

Learning from Administration Data (LAD) Project

Analysis Plan for the NEWS Evaluation

September 16, 2019

Project Background

More than twenty years have passed since the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) of 1996 dramatically altered the structure of the safety net in the United States, replacing the Aid to Families with Dependent Children program with Temporary Assistance for Needy Families (TANF). The new program included, among other things, work requirement after a certain period of benefit receipt and time limits on benefit paid by federal funds. Since TANF was a block grant to states, states were given significant latitude in the design of their programs, and many imposed stricter work requirements than required by federal law. Many also changed their benefit amounts to include more generous earnings disregards, to encourage work.

These changes, along with expansions to the Earned Income Tax Credit (EITC) over the past two decades, have meant that the safety net increasingly emphasizes and encourages work. And, in fact, research has documented that these policy changes led to notable declines in welfare caseload and increases in women's employment rates during the 1990s.¹ Questions remain, however, about whether the "work-first" welfare system raised family incomes enough to reduce poverty over the long-run or helped end the "cycle of poverty" by improving the outcomes of the children of welfare recipients.

The Learning from Administrative Data (LAD) initiative seeks to extend the evidence of the long-term effects of these types of welfare policies by assessing long-term impacts on children. The project will take advantage of multiple randomized controlled trials conducted by MDRC in the 1990s, including over 55,000 participants (mostly welfare recipients). Merging data for the study participants with the administrative data at Census provides a singular opportunity to study the long run (over 20-year) effects of these policies on outcomes such as employment and earnings, fertility, marriage, and mortality.

¹ See reviews by Grogger and Karoly (2009) and Ziliak (2015).

The programs studied included a range of policies to encourage work and/or reduce benefit use, such as earnings supplements, work mandates, time limits, childcare subsidies, and education services. Three examples include the following:

- **Minnesota Family Investment Program (MFIP):**
Period: 1994-1998
Target group: ongoing recipients and new applicants for public assistance.
Policies tested: more generous earnings disregard, mandatory participation in employment-related services
- **New Hope for Families and Children (New Hope):**
Period: 1994-1997
Target group: low-income families living in two Milwaukee neighborhoods.
Policies tested: earnings supplement, low-cost health insurance, subsidized child care, and community service job
- **National Evaluation of Welfare to Work Strategies (NEWWS):**
Period: 1991-1996
Target group: welfare recipients in 11 programs across the country
Policies tested: mandatory job search assistance and/or basic skills courses

This analysis plan focuses on the NEWWS evaluation. The NEWWS evaluation predates PRWORA and TANF but remains relevant in the ongoing debate about how best to move people into self-sufficiency without unintentionally jeopardizing the well-being of their children.

Intervention Description

NEWWS was designed to directly and reliably test the effects of alternative approaches to helping welfare recipients find jobs and leave public assistance. The evaluation examined the effects of 11 mandatory welfare-to-work programs on welfare recipients and their children. The interventions operated in seven sites around the country and took different approaches to helping welfare recipients find jobs, advance in the labor market, and leave public assistance. The timing of the random assignment intake varied across sites -- between June 1991 and December 1994. There were three approaches tested -- two primary preemployment approaches (one that emphasized short-term job search assistance and encouraged people to find jobs quickly and one that emphasized longer-term skill-building activities, primarily basic education, before entering the labor market) -- and a third approach that mixed elements of the other two. The programs were

operated in seven sites across the country: Atlanta, Georgia; Grand Rapids, Michigan; Riverside, California; Columbus, Ohio; Detroit, Michigan; Oklahoma City, Oklahoma; and Portland, Oregon.

The effects of the NEWWS programs were estimated based on a wealth of data on more than 40,000 single-parent families. The study was conducted as a randomized controlled trial, in which each parent was randomly assigned to a program group (in some sites, there were two program groups), whose members were eligible for program services and subject to the participation mandate, or a control group, whose members were not.

The study found that both the employment- and education-focused approaches substantially increased earnings during the five-year follow-up period, although the employment-focused approach produced larger increases in earnings and the effects emerged earlier in the follow-up period.² Most of the programs increased earnings during the second and third years of the follow-up period, and the effects generally faded by the fourth and fifth years. Only two programs continued to produce statistically significant earnings impacts at the end of the fifth year. A subsequent analysis found no effects on work or earnings 10 to 15 years after study entry.³

The programs reduced public assistance receipt over the five-year period. These reductions offset the increases in earnings, however, with the result that there were no significant impacts on income (measured as earnings plus public assistance and tax credits minus payroll taxes).

The evaluation examined effects on a range of other outcomes for parents and children. In sum:

- The programs had few effects on marriage, fertility or living arrangements of the participating adults, although it did appear to lead to a modest reduction in physical domestic abuse.
- There were few effects at the five-year mark on young children, or those who were pre-school aged at study entry. Outcomes examined included cognitive functioning, social skills and behavior, health, and safety.

² See for example, Hamilton et al. (2001) and Hamilton (2002).

³ Freedman (2008); Hamilton and Michalopoulos (2016).

- The programs led to small negative effects on academic outcomes for children who were adolescents at study entry (consistent with findings from other welfare-to-work programs).
- The programs increased the use of childcare, effects that mirrored the effects on employment and faded by year 5.
- The programs increased participation in job search, adult basic education, and, to a lesser extent, vocational training. They increased GED receipt among those without a diploma/GED at study entry. However, there were no measured effects on reading or math skills.

Theory of Change

Theoretically, there are various channels through which welfare policies, such as work requirements, time limits on welfare receipt, earnings supplements, reduced benefit reduction rates, and childcare subsidies can affect children. The mediators include financial resources expended on children; parental employment, education and behaviors; the quality and quantity of childcare arrangements; parental stress and mental health; and the quality of the home environment provided to the child.

NEWWS had effects on several of these mediators, such as notable increases in employment and earnings among parents, an increase in the use of childcare, and an increase in parents' education levels (GED receipt). Although in most sites the effects faded by Years 4 and 5, it is possible, and recent research supports the idea, that even a temporary change in family circumstances and childhood experiences can have long-term effects on children.⁴

Theory would also predict potential long-term effects on multiple outcomes for adults, owing to early increases in work, earnings, and education. Although, as noted earlier, a longer-term analysis from the NEWWS evaluation found no effects on adults' work or earnings 10 to 15 years after study entry, it is possible that effects might emerge as the economy changes or as the parents age. The 10- to 15-year follow-up, for example, took

⁴ Prominent recent examples include the Perry Preschool (early childhood education), Head Start, and Tennessee STAR (elementary school classroom) interventions (Heckman and Karapakula (2019), Chetty et al. (2011)). For a summary of the evidence on the long run effects of childhood receipt of social safety net programs see Hoynes and Schanzenbach (2018).

place for the most part before the Great Recession, so it is possible that more work experience or educational gain in earlier years may have affected how participants weathered the downturn.⁵ In addition, about a third of study participants were age 35 or older at study entry, and many of them will be nearing or at retirement age during the long-term follow-up period. Early work experience and basic education gains might affect patterns of employment over this key transition period. Less-educated workers, for example, retire at younger ages than more educated workers.⁶ Finally, research has documented the effects of education, unemployment, and job type on health and mortality, suggesting the potential for long-term effects in NEWWS.⁷

Confirmatory Research Questions

The key confirmatory research questions this study will address are the following:

1. Did the NEWWS program approaches generate long-run impacts on the employment and earnings of participants' children?
2. Did the cumulative long-run impacts on participants' children differ by program focus, e.g., for employment-focused programs versus education-focused programs?
3. Did the NEWWS program approaches lead to effects on later fertility, marriage, and mortality, for individuals who were children during the study period?
4. Do the program approach effects vary by children's gender and age at exposure to policy changes?

Exploratory Research Questions

Additional questions examined by the study include:

⁵ Hoynes, et al. (2012).

⁶ Venti and Wise (2015).

⁷ Sullivan and von Wachter (2009); Antonisse and Garfield (2008); Cutler and Lleras-Muney (2010).

1. Did the NEWWS program approaches generate long-run impacts on employment, earnings and mortality of adult participants?
2. Does NEWWS generate long-run impacts on participation in government assistance programs (e.g., Medicaid, subsidized housing, TANF, SNAP) among adult participants and their children?
3. Does NEWWS impact geographic mobility and the type of neighborhood (e.g., as characterized by crime rates and average educational attainment) where recipients and their children reside?
4. Do the long-term effects vary across subgroups, including those defined by parents' employment history, welfare receipt history, education level, housing status, marital status, and residential locations? Are the average effects estimated for the full sample similar across the distribution of outcomes, e.g., does the program lead to larger changes in earnings at the bottom of the earnings distribution?
5. Do the effects of different program approaches (employment- vs. education-focused) vary by family characteristics, such as parents' education level or employment history?
6. What are the potential mechanisms for the long-run impacts? To what extent do short-term impacts (e.g., in the 5-year follow-up study) on parental employment, income, educational attainment, or childcare use explain variation in long-term impacts across sites and interventions?⁸

Study Design

Main design: Random assignment

Secondary design: Instrumental variable-based mediation analysis

Eligible Population of Individuals and Sites

⁸ We observe educational attainment and childcare use in a subsample of about 5,400 adult participants (in the five-year client survey sample). Due to the smaller sample size, we recognize that these results may be more speculative.

For the analysis, we will use the full sample that was used in the main evaluation. This sample includes 41,715 adult welfare recipients and more than 81,000 children across all 11 sites.

Data Sources (Collection Methods and Instruments)

This study will rely on data from the original study and from the rich holding of Census data accessible to us via the Census Federal Statistical Research Data Centers.

The original study data includes the following information (see the Appendix for additional information):

1. **Client characteristics data** (N_Y2i=44,569 and N_Y5i=41,715 are the report samples for the two- and five-year follow-ups⁹ the full sample for Y5=44,569) were collected by welfare staff during interviews at time of random assignment.
2. **Private Opinion Survey** (N_Y2i=18,461) was a brief self-administered survey completed at random assignment at four sites (Atlanta, Grand Rapids, Riverside, and Portland).
3. **Reading and math tests** (N_Y2i=20,577) were administered at random assignment at four sites (Atlanta, Grand Rapids, Riverside, and Portland).
4. **Unemployment Insurance, welfare, and food stamps** records (N_Y2i=44,569) were collected from automated county and state systems for a 5-year follow-up period (except, TANF and FS participants in Oklahoma).¹⁰
5. **Two-year client survey** (N_Y2i=9,675) in 7 sites **and five-year client survey** (N_Y5i=5,463) in 7 sites, provided information about sample members. participation in training and education activities, attainment of education credentials, views of work and welfare, employment history, income, receipt of noncash benefits such as health coverage, childcare use, living situation and children's well-being. In addition, reading and math tests were administered to a subset of adults as part of each survey, to assess effects on adult literacy.

⁹ For the Final Report, MDRC decided to exclude two subgroups whose members had been included in the analysis for the 2-Year Impact Report. **Atlanta:** All sample members randomly assigned during July 1993 through January 1994, because they lacked five years of follow-up data for welfare and Food Stamp payments. **Portland:** The 75 percent of control group members who became eligible for program services after year 3. In addition, **Oklahoma City** sample members with fewer than 5 years of follow-up data for welfare and Food Stamp payments were excluded from calculations of program impacts on welfare, Food Stamps, and combined income.

¹⁰ All sample members have 5 years of follow-up data from UI earnings records, but some sample members in Atlanta and Oklahoma City have fewer than 5 years of welfare and Food Stamp data.

6. **Child Outcomes Study (COS) survey** (N_Y2i=3,018/N_Y5i=2,332) in 3 sites, provided information on focal children’s academic functioning, social skills, and health and safety. There was also a self-administered survey for mothers and focal children. The survey to the mother included questions about domestic abuse and the focal child survey included questions about academic functioning and social skills.
7. **Teacher survey** (N_Y5i=1,472) for focal children focused on questions to assess them with respect to their academic standing, academic progress, school engagement, behaviors requiring disciplinary action, and social skills.

The Census data holdings, available for study participants and their children, include the following:

- Employment and earnings from the IRS; and from state UI agencies via the LEHD;
- Federal tax credit (e.g. EITC) receipt, and marital status (based on tax filing status) from the IRS;
- Medicaid enrollment and utilization from CMS;
- Social Security disability program participation from SSA;
- Public housing program participation from HUD;
- Cash (TANF) and food assistance (SNAP) program participation from selected states;
- Fertility and death records from SSA, for analyses on teen and out-of-wedlock pregnancy, and mortality.
- California Franchise Tax Board data (covers only recent residents of California)

Sample Size and Minimum Detectable Effect (Sizes)

The table below presents the sizes of various analysis samples and the minimum detectable effects for earnings and employment. Also shown in the final column is the minimum detectable effect size. Given the large samples, the analysis will be able to detect fairly small impacts on key outcomes of interest.

Sample	N	Annual earnings	Employment (0/1)	Effect size
Adults	41715	268	0.012	0.024
Children	81000	192	0.009	0.017
Children less than age 12	54000	235	0.011	0.021

Children ages 12 to 18	27000	333	0.015	0.030
HCD adults	10927	523	0.024	0.048
LFA adults	12676	486	0.022	0.044
HCD children	21000	377	0.017	0.034
LFA children	25000	346	0.016	0.031

Outcome Variables

Long-term earnings for individuals who were children during the study period will be calculated as average annual earnings between a given age range, such as ages 26 to 30. Focusing on the latest interval available for the child sample reduces the number of outcomes examined and gets closer to the age at which earnings begin to stabilize. The exact range will depend on the available data and the sample of children used.¹¹

For adults, earnings will be defined as average annual earnings between the final few years of available data, e.g., 2012-2015. The main analysis will focus on all adults younger than age 60 in the final follow-up years, although an exploratory analysis will examine the effects over the years surrounding the age of retirement.

Outcomes related to fertility, marriage and mortality for the children will be measured as follows: had a birth before marriage, had a birth before age 20, married as of age 30 (pending outcome data availability), and mortality as of age 30 (pending outcome data availability).

Subgroup Definitions

Confirmatory:

1. Age of child at study entry (<6, 6-11, 12+)
2. Gender of child
3. Atlanta, Grand Rapids, and Riverside sites, where participants were randomized into Human Capital Development, Labor Force Attachment, or control programs

¹¹ The IRS form 1040 returns cover tax years 1969, 1974, 1979, 1984, 1989, 1994, 1995, 1998-2016; The IRS information returns cover tax years 1996-2004 (1099-R), 2003-2015; The unemployment insurance records (LEHD) covers years 1990-2014; the CA state tax dataset covers tax years 2010-2015.

Exploratory:

1. Education level of parent (has HS diploma/GED versus not)
2. Employment history of parent (worked in year prior to study entry versus did not work)
3. Program Welfare history of parent
4. Number of employment barriers
5. Race/ethnicity
6. Age of parent
7. Local economy
8. Household composition
9. Other exploratory subgroups may include those defined using information from the Private Opinion Survey, reading and math tests, the Child Outcomes Study survey, and the Teacher survey.

Estimation strategy

Basic model for full sample

The basic estimation strategy is to compare average outcomes for the program and control groups. In particular, we will estimate the following regression model, over the pooled sample, of the effect of assignment to the treatment (T) on long run outcomes (Y) for each individual i in site j:

$$Y_{ij} = \alpha + T_{ij}\beta + X_{ij}\lambda + S_{ij}\gamma_j + \varepsilon_{ij}$$

We are interested in the estimate of β , which captures the effect of NEWWS on the long run outcome (for example, earnings in young adulthood). The vector X consists of baseline characteristics such as parental age, race, education, employment and welfare history. We are guided by prior research as to the relevant baseline characteristics to include. The vector S includes site dummies.

Given the random assignment of T, including the covariates X is not required for obtaining an unbiased estimate of the welfare experiments effects, though it may help to

improve the precision of the estimator.¹² The standard errors would be clustered at the family level, to account for correlation across multiple children.

Model for subgroups

To explore heterogeneity across subgroups, we will estimate the following model:

$$Y_{ij} = \alpha + T_{ij}\beta_1 + T_{ij}D_{ij}\beta_2 + X_{ij}\lambda + S_{ij}\gamma_j + \varepsilon_{ij}$$

In this example, D is an indicator variable equal to 1 for a particular subgroup (e.g., women) and equal to 0 for their counterparts (men). In this case, β_1 would capture the effects for men and β_2 would capture the difference in effects for women. Separate models would be estimated for each subgroup (e.g., age, education level of parents), with the subgroup by treatment interaction. In each case, the variable D (indicating the subgroup of interest) would need to be included in the control vector X to capture the main effect between D and Y in the control group.

In addition to heterogeneity of impacts by subgroups, we can also explore how experiments affected the distribution of earnings and income of parents and children. In the models above, the treatment effect β provides the long run effect of reform on average earnings (or other outcomes). Bitler, Gelbach and Hoynes (2006), however, found that in the Jobs First experiment, the average impacts mask important effects on the distribution: There were few effects on earnings at the bottom of the distribution but positive effects among families with relatively higher earnings. Following their methodology, we will use quantile regression models to estimate quantile treatment effects (QTE) to examine impacts on the distribution of earnings (or another continuous outcomes). If, for example, we find positive effects of a NEWWS program approach on long run earnings, the QTE could reveal whether this effect is the result of gains in earnings occurring throughout the distribution or concentrated at the middle or higher areas of the earnings distribution.

Mediation Analysis

Finally, we will explore the possibility of conducting mediation analysis to examine how the long-run effects of short-term changes in welfare policies are transmitted through their initial impact on outcomes such as family income, maternal employment, and

¹² The earlier NEWWS publications demonstrated program and control balance, indicating that randomization was valid.

educational attainment¹³ by combining the results across the different NEWS sites.

We will estimate models where we seek to identify the effect of mediators M (where M could be a vector of mediators) on outcome Y for individual i in experimental site j :

$$Y_{ij} = \alpha + M_{ij}\delta + X_{ij}\lambda + \varepsilon_{ij}$$

Given the findings in existing evaluations, we are particularly interested in mediators including family income (through two and four years post study entry), maternal employment (share of quarters with earnings through two and four years post study entry), maternal education and training, and use of childcare. M is endogenous but we will use (site \times treatment) dummies as instruments.¹⁴

Given the multiple sites, we have the potential to identify multiple mediating channels. This approach provides a natural extension of the existing MDRC Next Generation projects and other related work on welfare to work experiments.¹⁵ This IV approach is also used in other settings such as MTO and the growing methodological and applied research on *multisite multi mediator* settings in education research.¹⁶

As in all instrumental variable settings, the success of this approach depends on two factors. We need a strong first stage, with contrasting treatments across sites and independent effects on the different mediators. Our findings from the main analysis will reveal whether there is variation across sites in effects on key mediators. If not, we will postpone this analysis until additional studies, such as MFIP and New Hope, are brought into the analysis.

The first stage of this IV approach is essentially the evaluations that have already taken place. Thus, we have the advantage of knowing that they have predictive power as well as guidance about what upstream channels were affected by the intervention. Second, is the exclusion restriction – that the interventions have no relationship with the outcome except through the mediators we have specified. Additionally, the model assumes

¹³ Educational attainment and childcare use information is available for a subsample of about 5400 adult participants.

¹⁴ Some of the sites include more than one treatment arm. Thus, the instruments are at the site by treatment arm level.

¹⁵ See Gennetian et al. (2002), Duncan et al. (2011), and Bloom et al. (2003).

¹⁶ See Kling et al. (2007) for MTO and Raudenbush and Bloom (2015) for work in education.

linearity (and homogeneity) in M , which can be visually examined by producing a scatterplot with a point for each site-treatment combination, in which the first stage estimates are plotted along the x-axis and the reduced form effects plotted along the y-axis.

The results from this stage of the analysis will document the long-run effects of each intervention and provide a nuanced picture of the intergenerational impacts of the various program approaches.

Accounting for Multiple Inference (Multiple Hypothesis Testing)

As discussed above, in our research agenda we will examine a wide range of outcomes including earnings, employment, program participation, fertility, marriage, mortality, and neighborhood characteristics. The large number of possible outcome variables complicates inference, and we will use appropriate corrections for multiple hypothesis testing for our confirmatory analyses, e.g. the techniques reviewed by Anderson (2008).

One possibility to reduce the dimensionality of outcomes is to follow the literature and create indices representing separate outcome areas such as in Kling, Liebman, and Katz (2007) in their analysis of Moving to Opportunity and Hoynes et al. (2016) in their analysis of the long run effects of food stamps. In this case, the index is constructed as:

$$Y_i = \frac{\sum_{j=1}^J \frac{(Y_{ij} - \mu_j)}{\sigma_j}}{J}$$

Where each outcome is standardized using the mean (μ_j) and standard deviation (σ_j) in the control group and then averaged across all of the different outcomes j . Following the literature, we may consider constructing indices for human capital, economic self-sufficiency, family/demography, and neighborhood quality.

References

- Anderson, Michael L. "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American statistical Association* 103, no. 484 (2008): 1481-1495.
- Antonisse, Larisa, and Rachel Garfield. "The relationship between work and health: Findings from a literature review." *Kaiser Family Foundation* 7 (2018).
- Bitler, Marianne P., Jonah B. Gelbach, and Hilary W. Hoynes. "What mean impacts miss: Distributional effects of welfare reform experiments." *American Economic Review* 96, no. 4 (2006): 988-1012.
- Bloom, Howard S., Carolyn J. Hill, and James A. Riccio. "Linking program implementation and effectiveness: Lessons from a pooled sample of welfare-to-work experiments." *Journal of Policy Analysis and Management* 22, no. 4 (2003): 551-575.
- Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. "How does your kindergarten classroom affect your earnings? Evidence from Project STAR." *The Quarterly journal of economics* 126, no. 4 (2011): 1593-1660.
- Cutler, David M., and Adriana Lleras-Muney. "Understanding differences in health behaviors by education." *Journal of health economics* 29, no. 1 (2010): 1-28.
- Duncan, Greg J., Pamela A. Morris, and Chris Rodrigues. "Does money really matter? Estimating impacts of family income on young children's achievement with data from random-assignment experiments." *Developmental psychology* 47, no. 5 (2011): 1263.
- Freedman, Stephen, and Jared Smith. 2008. "Examining the Effectiveness of Different Welfare-to-Work Approaches: Extended Follow-Up of TANF and Employment Outcomes for the National Evaluation of Welfare-to-Work Strategies (NEWWS) Project." Unpublished memorandum to the U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation. New York: MDRC.
- Gennetian, Lisa A., Johannes M. Bos, and Pamela A. Morris. "Using instrumental variables analysis to learn more from social policy experiments." *Moving to the next level: Combining experimental and nonexperimental methods to advance employment policy research*. New York: Russell Sage Foundation (2002).
- Grogger, Jeff, Lynn A. Karoly, and Jeff Grogger. *Welfare reform: Effects of a decade of change*. Harvard University Press, 2009.

Hamilton, Gayle. 2002. *Moving People from Welfare to Work: Lessons from the National Evaluation of Welfare-to-Work Strategies*. Washington, DC: U.S. Department of Health and Human Services.

Hamilton, Gayle, Stephen Freedman, Lisa Gennetian, Charles Michalopoulos, Johanna Walter, Diana Adams-Ciardullo, Anna Gassman-Pines et al. "How effective are different welfare-to-work approaches." *Five-Year Adult and Child Impacts for Eleven Programs. National Evaluation of Welfare-to-Work Strategies*. Washington, DC: US Department of Health and Human Services (2001).

Hamilton, Gayle, and Charles Michalopoulos. 2016. "Job Search or Basic Education Participation First: Which Improves Welfare Recipients' Earnings More in the Long Term?." MDRC.

Heckman, James J., and Ganesh Karapakula. *The Perry Preschoolers at late midlife: A study in design-specific inference*. No. w25888. National Bureau of Economic Research, 2019.

Hoynes, Hilary, Douglas L. Miller, and Jessamyn Schaller. "Who suffers during recessions?" *Journal of Economic perspectives* 26, no. 3 (2012): 27-48.

Hoynes, Hilary W., and Diane Whitmore Schanzenbach. *Safety net investments in children*. No. w24594. National Bureau of Economic Research, 2018.

Hoynes, Hilary, Diane Whitmore Schanzenbach, and Douglas Almond. "Long-run impacts of childhood access to the safety net." *American Economic Review* 106, no. 4 (2016): 903-34.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. "Experimental analysis of neighborhood effects." *Econometrica* 75, no. 1 (2007): 83-119.

Raudenbush, Stephen W., and Howard S. Bloom. "Learning about and from a distribution of program impacts using multisite trials." *American Journal of Evaluation* 36, no. 4 (2015): 475-499.

Sullivan, Daniel, and Till Von Wachter. "Job displacement and mortality: An analysis using administrative data." *The Quarterly Journal of Economics* 124, no. 3 (2009): 1265-1306.

Venti, Steven, and David A. Wise. "The long reach of education: early retirement." *The Journal of the Economics of Ageing* 6 (2015): 133-148.

Ziliak, James P. "Temporary assistance for needy families." In *Economics of Means-Tested Transfer Programs in the United States, Volume 1*, pp. 303-393. University of Chicago Press, 2015.

Appendix

Description of Data Sources for the Five-Year Impact Study

The outcomes and impacts presented in this report are drawn from four primary data sources: unemployment insurance, welfare, and Food Stamp administrative records; surveys of sample members that were conducted at the two-year and five-year follow-up points (the Two-Year Client Survey and the Five-Year Client Survey); a survey of sample members focused on outcomes for children (the Child Outcomes Study survey); and a teacher survey.

Client characteristic data. Standard personal data, such as educational background and welfare history, were collected by welfare staff during interviews at the time of random assignment and are available for all 41,715 heads of the single-parent families in the full impact sample.

Private Opinion Survey. Data on attitudes and opinions about welfare-to-work programs and employment prospects were collected through the Private Opinion Survey (POS), a brief, self-administered survey that was completed at random assignment in four of the sites (Atlanta, Grand Rapids, Riverside, and Portland), and are available for 18,461 respondents in these sites. These sample members represent 93 percent of those randomly assigned in the four sites during the periods when the POS was being administered.

Reading and math tests. Reading and math achievement tests were administered in four sites (Atlanta, Grand Rapids, Riverside, and Portland) at random assignment. Test scores are available for 20,577 sample members. These sample members represent about 93 percent of those randomly assigned in the four sites during the period when the tests were administered.²³

Field research. MDRC staff observed all 11 programs in operation and interviewed enrollees, case managers, service providers, and program administrators in each site. Information was collected about a range of issues, such as management philosophy and structure, the degree to which the participation mandate was enforced, the nature of interactions between caseworkers and program participants, the extent to which the program was able to work with all those mandated to participate in it, the availability of services, and the relationships that program staff had established with outside service providers and income maintenance staff in the sites.

Unemployment insurance, welfare, and Food Stamp administrative records data. Most employment, earnings, and public assistance impacts were computed using automated county and state unemployment insurance (UI), welfare, and Food Stamp administrative records data.

Five years of follow-up data from the UI system are available for all members of the full impact sample; five years of follow-up data from welfare and Food Stamp administrative records are available for all sample members in all sites except Oklahoma. UI earnings, which are recorded statewide, provide unbiased measures of program impacts on employment and earnings. These data, however, do not include earnings from out of state; from jobs not usually covered by the UI system, such as self-employment, federal employment, or informal child care (all types of work that may have been “off the books”); or from employers who do not report earnings. Some of the earnings missed by the UI system may be captured by earnings and employment data collected through the two-year and five-year surveys.

In all sites except Riverside, welfare and Food Stamp payments were also recorded statewide, and payments are captured for all sample members except those who moved out of state. In Riverside (as everywhere in California), welfare and Food Stamp payments were recorded only within each county, which means that payments received by sample members who moved outside the county were not included in the analysis. Although this could lead to an underestimate of the payments received in the Riverside sample, it should not bias the impact estimates because there is no reason to expect the program and control groups to show different patterns of moving between counties.

UI earnings data are collected by calendar quarter: January through March, April through June, and so forth. For purposes of the evaluation, these data were reorganized so that the quarter during which a sample member was randomly assigned is always designated quarter 1, followed by quarter 2, and so forth. These quarters are then grouped into “years.” Quarter 1 is not included in year 1 because it includes some income earned before random assignment, especially for sample members randomly assigned near the end of a calendar quarter. Thus, year 1 covers quarters 2 through 5, year 2 covers quarters 6 through 9, and so forth.

Welfare and Food Stamp payments were recorded monthly but were grouped into quarters and years to align with the earnings data.

Two-Year Client Survey and Five-Year Client Survey. As noted in a previous section, this report includes the results of a survey administered at the five-year follow-up point and some results of a survey administered at the two-year follow-up point. Both the two-year and five-year surveys provide information about sample members' participation in training and education activities, attainment of education credentials, views of work and welfare, employment history, income, receipt of noncash benefits such as health coverage, childcare use, living situations, and children's well-being. Survey responses are the only source of information about many key outcomes, such as participation patterns for control group members, work hours and wages, income from other people in the household, and outcomes for children. For some outcomes, such as employment, respondents provided information that was also recorded from administrative data. It is possible for data from these two sources to differ. Because the five-year survey respondents represent a subsample of the full impact sample that was selected during a shorter period of random assignment months, the impact and survey samples may differ with respect to observed characteristics (such as educational attainment or prior work history) or with respect to unmeasured characteristics (such as assertiveness or learning style) that might have affected their ability to find and retain employment. (For more information on survey response bias and the degree to which the survey sample and full impact samples differ, see Appendix G.)

In some cases, administrative records data may be more accurate than the survey data. The client survey depends on people's ability to recall information about events or jobs that they may have held up to five years prior to being interviewed, and failures of memory can give rise to discrepancies between the dates of employment or amounts of earnings reported in the survey and reflected in administrative records. In addition, some respondents may have been reluctant to provide information on employment and income that could be found in administrative records or, alternatively, may have exaggerated their earnings and income. In other cases, however, survey data may be more accurate, such as when respondents were working off the books or in short-term employment. The survey may also have captured earnings that employers failed to report or reported inaccurately to the UI system. (For more information on the differences between UI reported and survey-based measures of earnings, see Appendix H.)

Additional COS survey data. COS respondents provided information on focal children's academic functioning, social skills, and health and safety. In addition, mothers and the focal children themselves completed a Self-Administered Questionnaire (SAQ). Mothers' SAQ included questions about domestic abuse; children's SAQ included questions about academic functioning and social skills.

Teacher survey. Current teachers of focal children in the COS were asked to assess them with respect to their academic standing, academic progress, school engagement, behaviors requiring disciplinary action, and social skills. The teacher survey complements the data collected from mothers and the children themselves. Reports from teachers and mothers sometimes differ. Possible explanations include the following: The children behaved differently in the presence of mothers and teachers, mothers and teachers perceived the children's behavior differently, or mothers and teachers based their reports on different criteria.

Cost data. The cost analysis used data drawn from state, county, and local fiscal records, supportive service payment records, administrative records, the Two-Year Client Survey, the Five-Year Client Survey, and case file participation records.

Benefit-cost data. The benefit-cost analysis is based on administrative records data (UI reported earnings, welfare, and Food Stamp payments), Two-Year Client Survey data, Five-Year Client Survey data, and published data.

Published data and agency reports. Published data and reports from government agencies were used to gather additional information about the environments in each of the sites, including unemployment rates, welfare caseloads, and welfare grant levels.