What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments

Marianne P. Bitler
RAND Corporation

Jonah B. Gelbach
University of Maryland

Hilary W. Hoynes
University of California, Davis and NBER*

First version: January 2003
This version: June 1, 2004

Abstract

Labor supply theory predicts systematic heterogeneity in the impact of recent welfare reforms on earnings, transfers, and income. Yet most welfare reform research focuses on mean impacts. We investigate the importance of heterogeneity using random-assignment data from Connecticut's Jobs First waiver, which features key elements of post-1996 welfare programs. Estimated quantile treatment effects exhibit the substantial heterogeneity predicted by labor supply theory. Thus mean impacts miss a great deal. Looking separately at samples of dropouts and other women does not improve the performance of mean impacts. We conclude that welfare reform's effects are likely both more varied and more extensive than has been recognized.

*Correspondence to Hoynes at UC Davis, Department of Economics, 1152 Social Sciences and Humanities Building, One Shields Avenue, Davis, CA 95616-8578, phone (530) 752-3226, fax (530) 752-9382, or hwhoynes@ucdavis.edu; Gelbach at gelbach@glue.umd.edu; or Bitler at bitler@rand.org. Bitler gratefully acknowledges the financial support of the National Institute of Child Health and Human Development and the National Institute on Aging. This work has not been formally reviewed or edited. The views and conclusions are those of the authors and do not necessarily represent those of the RAND Corporation. The data used in this paper are derived from data files made available to researchers by MDRC. The authors remain solely responsible for how the data have been used or interpreted. We are very grateful to MDRC for providing the public access to the experimental data used here. We would also like to thank Dan Bloom, Mary Daly, Jeff Grogger, Richard Hendra, Guido Imbens, Sanders Korenman, Chuck Michalopoulos, Lorien Rice, Susan Simmat, Jeff Smith, Till von Wachter, Arthur van Soest, and Johanna Walter for helpful conversations, as well as seminar participants from Berkeley, Chicago–Harris School, Cornell, Davis, Delaware, GW, the IRP summer workshop, Johns Hopkins, Maryland, Michigan, the NBER, PAA, PPIC, SOLE, the RAND Corporation, and Syracuse.

1 Introduction

Several years have now passed since the elimination of Aid to Families with Dependent Children (AFDC), the principal U.S. cash assistance program for six decades. In 1996, enactment of the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) required all 50 states to replace AFDC with a Temporary Assistance for Needy Families (TANF) program. State TANF programs differ from AFDC in many fundamental ways. Key examples include lifetime limits on program participation, enhanced work incentives through expanded earnings disregards, stringent work requirements, and financial sanctions for failure to comply with these requirements. Theory makes heterogeneous predictions concerning the sign and magnitude of labor supply and welfare use responses to such reforms. Thus mean impacts will tend to average together positive and negative labor supply responses, possibly obscuring the extent of welfare reform's effects. Thus a critical element in evaluating the dramatic changes in welfare policy is to measure the impact of TANF on earnings and income in a way that allows for heterogeneous treatment effects.

The welfare reform literature that has developed in the last several years is enormous. We confine our discussion of this literature to a few particularly relevant papers; excellent comprehensive summaries of this research appear in reviews by Blank (2002), Moffitt (2002), and Grogger, Karoly & Klerman (2002). Nonexperimental studies (*e.g.*, Moffitt (1999) and Grogger (Forthcoming)) have found mixed results concerning the impact of welfare reform on income. Experimental studies examining pre-PRWORA state reforms suggest that generous increases in earnings disregards are important for generating mean income gains, but these gains disappear after time limits take effect (*e.g.*, Bloom & Michalopoulos (2001), Grogger et al. (2002)). With respect to treatment effect heterogeneity, Schoeni & Blank (2003) compare the full distribution of the income-to-needs ratio before and after TANF, finding increases at all but the very lowest percentiles. However, as they discuss, their simple before-and-after methods cannot distinguish impacts of TANF from the effects of strong labor markets. The most common way to address distributional concerns is to estimate mean impacts for subgroups of the population (defined using education, race, and welfare and employment history) thought to be particularly at risk for welfare dependence.[1] Michalopoulos &

[1]Schoeni & Blank (2000) compare the 20th and 50th percentiles of the CPS family income distribution before and after implementation of TANF. They find negative (but insignificant) impacts of TANF on the 20th percentile, and positive and significant impacts on the 50th percentile for a sample of women with less than a high school education.

Schwartz (2000) review 20 randomized experiments and conclude that "Although the programs did not increase [mean] income for most subgroups they also did not decrease [mean] income for most subgroups" (p. ES-10). Grogger et al. (2002) summarize both nonexperimental and experimental evidence concerning mean impacts as follows: "the effects of reform do not generally appear to be concentrated among any particular group of recipients" (p. 231).

In this paper, we address heterogeneous theoretical predictions by estimating quantile treatment effects (QTE) across the distributions of earnings, transfer payments (cash welfare plus Food Stamps), and total measurable income (the sum of earnings and transfers).[2] We do so using public-use data files from the Manpower Demonstration and Research Corporation's (MDRC) experimental evaluation of Connecticut's Jobs First waiver from AFDC rules. The Jobs First program, which we discuss in detail below, has both the most generous earnings disregard in the nation and the strictest time limit. It thus provides ideal terrain for investigating whether theoretically predicted treatment effect heterogeneity actually occurs. As discussed in Blank (2002) and formalized in Bitler, Gelbach & Hoynes (2003*b*), identifying the impact of TANF using nonexperimental methods is difficult given that TANF was implemented in all states within a very short period and during the strongest economic expansion in decades. Having access to experimental data is particularly useful in the present context.

Our empirical findings may be summarized with four important conclusions. First, we find evidence of substantial heterogeneity in response to welfare reform. Second, the heterogeneity is broadly consistent with the predictions of static labor supply theory. We find that Jobs First had no impact on the bottom of the earnings distribution, it increased earnings in the middle of the distribution, and—before time limits took effect—it reduced earnings at the top of the distribution.

Some of the MDRC waiver evaluations (e.g., Bloom, Scrivener, Michalopoulos, Morris, Hendra, Adams-Ciardullo & Walter (2002) and Bloom, Kemple, Morris, Scrivener, Verma & Hendra (2000)) include estimates comparing the fraction of treatment and control group members with income in broad categories. This approach, essentially a tabular form of histogram plots, is similar in spirit to ours. With respect to subgroups, Schoeni & Blank (2000) find that welfare reforms led to increases (insignificant in the case of TANF) in mean family income for female dropouts in the CPS. Using similar data, Bennett, Lu & Song (2002) find that TANF is associated with reductions in the income-to-needs ratio for poor children who live with a single parent having less than a high school education.

[2]QTE have been used in previous experimental evaluations. Examples of their use in evaluating the Job Training and Partnership Act include Heckman, Smith & Clements (1997), Firpo (2003), and Abadie, Angrist & Imbens (2002); Friedlander & Robins (1997) estimate QTE in evaluating effects of job training in earlier welfare reform experiments. However, the source of heterogeneous treatment effects in these cases is difficult to identify, since they mostly involve changes to training programs or job search assistance. Unlike such black-box reforms, in the present context it is clear why theoretical predictions are heterogeneous.

Third, contrary to much recent discussion among policymakers and researchers, our results suggest the possibility that welfare reform reduced income for a nontrivial share of the income distribution after time limits take effect. Fourth, we find that the essential features of our empirical findings could not have been revealed using mean impact analysis on judiciously chosen subgroups. In particular, the intra-group variation in QTE greatly exceeds the inter-group variation in mean impacts.

The remainder of the paper is organized as follows. In section 2 we provide a overview of the Jobs First program and its theoretically predicted effects. We then discuss our data in section 3. In section 4, we present empirical evidence that strongly suggests the time limit was an important program feature, and we present mean treatment effects in section 5. Our main QTE results appear in section 6, and we conclude in section 7.

2 The Jobs First Program and Its Economic Implications

Table 1 summarizes the major features of Connecticut's Jobs First waiver program. The Jobs First waiver contained each of the key elements in PRWORA: time limits, work requirements, and financial sanctions. For comparison, the table also includes a summary of the pre-existing AFDC program. Jobs First's earning disregard policy is quite simple: every dollar of earnings below the federal poverty line (FPL) is disregarded for purposes of benefit determination. This policy is very generous by comparison to AFDC's. The statutory AFDC policy was to disregard the first $120 of monthly earnings during a woman's first 12 months on aid, and $90 thereafter. In the first four months, benefits were reduced by two dollars for every three dollars earned, and starting with the fifth month on aid, benefits were reduced dollar-for-dollar, so that the long-run statutory implicit tax rate on earnings above the disregard was 100%. In addition to these basic flat and marginal disregards, AFDC also had child care and work-expense disregards.[3]

[3] In addition to the work expense and child care disregards, there are two reasons why the AFDC effective tax rate can be lower than 100%. First, AFDC eligibility redetermination occurs less frequently than monthly, so there can be a long lag between the month when an AFDC participant earns income and the date when benefits are reduced. Second, the EITC provides a 40% wage subsidy in its phase-in region, which generally ends above Connecticut's maximum benefit level; since the EITC is received in the form of a refundable subsidy payed by the IRS, it changes the effective benefit reduction rate but would not be observable in our data. We note that the EITC is available to both experimental groups in our data, so it raises the net wage above its before-tax level for both groups. In Bitler, Gelbach & Hoynes (2003*a*), we present local nonparametric regressions of transfer payments on earnings to give a sense of the empirical benefit reduction rate applied to AFDC recipients in the experiment. We found an effective

As shown in Table 1, the Jobs First time limit is 21 months, which is currently the shortest in the U.S. (Office of Family Assistance (2003, Table 12:10)). By contrast, there are no time limits in the AFDC program. In addition, work requirements and financial sanctions were strengthened in the Jobs First program relative to AFDC. For example, the Jobs First work requirements moved away from general education and training, focusing instead on "work first" training programs. Further, Jobs First exempts from work requirements only women with children under the age of 1, and financial sanctions are levied on parents who do not comply with work requirements. While Jobs First's sanctions are more stringent than AFDC's, the available evidence suggests that they were rarely used. For more information on these and other features of the Jobs First program see our earlier working paper (Bitler et al. (2003*a*)) and MDRC's final report on the Jobs First evaluation [Bloom et al. (2002), henceforth "the final report"].

Basic labor supply theory makes strong and heterogeneous predictions concerning welfare reforms like those in Jobs First. In the rest of this section, we discuss the economic impacts of Jobs First on the earnings, transfers, and income distributions. We focus on earnings disregards and time limits, since they are the salient features for examining heterogeneous treatment effects.

2.1 Economic Impacts of Earnings Disregards

To begin, Figure 1 shows a stylized budget constraint in income-leisure space before and after Jobs First. The AFDC program is represented by line segment AB while Jobs First is represented by AF. The Jobs First program dramatically affects the budget constraint faced by welfare recipients—lowering the benefit reduction rate to 0% and raising the breakeven earnings level to the FPL.[4] The effective AFDC benefit reduction rate in this figure is below the statutory rate long-run rate of 100% (see footnote 3 for a discussion).

What is the impact of this transformation of the on-welfare budget segment from AFDC's AB to Jobs First's AF? To begin, we make the usual static labor supply model assumptions: the

---

benefit reduction rate of about one-third, similar to earlier studies of the national caseload in McKinnish, Sanders & Smith (1999) and Fraker, Moffitt & Wolf (1985).

[4]Under AFDC rules, eligibility for AFDC conferred categorical eligibility for Food Stamps, with a 30% benefit reduction rate applied to non-Food Stamps income. Under Jobs First, Food Stamps rules mirror those for cash assistance: Food Stamps benefits are determined after disregarding all earnings up to the poverty line (though this Food Stamps disregard expansion operates only while a woman assigned to Jobs First is receiving cash welfare payments). However, losing eligibility for welfare benefits under Jobs First assignment (e.g., through time limits) need not eliminate Food Stamp eligibility, since one could still satisfy the Food Stamps need standard.

woman can freely choose hours of work at the given offered wage, and offered wages are constant. In particular, we ignore any human capital, search-theoretic, or related issues. We also assume that there is no time limit. Later we relax these assumptions.

Consider first the case in which an AFDC-assigned woman locates at point A, working zero hours and receiving the maximum benefit payment G. Depending on her preferences (e.g., the steepness of her indifference curves) assignment to Jobs First could lead to either of two outcomes. First, she might continue to work zero hours and receive the maximum benefit. Second, she might enter the labor market, moving from A to some point on AF; transfer income remains at the maximum benefit level, while total income rises. This labor supply prediction—together with others discussed below—is summarized in Table 2, which indicates whether Jobs First changes the after-tax wage (in this case yes) and non-labor income (in this case no). Table 2 then indicates the predicted location on the Jobs First budget set and the impact of Jobs First assignment on earnings, transfers, and income.[5]

We next consider points such as C, where women work positive hours and receive welfare when they are assigned to AFDC. For such women, assignment to Jobs First has only a price effect: the benefit reduction rate is lower, but there is no change in nonlabor income at zero hours of work. As long as substitution effects dominate income effects when only the net wage changes, Jobs First will cause an increase in hours, earnings, transfers, and income.

Now imagine that a woman's preferences are such that she would not participate in welfare if assigned to AFDC, instead locating at a point like D. At this point, her earnings would be between the maximum benefit amount and the FPL. Assignment to Jobs First would make this woman income-eligible for welfare even if she did not change her behavior; this is the case of Ashenfelter's (1983) "mechanical" induced eligibility effect. However, Jobs First assignment leads to an increase in non-labor income, which is predicted to reduce hours of work. We thus expect Jobs First to induce moves from D to points on AF right of D and left of A. Earnings will fall, transfer payments will rise, and the change in their sum is ambiguous.

At this point it is important to recall that we are interested in the labor supply choices of women under counterfactual assignment to Jobs First and AFDC. As discussed more fully below, all women

[5]Note that labor supply theory makes predictions about hours worked. Assuming no change in offered wages, this implies a prediction about earnings. Thus the table includes a single prediction for hours/earnings. This is important, since we observe earnings but not hours in our data set.

in the Jobs First experiment have applied for public assistance, and most are not working. As is well established in the literature on welfare dynamics, women leave welfare at differing rates (Bane & Ellwood (1994)). Since we observe each woman for four years after random assignment, the same woman may have different offered wages, fixed costs of work, or preferences at different times in our data. Thus her counterfactual budget set location under AFDC assignment will vary over time, and in the preceding paragraph's discussion, we have in mind simply that at some point after random assignment, a choice to locate at point D might be realized. It is very important to understand that the increase in welfare participation due to Jobs First assignment in such cases can be due either to reduced exit or increased re-entry. In the static context, there is no way (or need) to distinguish between entry and non-exit. However, since inclusion in the experiment requires that a woman has (at least) applied for welfare, we expect that the depicted increase in welfare participation is more likely to occur from reduced exit than from increased re-entry.[6]

Next consider a woman who would locate at a point like E if assigned to AFDC. At E, earnings are between the poverty line and the sum of the maximum benefit and the poverty line. Such points are clearly dominated under Jobs First assignment: the woman can increase income by reducing hours of work and claiming welfare. Because of the large income effect due to Jobs First eligibility (and the substitution effect arising from the notch at the FPL), the woman may in principle reduce earnings to any level above zero. In other words, she may again locate at any point on AF left of A; this is an example of Ashenfelter's behavioral induced eligibility effect.

Lastly, consider a woman who under AFDC assignment would locate at points like H, where earnings exceed the sum of the poverty line and the maximum benefit (above the notch). Depending on her preferences, Jobs First assignment will be associated with either of two possible outcomes. First, the woman might reduce hours of work so that her earnings fall to or below the poverty line; transfers increase and total income decreases. Reduced income in this case is compensated for by reduced disutility from labor; this is another example of Ashenfelter's (1983) behavioral induced

[6]In part to mitigate possible entry effects, the Jobs First program has "dual eligibility" rules. While the FPL is used to determine continuing eligibility for current *recipients*, successful *applicants* must have monthly earnings no greater than $90 plus the state welfare needs standard (which was $745 for a family of three in 1999) leading to a considerably more stringent earnings test for applicants than for recipients (whose earnings need only be below the poverty line of $1,138). This dual eligibility policy will tend to reduce the earnings level at which any actual entry effects occur, but it will not eliminate entry incentives, nor does it mitigate the deterred-exit effects we discuss above. Since static labor supply analysis is qualitatively unaffected by the dual eligibility rule for applicants, we do not address it separately here.

eligibility effect;[7] Blinder & Rosen (1985) discuss the positive and normative implications of notches in budget constraints in more detail. Second, Jobs First assignment might have no effect for such women: if disutility of labor were sufficiently low, reduced labor hours would not fully compensate for the income lost in moving from H to AF,[8] so the woman would stay at point H.

The set of points $\{A, C, D, E, H\}$ exhausts all qualitatively possible earnings-hours combinations under AFDC assignment. Thus, we use the final columns of Table 2 to summarize the impact of Jobs First on earnings, transfers, and income. For some part of the bottom of the distribution, the Jobs First earnings effect will be zero. At the very top of the earnings distribution, Jobs First will also have no effect on earnings, since top earners will choose to participate in neither AFDC nor Jobs First. In between these extremes, we expect the Jobs First earnings distribution to be higher at lower earnings levels, primarily due to increased labor force participation under Jobs First. Income effects for newly mechanically eligible women will tend to mitigate this prediction; which effect predominates is an empirical question. Lastly, there will be a range of earnings toward the top of the distribution where Jobs First earnings are lower than AFDC earnings due to behavioral induced eligibility effects.[9]

Before time limits take effect, static labor supply theory makes very simple predictions concerning the transfer-payments distribution: no one's transfers will fall, while some women will receive an increase in welfare payments (from zero to the maximum benefit). Lastly, the effects at the bottom of the transfer payments distribution will be zero under either program, since some women will not receive welfare under either program assignment.

[7] Again, it is important to note that behavioral induced eligibility effects may result either from re-entry or non-exit.

[8] One might reasonably ask how a woman whose earnings under AFDC assignment would have reached point H could ever wind up in the Jobs First demonstration in the first place. Recall that static analysis obscures the dynamic process governing earnings of welfare participants before and after they apply for assistance. Shortly before application, one expects to (and we do) observe that earnings fell among applicants; this is the well-known Ashenfelter dip. After a period of time, mean reversion will tend to cause earnings to rise toward a more normal level.

[9] We are not the first to point out that changes in the earnings disregard can lead to heterogeneous impacts on labor supply. The AFDC literature makes this point when discussing changes in the benefit reduction rate (see Moffitt's (1992) review for a discussion), and it is also discussed with varying emphasis in the recent welfare reform literature. For a useful summary of different policies for changing earnings disregards in welfare reforms see Blank, Card & Robins (2000).

2.2 Economic Impacts of Time Limits

Jobs First features a 21-month time limit. Once the time limit takes effect, some women will no longer be eligible for any welfare benefits. For women who would have left AFDC anyway, the time limit's effect on welfare payments is zero. However, among those for whom the time limit is binding, Jobs First assignment will simply remove the segment AB from the budget set. This change will increase labor supply due to the fall in nonlabor income and the rise in the net wage. No one's labor supply should fall as a result of the time limit's imposition, and the behavioral induced-eligibility effect will disappear for time-limited women. Thus we expect the time limit to reinforce predicted positive earnings effects while eliminating predicted negative earnings effects.

With forward-looking behavior, the time limit may also have effects on women who have not yet exhausted eligibility. For example, participants may conserve their eligibility by reducing welfare use and increasing labor supply even before the time limit binds, as discussed in Grogger & Michalopoulos (2003). This "banking effect" would reduce the expected increase in transfer income resulting from Jobs First assignment. It would also have two opposing effects on the earnings distribution, both caused by a reduction in a given woman's reservation wage for entering employment. First, the banking effect will cause more women to work under Jobs First, increasing earnings relative to AFDC. Second, if there are search frictions, then reduced search durations likely will be coupled with lower accepted wages among Jobs First women (though increased search intensity could also drive reduced durations). Reduced wages will reduce earnings at any given number of hours, and they will also reduce desired hours of labor supply, further reducing earnings.

With scarce jobs and search frictions, one might expect the time limit to cause women to accept lower offered wages (or cease human capital investment earlier than planned) even in the absence of any banking effect. This could happen if women are worried about having to wait in a "job queue" after the time limit hits.[10] Such a queuing effect would have the same effects on the earnings distribution as the banking effect. However, it would not necessarily change the transfers distribution, since employed Jobs First women might choose to stay on welfare until time limits hit. We return to the banking and queuing effects in section 6.

[10] We are grateful to an anonymous referee for suggesting this possibility.

2.3 Economic Impacts of Other Changes in Jobs First

Jobs First brought a number of other reforms, including increased job search assistance, work requirements, sanctions for non-compliance, more generous child support pass-through, more generous asset limits, child care and medical insurance expansions, and family caps. These changes are less important in the current context either because they were relatively minor policy changes, or because they were not enforced stringently. With the exception of the child support pass-through, these changes all lead to the prediction that labor supply should rise, though perhaps at lower reservation wages.[11] Thus their predicted effects on earnings are ambiguous for the same reasons as with the banking and queuing effects. All these changes should reduce welfare payments for some women while increasing them for no women. Due to space considerations, we will not discuss these reforms here; interested readers may consult Bitler et al. (2003*a*) as well as the final report for more detail.

3 Data

Under federal law, states were required to conduct formal evaluations when they implemented AFDC waivers. Connecticut fulfilled this requirement by hiring MDRC to conduct a random-assignment study of Jobs First. We use data made available by MDRC to outside researchers on completion of an application process. Random assignment in the Jobs First evaluation took place between January 1996 and February 1997. Data collection continued through the end of December 2000. The experimental sample includes cases that were ongoing (the recipient, or stock, sample) or opened (the applicant, or flow, sample) in the New Haven and Manchester welfare offices during the

[11] Jobs First disregarded the first $100 of child support per month while AFDC disregarded only the first $50; the resulting increase in nonlabor income should lead to reduced labor supply. Also, the entire child support payment was passed on to women on Jobs First, with Jobs First welfare payments reduced one-for-one by the amount of the child support payments above $100. Thus the sum of child support and transfers received is unaffected by this "legal incidence" issue; aside from possible impacts of the increased pass-through, the change in the composition of child support and welfare payments should not change labor supply. However, our QTE results for the transfers distribution could be affected since we do not have administrative data on child support payments: the compositional change could make it look like Jobs First causes a reduction in a woman's transfer payments, even when this reduction is offset by increased child support receipts. We are able to use the three-year survey's data on child support payments to assess the importance of this issue, and we find no empirical evidence that impacts on received child support payments are enough to importantly affect our QTE findings for the transfers distribution. Details are available on request.

random assignment period.[12] Assignment for recipients took place when they received an annual AFDC eligibility redetermination.

MDRC's evaluation and public-use samples include data on a total of 4,803 cases. Of these, 2,396 were assigned to Jobs First and 2,407 to AFDC. Quarterly earnings data and monthly data on welfare and Food Stamps income are available for most of the two years preceding program assignment and for at least 4 years after assignment.[13] Demographic data—including information on number of children, educational attainment, age, race and ethnicity, marital status, and work history of the sample member—were collected at an interview prior to random assignment.[14] During the evaluation period, Connecticut's non-experimental caseload was moved to Jobs First; with only a few exceptions, only the experimental control group continued under the AFDC rules.[15]

The first column of Table 3 provides means for a national sample of AFDC recipients in 1996,[16] while the next two columns provide means for the same characteristics among women in our experimental data. The Jobs First experimental sample generally mirrors the characteristics of the national sample, with the exceptions that the experimental sample has substantially greater fractions of never married and less-educated women compared to the national caseload.

The bottom part of the table reports statistics for the experimental treatment and control samples concerning pre-treatment earnings, employment, and welfare use, as well as whether women

[12] In a previous version of this paper, we incorrectly stated that all cases were part of the experiment. In fact, between January 1996 and July 1996, (a random) half of all ongoing or newly opened cases were included in the random assignment process. Beginning in July 1996 and continuing through the remainder of random assignment, all ongoing and open cases were included in the random assignment process. According to conversations with MDRC staff, this change was made to increase sample sizes.

[13] Data are available for earnings (transfers) for 8 (7) quarters preceding random assignment. After random assignment, there are 16 quarterly observations on Connecticut earnings for every sample member except 30 people who entered the sample in January or February of 1997. Earnings data come from Connecticut's Unemployment Insurance (UI) system, so earnings not covered by UI are missed; fortunately the vast majority of employment is covered by UI. Data on Food Stamps and welfare payments come from Connecticut's Eligibility Management System (EMS), which warehouses information about welfare use. To preserve confidentiality, MDRC rounded several key variables before releasing the public-use data (e.g., they rounded quarterly earnings data to the nearest $100 and Jobs First, AFDC, and Food Stamps payments to the nearest $50). For cases with true amounts between 0 and the lowest reported nonzero value ($50 or $100), true values are rounded up, so that there are no false zeroes in the data.

[14] MDRC also conducted a survey on a subset of the sample about three years after random assignment. These data have been used by others to analyze impacts on other measures of family and child well-being (e.g., Loeb, Fuller, Kagan & Carrol (2003)).

[15] At the time Jobs First started, Connecticut was evaluating a prior welfare reform program. The experimental sample from the prior reform was not assigned to the Jobs First experimental sample, and the control group from that earlier experiment continued to face the AFDC program rules.

[16] The estimates for the national caseload are constructed using March 1997 CPS data. The sample includes all women aged 16–54 who have an own child in the household and whose family was reported to receive positive AFDC income in the prior calendar year.

came from the experimental recipients sample. The fourth column of the table reports unadjusted differences across the program groups. Overall, demographic characteristics are substantively similar across experimental program groups. However, there are important exceptions: the Jobs First group had lower earnings, greater cash welfare use, larger families, and a greater share of the sample coming from the recipients sample than did the AFDC group. In personal correspondence, MDRC staff indicate that over the many experiments that they conduct, they usually see (and would expect to see) statistically significant differences across the treatment and control groups in some of baseline comparisons. Unfortunately, a χ^2 test statistic of 36 reveals clear evidence of joint significance of the 19 differences in Table 3. Nonetheless, our extensive discussions and correspondence with MDRC and Connecticut welfare officials have convinced us that this experiment was well designed and executed, and that no feature of the random assignment process could have been expected to cause the observed differences across program groups.

Thus it appears that the differences in baseline characteristics are due only to bad luck in the assignment process. While it is comforting to know that there is not a more problematic explanation, we must still address the empirical differences in key baseline characteristics. A common approach is to control for pre-treatment variables in a linear regression, but a more theoretically appropriate approach is to use inverse-propensity score weighting. To implement this weighting procedure, we estimate the probability that person i is in the treatment group using predicted values from a logit model in which the treatment dummy is related to the following variables: quarterly earnings in each of the 8 pre-assignment quarters, separate variables representing quarterly AFDC and quarterly Food Stamps payments in each of the 7 pre-assignment quarters, dummies indicating whether each of these 22 variables is nonzero, and dummies indicating each of whether the woman was employed at all, on welfare at all in the year preceding random assignment, or in the applicant sample. We also include dummies indicating each of the following baseline demographic characteristics: being white, black, Hispanic, never married, or separated; having a high school diploma/GED, more than a high school education, and more than two children; being younger than 25 or aged 25–34; and dummies indicating whether baseline information is missing (as is the case for fewer than 200 observations) for completed education, the number of children, or marital status. Denoting the estimated propensity score for person i as $\hat{p}_i$ and the treatment dummy as T_i, the estimated inverse-propensity score weight for person i is

$$\hat{\omega}_i \equiv \frac{T_i}{\hat{p}_i} + \frac{1 - T_i}{1 - \hat{p}_i}. \quad (1)$$

The final column of Table 3 reports estimated differences after adjusting using inverse propensity score weighting. As statistical theory predicts, the differences are reduced to almost exactly zero. We use inverse-propensity score weights in the standard fashion for all estimators employed below.[17,18] It is important to point out that the propensity score adjustment does not alter any of our qualitative conclusions: all of our conclusions hold whether we weight or not, and whether or not our propensity score model includes demographic controls. Unweighted results are available on request.

4 Empirical Evidence on the Time Limit

As stated above, Jobs First's 21-month time limit is currently the shortest in the U.S. About 29% of the treatment group reached the time limit in the first 21 months of the evaluation period, and more than half reached the time limit within four years after random assignment (see the final report for discussion). Under certain circumstances, Jobs First caseworkers were empowered to provide both indefinite exemptions[19] from the time limit and to provide 6-month extensions. According to the final report, in the spring of 1998, 26% of the statewide (not just the experimental) caseload was exempt from the time limit. This number rose to 49% by March 2001, though this increase appears to largely reflect progressive exits from the caseload by more able (and time-limited) recipients. Extensions were granted to a non-exempt woman if her family income was below the applicable maximum benefit payment and she had made a good-faith effort to find and

[17] The formal literature on propensity scores began with Rosenbaum & Rubin (1983). DiNardo, Fortin & Lemieux (1996) use propensity score weights to examine the impact of institutional factors like the minimum wage across the wage distribution. Recent papers analyzing mean treatment effects using propensity score methodology include Heckman, Ichimura & Todd (1998) and Hirano, Imbens & Ridder (2003). Firpo (2003) has provided formal proof that inverse propensity-score weighting corrects for bias in estimating quantiles of the counterfactual treated and control distributions, with the simple differences of sample adjusted quantiles then serving as consistent estimates of the QTE. The weights given in (1) uncover treatment effects for the entire population represented by the experimental population. Alternative weights could be used to estimate the effects of treatment on the treated, but our objective is to estimate the effects of Jobs First under the assumption of generalizability.

[18] The variance of estimated treatment effects (mean or quantile) will depend partly on the variance of the estimated propensity score (and its covariance with treatment and with the unexplained part of the outcome of interest). We address this issue with the bluntest instrument possible, by simply bootstrapping all of our estimates.

[19] Those circumstances include physical or mental incapacitation, responsibility to care for a disabled relative, having a child aged younger than 1, and being deemed unemployable due to limited work history and human capital.

retain employment.[20] If no good-faith determination was made, then an extension was still possible if "there were circumstances beyond the recipient's control that prevent[ed] her from working" (final report, page 63).

In light of these statistics, it is critical to show that the time limit policy has *de facto* relevance. We do so using Figure 2. The solid line in the figure plots the treatment effect due to Jobs First on the first-spell survival function.[21] This series is calculated as the Jobs First group's first-spell survival function minus the AFDC group's survival function, with the latter plotted as the smooth, dashed line in the figure. Figure 2 has five key features. First, the treatment effect of Jobs First on first-spell survival is actually positive throughout the pre-time limit period, reflecting the increased generosity of the program before time limits take effect. Second, there is a sharp drop of 10 percentage points in the survival treatment effect at month 22: exactly the month when time limits can first bind. Third, the treatment effect on welfare participation is negative after this point. Fourth, the time limit was not binding for everyone. At month 22 the control group survival rate was 40%, and the month-22 treatment effect was -0.024. This is of course just another way of saying that exemptions and extensions were provided, as we knew. Fifth, there are (smaller) sharp drops at the six-month intervals when extensions expire. Overall, Figure 2 provides compelling evidence that the time limit policy was binding for a substantial number of women. This is the important fact for our purposes.

5 Mean Treatment Effects

The first column of Table 4 reports estimated mean levels among the Jobs First group for several variables, over the entire 16-quarter post-treatment period. The second column provides means for the AFDC group over the same period, and the third column provides the resulting mean

[20] Determination of good faith appears to have been somewhat complicated, often involving "extensive investigation, including talking with former employers, but staff reported that it often remains unclear why recipients left a job, reduced hours, and so on" (final report, page 69).

[21] We label women as being in their first spell at the beginning of this period if they have cash welfare income in either the month of random assignment or the following month. For those in the stock sample, some spells will have ended coincidentally in the month of random assignment; for those in the flow sample, not all applications will be accepted. For these reasons, only 85% of AFDC-assigned women and 88% of Jobs First-assigned women are in a first spell at the beginning of the analysis period. Note that this first-spell definition does not match the usual one, since we include all ongoing spells, whether or not they are left-censored. The time origin typically is the beginning of a fresh spell, whereas our time origin is the time of experimental assignment, which may or may not coincide with the beginning of a welfare spell.

impacts. The first three rows concern average quarterly values of total income (defined as the sum of earnings and total transfers), earnings, and total transfers (defined as cash payments plus Food Stamps). These results show that over the four years following random assignment, the impact of Jobs First on average total income was $136 (about 5% compared to estimated AFDC baseline quarterly income of $2,609). About two-thirds of this impact is due to an insignificant increase in earnings, with the remainder due to a significant increase in transfers of $40, about 4%.

The bottom three rows provide means and impacts for binary variables indicating the fraction of quarters for which the person had positive levels of income, earnings, and transfers in the full 16-quarter period. For example, the value of 0.852 for "Any income" means that among women assigned to Jobs First, 85.2% of all person-quarters had a positive value for at least one of UI earnings, cash assistance, or Food Stamps.[22] The results show that the probability of having any earnings was 7.1 percentage points greater among the Jobs First group than the control group, an effect of 14% relative to the control group baseline. The probability of having any income or any transfers is essentially identical across treatment status over the full 16-quarter period.[23]

Both theory and the above evidence on the time limit suggest that in the first 21 months after random assignment—before time limits bind for anyone—behavior induced by Jobs First is very different from behavior during the final 27 months. Thus, we separately estimate mean treatment effects for the pre- and post-time limit periods. The second set of columns concerns the first 7 quarters of data, while the third set concerns the last 9 quarters. The results suggest that average earnings increased 7% in the pre-time limit period and 6% in the post-time limit period; in each case this effect is insignificant. Jobs First significantly increases the fraction of person-quarters with any earnings in each period. Mean impacts for transfers are starkly different in the early and later periods. During the first 7 quarters, Jobs First members received $212—or 16%—more in transfers than did control group women. During the later period, Jobs First members received $98—or 12%—less in transfers. The same pattern is clear for the fraction of person-quarters with positive transfers.

The net result of these changes in earnings and transfers is that Jobs First increased mean total

[22]This also means that about 15% of person-quarters had no value in any quarter for any of these variables. This could mean that 15% of persons never have any income, that everyone has positive income for all but 15% of quarters, or something in between. We return to this issue below.

[23]The share having any earnings can increase even while the share having any income does not because women caused to work by Jobs First assignment would have had welfare income if assigned to AFDC.

income significantly—in both economic and statistical terms—in the pre-time limit period. Nearly three-fourths of this increase is due to increased transfer income, rather than earnings. By contrast, in the post-time limit period, mean income was virtually identical across treatment status. This is the result of nearly equal increases in mean earnings and reductions in mean transfers.

6 Quantile Treatment Effects

Before presenting our QTE estimates, it will be helpful to briefly review quantiles and QTE. For any variable Y having *cdf* $F(y) \equiv Pr[Y \leq y]$, the q^{th} quantile of F is defined as the smallest value y_q such that $F(y_q) = q$. If we consider two distributions F_{JF} and F_{AFDC}, we may define the QTE as $\Delta_q = y_q(JF) - y_q(AFDC)$. This treatment effect may be seen to equal the horizontal distance between the graphs of F_{JF} and F_{AFDC} at probability value q; equivalently, it is the vertical distance between the graphs of the inverse *cdf*s.[24] The QTE estimates we report below are constructed in exactly this fashion: to estimate the QTE at the q^{th} quantile, we calculate the q^{th} quantile of the given Jobs First distribution and then subtract the q^{th} quantile of the given AFDC distribution. As a simple example, estimating the QTE at the 0.50 quantile simply involves taking the sample median for the treatment group and subtracting the sample median for the control group.[25] Appendix Table 1 reports the deciles of the Jobs First and AFDC group distributions for each outcome variable and time period we consider below; readers may thus calculate the QTE at these deciles. This table is also useful for assessing the magnitude of the estimated QTE relative to the baseline.

At this point, we want to emphasize that QTE do not generally identify quantiles of the treatment effect distribution. For example, if Jobs First causes rank reversals in the earnings distribution, then knowing the difference of quantiles in the two distributions is not enough to calculate the Jobs First treatment effect given knowledge of someone's earnings under AFDC assignment. It is easy, however, to see that if any QTE is negative (positive), then the treatment effect must

[24]Inverse *cdf* plots for all the variables we consider below are available on request from the authors.

[25]The only complication is to account for propensity score weighting. To do so, observe that the weight $\hat{\omega}_i$ exceeds unity for each i, so it must be normalized. Define the normalized weight $\hat{v}_i \equiv \hat{\omega}_i / \sum_i \hat{\omega}_i$, and let $T_i = JF$ if a woman is assigned to Jobs First and $T_i = AFDC$ if assigned to AFDC. Let $Y_i(t)$ be woman i's counterfactual outcome value if she has $T_i = t$. Then the weighted empirical *cdf* for program group t is $\hat{F}_t(y) \equiv \sum_i \hat{v}_i 1(T_i = t \text{ and } Y_i(t) \leq y)$, which is the total mass in group t such that the value of Y does not exceed y.

also be negative (positive) for some nondegenerate interval of the counterfactual AFDC earnings distribution. We also note that, like QTE estimates, classical social welfare function analysis would require only the empirical distributions of the two program groups. These and related issues are discussed in more detail in Bitler et al. (2003*a*).[26]

6.1 QTE for Earnings

We now turn to our main results: QTE for 98 centiles in graphical form.[27] Since we use the person-quarter as the unit of analysis, there are $7 \times 4,803 = 33,621$ observations for the first seven quarters. For the last nine quarters, there are $9 \times 4,773 = 42,957$ observations (as discussed in footnote 13, we lack quarter-16 data on 30 experimental participants).[28] To deal with within-person statistical dependence, our bootstrap procedure uses nonoverlapping person-level blocks (i.e., we re-sample persons in an *iid* fashion and then use the full profiles of each re-sampled woman's earnings to calculate the QTE estimates). We use 1,000 nonparametric bootstrap replications and then calculate standard errors using the empirical standard deviation of the bootstrap distribution for each QTE;[29] our confidence intervals are then computed using the normal distribution.

We plot the earnings QTE (as a solid line) for the first 7 quarters after assignment in Figure 3. For comparison purposes, the mean treatment effect is plotted as a horizontal (dashed) line, and the 0-line is provided for reference. Dotted lines represent two-sided 90% confidence intervals. This

[26]Heckman et al. (1997) provide a more general discussion of treatment effect heterogeneity and associated normative analysis issues. For a discussion of the potential outcomes framework undergirding our work, see for example that paper and Imbens & Angrist (1994).

[27]We computed the QTE at the 99^{th} quantile but do not include it in the figures below because its variance is frequently large enough to distort the scale of the figures. The extreme variance at high quantiles for unbounded distributions is well known; we do not have the same problem at the bottom of the distributions because they are all bounded below by zero.

[28]In Bitler et al. (2003*a*) we also present QTE generated using values of earnings, transfers and income averaged over the first 7 and last 9 quarters. The results are qualitatively similar to those presented here.

[29]Researchers rarely use formal methods to choose the number of bootstrap replications. Andrews & Buchinsky (2000) provide a three-step procedure that ensures bootstrapped standard errors will come within a specified percentage deviation (their *pdb*) from the ideal bootstrap estimated standard error, with a given pre-specified probability $1 - \tau$ (the ideal bootstrap estimate is what one would get using an infinite number of bootstrap replications). We found that 1,000 replications is more than sufficient to satisfy $(pdb, \tau) = (10, .05)$ for all but a few QTE in each of our six outcome-by-time-period groups. The few quantiles where more replications would be needed all occur in one of two cases. The first involves the very top of the distributions we consider (typically percentiles 98 and 99). The second involves quantiles where the QTE are constant (typically 0) for multiple quantiles either just above or just below the given quantile. The empirical bootstrap distribution shows that at these quantiles, the finite-sample distribution of the QTE estimates is very kurtotic. As Andrews & Buchinsky (2000) show, for bootstrapped standard errors the necessary number of bootstrap replications for given values of (pdb, τ) is governed entirely by the finite-sample excess kurtosis of the estimator in question.

figure shows that for quarterly earnings in the pre-time limit period, the QTE is identically zero for nearly half of all person-quarters. This result occurs because quarterly earnings are identically 0 for 48% of person-quarters in the Jobs First group over the first 7 quarters and 55% of corresponding AFDC group person-quarters. For quantiles 49–82, Jobs First group earnings are greater than control group earnings, yielding positive QTE estimates. Between quantiles 83–87, earnings are again equal (though non-zero). Finally, for quantiles 88–98, AFDC group earnings exceed Jobs First group earnings, yielding negative QTE estimates. For quantiles 89–92, these negative estimates are statistically significantly different from zero based on individually applied tests.[30] These results are exactly what basic labor supply theory, discussed above, predicts. That is, the QTE at the low end are zero, they rise, and then they eventually become negative. The negative effects at the top of the earnings distribution are particularly interesting given that they have typically not been found in other programs (*e.g.*, Eissa & Liebman's (1996) study of the EITC).

The range of the QTE point estimates is quite large: $[-\$300, \$500]$. The variation in the impact across the quantiles of the distributions is unmistakably significant, both statistically and substantively; these results suggest that the mean treatment effect is far from sufficient to characterize Jobs First's effects on earnings.[31]

Figure 4 plots the earnings QTE results in quarters 8–16, after the time limit takes effect for at least some women. For the first 80 quantiles, these results are broadly similar to those for the pre-time limit period (though they have a somewhat wider range). For quantiles 80–98, we again find negative treatment effects (with a few being zero), but none is individually significant.[32]

[30] We can use the empirical bootstrap distribution to construct the p-value for testing the null hypothesis that fewer than k of these 11 QTE estimates are simultaneously negative. Doing so yields p-values of 0.009, 0.020, 0.032, 0.043, 0.063, 0.088, and 0.119 for k =1–6, respectively. Thus the evidence is overwhelming that there is a small range of negative QTE at the top.

[31] Under the null of constant treatment effects, the true value of all QTE must equal the mean treatment effect. Thus, a test for the significance of any QTE is a test for the existence of heterogeneous treatment effects. Similarly, the largest QTE estimate provides a measure of how poorly the assumption of constant treatment effects performs. However, the maximum QTE has been chosen because of its large realized value, so one might wonder whether the difference between the maximum QTE and the mean treatment effect is large simply because of sampling variation. To get a sense of the relative sampling variations, we used the empirical bootstrap distribution for the pre-time limit transfers distribution to estimate the distribution of the difference in the maximum QTE and the mean treatment effect. In the 6,000 replications we considered (1,000 replications for each of three outcome variables and two time periods), the mean treatment effect was never greater than the maximum QTE. In fact, it was never even close: the maximum-minus-mean difference was never lower than 120% of the mean treatment effect (in the one case with a negative mean treatment effect, we used the largest negative QTE estimate in lieu of the maximum). Thus the range of the quantile treatment effects is statistically large relative to the mean in all cases.

[32] Tests based on the empirical bootstrap distribution yield a p-value of 0.099 for the null that none of these 19

Given that time limits are likely to bind for women with moderate to high earnings capacity, static labor supply theory predicts that we should find essentially no effect at the top of the earnings distribution—as actually occurs.

The expanded disregard can reduce earnings via entry or non-exit only while women retain welfare eligibility. There are two sets of women who can be eligible for Jobs First welfare even after month 21: those who left welfare before month 21, and those who receive exemptions or extensions. Women in the first group are unlikely candidates for behavioral induced eligibility effects after the seventh quarter given the fact that they have already left once, together with the more stringent earnings test for re-entry (see footnote 6 above). Getting an extension or exemption generally requires having earnings below the maximum benefit level, which is typically substantially below the poverty line (the difference depends on family size). It seems particularly unlikely that the Jobs First notch would cause entry or non-exit effects for these women. Thus static labor supply theory predicts significant behavioral induced eligibility effects in the first seven—but not the last nine—quarters of the Jobs First experiment. This is exactly the pattern we see.

Consequently, we suspect that the reduction in earnings at the top of the distribution caused by Jobs First is most likely due to behavioral induced eligibility effects of the disregard expansion. However, section 2's discussion of search frictions provides two alternative explanations: the banking and queuing effects. Each would cause a reduction in the reservation wage for entering employment, and thus lower earnings. In fact, data from the three-year follow-up survey do suggest that employed Jobs First women have lower wages throughout much of the top half of the wage distribution. However, unless wage growth is correspondingly greater among those who take lower starting wages, we would expect negative banking and queuing effects on earnings to persist throughout the study period, even after women leave welfare. Thus behavioral induced eligibility effects appear more consistent with the observed pattern of negative QTE at the top of the earnings distribution than do banking or queuing effects.

We can offer additional evidence to distinguish the behavioral induced eligibility effect and the banking effect. The banking effect implies not only that women should enter employment at lower wages, but also that they should exit welfare at lower wages. If lower reservation wages for exiting welfare were the only cause of reduced earnings at the top, then welfare participation rates at

QTE estimates is negative and a p-value of 0.149 for the null that no more than one is negative.

higher earnings levels should be lower among Jobs First than among AFDC women.[33] To examine this hypothesis, we first sort person-quarter observations on earnings into 10 bins corresponding to deciles of the AFDC group earnings distribution; we do this separately for the first 7 and last 9 quarters. We then define an indicator variable equal to 1 when a woman has cash welfare income each month of a quarter, and 0 otherwise. The banking effect suggests that in the pre-time limit period, the fraction with welfare income each month should be lower in the Jobs First group among women with relatively high quarterly earnings. We find the opposite to be true: for the AFDC group, 24% of women in decile 9 and 10% of women in decile 10 have welfare income each month of the quarter; for the Jobs First group, the corresponding numbers are 62% and 26%.

Further evidence on the queuing effect is more difficult to provide. Like the behavioral eligibility and banking effects, it implies that earnings should fall at the top of the distribution. Unlike the banking effect, it does not imply that welfare participation should also fall. The only queuing effect prediction that would allow us to distinguish between queuing and behavioral eligibility involves "bunching at the kinks" of the budget set. In particular, the large notch in the Jobs First budget set should lead to a mass point in the earnings distribution at the poverty line. This implies a spike in the density at that point, with a discontinuous drop occurring right above the poverty line. In fact, this prediction of precise bunching at the poverty line holds only when a substantial fraction of women can perfectly choose their hours. If many women can not, we would no longer expect to see mass points. Instead, we would expect increased density over a nondegenerate earnings range below the poverty line, and we would still expect a discontinuous drop in the density at the poverty line.[34]

Unfortunately, there are several reasons why our data are not ideal for this exercise. First, while Jobs First welfare payments are determined using monthly earnings levels, our earnings data are quarterly. Unless a woman has welfare income each month in a quarter, bunching could be consistent with having quarterly earnings above three times the monthly poverty threshold (since

[33] Of course, AFDC's rules imply that few women with high earnings should be eligible; those who are receiving benefits must either have large disregards or be evading the rules. But we can still test the null hypothesis that there is no behavioral eligibility effect due to Jobs First, in which case the participation rate at the top of the earnings distribution should be lower under Jobs First.

[34] As noted above, these predictions about features of the density around the poverty line assume that women understand the Jobs First disregard policy. As also noted, our discussions with Connecticut welfare officials indicate that the disregard policy was chosen in part because of its simplicity so that both recipients and caseworkers would understand the benefit formula.

the woman's earnings might exceed the poverty line in some months of a quarter but not others). Second, our only data on the number of children in a welfare case are compiled at the time of intake and do not vary over time, so in some cases we assign the wrong poverty line to each woman. Third, the variable number of children at intake is censored for women with more than two children, so these women cannot be used to study bunching. Given the importance of measurement when looking for discontinuities, the power of our discontinuity-based approach will be limited.

Nonetheless, to investigate these predictions, we used McCrary's (2004) two-step procedure for density estimation when a discontinuity is possible. We find no evidence of a discontinuous drop in the density at the poverty line. But we do see an increase in the density over a quarterly earnings range between the poverty line and about $2,000 below it. This hump in the density function is especially pronounced among women who receive welfare income every month of a quarter, where we would most expect it. Moreover, in the first seven quarters after assignment, the shapes (if not the scales) of the AFDC and Jobs First densities are remarkably similar except for this hump. The lack of precise bunching has been found elsewhere (*e.g.*, see Saez's (2002) study of tax rates).[35]

In sum, the evidence is consistent with behavioral responses to the Jobs First disregard policy. We cannot completely rule out the possibility that banking or queuing effects drive part of the observed negative earnings effect at the top of the earnings distribution. However, evidence of such effects would be interesting and important in its own right, since both effects would suggest that time limits lead to lower-quality job matches.

6.2 QTE for Transfers

Figure 5 presents results for transfer income in the first seven quarters, and Figure 6 presents results for the last nine quarters. The most notable feature of these results is the radical difference in the treatment effects of Jobs First across the pre- and post-time limit period. In the first seven quarters, the QTE are identically 0 for the bottom 20 quantiles, reflecting the fact that for 20% of person-quarters, both the treatment and control group have zero transfer income. For all quantiles except two above the 20th, transfer income in the pre-time limit period is greater among Jobs First women than among AFDC women. This finding greatly extends the result for mean treatment effects presented in section 5. Moreover, the range of QTE in this period is very large, with the

[35] One author who does find clear evidence of bunching is Friedberg (2000).

largest QTE reaching \$700. As a basis of comparison, this is nearly a third of the maximum quarterly value of Connecticut's combined AFDC-Food Stamps payment for a family of three. Thus in the pre-time limit period, Jobs First clearly is associated with a substantial upward shift in transfers for most of the distribution, as would be expected either from the simple mechanical effect of a more generous benefit schedule or from behavioral responses. Furthermore, the pattern of the QTE is consistent with theoretical predictions: no increase at the very top of the transfer distribution (which is both theoretically and empirically likely to be the bottom of the earnings distribution) or the very bottom (where no one participates) and increases in transfers everywhere in between.

The graph for quarters 8–16 is much different. For the lowest 48 quantiles, the Jobs First and AFDC transfer distributions are equal, with both showing zero transfer income at all these quantiles. However, at essentially all quantiles between 49–96, the Jobs First group receives less transfer income. The size of the reductions in transfer income can be quite large: the largest quarterly reduction is \$550, and the reduction is at least \$300 for all quantiles from 64–76. Results not reported here show that most of this result is due to the smaller fraction of Jobs First than AFDC women who receive any cash assistance in quarters 8–16 (the difference is 9%, compared to 4% for total transfers). When we estimate QTE results for cash assistance ignoring Food Stamps and including only those person-quarters having positive cash assistance, the QTE estimates are actually almost all positive. This result reflects the more generous Jobs First disregard, given eligibility. Thus the negative QTE results for transfer payments in Figure 6 are primarily driven by reductions in the rate of cash assistance, which shifts the entire transfer *cdf* leftward (meaning the inverse *cdf* shifts downward).

6.3 QTE for Total Income

We plot QTE results for total income in the pre-time limit period in Figure 7. These results again suggest a large degree of treatment effect heterogeneity: they range from 0 for the bottom 10 quantiles—where total income from administratively measurable sources is 0 in both groups—to \$800 at the top of the range. The mean treatment effect for this period is \$296, so again the range of quantile treatment effects is large compared to the mean treatment effect.

Because total income as we observe it is the sum of earnings and transfers, there need not be any particular relationship between QTE for total income and its components. To investigate the nature of changes in the total income distribution, we used local nonparametric regression techniques to estimate average earnings and average income at each quantile y_q of the total income distributions for Jobs First and AFDC women. We then plotted the treatment effect on these averages (the Jobs First average minus the AFDC average) against each $q \in \{1, 2, \ldots, 98\}$. The results (available on request) show that over the first 50 total income quantiles, the pre-time limit treatment effect on earnings is essentially zero. Thus these total income QTE are driven by increases in transfer payments (which range between zero and about \$200). Over the next 24 quantiles, the average difference in earnings is positive and generally increasing; at the 74$^{\text{th}}$ income quantile, the treatment effect on average earnings is about \$700. Over this same range, the treatment effect on average transfers falls to essentially zero. Between quantiles 75 and 95 of the income distributions, the treatment effect on local average earnings then falls precipitously, to a low of −\$600 at the 95$^{\text{th}}$ quantile. At the same time, the treatment effect on local average transfers rises quickly, to \$1,100. These effects are very consistent with our interpretation that the negative QTE estimates at the top of the earnings distribution are due to behavioral induced eligibility effects: positive QTE estimates at the top of the income distribution are driven by a combination of increased transfers and reduced earnings.[36]

Figure 8 plots QTE results for the post-time limit period. The QTE results clearly show that Jobs First has considerable effects on the distribution of total income, in stark contrast to the trivial mean treatment effect of \$14. QTE estimates for total income are zero for the first 18 quantiles and are actually negative for about 25 quantiles; the largest precisely estimated reduction in quarterly total income is \$300 (the reduction at quantile 99 is \$400, but it is very imprecisely estimated). As above, we computed local mean treatment effects on earnings and transfers at each quantile of the income distribution. We found that the reductions in total income below quantile 40 are driven by moderate to large reductions in transfers coupled with either small earnings reductions or small earnings increases. Increases in income through about quantile 80 are driven by relatively large increases in earnings (as much as \$827) that exceed frequently substantial reductions (as much as \$527) in transfers. Above quantile 80, the QTE results for income are driven by small-to-moderate reductions in earnings (generally between 0 and −\$200), coupled with moderately large increases

[36]Queuing effects would also be consistent with this result, though banking effects would not.

in transfers (generally above \$200).

Before the adoption of PRWORA, many welfare advocates expressed great concern that welfare reform would harm many (actual or potential) welfare recipients. Yet a common conclusion in the welfare reform literature is that few if any welfare recipients have been harmed. Given relatively short lifetime time limits, our results for quarters 8–16 are more appropriate than the pre-time limit results for addressing this issue. Treated independently, twelve of our total-income QTE estimates for quarters 8–16 are significantly negative (most highly so), while 29 are significantly positive. Considered jointly, an empirical bootstrap test yields a p-value of 0.085 for the null hypothesis that no more than 16 QTE estimates are simultaneously negative (the p-value for 17 is 0.101). The p-value for testing the null that no more than 32 QTE estimates are simultaneously positive is 0.091 (the p-value for 33 is 0.113).

We draw two conclusions from this analysis. First, once time limits take effect, there are definitely negative effects on some women. Second, there is evidence of positive effects for a larger range of women. As usual when there are both winners and losers, resolving these opposing results would require the use of some normative metric, which is beyond the scope of this paper.[37]

6.4 Robustness Checks Related to Exits from Administrative Data

One concern in interpreting the above QTE results involves women who have zero total income in some quarters. For these women to survive, they must have some way to finance consumption other than UI-covered Connecticut earnings, cash assistance through Jobs First or AFDC, and Food Stamps. Such women could have some other source of earnings (UI-noncovered or under-the-table), they could have support (cash or in-kind) from family members or absent non-custodial parents, or they could have moved out of Connecticut. A substantial amount of discussion in the final report, mostly using the three-year followup survey, suggests that neither marriage nor migration rates were systematically affected by welfare policies and that child support payments were only slightly

[37]As noted above, different assumptions on how ranks in the Jobs First income distribution correspond to ranks in the counterfactual AFDC distribution will lead to very different conclusions about the fractions of the population made better or worse off. However, if rank reversals occur in such a way as to minimize the number of losers, then the losses of these losers will be particularly large. Equity concerns are thus not necessarily mitigated in such cases. In Bitler et al. (2003*a*) we provide a normative analysis using a class of traditional social welfare functions, with the functions' parameters allowed to vary; such an approach is not sensitive to rank reversals. We believe this issue is ripe for further study, perhaps using data from other state experiments as well.

impacted. That is not enough for our purposes, however, because it is always possible (for example) that high-earnings women were systematically caused by Jobs First to stay in Connecticut, while low-earnings women systematically moved out; this could affect QTE estimates systematically. To deal with this issue, we consider the sample of women with zero total income in any quarter and find the last chronological quarter in which each had nonzero total income. We then exclude all subsequent quarters for such women from the analysis, which eliminates slightly more than a fifth of the sample of person-quarters. There is virtually no variation across treatment status in the overall probability of such attrition the administrative data. Furthermore, at *each* quarter in the followup period, there are no statistically significant differences in the probability of exiting the sample between the treatment and control group. Nonetheless, we recalculated the QTE excluding our synthetic "movers." With the (expected) exception of parts of the distribution having zero income, the results estimated on this sample of nonmovers are qualitatively identical to the figures presented above.

6.5 Subgroups

As noted in section 1, the mean impacts literature has drawn the conclusion that there is little heterogeneity in treatment effects. However, some authors, e.g., Grogger et al. (2002, p. 231), have suggested that the common approach of using ad hoc subgroups would be unlikely to consistently reveal treatment effect heterogeneity even where it exists. To examine this issue, we followed a common approach in the welfare reform literature, considering separately high school dropouts and women with at least a high school diploma. Nondropouts are often used as a comparison group: given nondropouts' lower welfare participation rates, reforms are often thought to affect them less than they do dropouts. To be part of the Jobs First experiment, all women in our sample had to apply for welfare, so this argument is less clearcut than is typical. Nonetheless, this is a logical way to consider the subgroups question.[38] We report detailed QTE results for dropouts and nondropouts in our earlier working paper; here we simply summarize the main findings. First, differences in mean effects across dropout status are trivial. Second, heterogeneity in the QTE within dropout status appears to be no less than the heterogeneity when we pool observations.

[38] Various parts of the final report (especially Appendix I) contain analyses of a wide array of subgroups. MDRC's focus is on groups labeled "most disadvantaged" and "least disadvantaged", which are defined using dropout status and employment and welfare use histories. We discuss these definitions in more detail in Bitler et al. (2003*a*).

Thus the most common mean impacts-based subgroup approach misses the entire heterogeneity story.

7 Conclusion

Our results establish several clear conclusions. First, mean treatment effects miss a lot: estimated quantile treatment effects for earnings, transfers, and income show a great deal of heterogeneity. Theory predicts that mean treatment effects will average together opposing effects, and our results clearly confirm this prediction. Second, results for earnings are clearly consistent with predictions from labor supply theory that effects at the bottom should be zero, those in the middle should be positive, and those at the top should be negative (before time limits). Third, the effects of Jobs First are very different in the pre- and post-time limit period, especially with respect to the transfers distribution. Negative effects at the top of the earnings distribution appear only in the pre-time limit period, as we would expect. This fact suggests a role for behavioral induced eligibility effects, most likely through reduced exit rather than increased entry. Banking and queuing effects are complementary explanations. Fourth, it is not unreasonable to believe that Jobs First led to substantial increases in income for a large group of women. On the other hand, once time limits take effect, it likely had at best no impact, and perhaps a negative one, on another sizable group of women. This finding is at odds with results in Schoeni & Blank (2003), who find positive effects throughout the distribution except in the very lowest percentiles. Moreover, we find that most of the shift in the income distribution occurs at above-median quantiles. Fifth, our results are robust to dropping observations from women who may have moved out of state or otherwise left the public assistance system while having no earnings (e.g., gotten married). Sixth, focusing on differences in mean treatment effects between dropouts and nondropouts—perhaps the most common comparison-group approach—is virtually useless in uncovering the treatment effect heterogeneity we demonstrate. In sum, our results show that QTE methodology can play a very useful role in assessing the effects of welfare reform when theory predicts heterogeneous treatment effects of opposing signs. We hope that this methodology will be used more often to address and analyze heterogeneous effects of welfare and other reforms.

References

Abadie, A., Angrist, J. D. & Imbens, G. (2002), 'Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings', *Econometrica* **70**(1), 91–117.

Andrews, D. W. & Buchinsky, M. (2000), 'A three-step method for choosing the number of bootstrap repetitions', *Econometrica* **68**(1), 23–51.

Ashenfelter, O. (1983), 'Determining participation in income-tested social programs', *Journal of the American Statistical Association* **78**(383), 517–525.

Bane, M. J. & Ellwood, D. T. (1994), *Welfare Realities: From rhetoric to reform*, Harvard University Press, Cambridge and London.

Bennett, N., Lu, H.-H. & Song, Y. (2002), Welfare reform and changes in the economic well-being of children, Working Paper 9399, NBER.

Bitler, M., Gelbach, J. B. & Hoynes, H. (2003*a*), What mean impacts miss: Distributional effects of welfare reform experiments, Working Paper 10121, NBER.

Bitler, M. P., Gelbach, J. B. & Hoynes, H. W. (2003*b*), 'Some evidence on race, welfare reform and household income', *American Economic Review* **93**(2), 293–8. (Papers and Proceedings).

Blank, R. M. (2002), Evaluating welfare reform in the United States, Working Paper 8983, NBER.

Blank, R. M., Card, D. E. & Robins, P. K. (2000), Financial incentives for increasing work and income among low-income families, *in* D. E. Card & R. M. Blank, eds, 'Finding Jobs: Work and Welfare Reform', Russell Sage Foundation, New York, pp. 373–419.

Blinder, A. S. & Rosen, H. S. (1985), 'Notches', *American Economic Review* **75**(4), 736–47.

Bloom, D., Kemple, J. J., Morris, P., Scrivener, S., Verma, N. & Hendra, R. (2000), *The Family Transition Program: Final Report on Florida's Initial Time-Limited Welfare Program*, Manpower Demonstration Research Corporation, New York, NY.

Bloom, D. & Michalopoulos, C. (2001), How did welfare and work policies affect employment and income: A synthesis of research, Working paper, Manpower Demonstration Research Corporation.

Bloom, D., Scrivener, S., Michalopoulos, C., Morris, P., Hendra, R., Adams-Ciardullo, D. & Walter, J. (2002), *Jobs First: Final Report on Connecticut's Welfare Reform Initiative*, Manpower Demonstration Research Corporation, New York, NY.

DiNardo, J., Fortin, N. M. & Lemieux, T. (1996), 'Labor market institutions and the distribution of wages, 1973–1992: A semiparametric approach', *Econometrica* **64**(5), 1001–44.

Eissa, N. & Liebman, J. B. (1996), 'Labor supply response to the Earned Income Tax Credit', *Quarterly Journal of Economics* **111**(2), 605–37.

Firpo, S. (2003), Efficient semiparametric estimation of quantile treatment effects. Typescript, UC Berkeley Department of Economics.

Fraker, T., Moffitt, R. & Wolf, D. (1985), 'Effective tax rates and guarantees in the AFDC program, 1967–1982', *Journal of Human Resources* **20**(2), 251–63.

Friedberg, L. (2000), 'The labor supply effects of the Social Security earnings test', *Review of Economics and Statistics* **92**(1), 48–63.

Friedlander, D. & Robins, P. K. (1997), 'The distributional impacts of social programs', *Evaluation Review* **21**(5), 531–553.

Grogger, J. (Forthcoming), 'The effect of time limits, the EITC, and other policy changes on welfare use, work, and income among female-headed families', *Review of Economics and Statistics* .

Grogger, J., Karoly, L. A. & Klerman, J. A. (2002), Consequences of welfare reform: A research synthesis, Working Paper DRU-2676-DHHS, RAND.

Grogger, J. & Michalopoulos, C. (2003), 'Welfare dynamics under time limits', *Journal of Political Economy* **111**(3), 530–54.

Heckman, J., Ichimura, H. & Todd, P. (1998), 'Matching as an econometric evaluation estimator', *Econometrica* **64**, 605–54.

Heckman, J. J., Smith, J. & Clements, N. (1997), 'Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts', *Review of Economic Studies* **64**, 487–535.

Hirano, K., Imbens, G. W. & Ridder, G. (2003), 'Efficient estimation of average treatment effects using the estimated propensity score', *Econometrica* **71**(4), 1161–1189.

Imbens, G. W. & Angrist, J. D. (1994), 'Identification and estimation of local average treatment effects', *Econometrica* **62**(2), 467 – 75.

Loeb, S., Fuller, B., Kagan, S. L. & Carrol, B. A. (2003), 'How welfare reform affects young children: Experimental findings from Connecticut—a research note', *Journal of Policy Analysis and Management* **22**(4), 537–50.

McCrary, J. (2004), Estimating a discontinuous density function. Typescript, University of Michigan.

McKinnish, T., Sanders, S. & Smith, J. (1999), 'Estimates of effective guarantees and tax rates in the AFDC program for the post-OBRA period', *Journal of Human Resources* **34**(2), 312–45.

Michalopoulos, C. & Schwartz, C. (2000), What works best for whom: Impacts of 20 welfare-to-work programs by subgroup, Working paper, Manpower Demonstration Research Corporation, U.S. Department of Health and Human Services, and U.S. Department of Education.

Moffitt, R. (1992), 'Incentive effects of the U.S. welfare system: A review', *Journal of Economic Literature* **XXX**, 1–61.

Moffitt, R. (1999), The effect of pre-PRWORA waivers on welfare caseloads and female earnings, income and labor force behavior, *in* S. Danziger, ed., 'Economic Conditions and Welfare Reform', W. E. Upjohn Institute for Employment Research, Kalamazoo, MI, pp. 91–118.

Moffitt, R. (2002), Welfare programs and labor supply, Working Paper 9168, NBER.

Office of Family Assistance (2003), Temporary Assistance for Needy Families (TANF) fifth annual report to Congress, Working paper. http://www.acf.hhs.gov/programs/ofa/annualreport5/index.htm.

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

Saez, E. (2002), Do taxpayers bunch at kink points? Typescript.

Schoeni, R. & Blank, R. (2003), 'Changes in the distribution of child well-being over the 1990s', *American Economic Review* **93**(2), 304–8. (Papers and Proceedings).

Schoeni, R. F. & Blank, R. M. (2000), What has welfare reform accomplished? Impacts on welfare participation, employment, income, poverty, and family structure, Working Paper 7627, NBER.

Figure 1: Stylized Connecticut budget constraint under AFDC and Jobs First

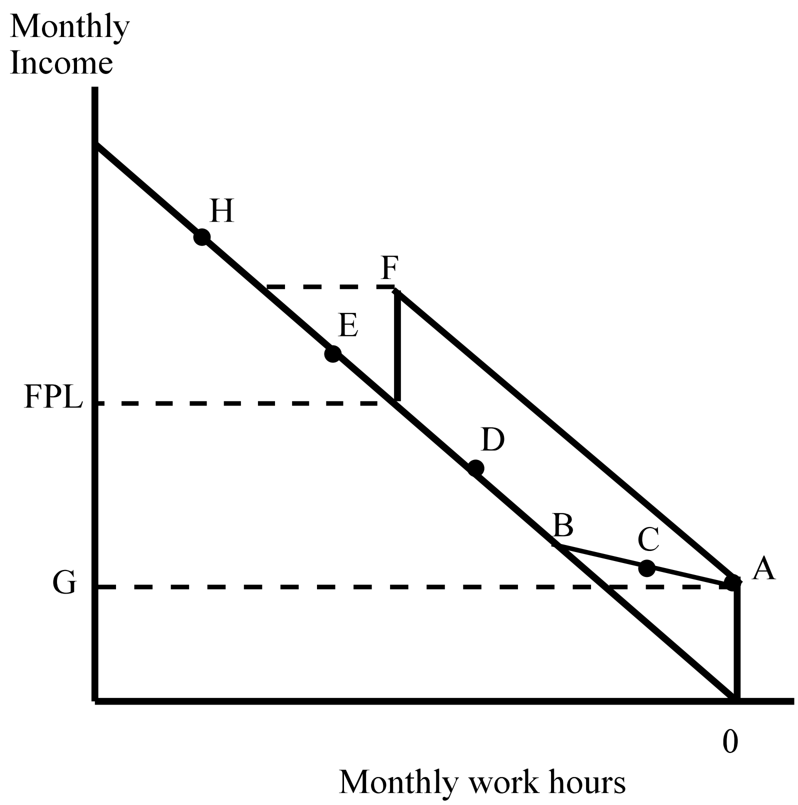
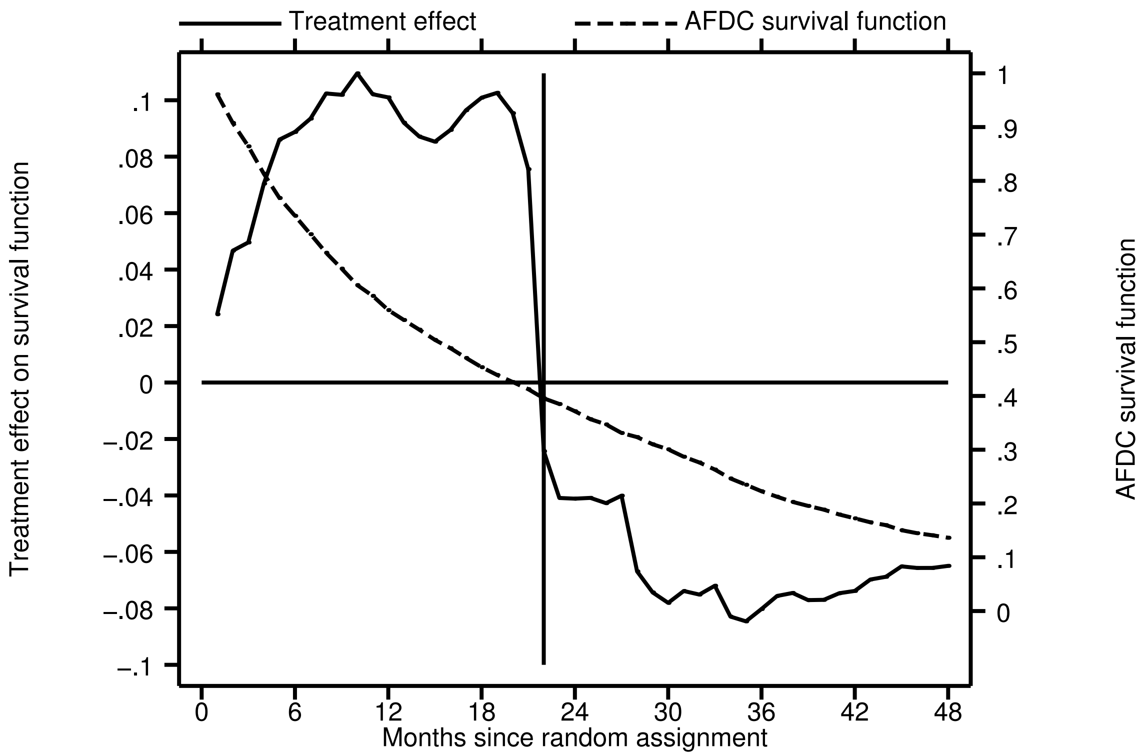
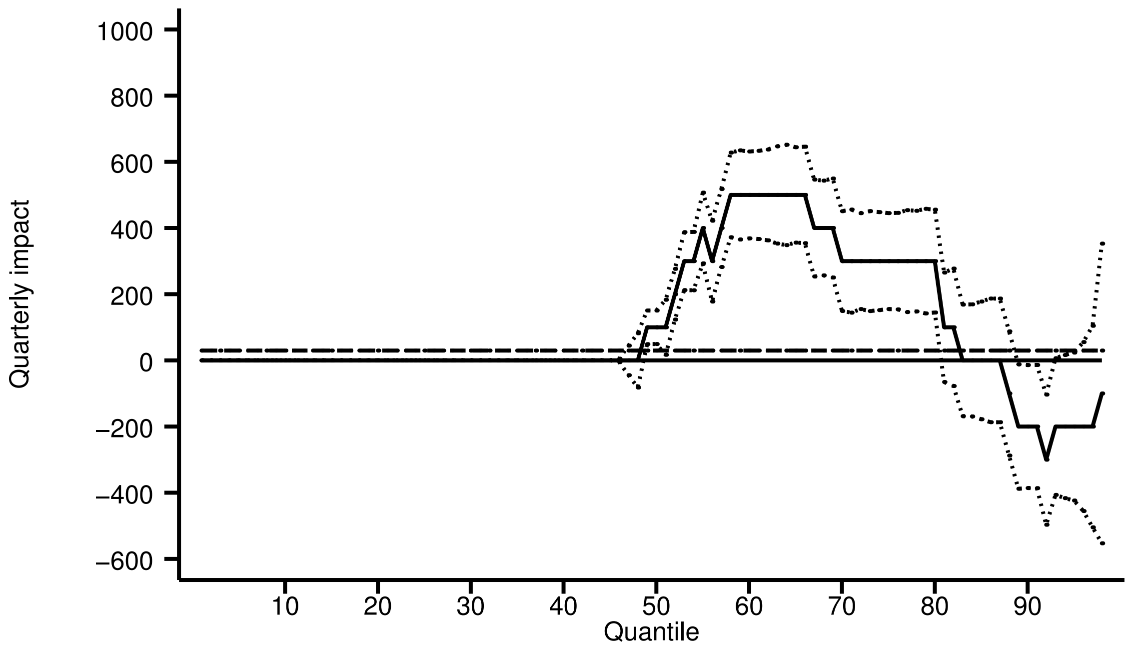


Figure 2: First-spell monthly survival function: AFDC group and treatment effect



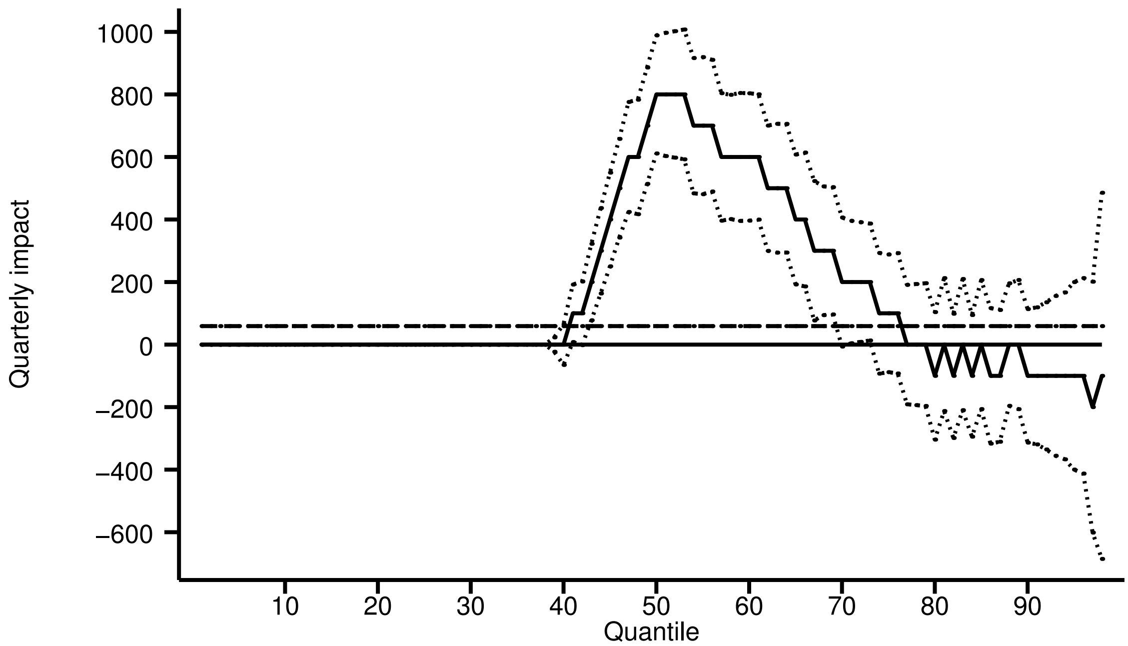
Notes: All statistics computed using inverse propensity-score weighting. See text for more details.

Figure 3: Quantile treatment effects on the distribution of earnings, quarters 1–7



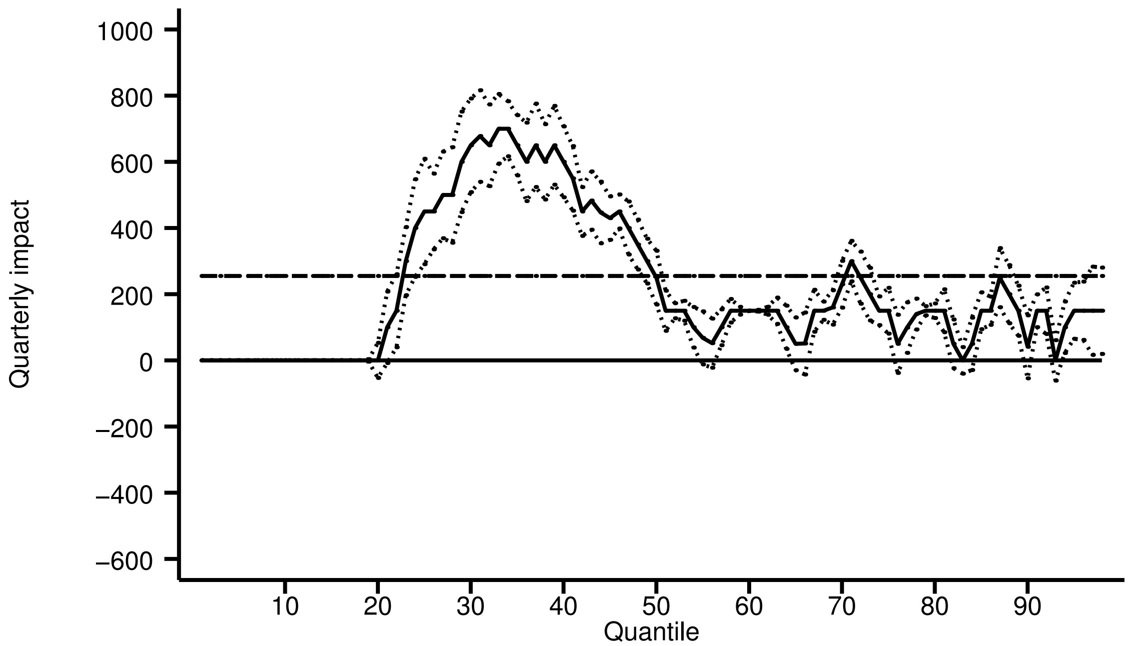
Notes: (*i*) Solid line is QTE. (*ii*) Dotted lines provide bootstrapped 90 percent confidence intervals. (*iii*) Dashed line is mean impact. (*iv*) All statistics computed using inverse propensity-score weighting. See text for more details.

Figure 4: Quantile treatment effects on the distribution of earnings, quarters 8–16



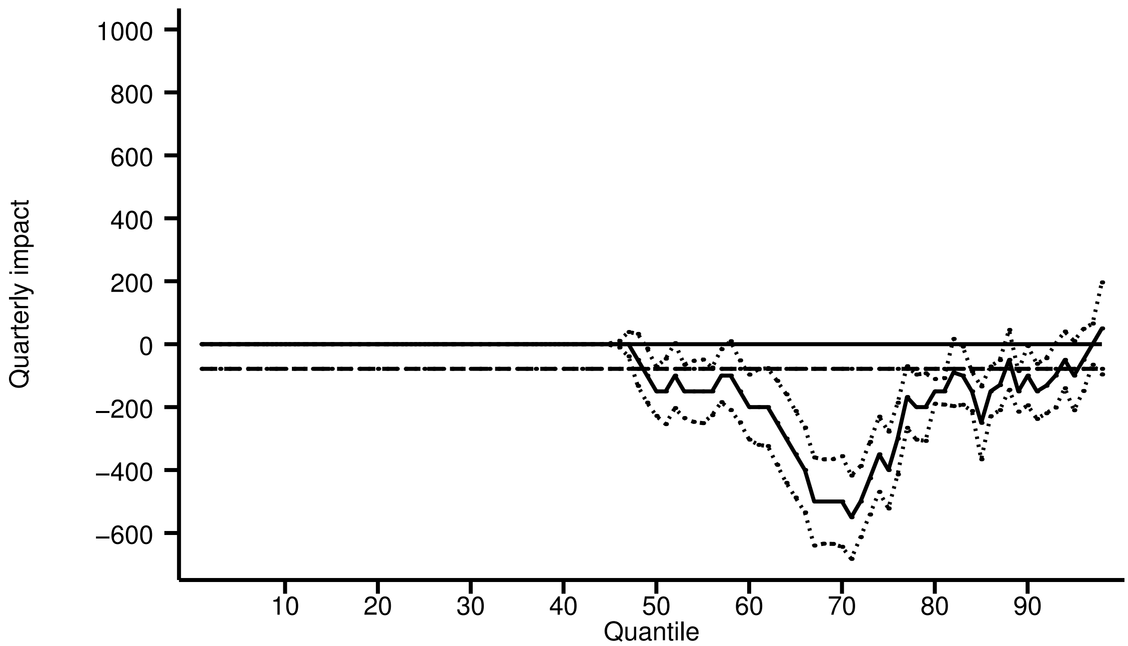
Notes: (*i*) Solid line is QTE. (*ii*) Dotted lines provide bootstrapped 90 percent confidence intervals. (*iii*) Dashed line is mean impact. (*iv*) All statistics computed using inverse propensity-score weighting. See text for more details.

Figure 5: Quantile treatment effects on the distribution of transfers, quarters 1–7



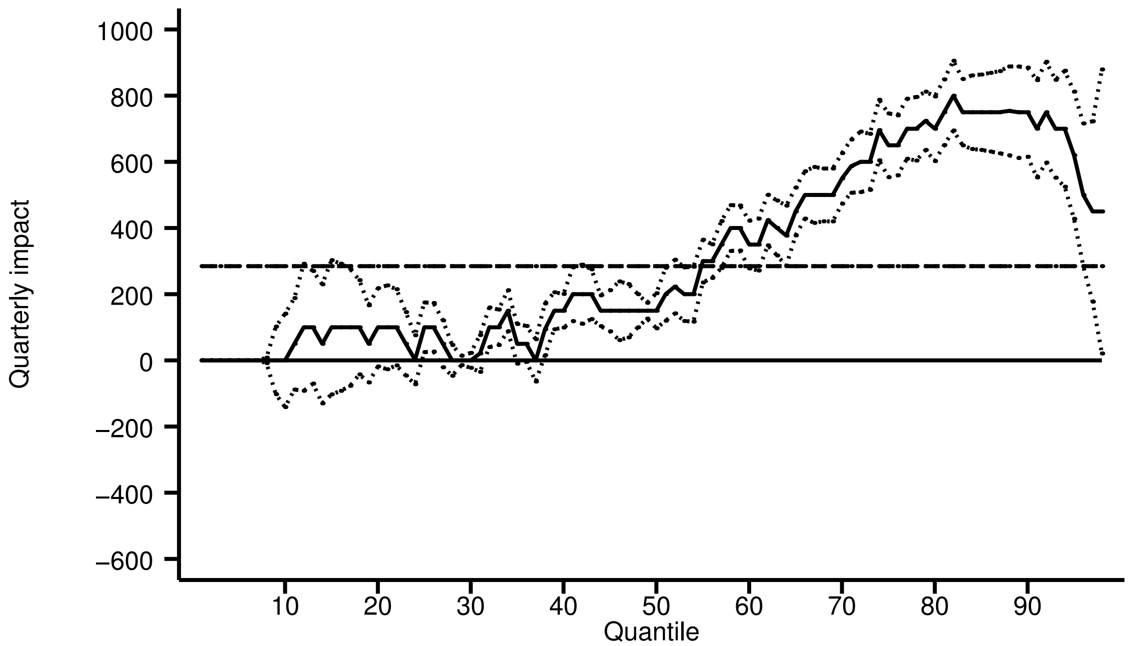
Notes: (*i*) Solid line is QTE. (*ii*) Dotted lines provide bootstrapped 90 percent confidence intervals. (*iii*) Dashed line is mean impact. (*iv*) All statistics computed using inverse propensity-score weighting. See text for more details.

Figure 6: Quantile treatment effects on the distribution of transfers, quarters 8–16



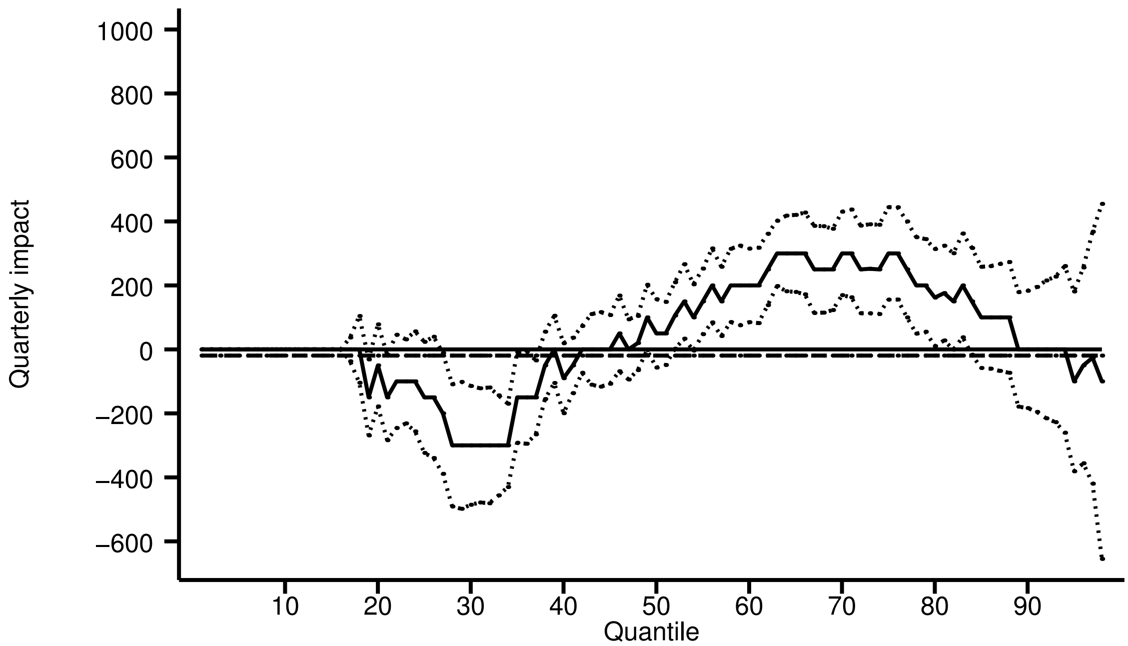
Notes: (*i*) Solid line is QTE. (*ii*) Dotted lines provide bootstrapped 90 percent confidence intervals. (*iii*) Dashed line is mean impact. (*iv*) All statistics computed using inverse propensity-score weighting. See text for more details.

Figure 7: Quantile treatment effects on the distribution of income, quarters 1–7



Notes: (*i*) Solid line is QTE. (*ii*) Dotted lines provide bootstrapped 90 percent confidence intervals. (*iii*) Dashed line is mean impact. (*iv*) All statistics computed using inverse propensity-score weighting. See text for more details.

Figure 8: Quantile treatment effects on the distribution of income, quarters 8–16



Notes: (*i*) Solid line is QTE. (*ii*) Dotted lines provide bootstrapped 90 percent confidence intervals. (*iii*) Dashed line is mean impact. (*iv*) All statistics computed using inverse propensity-score weighting. See text for more details.

Table 1: Key differences in Jobs First and AFDC programs

	Jobs First	AFDC
Earnings Disregard	All earned income disregarded up to poverty line (policy also applied to Food Stamps)	Months 1–3: \$120+1/3 Months 4–12: \$120 Months $>$ 12: \$90
Time limit	21 months (6-month extension if in compliance and non-transfer income less than maximum benefit)	*None*
Work Requirements	Mandatory work first, exempt if child <1	Education/training, exempt if child<2
Sanctions	1^{st} violation: 20% cut for 3 months 2^{nd} violation: 35% cut for 3 months 3^{rd} violation: grant cancelled for 3 months	*(Rarely enforced)* 1^{st}: adult removed from grant until compliant 2^{nd}: adult removed ≥ 3 months 3^{rd}: adult removed ≥ 6 months
Other policies	•Asset limit \$3000 •Partial family cap (50%) •2 years transitional Medicaid •Child care assistance •Child support pass-through	•Asset limit \$1000 •100-hour rule and work history requirement for 2-parent families •1-year transitional Medicaid

Sources: Bloom et al. (2002).

Table 2: Pre-time limit predicted effects of Jobs First assignment, by optimal choice given AFDC assignment

Location if assigned to AFDC	Compared to this point, does Jobs First assignment change: After-tax wage?	Non-labor income?	Location on Jobs First Budget set	Effect on distribution of: Hours/Earnings	Transfers	Income
A	Yes	No	A	0	0	0
	Yes	No	On AF, left of A	+	0	+
C	Yes	No	On AF, left of C	+	+	+
D	No	Yes	On AF, right of D	–	+	Either
E	No	Yes	On AF, left of A	–	+	Either
H	No	Yes	On AF, left of A	–	+	–
	No	No	H	0	0	0

Notes: Table contains predictions of static labor supply model for women facing AFDC and counterfactual Jobs First disregard rules (assuming all other rules are the same). Points are those labeled in Figure 1. There are two predictions for women at points A and H depending on those women's preferences.

Table 3: Characteristics of national caseload and experimental sample

	National Caseload (CPS)	Experimental sample			
		Levels		Differences	
		Jobs First	AFDC	Unadjusted	Adjusted
Demographic characteristics					
White	0.407	0.362	0.348	0.014	0.001
Black	0.355	0.368	0.371	-0.003	-0.000
Hispanic	0.202	0.207	0.216	-0.009	-0.001
Never married	0.473	0.654	0.661	-0.007	-0.000
Div/wid/sep/living apart	0.329	0.332	0.327	0.005	0.000
HS dropout	0.422	0.350	0.334	0.017	-0.000
HS diploma/GED	0.328	0.583	0.604	-0.021	0.001
More than HS diploma	0.250	0.063	0.058	0.004	0.000
More than two children	0.282	0.235	0.214	0.021*	-0.000
Mother younger than 25	0.242	0.289	0.297	-0.007	-0.000
Mother aged 25–34	0.435	0.410	0.418	-0.007	0.000
Mother older than 34	0.323	0.301	0.286	0.015	0.000
Recipient (stock) sample		0.624	0.593	0.031**	-0.001
Average quarterly pre-treatment values					
Earnings		679	786	-107***	-1
		(1,304)	(1,545)	(41)	(32)
Cash welfare		891	835	56**	-1
		(806)	(785)	(23)	(2)
Food Stamps		352	339	13	0
		(320)	(304)	(9)	(1)
Fraction of pre-treatment quarters with					
Any earnings		0.322	0.351	-0.029***	0.000
		(0.363)	(0.372)	(0.011)	(0.001)
Any cash welfare		0.573	0.544	0.029**	-0.001
		(0.452)	(0.450)	(0.013)	(0.001)
Any Food Stamps		0.607	0.598	0.009	0.000
		(0.438)	(0.433)	(0.013)	(0.001)

Notes: Standard errors in parentheses: for all but last column, these are estimated conventionally; for last column, standard errors are computed using 1,000 nonparametric bootstrap replications.
***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively (significance indicators provided only for difference estimates). National caseload statistics were constructed using all females aged 16–54 in the 1997 March CPS who had an own child in the household and whose family was reported to have positive AFDC income for calendar year 1996. All national caseload statistics are computed using March supplementary weights. Standard deviations omitted because all variables are binary. For earnings, 8 quarters of pre-treatment data are used. For cash welfare and Food Stamps, only 7 quarters are available for all observations.
Baseline data on a small number of observations for some variables are missing.

Table 4: Mean outcomes and impacts

	All quarters			Quarters 1–7			Quarters 8–16		
	Jobs First	AFDC	Adjusted Difference	Jobs First	AFDC	Adjusted Difference	Jobs First	AFDC	Adjusted Difference
Average quarterly level:									
Income	2,745	2,609	136**	2,744	2,450	294***	2,748	2,733	14
	(35)	(57)	(64)	(31)	(48)	(53)	(44)	(67)	(78)
Earnings	1,658	1,561	97	1,195	1,113	82	2,020	1,908	112
	(35)	(58)	(64)	(29)	(49)	(52)	(45)	(68)	(78)
Transfers	1,088	1,048	40**	1,550	1,337	212***	728	825	-98***
	(15)	(16)	(20)	(17)	(17)	(22)	(17)	(18)	(23)
Fraction of quarters with:									
Any Income	0.852	0.857	-0.005	0.865	0.862	0.003	0.809	0.820	-0.010
	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)	(0.010)	(0.007)	(0.006)	(0.009)
Any Earnings	0.561	0.490	0.071***	0.570	0.485	0.085***	0.593	0.527	0.066***
	(0.007)	(0.007)	(0.009)	(0.010)	(0.009)	(0.013)	(0.008)	(0.008)	(0.011)
Any Transfers	0.626	0.622	0.004	0.684	0.644	0.040***	0.496	0.519	-0.023**
	(0.007)	(0.007)	(0.009)	(0.009)	(0.010)	(0.013)	(0.008)	(0.009)	(0.011)
	2,381	2,392	4,773	2,396	2,407	4,803	2,381	2,392	4,773

Note: Standard errors in parentheses calculated using 1,000 nonparametric bootstrap replications.
***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively (significance indicators provided only for impact estimates). All statistics computed using inverse propensity-score weighting.

Appendix Table 1: Percentiles of the distribution of quarterly earnings, transfers, and total income for AFDC and Jobs First groups

	Percentiles of the Distribution								
	10	20	30	40	50	60	70	80	90
Earnings, Q1–7									
Jobs First Group	0	0	0	0	100	800	1,500	2,500	3,700
AFDC Group	0	0	0	0	0	300	1,200	2,200	3,900
Earnings, Q8–16									
Jobs First Group	0	0	0	0	1,000	2,000	3,000	4,000	5,600
AFDC Group	0	0	0	0	200	1,400	2,800	4,100	5,700
Transfers, Q1–7									
Jobs First Group	0	0	1,050	1,650	1,800	1,950	2,180	2,400	2,741
AFDC Group	0	0	400	1,050	1,550	1,800	1,950	2,250	2,700
Transfers, Q8–16									
Jobs First Group	0	0	0	0	0	450	1,000	1,800	2,300
AFDC Group	0	0	0	0	150	650	1,500	1,950	2,400
Total income, Q1–7									
Jobs First Group	150	1,450	1,800	2,100	2,400	2,850	3,400	4,100	5,150
AFDC Group	150	1,350	1,800	1,950	2,250	2,500	2,850	3,400	4,400
Total income, Q8–16									
Jobs First Group	0	150	1,000	1,810	2,400	3,050	3,800	4,662	5,900
AFDC Group	0	200	1,300	1,900	2,350	2,850	3,500	4,500	5,900

Note: All percentiles calculated using inverse propensity-score weighting. Difference between Jobs First Group percentile and AFDC Group percentile is the QTE for the given percentile, graphed in Figures 3–8.