure indeed suggest that the newly added boys come from richer families compared to the old. The boys in the old sample had a predicted family income of ~779, whereas the newly added boys come from families with predicted incomes of ~816, a 5% difference. Finally, the newly added boys are more likely to come from rural counties, and come from less generous states (that spend less on schools and social programs, and have less generous MP laws, **Table A8**).

Importantly the rejected and accepted children may differ in the new and old samples. To assess this, we investigate differences in pre-determined characteristics by accepted status within each sub-sample (**Table A.9**). This table reveals that the gap in predicted income between the accepted and rejected families is larger in the new sample of boys (about $100) compared to the old sample ($19). This suggests that the bias against finding positive effects of cash transfers is potentially larger in the new data.

Finally, we check to see if we can observe boys' outcomes at the same rate in the new and old data. We find that once we add all controls then there is no difference in the match rates for lifespan, education or income in the new or the old sample of boys (**Table A.12**). the differences in the match rates across accepted and rejected groups is not only statistically insignificant but small. So differential match rates are unlikely to result in different results across the new and old samples.

In sum while the differences in the old and new data sets are not large, they suggest the newly added families likely stem to benefit less from transfers: they are from slightly richer families living more rural counties with less generous benefits de jure and de facto. We also find that in the newly added families the gap in income between accepted and rejected is larger creating a larger bias against finding effects. Finally in the case of lifespan, we observe a substantial increase in the standard deviation lowering our power to detect statistically significant differences.

b. Comparing estimates of the effects of transfers on lifespan across samples

Our original study focused on lifespan as the primary outcome. **Table 6** shows the results for lifespan across different samples. In Column 1 we replicate the results using the DMF data for reference. Then we show the results using the new FS data for all the boys in the new data, for the old sample of boys, and the newly added boys. In Panel A we report OLS coefficients without

controls, in Panel B we report results controlling for the predetermined characteristics included in the original study, in Panel C we control for all available covariates, and finally in Panel D we report the ATEs from the GRF approach.

Comparing the DMF column with the "All FS boys" column we see that without controls the FS estimates are larger (0.018) than the DMF ones (0.013) but the FS estimates are statistically insignificant whereas the DMF ones are not, primarily because the FS standard errors are more than twice as large. If we control for the old set of predetermined characteristics the OLS estimates increase to 0.018 in the DMF sample, but fall substantially in the FS sample (0.010). Panel C shows that the FS estimates remain the same (0.009) if we add the new controls available in the FS sample only. Finally Panel D shows that the GRF estimates using the DMF data and the full FS data are in fact identical (0.015 vs. 0.016) but again the FS ones are not statistically significant because the standard errors are twice as large. We cannot reject the null that the DMF estimates and the GRF estimates are the same for the log of lifespan.

Two conclusions emerge from these comparisons. First, the results are somewhat sensitive to how we estimate effects, particularly in the FS data, which is perhaps not surprising since the RATE test for boys rejected homogenous effects in the FS data but not in the DMF data (**Table A19**). Second, the standard errors are much larger in the FS sample. Our hypothesis based on the summary statistics is that this is due to the increased window over which we observe mortality. We investigate this by restricting the deaths in the FS analysis to 1965 to 2012 to roughly match the DMF dates. Indeed, if we make this restriction, we find that the standard errors in the DMF and FS data are now identical (roughly 0.005-0.007) and the GRF ATE in the FS sample becomes statistically significant (**Table A11**).

Table 6 also replicates the analysis for the old and new subsamples in the DMF. Note that in the "old" sample the age at death has changed even though the children are the same because we use more data sources to ascertain the validity of a match. It shows that this similarity of results across samples masks significant differences. In the old sample we estimate larger and statistically significant effects of the transfers on lifespan ranging from 0.029 (raw gaps) to (0.042) in the GRF approach. In contrast the treatment effects are negative and insignificant in the new sample across all specifications.

Also noteworthy is that the ATE is in fact still positive and statistically significant for the DMF data and it is very similar to the OLS estimate, in contrast to what is reported by Słoczyński (2022)

who used our DMF data and a linear propensity score and found that the ATE was lower than the OLS. We further assess the sensitivity of the results to how we estimate the ATE (**Table A.13**). If we estimate the ATE using a linear propensity score, we can replicate the results reported by Słoczyński (2022) exactly: we find the ATE to be 0.011, with the ATT being 0.010 and the ATU being larger at 0.021. But this result is different if we use the Athey and Wager methods. In the same DMF data, we find that the ATE is 0.015, with the ATT and the ATU being essentially identical (0.0014-5). In the FS data we always find that the ATT is larger than the ATU. Thus, the results Słoczyński (2022) appear to be driven by his choice to estimate the ATE using a linear propensity score which gives a different result in the DMF data. This shows that the differences between GRF and OLS estimates are not only due to the fact that treatment effects are heterogenous and weighted differently by each approach, but also that how one controls for covariates is also important. Since the GRF methods are more non-parametric, we prefer these to OLS ones.

c. Estimates for education and income

We re-estimate effects on education and income in **Table 7**. In Panel A we report the results using the DMF (original data). We find an increase in education of about 0.3-0.4 years in either OLS or GRF, and this is significant with the GRF method. We also find a statistically significant increase in income of 100 dollars or more using both methods. Notice however that the samples used to estimate these are small, including fewer than 2,000 boys. Panel B shows the results using the new FS data for boys, which includes more than 9,000 boys. We do not find effects on education in the new sample and the effects for income are of around $40, a third as large as in the original study. For education we can reject the null that the new and the old results are the same. To understand why the results, differ, Panels C and D show the results in the old and the new FS samples. In both samples the effects of education are small and statistically insignificant, though the estimates are larger (0.14 v -0.06) in the old sample of boys. Thus, for education the new results show that the original data overestimated the effects due to having small samples.

On the other hand, the results for labor market outcomes persist in the old sample of boys, though the effects are smaller in magnitude. The effects for income are a quarter of the original size. But the ATEs are similar in the old and the new samples (though not in the OLS) though the effects are again a bit smaller in the new sample.

We did not collect work and occupation score in the old study. But we can investigate differences in these outcomes to confirm the findings for income. In the overall sample of boys, the GRF estimates significant increases in the fraction working and in occupation scores. These effects on work are entirely driven by the old sample of boys. The fraction working in this sample increases by a statistically significant 4.4-4.9 percentage points (a roughly 6% increase relative to the rejected mean) in the old sample but the point estimates are essentially 0 in the new sample. The point estimates for occupation scores are larger in the old sample compared to the new sample (1.1 versus 0.7) but they are similar in both.

In sum, in the new larger data we confirm the original estimates for lifespan (in percentage terms) but the new estimates are less precise because the distribution of lifespan in the new data is wider, and they are smaller in absolute terms because the baseline lifespan is lower. We find no impact on education in contrast to the earlier findings which appear to have resulted from the small sample size we had in the original study. We find a smaller but still meaningful impact on income and in other labor market outcomes. This is consistent with our findings that the treatment effects on income have substantial heterogeneity. The new estimates are smaller because the newly added observations for boys have null or smaller treatment effects. Newly added families are less poor and have a larger gap between accepted and rejected. Thus, the effects are more biased against finding a positive effect of the program – we always find that treatment effects across all outcomes are smaller in the new sample consistent with this hypothesis. Finally recall that we find smaller effects for girls in absolute terms for all outcomes, and this also contributes to our finding smaller estimates for lifespan and income than in the original study.

VII. Discussion

In this paper we estimate the effects of cash transfers to poor women with children on the lifetime outcomes of the children by investigating the long run impact of the Mothers' Pension program on children. Our original research found large impacts on boys' lifespan, education and income. Using newly collected data we find positive and statistically impacts on labor market outcomes among both boys and girls, and significant but noisy impacts on lifespan which are larger for boys than girls. But the effects on education disappear.

The effects in the original study were larger than in this replication. A finding that effects sizes are smaller in replications is common. Recent efforts in psychology and economics to replicate

noteworthy studies typically find that effect sizes fall by about 40-50% even among studies that replicate (Open Science Collaboration, 2015; Camerer et al 2016, 2018). In psychology the systematic decline in estimated effects as replications mount is referred to as the "decline effect". We confirm these observations in this replication. There are several reasons for this in our case: the data on lifespan has larger standard deviations and thus we have lower power than originally. Second the treatment effects are smaller for the newly added children, including girls for whom the estimates are smaller. Third the gap between accepted and rejected is larger in the new sample and thus our estimates are more biased towards zero. Finally some of the original effects (the effect for education) appear to have been driven by small samples.

The heterogeneity in treatment effects across sexes and populations suggest that not all cash transfers will have similar benefits to the ones we estimate here. In fact, we find there are individuals for whom the gains were negative. However, we found very little indication of how to improve the targeting of cash transfers using observable characteristics. But we document that targeting on the basis of poverty might not result in the largest gains. Thus, an important question for future research will be to understand what characteristics and circumstances are the best predictors of gains.

What do these new results imply about the value of cash transfer programs? Our estimates still imply that the marginal value of public funds (MVPF) is XX, greater than 1, even accounting for maternal behavioral responses, so the program still pays for itself (Aizer et al. 2023). Although the MVPF remains positive using these new estimates, they lower the MVPF we computed with the results of the original study. The findings suggest that other programs targeting poor families might have larger returns than cash transfers. For example, the MVPF for the CCC program, which provided employment to poor young males during the great recession, is around 6. Thus going forward the key question for policy is not whether cash transfers benefit children and are cost-effective, though clearly more evidence is also needed to best understand the effects of cash transfers. Instead, more research is needed to establish whether cash transfers are the most effective and efficient programs or whether other programs would yield greater social benefits.

References

Aizer, A., Eli, S., Ferrie, J. and Lleras-Muney, A., 2016. The long-run impact of cash transfers to poor families. American Economic Review, 106(4), pp. 935-971.

Akee, R. K., Copeland, W. E., Keeler, G., Angold, A., & Costello, E. J. (2010). Parents' incomes and children's outcomes: a quasi-experiment using transfer payments from casino profits. *American Economic Journal: Applied Economics*, *2*(1), 86-115.

Athey, S., Tibshirani, J. and Wager, S., 2019. Generalized Random Forests. *The Annals of Statistics*, *47*(2), pp.1148-1178.

Athey, S., & Wager, S. (2019). Estimating treatment effects with causal forests: An application. *Observational studies*, *5*(2), 37-51.

Bailey, M. J., Hoynes, H., Rossin-Slater, M., & Walker, R. (2024). Is the social safety net a long-term investment? Large-scale evidence from the food stamps program. *Review of Economic Studies*, *91*(3), 1291-1330.

Bulman, G., Fairlie, R., Goodman, S., & Isen, A. (2021). Parental resources and college attendance: Evidence from lottery wins. *American Economic Review*, *111*(4), 1201-1240.

Burns, D. M., Lee, L., Shen, L. Z., Gilpin, E., Tolley, H. D., Vaughn, J., & Shanks, T. G. (1997). Cigarette smoking behavior in the United States. *Changes in cigarette-related disease risks and their implication for prevention and control*, 13-42.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T. H., Huber, J., Johannesson, M., ... & Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, *351*(6280), 1433-1436.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., ... & Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature human behaviour*, *2*(9), 637-644.

Christensen, G., & Miguel, E. (2018). Transparency, reproducibility, and the credibility of economics research. *Journal of Economic Literature*, *56*(3), 920-980.

Drevenstedt, G. L., Crimmins, E. M., Vasunilashorn, S., & Finch, C. E. (2008). The rise and fall of excess male infant mortality. *Proceedings of the National academy of Sciences*, *105*(13), 5016-5021.

DeWitte, S.N., 2017. Sex and frailty. *Exploring Sex and Gender in Bioarchaeology; Agarwal, SC, Wesp, JK, Eds*, pp.189-220.

Duvendack, M., Palmer-Jones, R., & Reed, W. R. (2017). What is meant by "replication" and why does it encounter resistance in economics?. *American Economic Review*, *107*(5), 46-51.

Eli, S., P. Fishback, N. Kannan, A. Lleras-Muney and J. Uguccioni (2025). The Emergence of the Modern Welfare System: Evidence from the Mothers' Pension Program, Mimeo.

Goldin, C. D., & Katz, L. F. (2008). *The race between education and technology*. harvard university press.

Goldin, C., & Lleras-Muney, A. (2019). XX> XY?: The changing female advantage in life expectancy. *Journal of health economics*, *67*, 102224.

Goldstein, J. R., Osborne, M., Atherwood, S., & Breen, C. F. (2023). Mortality modeling of partially observed cohorts using administrative death records. *Population Research and Policy Review*, *42*(3), 36.

Gray-Lobe, G., Pathak, P. A., & Walters, C. R. (2023). The long-term effects of universal preschool in Boston. *The Quarterly Journal of Economics*, *138*(1), 363-411.

Haushofer, J., Niehaus, P., Paramo, C., Miguel, E., & Walker, M. (2025). Targeting impact versus deprivation. *American Economic Review*, *115*(6), 1936-1974.

Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour economics*, *19*(4), 451-464.

Jayachandran, S. (2015). The roots of gender inequality in developing countries. *Annual review of economics*, *7*(1), 63-88.

Lleras-Muney, A., Price, J., & Yue, D. (2022). The association between educational attainment and longevity using individual-level data from the 1940 census. *Journal of Health Economics*, *84*, 102649.

Lundberg, E. O. (1928). Progress of Mothers' Aid Administration. *Social Service Review*, *2*(3), 435-458.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.

Page, M. E. (2024). New Advances on an Old Question: Does Money Matter for Children's Outcomes?. *Journal of Economic Literature*, *62*(3), 891-947.

Shah, H., & Gennetian, L. A. (2024). Unconditional cash transfers for families with children in the US: a scoping review. *Review of Economics of the Household*, *22*(2), 415-450.

Słoczyński, T., 2022. Interpreting OLS estimands when treatment effects are heterogeneous: Smaller groups get larger weights. The review of economics and statistics, 104(3), pp. 501-509.

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *113*(523), 1228-1242.

Tables

Table 1: Summary Statistics

	All	Accepted	Rejected	p-value
Panel A. Outcomes				
Age at Death	70.922	71.039	70.037	0.010
	(19.451)	(19.41)	(19.736)	
% Observed	0.819	0.823	0.795	
Education (1940 Census)	9.967	9.982	9.851	0.017
	(2.312)	(2.298)	(2.422)	
% Observed	0.603	0.608	0.562	
$\mathbb{1}$(Education $\geq$ 12)	0.339	0.341	0.320	0.067
	(.473)	(.474)	(.467)	
Income (1940 Census)	462.812	459.701	487.300	0.078
	(634.83)	(628.98)	(678.789)	
% Observed	0.559	0.564	0.522	
$\mathbb{1}$(Income $>$ 0)	0.555	0.556	0.549	0.538
	(.497)	(.497)	(.498)	
Work	0.505	0.506	0.497	0.468
	(.5)	(.5)	(.5)	
Occupation Score	12.653	12.654	12.647	0.983
	(12.774)	(12.782)	(12.713)	
Panel B. Program Characteristics				
% Accepted	0.879	1.000	0.000	.
Duration	3.353	4.241	0.000	0.000
Amount	274.828	319.867	0.000	0.000
Panel C. Predetermined Characteristics				
Year of Birth	1913.236	1913.427	1911.844	0.000
Child Age	8.366	8.323	8.680	0.000
Child Gender Male	0.486	0.483	0.504	0.019
Widow	0.525	0.528	0.501	0.003
Year of Application	1921.606	1921.750	1920.556	0.000
Number of Children in Application	3.409	3.462	3.025	0.000
Age of Oldest Child in Application	11.310	11.318	11.250	0.302
Age of Youngest Child in Application	5.230	5.131	5.955	0.000
Predicted Income	787.140	780.626	836.252	0.004
Panel D. Additional Predetermined Characteristics				
Mom Age	37.082	37.036	37.419	0.004
Mom Education	7.638	7.656	7.502	0.000
Number of Kids Over 14 (Not in App.)	1.539	1.518	1.686	0.000
Number of Kids Died Pre-Application	0.227	0.228	0.216	0.251
Number of Mom's Siblings	4.326	4.324	4.340	0.836
Panel E. County Characteristics				
Population	49332.5	50708.5	39303.1	0.000
% Rural	0.538	0.534	0.565	0.000
Sex Ratio	1.097	1.098	1.092	0.005
% White Fem. Married w. Kids in LF	0.052	0.053	0.046	0.000
% Black	0.014	0.014	0.014	0.412
Observations				
Number of Individuals	29417	25868	3549	
Number of Families	12045	10359	1693	
Number of Counties	77	76	77	

Notes: Predicted Income is the estimated income based on observable characteristics of the family using 1915 Iowa Census. The characteristics used include family size, the ages of all siblings, maternal marital status, foreign-born status, race, educational attainment, occupational score, and their interactions. The reported p-value comes from a two-sample t test of each characteristic.

Table 2: Estimates of the Effect of Receiving MP Transfers

	ATE	ATT	ATU	ATE Overlap	OLS	Mean Rejected	N
Panel A. Age at Death							
Log (Age at Death)	0.012	0.012	0.010	0.011	0.009	4.179	24089
	(0.009)	(0.011)	(0.010)	(0.010)	(0.010)		
Age at Death	0.503	0.504	0.465	0.480	0.409	70.037	24098
	(0.409)	(0.521)	(0.427)	(0.445)	(0.443)		
Panel B. Education							
Years of Education	0.041	0.052	0.017	0.017	-0.056	9.851	17727
	(0.061)	(0.076)	(0.070)	(0.072)	(0.067)		
$\mathbb{1}$(Education $\geq$ 12)	0.009	0.011	0.002	0.002	-0.012	0.320	17727
	(0.012)	(0.015)	(0.013)	(0.013)	(0.013)		
Panel C. Labor Market							
Income	30.971**	33.070**	28.125*	26.481*	15.301	487.300	16430
	(12.234)	(14.991)	(15.271)	(15.587)	(16.223)		
Log(Income)	0.075**	0.090**	0.035	0.036	0.011	6.393	9123
	(0.034)	(0.042)	(0.034)	(0.034)	(0.032)		
$\mathbb{1}$(Income $>$ 0)	0.018*	0.016	0.025**	0.022**	0.015	0.549	16430
	(0.011)	(0.014)	(0.011)	(0.011)	(0.012)		
Work	0.028***	0.026**	0.033***	0.031***	0.028**	0.497	17937
	(0.010)	(0.013)	(0.010)	(0.011)	(0.011)		
Occupation Score	0.480**	0.373	0.804***	0.705***	0.612**	12.647	17937
	(0.243)	(0.308)	(0.255)	(0.264)	(0.279)		

Notes: ATE, ATT, and ATU are estimated using augmented inverse-propensity weighting (AIPW), a doubly robust method for average treatment effect estimation. ATE Overlap reports an estimand that avoids dividing by estimated propensity scores. We cluster at the family-level. * p<0.1,** p<0.05, *** p<0.01.

Table 3: Summary Statistics by Gender

	Boys	Girls	p-value
Panel A. Outcomes			
Age at Death	67.817	74.326	0.000
	(19.043)	(19.324)	
% Observed	0.882	0.760	
Education (1940 Census)	9.747	10.186	0.000
	(2.387)	(2.214)	
% Observed	0.618	0.588	
$\mathbb{1}$(Education $\geq$ 12)	0.298	0.379	0.000
	(.457)	(.485)	
Income (1940 Census)	737.494	173.684	0.000
	(710.176)	(365.177)	
% Observed	0.590	0.529	
$\mathbb{1}$(Income $>$ 0)	0.790	0.308	0.000
	(.407)	(.462)	
Work	0.755	0.257	0.000
	(.43)	(.437)	
Occupation Score	19.595	5.775	0.000
	(11.628)	(9.790)	
Panel B. Program Characteristics			
% Accepted	0.875	0.884	0.019
Duration	3.315	3.402	0.157
Amount	270.161	280.857	0.000
Panel C. Predetermined Characteristics			
Year of Birth	1913.267	1913.207	0.409
Child Age	8.336	8.394	0.226
Widow	0.521	0.528	0.258
Year of Application	1921.624	1921.589	0.565
Number of Children in Application	3.402	3.416	0.484
Age of Oldest Child in Application	11.283	11.334	0.238
Age of Youngest Child in Application	5.219	5.241	0.596
Predicted Income	795.511	779.134	0.187
Predicted Income (Accepted)	788.451	773.198	0.247
Predicted Income (Rejected)	847.036	825.336	0.552
Panel D. Additional Predetermined Characteristics			
Mom Age	37.040	37.122	0.345
Mom Education	7.616	7.658	0.080
Number of Kids Over 14 (Not in App.)	1.536	1.541	0.862
Number of Kids Died Pre-Application	0.226	0.227	0.837
Number of Mom's Siblings	4.289	4.361	0.146
Panel E. County Characteristics			
Population	49001.5	49645.0	0.421
% Rural	0.538	0.538	0.798
Sex Ratio	1.097	1.097	0.783
% White Fem. Married w. Kids in LF	0.052	0.052	0.734
% Black	0.014	0.014	0.691
Observations			
Number of Individuals	14286	15131	
Number of Families	8698	9010	
Number of Counties	74	77	

Notes: Predicted Income is the estimated income based on observable characteristics of the family using 1915 Iowa Census. The characteristics used include family size, the ages of all siblings, maternal marital status, foreign-born status, race, educational attainment, occupational score, and their interactions. The reported p-value comes from a two-sample t test of each characteristic.

Table 4: Estimates of the Effect of Receiving MP Transfers by Gender

	ATE	ATE Overlap	OLS	Mean Rejected	Z-test (p-value)	N
Panel A. Boys						
Age at Death						
Log (Age at Death)	0.016	0.012	0.009	4.135	0.813	13026
	(0.013)	(0.014)	(0.013)			
Age at Death	0.586	0.359	0.401	67.280	0.776	13033
	(0.534)	(0.579)	(0.573)			
Education						
Years of Education	-0.005	0.045	-0.030	9.638	0.624	9099
	(0.078)	(0.091)	(0.087)			
$\mathbb{1}$(Education ≥ 12)	0.000	0.007	-0.004	0.282	0.663	9099
	(0.015)	(0.017)	(0.016)			
Labor Market						
Income	43.765**	37.748	14.129	752.954	0.420	8667
	(19.810)	(24.596)	(24.030)			
Log(Income)	0.062*	0.028	0.001	6.550	0.686	6717
	(0.034)	(0.036)	(0.033)			
$\mathbb{1}$(Income > 0)	0.018	0.018	0.004	0.775	0.659	8667
	(0.013)	(0.014)	(0.014)			
Work	0.033**	0.029**	0.016	0.732	0.875	9205
	(0.013)	(0.014)	(0.013)			
Occupation Score	0.810**	0.956***	0.525	18.975	0.330	9205
	(0.327)	(0.367)	(0.351)			
Panel B. Girls						
Age at Death						
Log (Age at Death)	0.007	0.008	0.006	4.237		11063
	(0.012)	(0.014)	(0.014)			
Age at Death	0.404	0.608	0.351	73.696		11065
	(0.563)	(0.657)	(0.647)			
Education						
Years of Education	0.089	-0.019	-0.099	10.098		8628
	(0.079)	(0.093)	(0.088)			
$\mathbb{1}$(Education ≥ 12)	0.018	-0.004	-0.022	0.365		8628
	(0.018)	(0.019)	(0.018)			
Labor Market						
Income	16.687	13.746	14.492	163.744		7763
	(12.856)	(16.716)	(16.771)			
Log(Income)	0.110	0.066	0.046	5.851		2406
	(0.082)	(0.086)	(0.082)			
$\mathbb{1}$(Income > 0)	0.018	0.028	0.027	0.273		7763
	(0.018)	(0.018)	(0.018)			
Work	0.023	0.033**	0.036**	0.225		8732
	(0.016)	(0.017)	(0.017)			
Occupation Score	0.133	0.439	0.585	5.318		8732
	(0.364)	(0.384)	(0.383)			
Joint Test						
$\beta_{Boys} = \beta_{Girls}$	0.860					

Notes: ATE is estimated using augmented inverse-propensity weighting (AIPW), a doubly robust method for average treatment effect estimation. ATE Overlap reports an estimand that avoids dividing by estimated propensity scores. We cluster at the family-level. The Z-test reports p-values from Wald tests assessing whether the ATE Overlap coefficient in the boy sample equals that in the girl sample, under the assumption of independent samples. Joint Test panel reports p-value from a Wald test assessing whether the ATE Overlap coefficients are equal for Log (Age at Death), Years of Education, Income, Work and Occupation Score. * p<0.1,** p<0.05, *** p<0.01.

Table 5: Summary Statistics - Original and New Data

	DMF	FS Boys		
		(All)	(Old Sample)	(New Sample)
Panel A. Outcomes				
Age at Death	73.312	67.817	72.115	62.275
	(11.587)	(19.043)	(14.237)	(22.68)
% Observed	1.000	0.882	0.961	0.798
Education (1940 Census)	9.776	9.747	9.811	9.662
	(2.589)	(2.387)	(2.352)	(2.43)
% Observed	0.980	0.618	0.684	0.547
$\mathbb{1}$(Education ≥ 12)	0.318	0.298	0.308	0.285
	(.466)	(.457)	(.462)	(.452)
Income (1940 Census)	733.124	737.494	744.590	727.966
	(741.797)	(710.176)	(704.598)	(717.585)
% Observed	0.934	0.590	0.654	0.521
$\mathbb{1}$(Income > 0)	0.782	0.790	0.791	0.789
	(.413)	(.407)	(.406)	(.408)
Work		0.755	0.763	0.744
		(.43)	(.425)	(.436)
Occupation Score		19.595	19.794	19.330
		(11.628)	(11.563)	(11.711)
Panel B. Program Characteristics				
% Accepted	0.875	0.875	0.885	0.864
Duration	4	3.315	3.539	3.053
Amount	272	270.161	274.103	265.943
Panel C. Predetermined Characteristics				
Year of Birth	1913.463	1913.267	1913.445	1913.076
Child Age	8.373	8.336	8.402	8.265
Widow	0.547	0.521	0.539	0.502
Year of Application	1921.795	1921.624	1921.839	1921.393
Number of Children in Application	3.707	3.402	3.391	3.414
Age of Oldest Child in Application	11.515	11.283	11.294	11.272
Age of Youngest Child in Application	4.952	5.219	5.227	5.210
Predicted Income	391.770	795.511	778.784	815.703
Predicted Income (Accepted)	387.258	788.451	776.647	802.916
Predicted Income (Rejected)	423.331	847.036	795.306	903.046
Panel D. Additional Predetermined Characteristics				
Mom Age		37.040	36.926	37.161
Mom Education		7.616	7.688	7.539
Number of Kids Over 14 (Not in App.)		1.536	1.509	1.566
Number of Kids Died Pre-Application		0.226	0.226	0.226
Number of Mom's Siblings		4.289	4.468	4.098
Panel E. County Characteristics				
Population		49001.5	51539.6	46283.8
% Rural		0.538	0.524	0.552
Sex Ratio		1.097	1.088	1.106
% White Fem. Married w. Kids in LF		0.052	0.053	0.052
% Black		0.014	0.015	0.014
Observations				
Number of Individuals	7860	14286	7387	6899
Number of Families	5738	8698	5420	5160
Number of Counties	75	74	74	74

Notes: The DMF column refers to boys in the original data based on the Death Mortality Files. The FS columns use data derived from FamilySearch records, with "Old Sample" indicating boys from the original sample and "New Sample" indicating newly added boys. Predicted Income is the estimated income based on observable characteristics of the family using 1915 Iowa Census. The characteristics used include family size, the ages of all siblings, maternal marital status, foreign-born status, race, educational attainment, occupational score, and their interactions.

Table 6: Estimates of the Effect of Receiving MP Transfers on Age at Death - Original and New Data

	DMF	FS Boys		
		(All)	(Old Sample)	(New Sample)
Panel A. OLS without Controls				
	0.013**	0.018	0.029**	-0.012
	(0.006)	(0.013)	(0.012)	(0.023)
Panel B. OLS with Old Controls				
	0.018***	0.010	0.023*	-0.008
	(0.006)	(0.014)	(0.012)	(0.024)
Panel C. OLS with Full Controls				
		0.009	0.023*	-0.010
		(0.013)	(0.013)	(0.024)
Panel D. Generalized Random Forest				
ATE	0.015**	0.016	0.042***	-0.016
	(0.006)	(0.013)	(0.010)	(0.026)
ATE Overlap	0.015**	0.012	0.027**	-0.019
	(0.006)	(0.014)	(0.012)	(0.025)
Observations	7859	13026	7138	5791
Fraction Accepted	0.875	0.877	0.885	0.867
Mean of Accepted	73.437	68.079	72.335	62.698
Mean of Rejected	72.440	67.280	70.633	63.648

Notes: The DMF column refers to boys in the original data based on the Death Mortality Files. The FS columns use data derived from FamilySearch records, with "Old Sample" indicating boys from the original sample and "New Sample" indicating newly added boys. ATE is estimated using augmented inverse-propensity weighting (AIPW), a doubly robust method for average treatment effect estimation. ATE Overlap reports an estimand that avoids dividing by estimated propensity scores. "Old controls" are child age, length of family name, family size, age of youngest and oldest child in family, maternal marital status. "Full controls" add mother's age, mother's number of siblings, foreign status, number of kids over 14, number of kids died pre-application, and husband's lifespan. We cluster at the family-level. * p<0.1,** p<0.05, *** p<0.01.

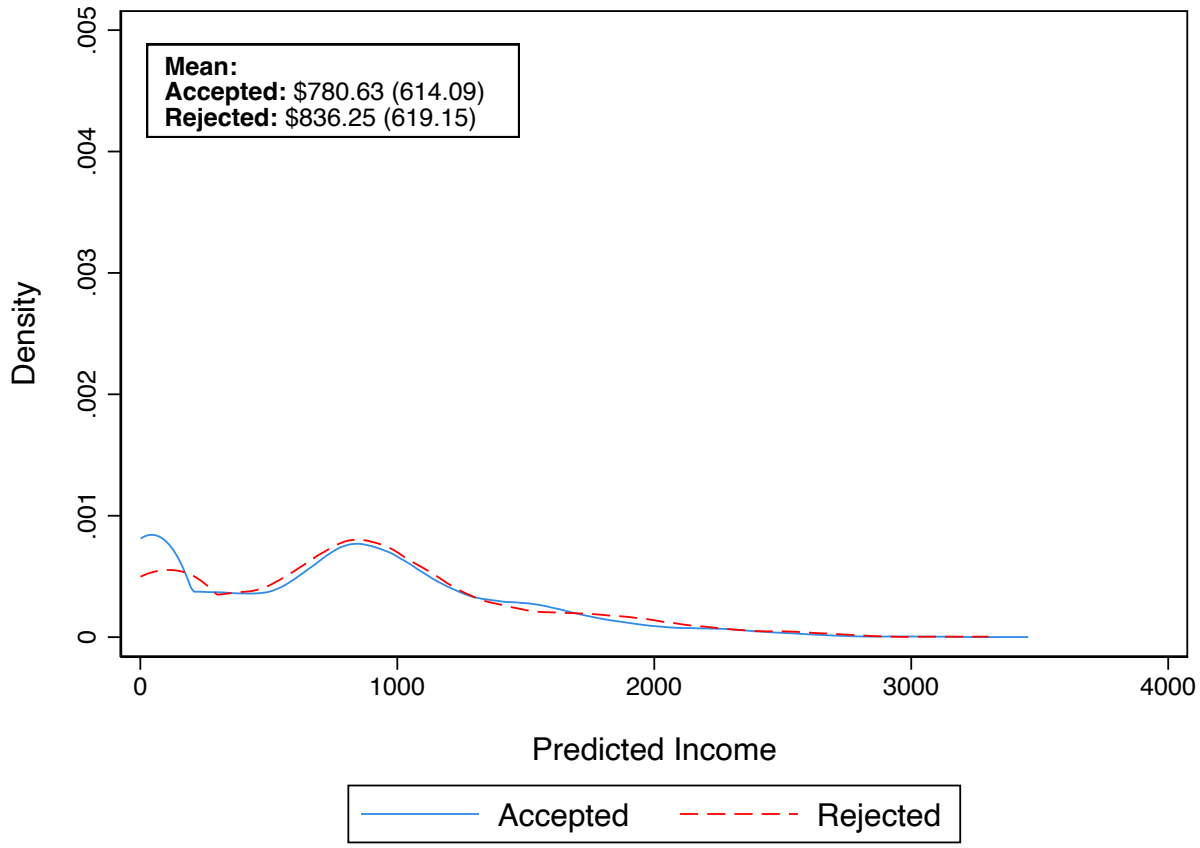
Table 7: Estimates of the Effect of Receiving MP Transfers - Original and New Data

	ATE	ATE Overlap	OLS	Mean Rejected	Z-test (p-value)	N
Panel A. DMF						
Years of Education	0.358*	0.431**	0.310	9.366		1975
	(0.193)	(0.192)	(0.201)			
Income	141.082***	121.951**	101.261**	670.658		1881
	(42.449)	(48.507)	(50.494)			
Panel B. FS Boys						
Years of Education	-0.005	0.045	-0.030	9.638	0.070	9099
	(0.078)	(0.091)	(0.087)			
Income	43.765**	37.748	14.088	752.954	0.122	8667
	(19.810)	(24.596)	(24.032)			
Work	0.033**	0.029**	0.016	0.732	.	9205
	(0.013)	(0.014)	(0.013)			
Occupation Score	0.810**	0.956***	0.525	18.975	.	9205
	(0.327)	(0.367)	(0.351)			
Panel C. FS Boys Old						
Years of Education	0.069	0.141	0.067	9.583	0.201	5074
	(0.095)	(0.121)	(0.114)			
Income	47.551*	44.677	38.665	763.846	0.192	4849
	(24.776)	(34.030)	(32.482)			
Work	0.049***	0.044**	0.032	0.745	.	5128
	(0.017)	(0.020)	(0.020)			
Occupation Score	1.110***	1.110**	0.615	19.386	.	5128
	(0.391)	(0.506)	(0.484)			
Panel D. FS Boys New						
Years of Education	-0.098	-0.060	-0.120	9.674	0.033	3972
	(0.117)	(0.127)	(0.124)			
Income	40.505	27.586	-0.277	750.105	0.116	3773
	(31.114)	(35.252)	(35.748)			
Work	0.012	0.009	-0.000	0.727	.	4024
	(0.021)	(0.020)	(0.019)			
Occupation Score	0.435	0.685	0.408	18.790	.	4024
	(0.528)	(0.522)	(0.506)			

Notes: The DMF column refers to boys in the original data based on the Death Mortality Files. The FS columns use data derived from FamilySearch records, with "Old" indicating boys from the original sample and "New" indicating newly added boys. ATE is estimated using augmented inverse-propensity weighting (AIPW), a doubly robust method for average treatment effect estimation. ATE Overlap reports an estimand that avoids dividing by estimated propensity scores. We cluster at the family-level. The Z-test reports p-values from Wald tests assessing whether the ATE Overlap coefficient equals that in the DMF sample, under the assumption of independent samples. * p<0.1,** p<0.05, *** p<0.01.

Figures

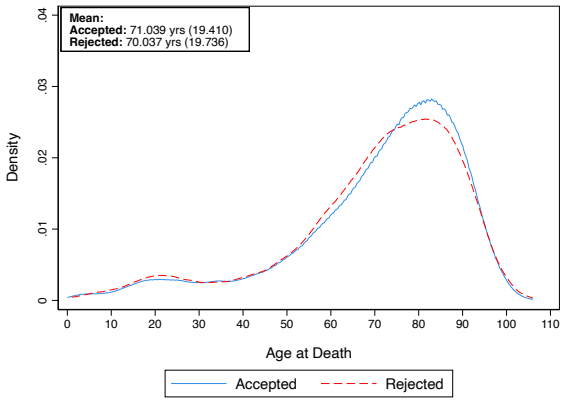
Figure 1: Distribution of Predicted Income by Accepted Status



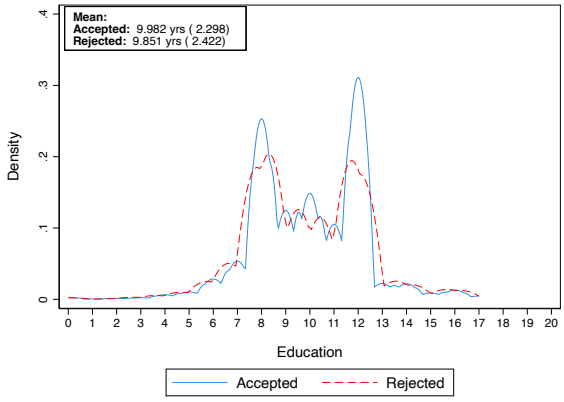
Notes: Predicted Income is the estimated income based on observable characteristics of the family using 1915 Iowa Census. The characteristics used include family size, the ages of all siblings, maternal marital status, foreign-born status, race, educational attainment, occupational score, and their interactions. The kernel density estimate is plotted. Standard deviation is in parentheses.

Figure 2: Distribution of the Outcomes by Accepted Status

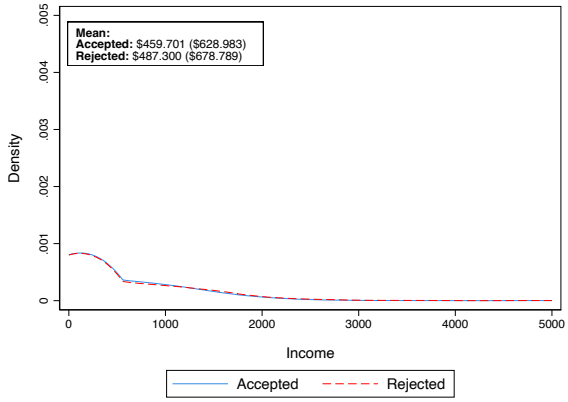
A. Age at Death



B. Education



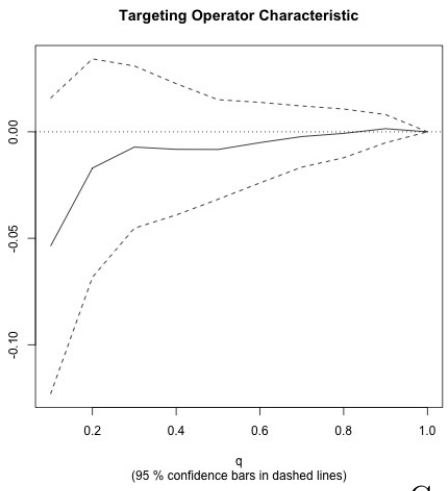
C. Income



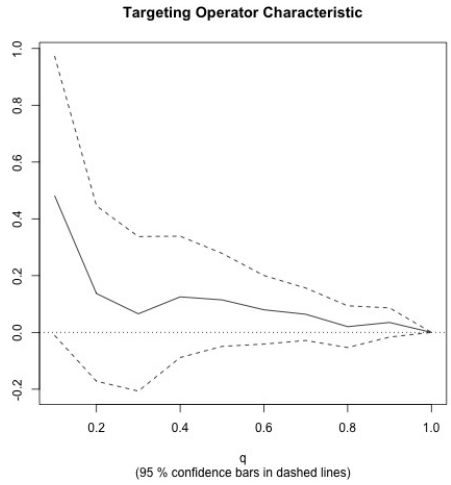
Notes: The kernel density estimate is plotted. Standard deviations are in parentheses.

Figure 3: Target Operating Curves

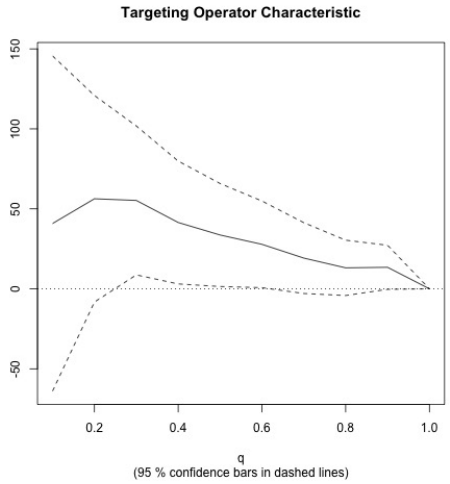
A. Age at Death



B. Education

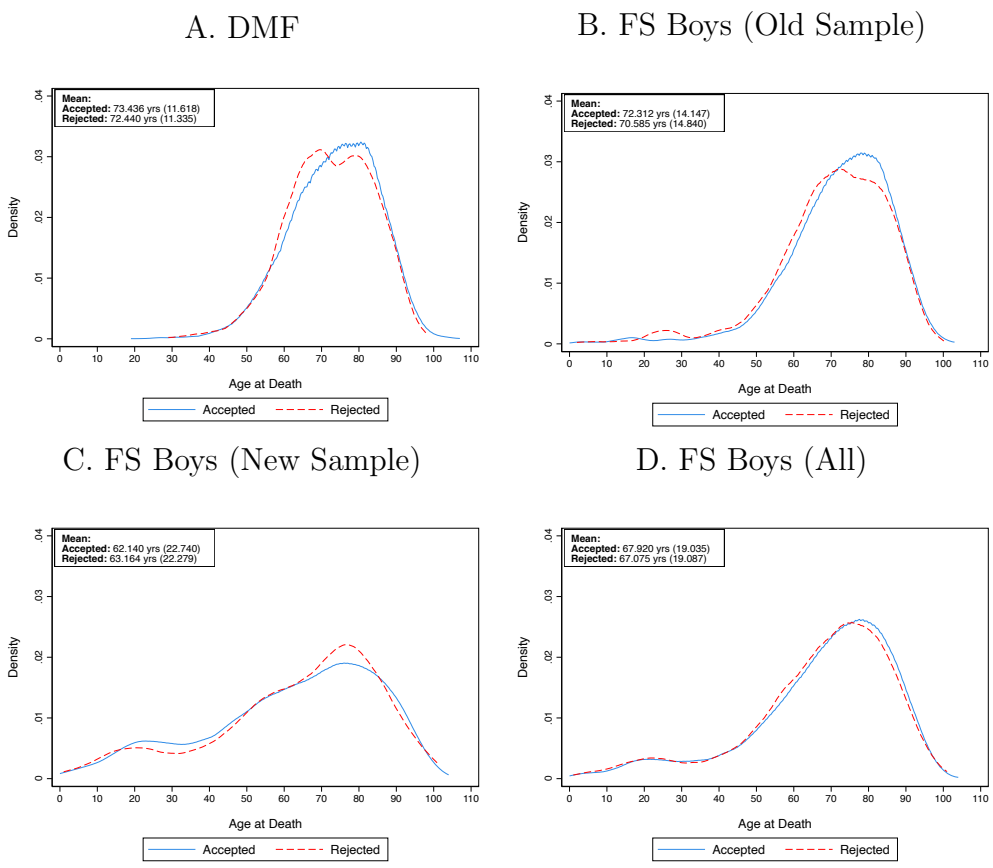


C. Income



Notes: The Target Operating Curve plots the estimated average treatment effect among individuals ranked by predicted CATEs. The x-axis is the the proportion of the population treated.

Figure 4: Distribution of the Outcomes - Original and New Data



Notes: The kernel density estimate is plotted. Standard deviations are in parentheses. The DMF refers to boys in the original data based on the Death Mortality Files. The FS panels use data derived from FamilySearch records, with "Old" indicating boys from the original sample and "New" indicating newly added boys.

A1 Appendix Tables

Table A1: Summary Statistics

	All	Accepted	Rejected	p-value
Panel F. State Characteristics				
$\frac{\text{Manufacturing Earnings}}{\text{U.S. Manufacturing Earnings}}$	1.111	1.109	1.129	0.000
Age of School Entry	7.652	7.661	7.588	0.000
Min. Age Child Labor	14.580	14.559	14.737	0.000
School Continuation Req.	0.347	0.364	0.224	0.000
Total Exp. on Social Programs	13.530	13.533	13.504	0.185
Money to Charities	14.818	14.819	14.813	0.721
Expenditures on Schools	15.584	15.598	15.481	0.000
Law Requires Work	0.549	0.537	0.634	0.000
Law Requires Re-App	0.465	0.456	0.525	0.000
Amount for First Child	17.902	17.746	19.039	0.000
Amount for Additional Child	7.826	7.819	7.879	0.269
Observations				
Number of Individuals	29417	25868	3549	
Number of Families	12045	10359	1693	
Number of Counties	77	76	77	

Notes: The reported p-value comes from a two-sample t test of each characteristic.

Table A2: Does MP Status Predict whether the Outcome is Observed?

	All Children	
	Original Controls	All Controls
Panel A. Age at Death Observed		
Accepted	0.023**	0.014*
	(0.010)	(0.007)
Mean for Rejected	0.795	0.795
N	29417	29417
Panel B. Education Observed		
Accepted	0.031***	0.019
	(0.011)	(0.012)
Mean for Rejected	0.562	0.562
N	29417	29417
Panel C. Income Observed		
Accepted	0.032**	0.021
	(0.012)	(0.013)
Mean for Rejected	0.522	0.522
N	29417	29417

Notes: "Original controls" are child age, length of family name, family size, age of youngest and oldest child in family, maternal marital status and state-level characteristics. "All controls" add mother's age, mother's number of siblings, mother's education foreign status, number of kids over 14, number of kids died pre-application, husband's lifespan, family composition and county-level characteristics. Standard errors are clustered at the county-level. * p<0.1,** p<0.05, *** p<0.01

Table A3: Effect of Different Clustering Level

	Age at Death		Education		Income	
	ATE	ATE Overlap	ATE	ATE Overlap	ATE	ATE Overlap
None	0.016***	0.013	-0.010	-0.019	30.362***	28.082*
	(0.006)	(0.011)	(0.037)	(0.064)	(9.142)	(16.422)
Family	0.012	0.011	0.041	0.017	30.971**	26.481*
	(0.009)	(0.010)	(0.061)	(0.072)	(12.234)	(15.587)
County-Year	0.010	0.010	0.088	0.063	26.375*	22.093
	(0.008)	(0.009)	(0.067)	(0.075)	(13.562)	(15.917)
County	0.011	0.012*	0.157	0.137	25.923*	17.736
	(0.007)	(0.007)	(0.099)	(0.094)	(14.988)	(16.562)

Notes: ATE is estimated using augmented inverse-propensity weighting (AIPW), a doubly robust method for average treatment effect estimation. ATE Overlap reports an estimand that avoids dividing by estimated propensity scores. * p<0.1,** p<0.05, *** p<0.01

Table A4: Summary Statistics by Gender

	Boys	Girls	p-value
Panel F. State Characteristics			
$\frac{\text{Manufacturing Earnings}}{\text{U.S. Manufacturing Earnings}}$	1.111	1.111	0.976
Age of School Entry	7.651	7.653	0.761
Min. Age Child Labor	14.579	14.582	0.731
School Continuation Req.	0.349	0.345	0.404
Total Exp. on Social Programs	13.531	13.528	0.820
Money to Charities	14.824	14.813	0.377
Expenditures on Schools	15.588	15.580	0.347
Law Requires Work	0.553	0.545	0.219
Law Requires Re-App	0.464	0.465	0.940
Amount for First Child	17.938	17.868	0.477
Amount for Additional Child	7.863	7.790	0.041
Observations			
Number of Individuals	14286	15131	
Number of Families	8698	9010	
Number of Counties	74	77	

Notes: The reported p-value comes from a two-sample t test of each characteristic.

Table A5: Does MP Status Predict whether the Outcome is Observed? By Gender

	Boys		Girls	
	Original Controls	All Controls	Original Controls	All Controls
Panel A. Age at Death Observed				
Accepted	0.021*	0.005	0.033**	0.027**
	(0.011)	(0.008)	(0.014)	(0.012)
Mean for Rejected	0.855	0.855	0.727	0.727
N	14860	14860	14557	14557
Panel B. Education Observed				
Accepted	0.037**	0.022	0.024	0.014
	(0.016)	(0.016)	(0.015)	(0.016)
Mean for Rejected	0.569	0.569	0.553	0.553
N	14860	14860	14557	14557
Panel C. Income Observed				
Accepted	0.037**	0.022	0.027*	0.018
	(0.018)	(0.018)	(0.016)	(0.016)
Mean for Rejected	0.541	0.541	0.501	0.501
N	14860	14860	14557	14557

Notes: "Original controls" are child age, length of family name, family size, age of youngest and oldest child in family, maternal marital status and state-level characteristics. "All controls" add mother's age, mother's number of siblings, mother's education foreign status, number of kids over 14, number of kids died pre-application, husband's lifespan, family composition and county-level characteristics. Standard errors are clustered at the county-level. * p<0.1,** p<0.05, *** p<0.01

Table A6: Correlation between Gender-Specific Estimates within Family of the Effect of Receiving MP Transfers

	Boy			
	Age at Death	Log Age at Death	Income	Education
Girl Outcomes				
Age at Death	0.963	0.859	-0.130	0.122
Log Age at Death	0.868	0.953	-0.134	0.246
Income	-0.153	-0.153	0.987	-0.176
Education	0.103	0.227	-0.132	0.994

Notes: Correlation is estimated using ATE Overlap that reports an estimand that avoids dividing by estimated propensity scores.

Table A7: Estimates of the Effect of Receiving MP Transfers by Family Structure

	Boys		Girls	
	Boys Only	Mixed Family	Girls Only	Mixed Family
Log (Age at Death)				
ATE	0.026	0.014	0.015	-0.001
	(0.025)	(0.015)	(0.023)	(0.013)
Mean Rejected	4.141	4.130	4.231	4.241
N	4271	8755	3391	7672
Tests:				
Mixed Family = Single Sex Family		0.680		0.551
Mixed Family Boy = Mixed Family Girl				0.451
Single Sex Family Boy = Single Sex Family Girl			0.744	
Education				
ATE	0.105	-0.054	-0.023	0.134
	(0.141)	(0.094)	(0.151)	(0.090)
Mean Rejected	9.754	9.549	10.289	9.973
N	2943	6156	2641	5987
Tests:				
Mixed Family = Single Sex Family		0.347		0.371
Mixed Family Boy = Mixed Family Girl				0.148
Single Sex Family Boy = Single Sex Family Girl			0.534	
Income				
ATE	-1.648	62.556***	44.277*	7.026
	(36.004)	(23.601)	(26.056)	(14.741)
Mean Rejected	823.455	699.834	151.378	171.946
N	2765	5902	2407	5356
Tests:				
Mixed Family = Single Sex Family		0.136		0.213
Mixed Family Boy = Mixed Family Girl				0.046
Single Sex Family Boy = Single Sex Family Girl			0.301	
Work				
ATE	0.019	0.034**	0.035	0.023
	(0.024)	(0.016)	(0.031)	(0.018)
Mean Rejected	0.740	0.725	0.207	0.238
N	2980	6225	2672	6060
Tests:				
Mixed Family = Single Sex Family		0.599		0.756
Mixed Family Boy = Mixed Family Girl				0.665
Single Sex Family Boy = Single Sex Family Girl			0.683	
Occupation Score				
ATE	0.130	1.090***	0.695	-0.069
	(0.617)	(0.387)	(0.698)	(0.434)
Mean Rejected	19.591	18.503	4.606	5.780
N	2980	6225	2672	6060
Tests:				
Mixed Family = Single Sex Family		0.187		0.353
Mixed Family Boy = Mixed Family Girl				0.046
Single Sex Family Boy = Single Sex Family Girl			0.544	
Joint Tests				
Labor Market Outcomes				
Jointly = 0 (p-value)	0.890	0.000	0.164	0.571
Mixed Family = Single Sex Family		0.237		0.474
Mixed Family Boy = Mixed Family Girl				0.043
Single Sex Family Boy = Single Sex Family Girl			0.659	
All Outcomes				
Jointly = 0 (p-value)	0.801	0.001	0.350	0.519
Mixed Family = Single Sex Family		0.381		0.599
Mixed Family Boy = Mixed Family Girl				0.056
Single Sex Family Boy = Single Sex Family Girl			0.836	

Notes: The Mixed Family column refers to boys (or girls) from families that include children of both genders. ATE is estimated using augmented inverse-propensity weighting (AIPW), a doubly robust method for average treatment effect estimation. ATE Overlap reports an estimand that avoids dividing by estimated propensity scores. The Z-tests report p-values from specified Wald tests using the ATE coefficient, under the assumption of independent samples. Joint Tests panel reports p-values from respective Wald tests assessing whether the ATE coefficients are equal for specified set of outcomes. * p<0.1,** p<0.05, *** p<0.01.

Table A8: Summary Statistics - Original and New Data

	DMF	FS Boys		
		(All)	(Old Sample)	(New Sample)
Panel F. State Characteristics				
$\frac{\text{Manufacturing Earnings}}{\text{U.S. Manufacturing Earnings}}$	1.105	1.111	1.106	1.116
Age of School Entry	7.611	7.651	7.597	7.708
Min. Age Child Labor	14.609	14.579	14.628	14.526
School Continuation Req.	0.354	0.349	0.345	0.354
Total Exp. on Social Programs	13.599	13.531	13.620	13.436
Money to Charities	14.896	14.824	14.914	14.727
Expenditures on Schools	15.596	15.588	15.623	15.550
Law Requires Work	0.575	0.553	0.579	0.525
Law Requires Re-App	0.495	0.464	0.500	0.426
Amount for First Child	18.148	17.938	18.443	17.397
Amount for Additional Child	7.979	7.863	8.020	7.696
Observations				
Number of Individuals	7860	14286	7387	6899
Number of Families	5738	8698	5420	5160
Number of Counties	75	74	74	74

Notes: The DMF refers to boys in the original data based on the Death Mortality Files. The FS columns use data derived from FamilySearch records, with "Old" indicating boys from the original sample and "New" indicating newly added boys.

Table A9: Summary Statistics by Accepted Status - Original and New Data

	DMF			FS Boys								
				All			Old Sample			New Sample		
	(All)	(Accepted)	(Rejected)	(All)	(Accepted)	(Rejected)	(All)	(Accepted)	(Rejected)	(All)	(Accepted)	(Rejected)
Panel A. Outcomes												
Age at Death	73.312	73.436	72.440	67.817	67.920	67.075	72.115	72.312	70.585	62.275	62.140	63.164
	(11.587)	(11.618)	(11.335)	(19.043)	(19.035)	(19.087)	(14.237)	(14.147)	(14.84)	(22.68)	(22.74)	(22.279)
% Observed	1.000	1.000	1.000	0.882	0.886	0.859	0.961	0.962	0.955	0.798	0.802	0.773
Education (1940 Census)	9.776	9.827	9.396	9.747	9.765	9.615	9.811	9.840	9.570	9.662	9.661	9.666
	(2.589)	(2.576)	(2.662)	(2.387)	(2.376)	(2.463)	(2.352)	(2.343)	(2.418)	(2.43)	(2.418)	(2.514)
% Observed	0.980	0.980	0.984	0.618	0.624	0.575	0.684	0.689	0.645	0.547	0.552	0.512
1(Education $\geq$ 12)	0.318	0.322	0.286	0.298	0.301	0.277	0.308	0.312	0.267	0.285	0.285	0.288
	(.466)	(.467)	(.453)	(.457)	(.459)	(.448)	(.462)	(.464)	(.443)	(.452)	(.451)	(.453)
Income (1940 Census)	733.124	742.239	666.174	737.494	732.142	778.203	744.590	742.024	765.717	727.966	718.594	792.524
	(741.797)	(749.145)	(683.181)	(710.176)	(705.64)	(742.88)	(704.598)	(698.29)	(754.917)	(717.585)	(715.49)	(729.393)
% Observed	0.934	0.933	0.944	0.590	0.596	0.547	0.654	0.659	0.617	0.521	0.527	0.485
1(Income $>$ 0)	0.782	0.786	0.749	0.790	0.789	0.799	0.791	0.792	0.788	0.789	0.786	0.811
	(.413)	(.41)	(.435)	(.407)	(.408)	(.401)	(.406)	(.406)	(.409)	(.408)	(.41)	(.392)
Work				0.755	0.755	0.756	0.763	0.765	0.747	0.744	0.741	0.767
				(.43)	(.43)	(.43)	(.425)	(.424)	(.435)	(.436)	(.438)	(.423)
Occupation Score				19.595	19.592	19.626	19.794	19.836	19.448	19.330	19.258	19.826
				(11.628)	(11.651)	(11.458)	(11.563)	(11.563)	(11.564)	(11.711)	(11.764)	(11.346)
Panel B. Program Characteristics												
% Accepted	0.875	1.000	0.000	0.875	1.000	0.000	0.885	1.000	0.000	0.864	1.000	0.000
Duration	3.504	4.271	0.000	3.315	4.082	0.000	3.539	4.240	0.000	3.053	3.887	0.000
Amount	272.131	311.303	0.000	270.161	309.082	0.000	274.103	309.884	0.000	265.943	308.201	0.000
Panel C. Predetermined Characteristics												
Year of Birth	1913.463	1913.641	1912.218	1913.267	1913.446	1912.015	1913.445	1913.609	1912.180	1913.076	1913.267	1911.866
Child Age	8.373	8.308	8.827	8.336	8.291	8.648	8.402	8.358	8.743	8.265	8.218	8.562
Widow	0.547	0.549	0.532	0.521	0.524	0.499	0.539	0.541	0.521	0.502	0.506	0.478
Year of Application	1921.795	1921.909	1920.997	1921.624	1921.757	1920.697	1921.839	1921.960	1920.906	1921.393	1921.533	1920.509
Number of Children in Application	3.707	3.731	3.538	3.402	3.452	3.053	3.391	3.441	3.013	3.414	3.465	3.089
Age of Oldest Child in Application	11.515	11.466	11.861	11.283	11.288	11.253	11.294	11.291	11.323	11.272	11.285	11.190
Age of Youngest Child in Application	4.952	4.855	5.630	5.219	5.119	5.919	5.227	5.137	5.916	5.210	5.098	5.921
Predicted Income	391.770	387.258	423.331	795.511	788.451	847.036	778.784	776.647	795.306	815.703	802.916	903.046
Panel D. Additional Predetermined Characteristics												
Mom Age				37.040	36.986	37.417	36.926	36.905	37.089	37.161	37.074	37.713
Mom Education				7.616	7.632	7.501	7.688	7.708	7.538	7.539	7.550	7.468
Number of Kids Over 14 (Not in App.)				1.536	1.514	1.690	1.509	1.501	1.571	1.566	1.529	1.797
Number of Kids Died Pre-Application				0.226	0.225	0.233	0.226	0.227	0.217	0.226	0.223	0.248
Number of Mom's Siblings				4.289	4.296	4.240	4.468	4.446	4.634	4.098	4.131	3.885
Panel E. County Characteristics												
Population				49001.5	50371.7	39430.2	51539.6	53198.2	38750.1	46283.8	47269.5	40043.1
% Rural				0.538	0.534	0.566	0.524	0.519	0.563	0.552	0.550	0.569
Sex Ratio				1.097	1.097	1.094	1.088	1.089	1.085	1.106	1.106	1.102
% White Fem. Married w. Kids in LF				0.052	0.053	0.047	0.053	0.054	0.045	0.052	0.053	0.048
% Black				0.014	0.014	0.014	0.015	0.015	0.016	0.014	0.014	0.012
Observations												
Number of Individuals	7860	6877	983	14286	12497	1789	7387	6539	848	6899	5958	941
Number of Families	5738	4990	748	8698	7550	1152	5420	4767	654	5160	4451	710

Notes: The DMF refers to boys in the original data based on the Death Mortality Files. The FS panels use data derived from FamilySearch records, with "Old" indicating boys from the original sample and "New" indicating newly added boys. Predicted Income is the estimated income based on observable characteristics of the family using 1915 Iowa Census. The characteristics used include family size, the ages of all siblings, maternal marital status, foreign-born status, race, educational attainment, occupational score, and their interactions.

Table A10: Summary Statistics by Accepted Status - Original and New Data

	DMF			FS Boys								
				All			Old Sample			New Sample		
	(All)	(Accepted)	(Rejected)	(All)	(Accepted)	(Rejected)	(All)	(Accepted)	(Rejected)	(All)	(Accepted)	(Rejected)
Panel F. State Characteristics												
$\frac{\text{Manufacturing Earnings}}{\text{U.S. Manufacturing Earnings}}$	1.105	1.103	1.119	1.111	1.108	1.129	1.106	1.104	1.120	1.116	1.113	1.137
Age of School Entry	7.611	7.616	7.573	7.651	7.660	7.587	7.597	7.608	7.508	7.708	7.716	7.658
School Continuation Req.	0.354	0.370	0.245	0.349	0.365	0.238	0.345	0.363	0.212	0.354	0.368	0.261
Total Exp. on Social Programs	13.599	13.598	13.603	13.531	13.535	13.507	13.620	13.611	13.690	13.436	13.451	13.343
Money to Charities	14.896	14.893	14.919	14.824	14.825	14.814	14.914	14.906	14.975	14.727	14.736	14.669
Expenditures on Schools	15.596	15.610	15.502	15.588	15.601	15.495	15.623	15.634	15.543	15.550	15.566	15.451
Law Requires Work	0.575	0.561	0.678	0.553	0.541	0.630	0.579	0.565	0.682	0.525	0.515	0.583
Law Requires Re-App	0.495	0.484	0.574	0.464	0.457	0.518	0.500	0.490	0.578	0.426	0.420	0.463
Amount for First Child	18.148	17.949	19.535	17.938	17.765	19.147	18.443	18.236	20.037	17.397	17.248	18.345
Amount for Additional Child	7.979	7.963	8.093	7.863	7.862	7.873	8.020	7.997	8.193	7.696	7.713	7.585
Observations												
Number of Individuals	7860	6877	983	14286	12497	1789	7387	6539	848	6899	5958	941
Number of Families	5738	4990	748	8698	7550	1152	5420	4767	654	5160	4451	710

Notes: The DMF refers to boys in the original data based on the Death Mortality Files. The FS panels use data derived from FamilySearch records, with "Old" indicating boys from the original sample and "New" indicating newly added boys.

Table A11: Estimates of the Effect of Receiving MP Transfers on Age at Death - Original and New Data (Restricting Deaths)

	DMF	FS Boys		
		(All)	(Old Sample)	(New Sample)
Panel A. OLS without Controls				
	0.013**	0.005	0.012**	-0.005
	(0.006)	(0.005)	(0.006)	(0.008)
Panel B. OLS with Old Controls				
	0.018***	0.006	0.013**	-0.003
	(0.006)	(0.005)	(0.006)	(0.008)
Panel C. OLS with Full Controls				
		0.006	0.012**	-0.003
		(0.005)	(0.006)	(0.008)
Panel D. Generalized Random Forest				
ATE	0.015**	0.030***	0.047***	0.000
	(0.006)	(0.007)	(0.008)	(0.013)
ATE Overlap	0.015**	0.006	0.014**	-0.007
	(0.006)	(0.005)	(0.007)	(0.009)
Observations	7859	10595	6624	3897
Fraction Accepted	0.875	0.878	0.888	0.863
Mean of Accepted	73.437	74.495	74.647	74.169
Mean of Rejected	72.440	74.138	73.787	74.490

Notes: The DMF column refers to boys in the original data based on the Death Mortality Files. The FS columns use data derived from FamilySearch records, with "Old Sample" indicating boys from the original sample and "New Sample" indicating newly added boys. The sample is restricted to those who died between 1965 to 2012. ATE is estimated using augmented inverse-propensity weighting (AIPW), a doubly robust method for average treatment effect estimation. ATE Overlap reports an estimand that avoids dividing by estimated propensity scores. "Old controls" are child age, length of family name, family size, age of youngest and oldest child in family, maternal marital status. "Full controls" add mother's age, mother's number of siblings, foreign status, number of kids over 14, number of kids died pre-application, and husband's lifespan. We cluster at the family-level. * p<0.1,** p<0.05, *** p<0.01.

Table A12: Does MP Status Predict whether the Outcome is Observed? - Original and New Data

	DMF	FS Boys					
		All		Old Sample		New Sample	
	Original Controls	Original Controls	All Controls	Original Controls	All Controls	Original Controls	All Controls
Panel A. Age at Death Observed							
Accepted	0.029**	0.021*	0.005	0.004	0.001	0.026	0.007
	(0.014)	(0.011)	(0.008)	(0.008)	(0.007)	(0.016)	(0.011)
Mean for Rejected	0.477	0.855	0.855	0.956	0.956	0.769	0.769
N	16554	14860	14860	7433	7433	7300	7300
Panel B. Education Observed							
Accepted	0.003	0.037**	0.022	0.033	0.023	0.033	0.016
	(0.004)	(0.016)	(0.016)	(0.021)	(0.021)	(0.022)	(0.019)
Mean for Rejected	0.036	0.569	0.569	0.644	0.644	0.511	0.511
N	16554	14860	14860	7433	7433	7300	7300
Panel C. Income Observed							
Accepted	0.002	0.037**	0.022	0.035	0.026	0.034	0.018
	(0.004)	(0.018)	(0.018)	(0.022)	(0.022)	(0.022)	(0.021)
Mean for Rejected	0.035	0.541	0.541	0.616	0.616	0.482	0.482
N	16554	14860	14860	7433	7433	7300	7300

Notes: The DMF column refers to boys in the original data based on the Death Mortality Files. The FS columns use data derived from FamilySearch records, with "Old Sample" indicating boys from the original sample and "New Sample" indicating newly added boys. "Original controls" are child age, length of family name, family size, age of youngest and oldest child in family, maternal marital status and state-level characteristics. "All controls" add mother's age, mother's number of siblings, mother's education foreign status, number of kids over 14, number of kids died pre-application, husband's lifespan, family composition and county-level characteristics. Standard errors are clustered at the county-level. * p<0.1,** p<0.05, *** p<0.01

Table A13: Estimates of the Effect of Receiving MP Transfers on Age at Death Different Methods

	ATE	ATT	ATU	ATE Overlap	OLS	Mean of Rejected	N
Panel A. Słoczyński (2022)							
DMF	0.011	0.010	0.021***		0.018***	72.440	7859
	(0.008)	(0.009)	(0.007)		(0.006)		
FS Boys	0.016	0.017	0.007		0.009	67.280	13026
	(0.015)	(0.015)	(0.015)		(0.013)		
Panel B. Random Forest							
DMF	0.015**	0.014**	0.015**	0.015**	0.018***	72.440	7859
	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)		
FS Boys	0.016	0.020	0.011	0.012	0.009	67.280	13026
	(0.013)	(0.016)	(0.013)	(0.014)	(0.013)		

Notes: The DMF panel refers to boys in the original data based on the Death Mortality Files. The FS panel use data derived from FamilySearch records. Panel A uses diagnostics developed by Sloczyński (2022). For Panel B, ATE is estimated using augmented inverse-propensity weighting (AIPW), a doubly robust method for average treatment effect estimation. ATE Overlap reports an estimand that avoids dividing by estimated propensity scores. * p<0.1,** p<0.05, *** p<0.01.

Table A14: Estimates of the Effect of Receiving MP Transfers
Dropping Those Born after 1920

	ATE	ATT	ATU	ATE Overlap	OLS	Mean Rejected	Z-test (p-value)	N
Panel A. Education								
Years of Education	0.051	0.049	0.004	0.012	-0.073	9.946	0.059	14301
	(0.066)	(0.083)	(0.077)	(0.079)	(0.075)			
$\mathbb{1}$(Education $\geq$ 12)	0.017	0.016	0.002	0.004	-0.013	0.349	0.798	14301
	(0.013)	(0.017)	(0.015)	(0.015)	(0.014)			
Panel B. Labor Market								
Income	40.212***	34.857**	27.118	26.548	18.608	565.888	0.069	13531
	(14.226)	(17.739)	(17.771)	(17.995)	(18.611)			
Log(Income)	0.081**	0.078**	0.030	0.031	0.018	6.526	0.317	8200
	(0.032)	(0.040)	(0.033)	(0.034)	(0.033)			
$\mathbb{1}$(Income $>$ 0)	0.025**	0.017	0.022*	0.021*	0.018	0.596	0.708	13531
	(0.011)	(0.014)	(0.012)	(0.012)	(0.013)			
Work	0.039***	0.031**	0.034***	0.034***	0.032**	0.548	.	14471
	(0.011)	(0.014)	(0.011)	(0.012)	(0.013)			
Occupation Score	0.770***	0.503	0.793***	0.742**	0.682**	14.289	.	14471
	(0.265)	(0.338)	(0.285)	(0.292)	(0.313)			

Notes: The sample drops those born after 1920. ATE, ATT and ATU are estimated using augmented inverse-propensity weighting (AIPW), a doubly robust method for average treatment effect estimation. ATE Overlap reports an estimand that avoids dividing by estimated propensity scores. The Z-test reports p-values from Wald tests assessing whether the coefficient equals that in the DMF sample, under the assumption of independent samples. * p<0.1,** p<0.05, *** p<0.01.

Table A15: Best Linear Projection

	Log Age at Death		Education		Income		Work		Occupation Score	
	β	(SE)	β	(SE)	β	(SE)	β	(SE)	β	(SE)
Covariates										
Child Age	0.006**	(0.003)	-0.002	(0.017)	9.984***	(3.746)	0.004	(0.003)	0.159**	(0.077)
Child Gender Male	0.019	(0.017)	-0.075	(0.097)	16.297	(23.968)	-0.007	(0.021)	0.314	(0.513)
Widow	0.009	(0.019)	-0.060	(0.123)	3.993	(24.742)	-0.015	(0.021)	-0.311	(0.490)
Year of Application	0.003	(0.004)	-0.018	(0.027)	2.603	(5.766)	0.002	(0.004)	0.100	(0.101)
Number of Children in Application	-0.000	(0.011)	0.167**	(0.079)	-0.972	(13.496)	0.005	(0.013)	0.120	(0.299)
Age of Oldest Child in Application	-0.006	(0.005)	-0.067**	(0.032)	-0.764	(6.376)	-0.007	(0.007)	-0.178	(0.152)
Age of Youngest Child in Application	-0.001	(0.005)	0.051	(0.031)	-0.556	(6.922)	0.009	(0.006)	0.059	(0.133)
Mom Age	0.000	(0.002)	-0.007	(0.011)	-1.323	(2.632)	0.003	(0.002)	0.043	(0.049)
Mom Education	-0.002	(0.004)	0.052*	(0.028)	8.586*	(5.120)	0.006	(0.005)	0.039	(0.105)
Number of Kids Over 14 (Not in App.)	-0.001	(0.006)	0.025	(0.040)	9.891	(8.663)	-0.005	(0.006)	0.007	(0.152)
Number of Kids Died Pre-Application	-0.001	(0.016)	0.065	(0.107)	-45.457**	(17.946)	-0.009	(0.016)	-0.675*	(0.370)
Number of Mom's Siblings	0.003	(0.003)	0.008	(0.015)	-2.578	(3.147)	-0.002	(0.003)	-0.025	(0.060)
County % Rural	-0.017	(0.055)	0.319	(0.400)	-65.298	(74.996)	-0.098	(0.063)	-2.993**	(1.360)
County Sex Ratio	0.016	(0.129)	0.287	(0.816)	-98.721	(155.002)	0.043	(0.134)	0.684	(2.631)
County % White Fem. Married w. Kids in LF	-0.028	(0.915)	5.055	(6.795)	-390.560	(841.363)	-0.125	(1.147)	-14.884	(17.750)
County % Black	0.320	(0.395)	-0.333	(2.383)	273.506	(450.307)	-0.177	(0.393)	-3.729	(8.700)
County Population	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	-0.000	(0.000)	-0.000	(0.000)
State Manufacturing	0.202	(0.198)	-1.901	(1.382)	148.509	(292.680)	0.088	(0.206)	5.216	(5.063)
State Age of School Entry	0.021	(0.017)	0.023	(0.142)	32.542	(25.631)	0.029	(0.020)	1.182**	(0.469)
State Min. Age Child Labor	0.003	(0.025)	-0.268	(0.176)	52.023	(34.146)	0.012	(0.037)	0.675	(0.663)
State School Continuation Req.	0.012	(0.031)	-0.006	(0.224)	-39.111	(40.738)	0.003	(0.033)	-0.734	(0.832)
State Total Exp. on Social Programs	0.008	(0.012)	0.054	(0.102)	12.498	(22.251)	0.005	(0.016)	-0.038	(0.376)
State Money to Charities	-0.001	(0.032)	0.267	(0.245)	4.373	(47.419)	0.017	(0.037)	0.991	(0.860)
State Expenditures on Schools	-0.031	(0.033)	-0.446*	(0.229)	33.064	(42.254)	-0.021	(0.034)	-0.472	(0.825)
State Law Requires Work	-0.032	(0.036)	0.044	(0.280)	-6.549	(62.100)	-0.002	(0.042)	-0.710	(1.003)
State Law Requires Re-App	0.066	(0.041)	-0.115	(0.252)	-26.602	(61.085)	0.001	(0.045)	0.217	(1.069)
State Amount for First Child	-0.000	(0.003)	0.028	(0.023)	-8.073*	(4.901)	-0.001	(0.004)	-0.083	(0.082)
State Amount for Additional Child	-0.001	(0.005)	-0.033	(0.038)	14.751*	(7.779)	0.008	(0.006)	0.250*	(0.132)
Joint Tests										
Family Characteristics	0.733		0.123		0.082		0.592		0.432	
County & State Characteristics	0.943		0.614		0.530		0.944		0.130	
Number of Observations	24089		17727		16430		17937		17937	

Notes: The coefficients are obtained by regressing doubly robust scores derived from the forest on the covariates. The Joint Tests panel reports p-values from Wald tests assessing whether the respective coefficients are jointly equal to 0 assuming covariances are 0. * p<0.1,** p<0.05, *** p<0.01.

Table A16: Best Linear Projection - Boys

	Log Age at Death		Education		Income		Work		Occupation Score	
	β	(SE)	β	(SE)	β	(SE)	β	(SE)	β	(SE)
Covariates										
Child Age	0.007	(0.004)	0.006	(0.024)	21.630***	(6.512)	0.005	(0.005)	0.252**	(0.109)
Widow	0.008	(0.027)	-0.076	(0.161)	12.291	(40.117)	0.029	(0.027)	0.298	(0.667)
Year of Application	0.003	(0.007)	-0.030	(0.034)	4.601	(9.461)	0.004	(0.006)	0.195	(0.143)
Number of Children in Application	0.003	(0.014)	0.021	(0.089)	11.556	(19.829)	0.005	(0.014)	0.407	(0.326)
Age of Oldest Child in Application	-0.008	(0.007)	-0.018	(0.042)	-9.329	(9.708)	-0.006	(0.007)	-0.402**	(0.169)
Age of Youngest Child in Application	-0.003	(0.007)	-0.001	(0.038)	-3.639	(10.757)	0.008	(0.007)	0.028	(0.161)
Mom Age	-0.000	(0.003)	-0.018	(0.015)	-2.963	(4.549)	0.001	(0.003)	0.018	(0.066)
Mom Education	-0.010*	(0.006)	0.060*	(0.033)	17.352**	(7.520)	0.012**	(0.006)	0.082	(0.144)
Number of Kids Over 14 (Not in App.)	0.001	(0.009)	0.016	(0.053)	15.740	(15.626)	-0.008	(0.010)	0.076	(0.229)
Number of Kids Died Pre-Application	-0.006	(0.027)	0.041	(0.130)	-56.208*	(32.015)	0.006	(0.021)	-0.282	(0.493)
Number of Mom's Siblings	0.006	(0.005)	0.009	(0.018)	-9.442*	(5.106)	-0.001	(0.003)	-0.024	(0.080)
County % Rural	-0.053	(0.080)	-0.594	(0.491)	-58.360	(118.570)	-0.054	(0.083)	-1.656	(1.997)
County Sex Ratio	0.097	(0.177)	1.034	(1.046)	-247.756	(250.622)	0.036	(0.186)	-0.527	(3.811)
County % White Fem. Married w. Kids in LF	0.243	(1.603)	-0.660	(7.993)	-970.273	(1427.583)	0.019	(1.569)	-40.040	(25.502)
County % Black	0.080	(0.459)	-1.891	(2.924)	297.852	(707.483)	-0.006	(0.582)	1.117	(13.577)
County Population	0.000	(0.000)	-0.000	(0.000)	0.000	(0.001)	-0.000	(0.000)	0.000	(0.000)
State Manufacturing	0.354	(0.278)	0.279	(1.691)	299.074	(475.577)	0.430	(0.269)	13.214*	(7.698)
State Age of School Entry	0.030	(0.025)	0.347**	(0.154)	46.508	(40.071)	0.023	(0.025)	1.008	(0.657)
State Min. Age Child Labor	0.020	(0.031)	-0.382*	(0.219)	56.905	(58.031)	-0.003	(0.050)	-0.329	(1.090)
State School Continuation Req.	0.040	(0.049)	0.143	(0.292)	-69.877	(68.659)	0.047	(0.045)	1.071	(1.040)
State Total Exp. on Social Programs	-0.011	(0.016)	-0.111	(0.131)	19.028	(35.200)	0.013	(0.019)	0.149	(0.466)
State Money to Charities	0.017	(0.045)	0.681**	(0.294)	42.207	(73.612)	0.042	(0.046)	2.157*	(1.125)
State Expenditures on Schools	-0.035	(0.049)	-0.649**	(0.285)	9.422	(67.140)	-0.060	(0.044)	-2.204**	(1.123)
State Law Requires Work	0.008	(0.053)	-0.275	(0.340)	25.724	(96.114)	0.022	(0.050)	0.848	(1.343)
State Law Requires Re-App	0.081	(0.067)	-0.059	(0.319)	-88.529	(95.988)	0.003	(0.057)	-0.033	(1.360)
State Amount for First Child	-0.001	(0.004)	0.054*	(0.029)	-9.276	(8.431)	-0.002	(0.005)	-0.083	(0.130)
State Amount for Additional Child	0.004	(0.007)	0.011	(0.051)	18.192	(12.893)	0.014*	(0.008)	0.392*	(0.207)
Joint Tests										
Family Characteristics	0.588		0.830		0.006		0.487		0.158	
County & State Characteristics	0.944		0.042		0.802		0.787		0.128	
Number of Observations	13026		9099		8667		9205		9205	

Notes: The coefficients are obtained by regressing doubly robust scores derived from the forest on the covariates. The Joint Tests panel reports p-values from Wald tests assessing whether the respective coefficients are jointly equal to 0 assuming covariances are 0. * p<0.1,** p<0.05, *** p<0.01.

Table A17: Best Linear Projection - Girls

	Log Age at Death		Education		Income		Work		Occupation Score	
	β	(SE)	β	(SE)	β	(SE)	β	(SE)	β	(SE)
Covariates										
Child Age	0.005	(0.004)	-0.008	(0.025)	-3.387	(3.986)	0.003	(0.006)	0.051	(0.121)
Widow	0.010	(0.025)	-0.024	(0.155)	-7.687	(25.420)	-0.062*	(0.032)	-0.984	(0.724)
Year of Application	0.002	(0.005)	-0.003	(0.035)	-0.560	(5.371)	-0.001	(0.006)	-0.003	(0.130)
Number of Children in Application	-0.004	(0.013)	0.321***	(0.096)	-15.468	(13.725)	0.004	(0.021)	-0.206	(0.530)
Age of Oldest Child in Application	-0.003	(0.006)	-0.123***	(0.043)	8.438	(6.882)	-0.008	(0.012)	0.068	(0.270)
Age of Youngest Child in Application	0.001	(0.006)	0.104**	(0.042)	3.567	(6.690)	0.011	(0.009)	0.112	(0.192)
Mom Age	0.000	(0.002)	0.003	(0.015)	0.455	(2.332)	0.004	(0.003)	0.058	(0.073)
Mom Education	0.009*	(0.005)	0.044	(0.040)	-1.204	(6.461)	-0.001	(0.008)	0.002	(0.172)
Number of Kids Over 14 (Not in App.)	-0.003	(0.007)	0.040	(0.051)	2.866	(7.693)	-0.001	(0.010)	-0.069	(0.212)
Number of Kids Died Pre-Application	0.003	(0.018)	0.077	(0.141)	-34.015**	(16.709)	-0.023	(0.023)	-1.066**	(0.496)
Number of Mom's Siblings	-0.001	(0.003)	0.007	(0.019)	5.114	(3.256)	-0.002	(0.004)	-0.028	(0.091)
County % Rural	0.025	(0.069)	1.265**	(0.531)	-86.050	(81.896)	-0.147*	(0.088)	-4.703**	(1.940)
County Sex Ratio	-0.068	(0.176)	-0.494	(1.003)	73.119	(176.350)	0.063	(0.175)	2.289	(3.818)
County % White Fem. Married w. Kids in LF	-0.281	(0.925)	11.164	(8.204)	225.932	(883.811)	-0.290	(1.152)	9.671	(26.549)
County % Black	0.513	(0.651)	0.970	(3.262)	307.918	(476.835)	-0.352	(0.523)	-9.300	(10.573)
County Population	-0.000	(0.000)	0.000*	(0.000)	-0.000	(0.000)	-0.000	(0.000)	-0.000**	(0.000)
State Manufacturing	0.021	(0.253)	-4.081**	(1.766)	-49.067	(285.419)	-0.274	(0.320)	-3.301	(7.498)
State Age of School Entry	0.010	(0.021)	-0.343*	(0.198)	16.451	(28.215)	0.036	(0.032)	1.417**	(0.705)
State Min. Age Child Labor	-0.017	(0.035)	-0.100	(0.235)	44.792	(30.162)	0.028	(0.047)	1.811**	(0.835)
State School Continuation Req.	-0.018	(0.035)	-0.156	(0.262)	-8.136	(39.037)	-0.045	(0.049)	-2.632*	(1.419)
State Total Exp. on Social Programs	0.030*	(0.016)	0.264**	(0.126)	4.037	(22.730)	-0.005	(0.027)	-0.218	(0.664)
State Money to Charities	-0.023	(0.039)	-0.221	(0.299)	-42.257	(47.307)	-0.013	(0.057)	-0.401	(1.365)
State Expenditures on Schools	-0.026	(0.040)	-0.234	(0.284)	60.774	(44.219)	0.020	(0.053)	1.325	(1.221)
State Law Requires Work	-0.082*	(0.046)	0.409	(0.341)	-53.846	(64.907)	-0.037	(0.064)	-2.684*	(1.542)
State Law Requires Re-App	0.055	(0.044)	-0.117	(0.303)	52.304	(59.843)	0.012	(0.064)	0.975	(1.889)
State Amount for First Child	0.002	(0.003)	-0.006	(0.028)	-6.124	(3.863)	-0.000	(0.006)	-0.080	(0.112)
State Amount for Additional Child	-0.007	(0.006)	-0.075*	(0.043)	10.093	(7.703)	0.002	(0.009)	0.090	(0.191)
Joint Tests										
Family Characteristics	0.914		0.003		0.468		0.644		0.710	
County & State Characteristics	0.701		0.017		0.689		0.953		0.019	
Number of Observations	11063		8628		7763		8732		8732	

Notes: The coefficients are obtained by regressing doubly robust scores derived from the forest on the covariates. The Joint Tests panel reports p-values from Wald tests assessing whether the respective coefficients are jointly equal to 0 assuming covariances are 0. * p<0.1,** p<0.05, *** p<0.01.

Table A18: Rank-Weighted Average Treatment Effect

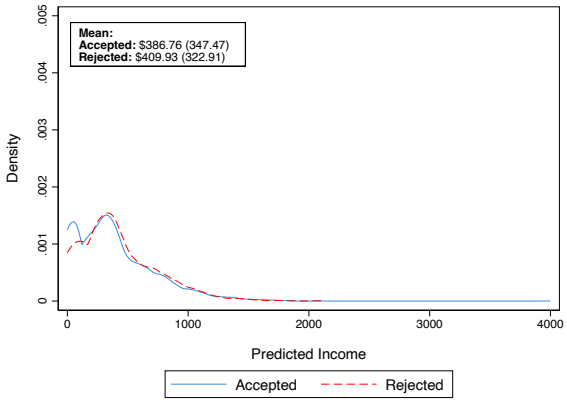
	DMF	FS		
		(All)	(Boys)	(Girls)
RATE				
Longevity	-0.003	-0.013	-0.028*	0.008
	(0.008)	(0.014)	(0.015)	(0.026)
Education	0.010	0.165**	0.088	-0.031
	(0.276)	(0.083)	(0.098)	(0.089)
Income	-69.672	35.448**	12.529	62.550***
	(64.030)	(14.967)	(24.729)	(17.829)

Notes: The Rank-Weighted Average Treatment Effect (RATE) is a scalar summary measure that evaluates how effectively a prioritization rule (e.g., ranking units by predicted treatment effect) targets individuals who benefit most from treatment. It is computed as a weighted average of the Targeting Operator Characteristic (TOC) curve, which plots the realized treatment effect among the top fraction q of ranked units. The DMF column refers to boys in the original data based on the Death Mortality Files. The FS columns use data derived from FamilySearch records. * p<0.1,** p<0.05, *** p<0.01.

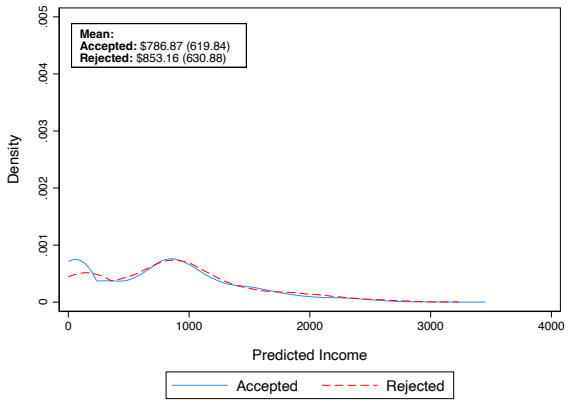
A2 Appendix Figures

Figure A1: Distribution of the Predicted Income - Original and New Data

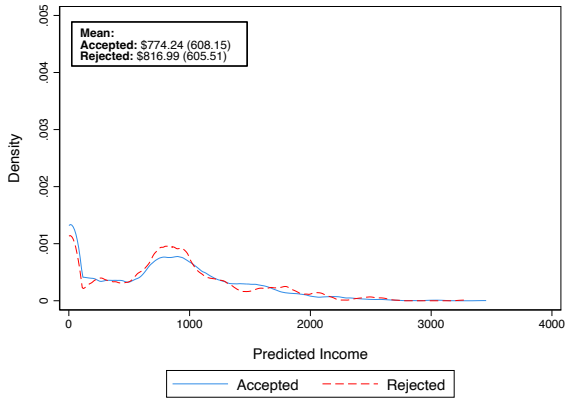
DMF



FS Boys

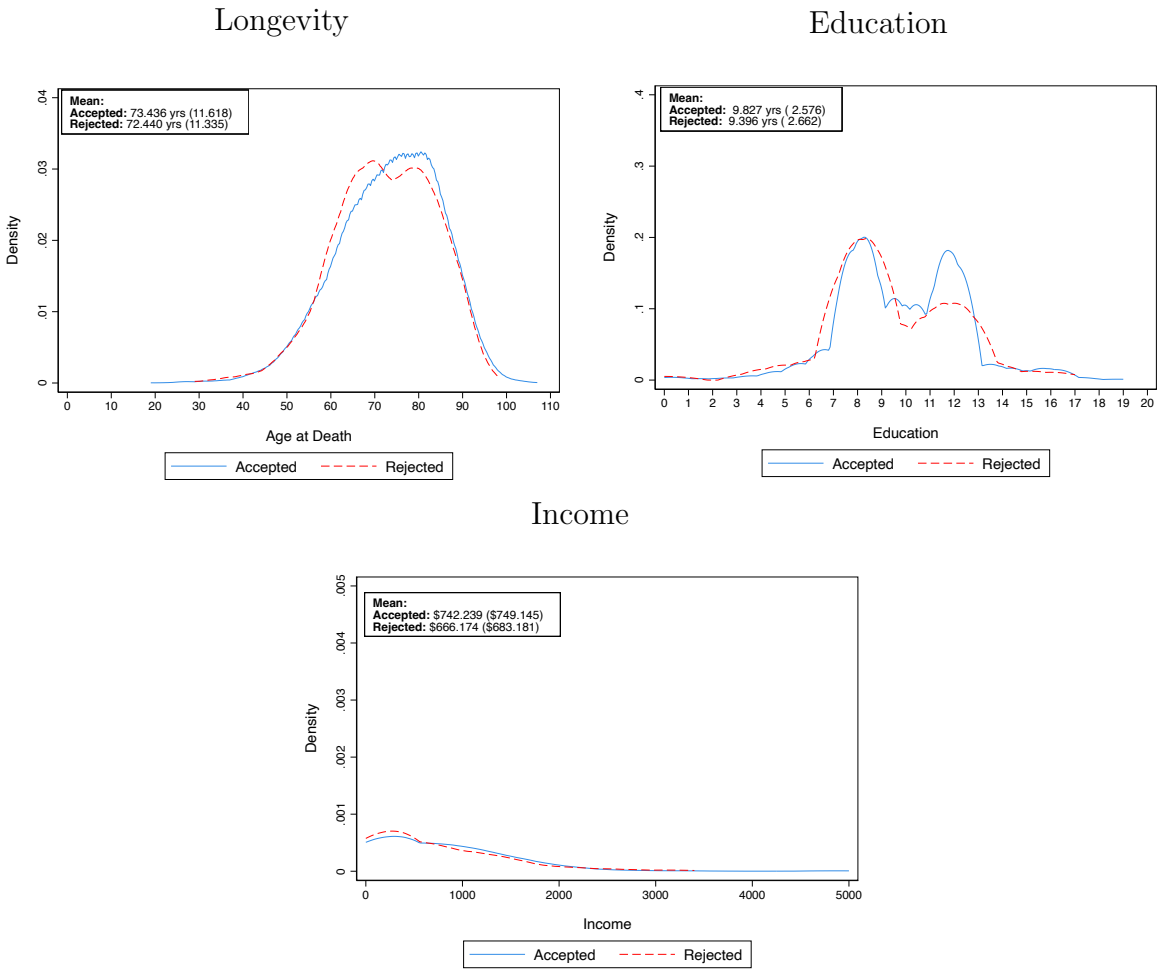


FS Girls



Notes: The kernel density estimate is plotted. Standard deviations are in parentheses. The DMF refers to boys in the original data based on the Death Mortality Files. The FS panels use data derived from FamilySearch records. Predicted Income is the estimated income based on observable characteristics of the family using 1915 Iowa Census. The characteristics used include family size, the ages of all siblings, maternal marital status, foreign-born status, race, educational attainment, occupational score, and their interactions.

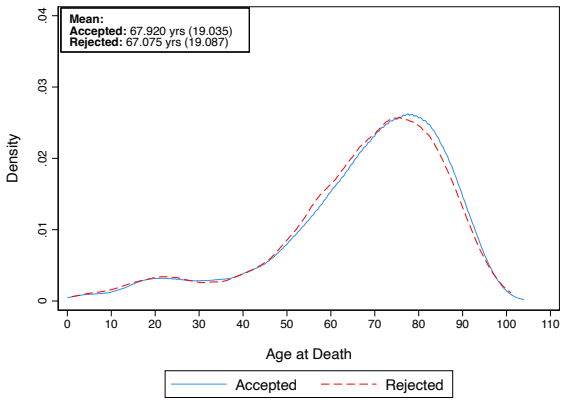
Figure A2: Distribution of the Outcomes - DMF



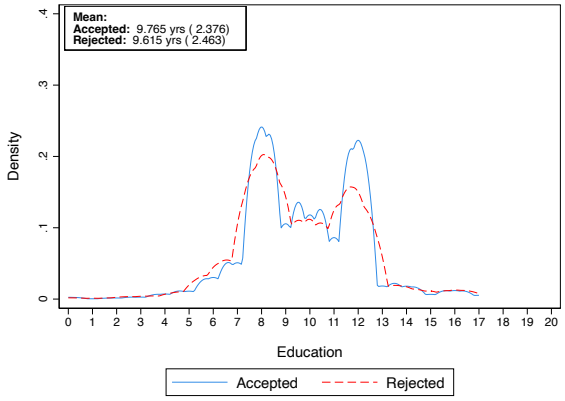
Notes: The kernel density estimate is plotted. Standard deviations are in parentheses. The DMF refers to boys in the original data based on the Death Mortality Files. The FS panels use data derived from FamilySearch records.

Figure A3: Distribution of the Outcomes - FS Boys

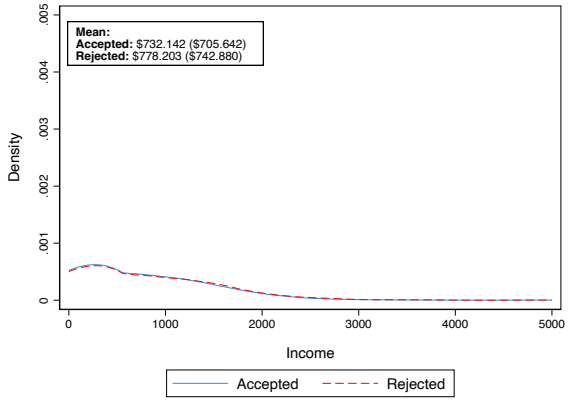
Longevity



Education

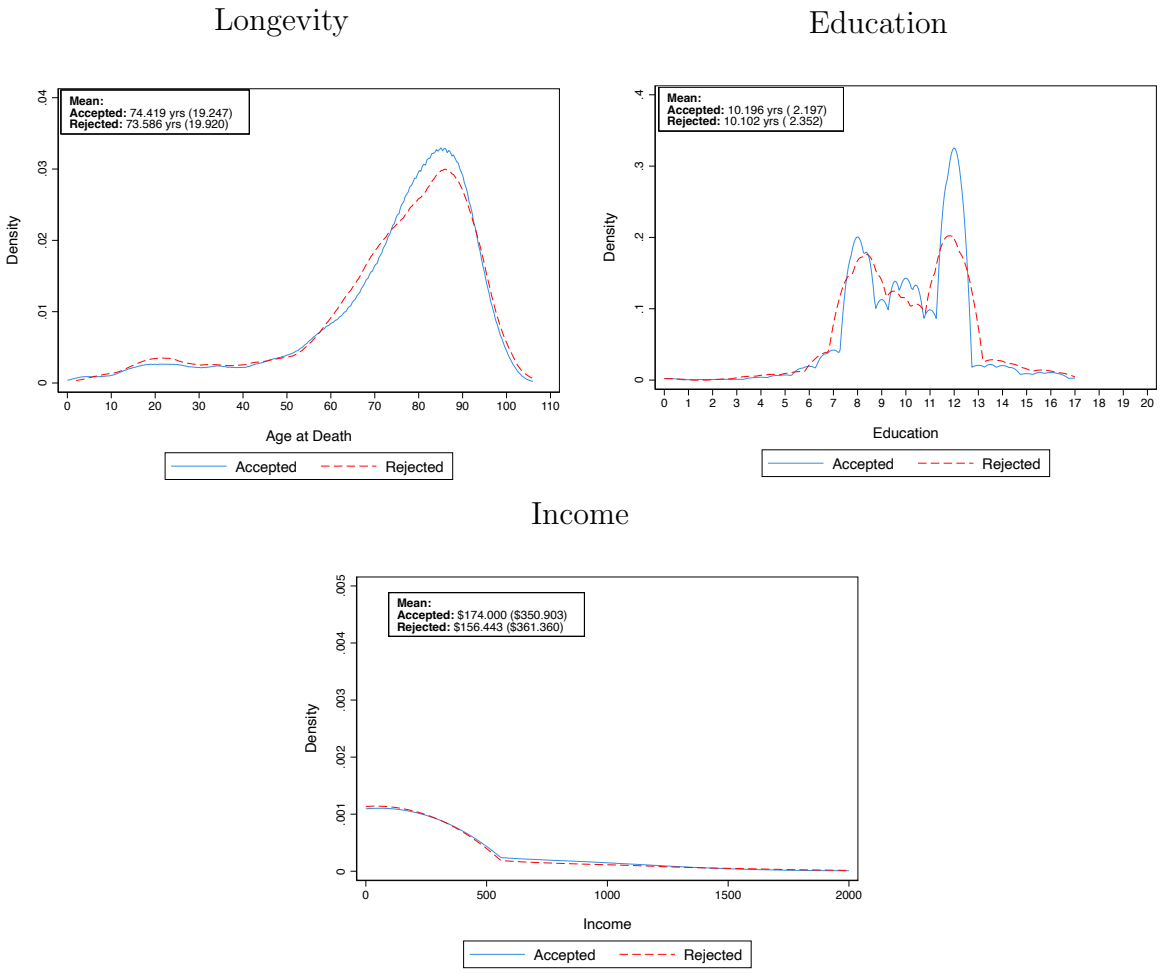


Income



Notes: The kernel density estimate is plotted. Standard deviations are in parentheses. The DMF refers to boys in the original data based on the Death Mortality Files. The FS panels use data derived from FamilySearch records.

Figure A4: Distribution of the Outcomes - FS Girls

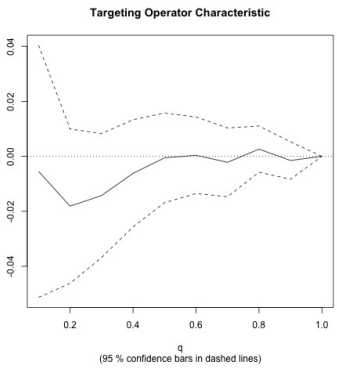


Notes: The kernel density estimate is plotted. Standard deviations are in parentheses. The DMF refers to boys in the original data based on the Death Mortality Files. The FS panels use data derived from FamilySearch records.

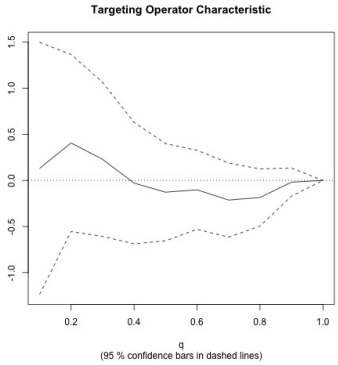
Figure A5: Target Operating Curves - Original and New Data

DMF

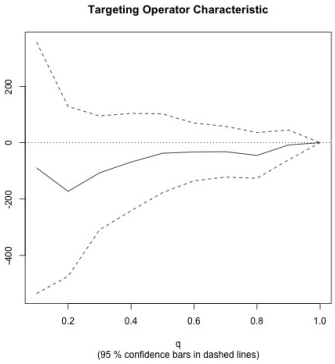
A. Age at Death



B. Education

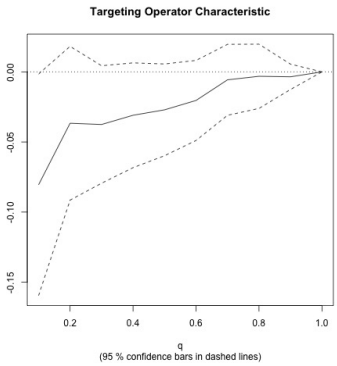


C. Income

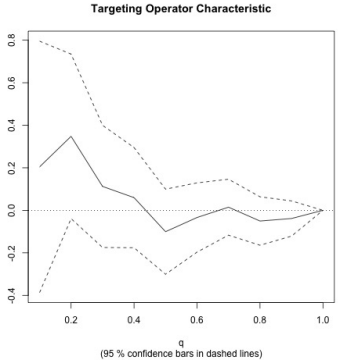


FS Boys

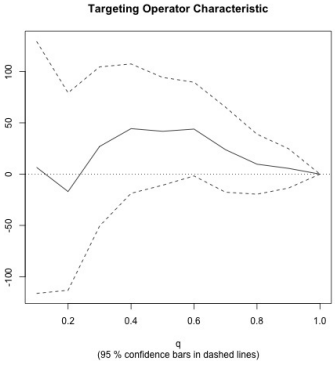
A. Age at Death



B. Education

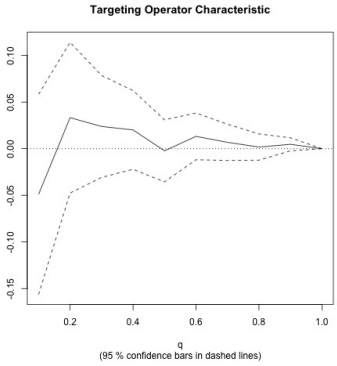


C. Income

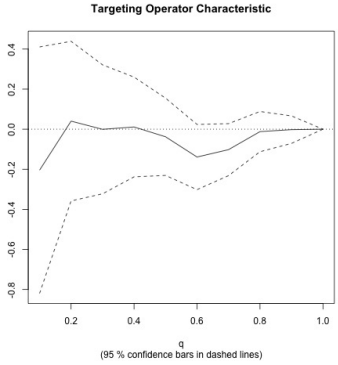


FS Girls

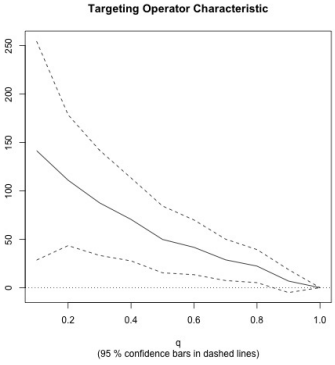
A. Age at Death



B. Education

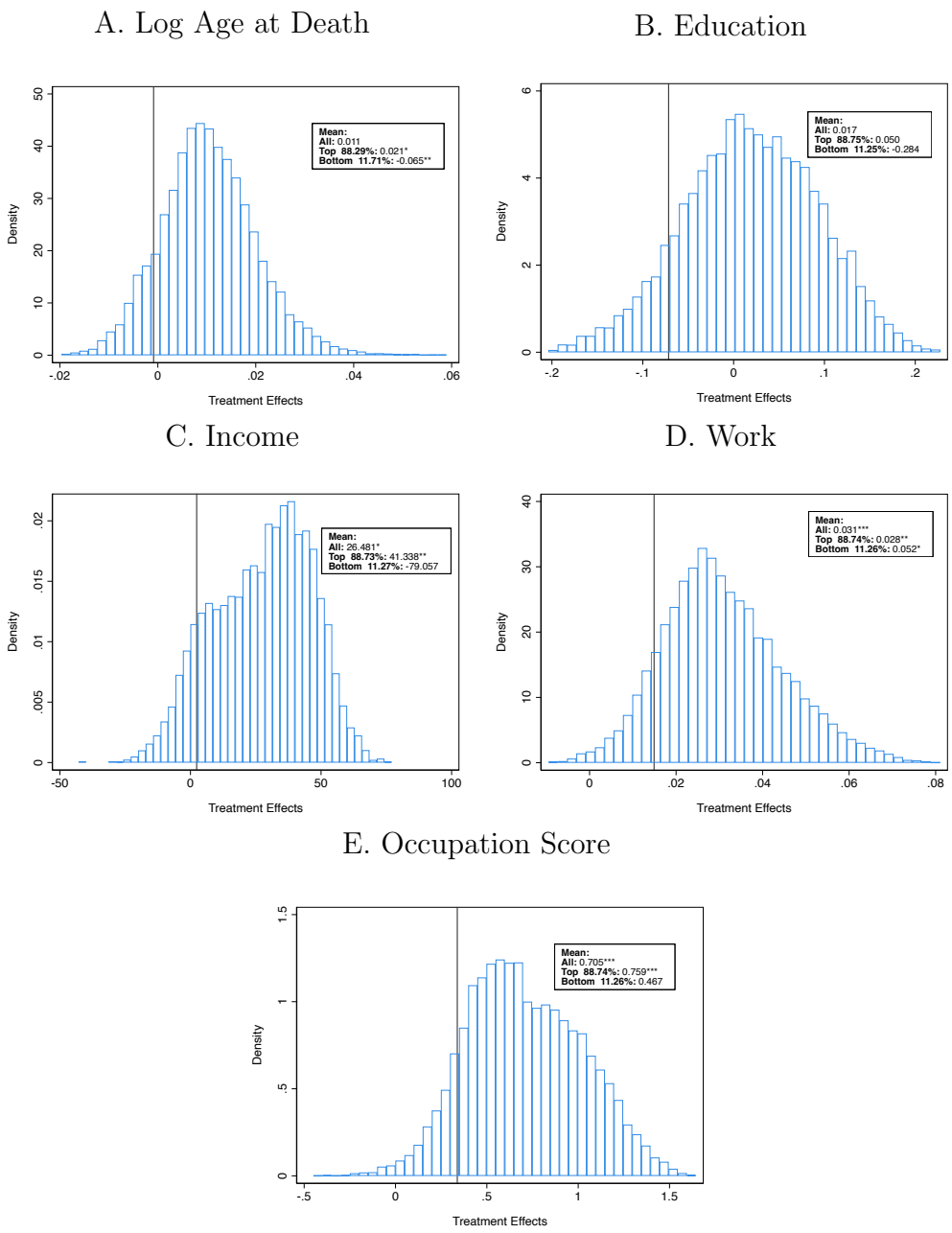


C. Income



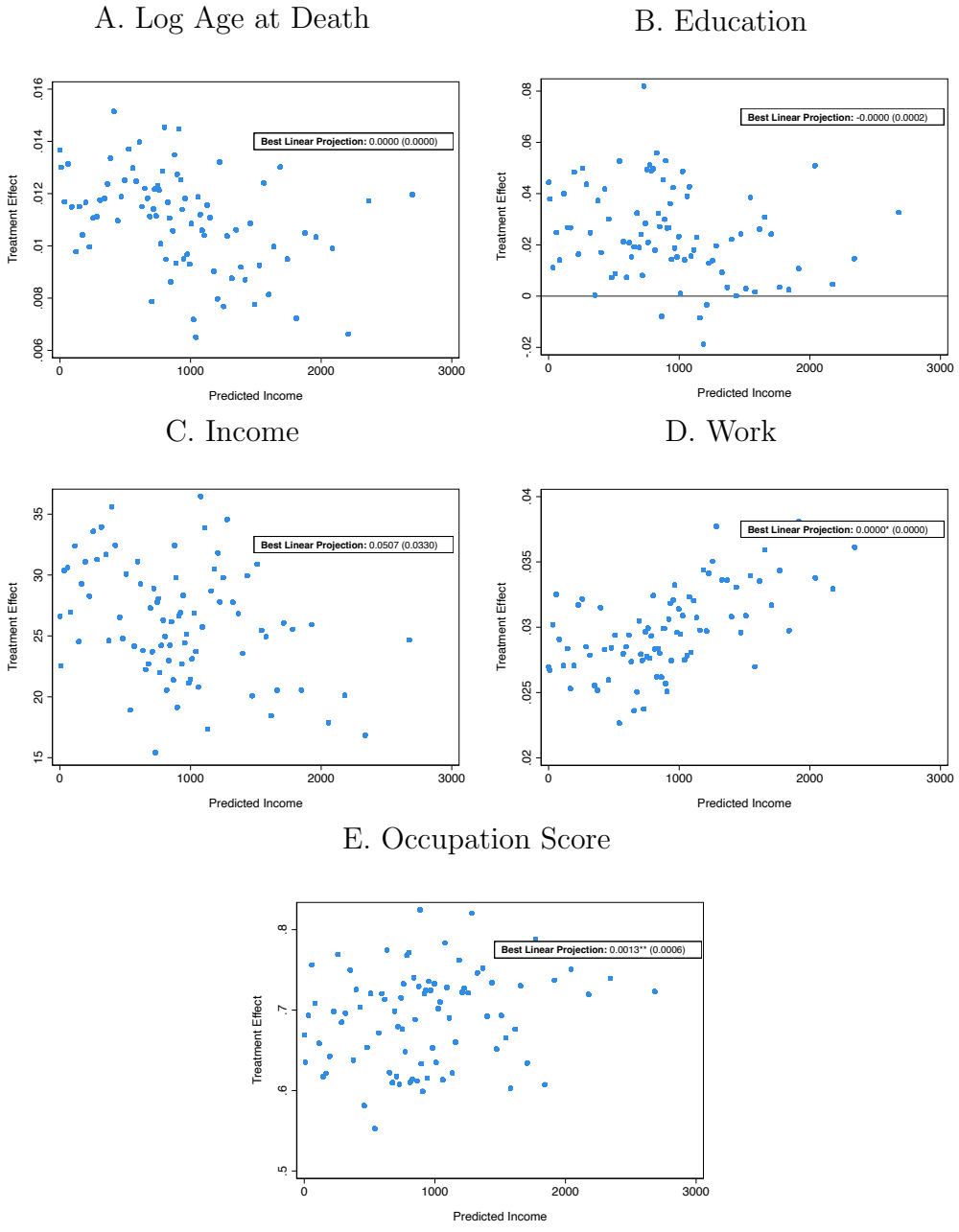
Notes: The Target Operating Curve plots the estimated average treatment effect among individuals ranked by predicted CATEs. The x-axis is the proportion of the population treated. The DMF refers to boys in the original data based on the Death Mortality Files. The FS panels use data derived from FamilySearch records.

Figure A6: Potential Gains



Notes: The figures plot distributions of individual-level treatment effects obtained from GRF. ATE Overlap estimates are provided for the full sample, for the top x% of the population, representing the accepted group, and for the remaining (100-x)%. Gray lines represent the cutoff for the top x%.

Figure A7: Treatment Effect vs. Predicted Income



Notes: The figure plots individual treatment effects binned into 100 groups by predicted income. A coefficient from projecting treatment effects on the predicted income is provided. Predicted Income is the estimated income based on observable characteristics of the family using 1915 Iowa Census. The characteristics used include family size, the ages of all siblings, maternal marital status, foreign-born status, race, educational attainment, occupational score, and their interactions.