Has Moving Welfare Recipients from Welfare to Work Influenced their Demographic Behavior? A Non-Experimental Analysis of Experimental Data from Four States

David J. Fein
Abt Associates Inc.

Rebecca A. London
University of California, Santa Cruz

Jane G. Mauldon
University of California, Berkeley

March 25, 2005

Paper presented at the 2005 annual meeting of the Population Association of America in Philadelphia, PA. We thank Christopher Baum, Dan Gubits, Ozan Jacquette, Steve Kennedy, and Mark Schaeffer for valuable suggestions and assistance. We also gratefully acknowledge financial support for the Welfare Reform and Family Formation Project from the Annie E. Casey and Smith Richardson Foundations.
Please direct correspondence to David Fein at david_fein@abtassoc.com; Rebecca London at rlondon@ucsc.edu; or Jane Mauldon at jmauldon@socrates.berkeley.edu .

Abstract

A fundamental goal of U.S. welfare reform in the 1990s was to reduce financial incentives to single parenting in the traditional welfare system. The reforms placed time limits on cash assistance and moved thousands of single parents from welfare to work, but they appear to have had at most small and mixed effects on living arrangements or fertility. Are demographic behaviors simply not very sensitive to financial incentives, or do the economic effects of welfare reform tend to be offsetting (e.g., decreased benefits, increased earnings)? We address this question using data on single welfare mothers from four state welfare reform waiver experiments, exploiting heterogeneity in economic impacts as instruments. Findings show that increased employment has little effect on marriage or cohabitation but does bring more doubling up with other adults. Benefit reductions have the expected effects of increasing marriage and decreasing births, particularly in the three states whose programs closely resembled TANF.

Introduction

A principle rationale for U.S. welfare reforms of the 1990s was the desire to stem the surge in single-parent families fueling rapid welfare caseload growth. Three of the landmark 1996 Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA)'s four goals emphasized strengthening marriage and curbing out-of-wedlock childbearing.[1] PRWORA assumed that by targeting financial assistance to single parents, the previous Aid to Families with Dependent Children (AFDC) program had discouraged marriage and encouraged non-marital childbearing. By curtailing benefits and promoting work—through time limits, work requirements, and work services—the new Temporary Assistance to Needy Families (TANF) program would reverse the tide.

To date, however, researchers have found that welfare reforms have had at most small and mixed effects on family formation (see reviews in Bitler et al., 2004a; Fein et al., 2002; Fitzgerald and Ribar, 2004; and Gennetian & Knox, 2003). As reducing single parenting was an important rationale for PRWORA's economic provisions and the relevant provisions remain in force, the reasons for the apparent lack of effects deserve scrutiny. To date, neither researchers nor policy makers have analyzed the matter carefully.

One possibility is that economic impacts did not reach the threshold needed for incentives to take hold (Murray, 2001) and for changes in individual attitudes to cascade, bringing effects on wider community norms (Ellwood and Jencks, 2001). Another possibility is that effects from decreased welfare benefits were offset by effects from increased employment and earnings. Finally, perhaps the veil of non-economic personal, family, and community attributes moderating demographic behaviors was simply too thick for financial incentives to penetrate.

In this paper, we analyze the relationship between welfare reforms' economic impacts and a range of demographic behaviors, including marriage, cohabitation, doubling up with other adults, and fertility. The analysis sample includes 7,310 initially unmarried mothers on welfare who participated in random assignment evaluations of welfare reforms in Delaware, Florida, Indiana, and Minnesota. Though begun just before the 1996 legislation, reforms in Delaware, Florida, and Indiana closely resembled TANF, and

[1] See U.S. House of Representatives (1996).

subsequently became those states' TANF programs. The Minnesota demonstration also anticipated that state's TANF program, but embodied more generous financial work incentives and did not include time limits.

Experimental impact estimates for the four states reveal statistically significant positive impacts on marriage (Delaware) and cohabitation (Florida) and negative impacts on doubling up (Minnesota) and births (Indiana). The findings are consistent with our general hypotheses that intensified financial pressures (Delaware, Florida, Indiana) would increase need for supportive living arrangements and diminish interest in childbearing, whereas reduced financial pressure (in Minnesota) would reduce need for support. The impacts' small magnitudes and inconsistencies across states with similar provisions echo the mixed findings from previous studies.

To assess whether and how welfare reform's economic impacts have affected demographic behaviors, we investigate the relationship between several economic outcomes, measured in the first two follow-up years, and a series of demographic outcomes measured in subsequent surveys. We compare two sets of estimates. The first is derived from ordinary probit regressions for each demographic outcome, with right-hand side variables including average levels of employment, earnings and welfare benefits for the first two follow-up years and a series of baseline control variables. Such estimates are vulnerable to endogeneity biases arising from omitted right-hand side variables and from reverse causation (i.e., if economic variables are affected by, as well as influence, demographic behaviors). Our second set of estimates addresses such biases by instrumenting for economic variables in equations for each family formation. Our instruments for employment, earnings and welfare benefits are interactions between random assignment status and a series of exogenous sample characteristics. Diagnostic tests indicate that the conditions for IV analysis are reasonably well met. Correlations between the instruments and economic regressors are statistically significant and large enough to avoid weak instruments bias, and the instruments are uncorrelated with residuals in the equation for each demographic outcome. As usual in IV analyses, the benefits of statistical consistency come at the price of a substantial loss in statistical precision. Accordingly, the IV estimates provide valuable guides to interpretation but are not satisfactory replacements for the un-instrumented estimates.

Non-experimental results indicate that welfare benefit reductions have a marked positive effect on marriage and a weaker negative effect on births. On closer inspection, we find that these effects are concentrated in the three states whose policies most closely resembled TANF. Another significant effect—of decreased likelihood of doubling up from increases in welfare benefits—occurs only in Minnesota. Employment and earnings do not appear to have consistent effects on marriage and cohabitation. In contrast, increased employment does appear to reduce the likelihood of doubling up and of having additional births. These findings support our general thesis—that subtle and opposing demographic responses underlie the small, mixed effects documented in prior research.

Theoretical Framework and Literature Review

We expect any demographic impacts of 1990s welfare reforms to arise mainly from impacts on women's economic status, as the reforms emphasized principally moving women from welfare to work. Principle policy tools involved provision of employment and supportive services and use of positive and negative financial incentives such as earnings disregards, sanctions, and time limits. To the extent that states addressed non-marital childbearing explicitly, they also relied on financial incentives—most notably, ending incremental benefit increases for additional children (so-called family cap policies) and eliminating special restrictions on welfare eligibility for two-parent families. TANF programs did not include services to promote marriage, and few took active steps to link recipients to family planning services.

Our conceptual framework thus accords a central role to changes in financial status as mediators of any impacts of welfare reform on family formation behaviors (see Figure 1). Specifically, impacts on employment, earnings and welfare benefits change women's economic situations, and thereby the potential financial motives for altering living arrangements and fertility (pathway 1). In addition to any effects of actual changes in economic status, awareness of the new welfare rules and requirements itself may affect the relevant motivations (represented in the dashed arrow to financial incentives). The model also identifies potential non-economic effects from increased employment (pathway 2) and from other services and any more general changes in the "culture of welfare" (pathway 3). Any influences through these three pathways depend on how they affect yet another set of intervening variables—the attitudes and opportunities that directly trigger changes in living arrangement and childbearing outcomes (labeled

"Proximate Influences" in the figure). Finally, the role proximate influences play in demographic behaviors is conditioned by a series of exogenous personal, family, and community characteristics ("Moderating Factors").

The potential complexity of any effects of TANF on family outcomes is thus evident in the possibilities for offsetting effects through different primary causal pathways, for primary influences to have varying effects on different proximate mechanisms, and for exogenous factors to moderate the latter. We now assess expectations and evidence on key elements in the framework in more detail.

Non-Economic Effects of Increased Employment

A "work first" approach—getting people into jobs as quickly as possible, regardless of how much the jobs pay—was the core philosophy in most states' welfare reforms in the 1990s. Key policies included providing services (e.g., work readiness, job search assistance, job development, and child care and transportation) and adding financial incentives and penalties to encourage work and discourage welfare reliance. There is strong evidence that such policies generated substantial increases in employment for single parents (Bloom & Michalopoulos, 2001; Grogger et al., 2002). Impacts on earnings were proportionate to impacts on employment, indicating that these programs generated increases in the types of low-wage jobs typically held by welfare recipients, rather than helping participants move to higher-wage jobs (ibid.).

Increased employment could have a variety of non-economic effects on family formation. It might improve marriage prospects by bringing welfare mothers into contact with a wider pool of employed men (Fein et al., 2002). If working increases self-esteem and reduces depression (Michalopoulos & Schwartz, 2001), it thereby also may foster improved relationships (Conger et al., 1999, 2002; Vinokur et al., 1998; Fox & Chancey, 1998). On the other hand, bad job experiences would tend to have negative psychological consequences, and reduced leisure and increased exhaustion and stress might preclude a more active social life. Because cohabitation typically involves less emotional and financial commitment than marriage (as does sharing living quarters with a relative or other adult), it may be a more viable response given increases in both social opportunities and time constraints (Clarkberg, 1999; Harknett & Gennetian, 2003). There is little good quantitative evidence on these hypotheses, particularly for low-income women (Fein et al., 2003). Qualitative studies find that poor single parents are reluctant to invest in relationships that interfere with responsibilities to their children and highly value their autonomy (Edin, 2000a, 2000b; Scott et al., 1999). In contrast to marriage and cohabitation, the increased employment has unambiguously negative implications for fertility, given the time and energy required to have and rear children.

Effects from Altered Financial Incentives

The standard economic model (Becker, 1991) predicts that increased earnings give women greater financial independence and reduce their economic incentives to marry, cohabit, and double up or to remain in such arrangements if they are dissatisfied. Countervailing incentives are likely, however, as women's earnings also may give them more bargaining power in relationships, improve their ability to meet perceived material requirements for marriage, lead men to see them as more viable partners, and help alleviate financial strains that can disrupt relationships (Edin, 2000a, 2000b; Conger et al., 1999, 2002). Empirical findings for marriage and cohabitation have been quite mixed (see reviews by Ellwood & Jencks, 2001; Fein et al., 2003), although there are indications of positive earnings effects on marriage and cohabitation for more recent cohorts and among low-income populations (e.g., Clarkberg, 1999; Sweeney, 2002).

There is less theoretical ambiguity in expectations that increased earnings will reduce doubling up with another relative or non-relative, given the high general valuation of autonomy and privacy. Similarly, it seems likely that earnings increases will raise the opportunity costs of childbearing to a degree exceeding any increases in perceptions of being better able to afford children due to greater earnings.

While welfare reforms increased employment and earnings, they also reduced average welfare benefits (Grogger et al., 2002). The benefit impacts varied across states due to differences in the size of earnings impacts, as well as in policies such as financial sanctions, time limits, family cap, earned income disregards, and state benefit levels.

The potential demographic consequences of benefit impacts are less ambiguous than employment and earnings impacts. Although many states ended special welfare eligibility restrictions for two-parent families, such families continued to be effectively penalized by income eligibility thresholds (Moffitt, 2002). Thus, because few married couples were eligible for welfare before or after reforms, the chief impact is likely to be on the relative value to single parents of continuing on welfare versus getting married. Benefit cuts reduce the value of maintaining welfare eligibility. The implications for cohabitation are less clear, as policies generally treat unmarried partners' incomes more liberally, and workers tend to enforce the rules more flexibly for unmarried couples (Moffitt et al., 1998). The foregoing leads to the expectation that reform-related benefit reductions generally encourage marriage, cohabitation, and doubling up, perhaps with weaker effects for cohabitation.

Many econometric analyses have measured the effects of state variation in benefit guarantee levels, but none have sought to identify the effects of reform-induced benefit reductions. A late-1990s meta-analysis (Moffitt, 1998) found that, although findings have been mixed, results from better-controlled analyses indicate higher state benefits are negatively related to marriage and positively associated with out-of-wedlock childbearing. There has been little econometric analysis of the effects of differing benefit levels on cohabitation or doubling up.[2]

The relevance of findings on state benefit guarantees to TANF impacts is open to question. Compared with infrequent adjustments to state benefit levels and longer-term changes accompanying inflation, the recent reforms imply more abrupt changes in individual families' incomes. Under TANF, welfare offices also emphasize connections between benefit amounts and personal behaviors (e.g., work, parenting). The one study that did isolate impacts of benefit changes was a test of financial work incentives that resulted, on net, in higher benefit amounts (Miller et al., 2000). Employing a two-treatment design, the Minnesota Family Investment Program (MFIP) demonstration found that the incentives, but not work requirements, were responsible for a small increase in the percent of long-term single-parent recipients in urban areas who were married. This impact was not significant for the overall sample, however.[3] London (2000) calculates that that benefit reductions from full-family sanctions imply decreased incentives for cohabitation and increased incentives for doubling up among single mother families, but this study does not provide empirical estimates of the effects.

---

[2] Manning and Smock (1995) find that welfare benefit receipt is negatively related to transitions from cohabitation to marriage for blacks but not for whites, but do not measure the effects of benefit amounts.

[3] The Canadian Self-Sufficiency Project, which also tested a combination of work services and incentives, found increased marriage in New Brunswick and decreased marriage in British Columbia (Harknett and Gennetian, 2003). The SSP design did not allow evaluators to isolate the effects of financial incentives, however.

In our model, we recognize the possibility that women receiving welfare may adjust their living arrangements or fertility in response to perceived financial incentives without necessarily experiencing changes in earnings or benefits (shown as dashed arrow to financial pressure in Figure 1). There is not a great deal of evidence on the degree to which such responses occur, but in general we think they are likely to be weak compared to the effects of actual financial changes. Pressures of daily survival may lead low-income women to ignore potential threats to income until they become imminent: research on attitudes towards time limits finds threats of benefit loss have little effect on behavior until recipients nearly exhaust their benefits (Bloom et al., 2002).

Non-Economic Influences

As noted earlier, welfare reforms of the 1990s put relatively little emphasis on influencing non-economic outcomes (represented by pathway 3 in Figure 1). To the extent that programs directly addressed other behaviors, provisions pertained more to parenting responsibilities and child well-being (e.g., child immunization and school attendance requirements, living arrangements of teen parents) than to adults' living arrangements or fertility. The two most common family formation provisions removed special eligibility restrictions on two-parent families (e.g., work history and 100-hour work requirements) and capped benefit increases for additional births.

Analysts calculate that ending special two-parent eligibility restrictions did little to increase financial marriage incentives (Moffitt, 2002). It therefore seems unlikely that removal of these restrictions had much by way of either economic or non-economic (e.g., message affirming two-parent families) effects. Most analyses have found family cap policies had no effect on fertility (Dyer & Fairlie, 2004; Fairlie & London, 1997; Joyce et al., 2004; Kearny, 2004; Turturo et al., 1997). A New Jersey experiment did find effects (Camasso et al., 1998), but this study was plagued by severe methodological problems (Rossi, 2000). The family cap research addresses the combined effects of any economic and non-economic impacts of these policies, however, and thus is not informative about possible impacts on family formation norms and values.

It also is unclear whether TANF has had any broader effects on the values presumed to underlie long-term economic dependency and single parenting (i.e., the "culture of welfare"). Implementation studies have documented changes in the business of welfare offices (Gais et al., 2001), but there has been surprisingly little rigorous research on the effects on family and work values and attitudes of recipients (Mauldon et al., 2004).

Other possible non-economic impacts might result from any increases in access to family planning and parenting services. To varying degrees TANF programs sought to increase provision of family planning services through TANF, Medicaid, and coordination with Title X family planning agencies (Burlingham et al., 1999; Hutson and Levin-Epstein, 2000). Additionally, some states have offered education, counseling, and case management intended to foster responsible parenting. Although not directly aimed at childbearing or marriage, parenting measures might increase awareness of the requirements of quality child rearing, thereby affecting attitudes toward when and with whom to have children.

Moderating Factors

If theory provides good reasons to expect impacts from welfare reforms on living arrangements and fertility, it does not provide a very sound basis for judging their strength. One substantial complication is the existence of a host of personal, family, and community factors that may moderate both economic impacts and demographic responses to economic impacts (shown as “Moderating Factors” in Figure 1).

Research on moderators of economic impacts has tended to focus on whether varying personal and family disadvantages make it difficult for welfare recipients to reap benefits from employment services and incentives. Michalopolous & Schwartz (2001) examine subgroup impacts for a wide range of indicators of disadvantage based on pooled data from 20 welfare reform experiments. They find smaller impacts on earnings associated with prior welfare receipt, lack of a high school diploma, having only one child, and high levels of depression at the start of the experiment. Reductions in benefits also generally are smaller in groups with smaller earnings gains.

In the presence of economic incentives, personal, family, and contextual characteristics also may affect propensities for change in demographic behavior. Younger, never married women may be more willing to turn to marriage than their older, more experienced counterparts, and their greater fecundity and lower average number of children may make them more attractive as marriage partners (Fein, 2001; Gennetian

& Knox, 2003; Bramlett and Mosher, 2002). These same qualities would seem to favor cohabitation and also may make it easier to find relatives and non-relatives willing to share living quarters. The fertility-moderating effects of these characteristics are harder to foresee. Older women who already have had more than one child may find economic rationales for curtailing births more compelling than younger ones who are less likely to have achieved their desired family sizes. On the other hand, younger, low-parity women also have more potential fertility to postpone and demographic adjustments thus may be more evident. About one-third of women age 35 and over are contraceptively sterile (Mosher et al., 2004). Beyond age 40, infecundity rapidly approaches 100 percent and precludes policy impacts on births.

Some of the most frequently-cited contextual moderators of demographic behavior include: social norms about marriage, cohabitation and non-marital childbearing; the number of economically viable single men in the community; and access to family planning services (Gennetian & Knox, 2003; Fein et al., 2003; Mellor, 1998). The first two of these factors are thought to underlie marked racial and ethnic differences in demographic behavior. Evidence implicates both norms and poor economic opportunities for men as causes of the low marriage and cohabitation rates and high out-of-union childbearing among African Americans (Edin, 2000a, 2000b; Wilson, 1987; Fein et al., 2003). In contrast, marriage is a central value in Latino culture (Gennetian & Knox, 2003). Cultural differences also may explain divergent demographic responses to the same intervention in different geographic areas (Harknett and Gennetian, 2003).

Expected Net Effects of Welfare Reforms on Living Arrangements and Fertility

Taken together, our review leads us to expect increased employment and earnings will have mixed effects on marriage and cohabitation, but generally negative effects on doubling up and births. Welfare benefit reductions should encourage marriage, cohabitation, and doubling up, and discourage further childbearing. It does not seem likely that reforms have had any substantial effects on demographic behavior through channels other than employment, earnings, and welfare benefits, but such effect pathways have not been well studied.

Reforms' overall impacts will be determined by the strength of policy influences on each major pathway, as well as the extent of any "interference" from personal, family, or community moderating factors. Evidence from both experiments and non-experimental analyses suggests that positive impacts on earnings and negative impacts on benefit amounts roughly offset one another, leading to minimal impact on total income (Grogger et al., 2002). Allowing for a certain amount of theoretical ambiguity in employment and earnings effects on marriage and cohabitation, it seems safe to extrapolate a similar canceling of effects on demographic outcomes—with possibilities for significant effects in varying directions depending on localized patterns for economic impact and variation in moderating conditions.

Findings on the demographic impacts of welfare reforms to date indeed do indicate mixed results. The most ambitious effort to provide such estimates is Gennetian & Knox's (2003) analysis of marriage and cohabitation impacts from 14 welfare reform experiments. Their analysis was of single welfare recipients who participated in demonstrations ranging from comprehensive, TANF-like reforms to narrower mandatory work programs characteristic of the preceding era. The few statistically significant impacts included findings of increased marriage in Vermont's incentives program, decreased marriage in Riverside, California's labor force attachment program, and increased cohabitation in a Portland work program that put extra emphasis on boosting wages.[4] Other research has reported marriage impacts for Delaware's ABC program (positive); Canada's SSP program (positive in one site, negative in another); and Iowa's (negative for applicants, insignificant for ongoing recipients) (Fein, 2001; Fraker et al., 2002; Harknett & Gennetian, 2003).

With respect to moderating factors, Gennetian & Knox (2003) estimate impacts for subgroups defined by mother's age, age of youngest child, number of children, whether she ever married, race, level of economic disadvantage, and welfare and employment histories. Their findings indicate a number of significant moderating effects, but little by way of consistent patterns across the experiments studied. Contradictory subgroup patterns have been reported in other studies as well (see review in Fein et al., 2002).

[4] Although Miller et al. (2000) reported positive impacts for marriage among long-term recipients in the Minnesota Family Investment Program (MFIP) demonstration, Gennetian & Knox (2003) show that MFIP did not have a statistically significant impact for the overall (short- and long-term) sample.

Non-experimental analyses also provide mixed results for welfare reform impacts on marriage, with some indicating positive (Schoeni & Blank, 2000; Ellwood, 2000) and others negative (Bitler et al., 2004a; Rosenbaum, 2000) marriage impacts from TANF and earlier waiver policies. Non-experimental analysis generally has paid less attention to cohabitation or to the effects of moderating factors. One recent study found impacts on the probability of children living with married parents were negative for blacks (TANF), positive for Hispanics (waivers), and negative (waivers) and positive (TANF) for whites (Bitler et al., 2004b).

Turning to doubling up with other adults, several experiments provide impact estimates. Tests of TANF-like programs in Connecticut and Florida found no statistically significant impacts on the proportion of single parents living with other adults (Bloom et al., 2000; 2002), while the Iowa welfare reform experiment found positive impacts for applicants and no effect for ongoing recipients (Fraker et al., 2002). The experiments did not report impacts on doubling up for subgroups. The only non-experimental analysis of doubling up is Bitler et al. (2004b), who find statistically significant increases in the probability of Hispanic children living with a parent and grandparent from waivers, but no effects for blacks or whites.

There has been relatively more study of welfare reforms' impacts on fertility. Most of the welfare reform experiments have published impacts on the percent of participants having births, with follow-up periods ranging from three to five years. The demonstrations found little evidence of significant effects either in the older JOBS programs or in more comprehensive, TANF-like reforms in Connecticut, Delaware, Florida, Indiana, or Iowa (Beecroft et al., 2003; Bloom et al., 2002; Fein, 2001; Bloom et al., 2000; Fraker et al., 2002; Hamilton et al., 2001). As mentioned earlier, tests of narrower programs emphasizing family cap were inconclusive due to control group contamination and faulty analysis procedures. There has not been very much subgroup analysis for fertility impacts. Non-experimental analyses also provide little indication that welfare waivers or TANF have influenced childbearing among adults in the general population (Grogger et al., 2002; Levine, 2002), although one analysis finds consistent reductions in births among teens (Kaestner et al., 2003). Non-experimental analyses of pre-TANF reforms also have not found evidence for impacts on fertility of family cap policies (Dyer & Fairlie, 2004; Joyce et al., 2004).

The Four Welfare Reform Demonstrations Analyzed in the Present Study

This paper uses data from four welfare reform experiments to provide detailed demographic impact estimates and assess the economic pathways linking welfare reform policies to demographic behavior. Although begun just prior to TANF, the demonstration programs embodied key features of TANF and provided the templates for the states' subsequent TANF programs. Delaware, Florida, and Indiana, in particular, made only minor modifications to the demonstration policies in fashioning their TANF programs.

Each demonstration randomly assigned welfare recipients to either a treatment group that was subject to the new program or a control group that remained under the traditional welfare program, AFDC. Under AFDC, control group members were not subject to time limits, and most families were exempted from employment and training participation requirements that were effectively voluntary. Grants were reduced nearly a dollar for each additional dollar of earnings and increased for each additional birth. There were no requirements or sanctions for personal or parenting responsibilities. Two-parent families had to meet a stringent work history test and could not work over 100 hours a month.

Delaware's A Better Chance (ABC) program, implemented in 1995, made the largest number of changes to AFDC rules and applied the largest financial penalties for non-compliance (Fein et al., 2000). Work-related provisions included mandatory participation in employment activities such as job search for nearly all families; permanent, full-family sanctions for failure to participate in work activities; and an enhanced disregard of earned income. Full-family time limits limited assistance to 48 cumulative months of receipt, and a "work trigger" time limit conditioned benefits on taking an unsubsidized or community service job after 24 months. The program required parents to obtain information from a family planning provider, attend parenting education classes, and ensure that their children's school attendance and immunization status were satisfactory. Financial sanctions progressed to case-closure on continued non-compliance. ABC also included a family cap provision (ending benefit increases for additional births) and removed the work history and 100-hour restrictions for two-parent families. The experiment operated in five local offices chosen to represent different areas of the state.

Florida's Family Transition Program (FTP), implemented in 1994, contained a number of the same provisions, but structured them somewhat differently (Bloom et al., 2000). Like ABC, FTP's employment services provided an enhanced earnings disregard, but FTP allowed more education and training and levied more moderate financial sanctions. FTP time limits also affected the entire grant, but limited job-ready recipients to 24 months of assistance within any 60-month period and limited other recipients to 36 months within any 72-month period.[5] FTP also included school attendance and immunization requirements for children and eliminated the 100-hour rule and work history requirement for two-parent families. The FTP experiment operated only Escambia County, whose main city is Pensacola.

Indiana's welfare reform program, implemented in 1995, created two tracks—a Placement Track requiring job search for more job-ready and a Basic Track requiring education and training for less job-ready clients—and also levied more modest sanctions than ABC (Beecroft et al., 2003). Time limits restricted assistance to 24 months within any 60-month period but entailed loss only of the adult portion of the cash grant.[6] Indiana did not increase earnings disregards. Parenting provisions included school attendance and immunization requirements. Indiana added a family cap but did not change the two-parent eligibility rules. Unlike the other three demonstrations, Indiana operated its experiment on a statewide basis.

The Minnesota Family Investment Program (MFIP), launched in 1994, required long-term recipients to participate in work-oriented activities immediately, but allowed recent applicants to participate on a voluntary basis until they accumulated 24 months of assistance (Miller et al., 2000). The program provided both groups an especially generous earnings disregard, and it consolidated welfare, food stamps and a state-funded assistance program into a single cash payment. Sanctions were small compared with the other three states. MFIP did not feature time limits, a family cap, or parenting responsibility requirements but did eliminate the two-parent family work history and 100-hour requirements. The demonstration operated in three urban and four rural counties.

---

[5] Formal criteria defined "job-ready" on the basis of age, welfare history, education, and work experience.
[6] Time limits originally applied only to the Placement Track but were extended to all clients mid-way through the follow-up period.

Data and Measures

Sample surveys conducted three-to-five years after random assignment are the source for principle demographic measures—marriage, cohabitation, doubling up, and births—used in this analysis. The surveys, which also assessed a wide range of other dimensions of family well-being, were conducted by professional survey organizations by telephone with field follow-up and achieved response rates ranging from 70-80 percent. In each state, the sample frame was restricted to participants randomly assigned early in the demonstration in order to provide sufficient follow-up. In Indiana and Minnesota, disproportionate stratified sampling necessitates use of weights. The present analysis includes only women who were not living with a spouse at random assignment. The total sample size is 7,310, with state samples ranging from 1,410 (Delaware) to 2,167 (Minnesota). Average follow-up durations vary from 41 (Minnesota) to 60 (Indiana) months (see Table 1).

Measures for living arrangement outcomes are defined on the basis of survey roster questions identifying the name, relationship to respondent, and characteristics for all persons living in the household as of the survey date. Marriage and cohabitation are coded "1" if respondents identified a spouse or unmarried partner, respectively, among current residents and "0" otherwise. Doubling up is coded "1" if respondents listed any other adult as in the household, irrespective of whether a spouse or partner also was present, and "0" otherwise. Birth is coded "1" for an affirmative answer to a direct question about whether the respondent had any babies since the month of random assignment and "0" otherwise.

Average outcomes for control group members provide a sense of demographic behaviors in the absence of welfare reform (Table 1). Marriage rates range from 11 percent (Delaware) to 21 percent (Indiana) and are higher than cohabitation rates except in Minnesota. Roughly comparable fractions report other adults in the household. The proportion having births ranges from 27 percent (Minnesota) to 34 percent (Indiana).

Three economic outcomes figure as dependent and independent variables at different stages of the analysis. Our data on employment, earnings, and welfare benefits are from automated information systems and thus provide an accurate time series for each individual.[7] In most of the paper, analysis variables represent averages over the first two follow-up years for: quarterly employment rates, quarterly earnings amounts (including quarters with zero earnings), and quarterly welfare benefit amounts (also including zero-benefit quarters). We restrict these measures to the first two follow-up years to inject a degree of temporal separation between the time recipients experienced economic impacts and the time we measure their demographic outcomes (41 to 60 months after random assignment). This separation helps to capture lags between economic experiences and demographic responses and to guard against any biases from unmeasured mediators in the IV analyses.[8]

Average employment rates in the control group ranged from 43 to 49 percent, and average earnings ranged from $786 to $980. Average benefits are considerably higher in Minnesota ($1,369) than in the other three states ($439-$483), primarily because the Minnesota benefit records do not distinguish cash welfare from food stamps and state cash assistance benefits.[9]

Personal and family characteristics of each sample member are measured at the time of random assignment. We use these characteristics mainly in estimating demographic and economic impacts for subgroups. They reflect a number of the dimensions discussed in our literature review (under moderators), including: stage in the life cycle (age), constraints from children (number and age of youngest child), relationship history (ever married), cultural norms (race), and economic disadvantage (education, new applicant). Table 1 (bottom panel) shows sample distributions by state according to categories that figure in our subgroup analyses. The distributions show some notable differences. In particular, black women account for a majority of sample members in the two Southeastern states (Delaware and Florida), whereas white women are in the majority in the two Midwestern states.

---

[7] Employment and earnings measures are based on employer-provided wage records stored in state Unemployment Insurance record systems. Welfare benefits are based on monthly records from automated systems maintained by each state welfare agency.

[8] The IV analyses omit any possible mediators of welfare policy impacts on demographic behavior other than impacts on women's employment, earnings and welfare benefits. As we discuss elsewhere, the nature of welfare reforms and of the experimental designs are such that unmeasured mediators seem unlikely to have much influence on the IV estimates. To the extent that they do, restricting measures for economic regressors to earlier years minimizes the chance that the instruments will capture any reverse effects of demographic behavior on economic situations.

[9] We did not have the data on these other programs to replicate this definition in the other states. Policy impacts affected mostly the cash welfare component of the Minnesota payments, and this analysis is concerned with the effects of welfare benefit impacts on demographic behavior. Thus, although the overall average benefit amount is higher in Minnesota than in the other states, benefit impacts are defined in roughly consistent terms. A small degree of inconsistency arises from the fact that food stamps formulas adjust benefits upwards (or downwards) by a small amount when AFDC benefits decrease (or increase).

Compared with their Midwestern counterparts, women in the Delaware and Florida samples also have more children, are less likely to have a high school degree or equivalent, and (due to longer demonstration intake periods) are more likely to be new applicants.

Analysis Methods

Random assignment generates two groups with highly similar characteristics whose outcomes differ subsequently only as a result of differential welfare reform exposure. We use the following model to measure experimental impacts in each state:

(1a) $$f_m^* = b_0T + \beta X_1 + \varepsilon,$$

where f_m^* is each woman's latent propensity to experiences family formation outcome m, measured using 0-1 indicators for marriage, cohabitation, doubling up and births as defined above. The treatment group indicator, T, is coded "1" for individuals randomly assigned to welfare reform in each state (treatment group) and "0" for those assigned to remain under AFDC rules (control group), and b_0 is the effect of the reform on the dependent variable. The vector X_1 includes a series of control variables, with coefficients β, consisting of dummy variables representing categories of the seven baseline characteristics shown in Table 1. The control variables help to guard against small baseline differences arising by chance at random assignment and slightly increase the precision of the impact estimates. We use the same model to estimate impacts on economic variables (i.e., replacing f_m^* in Equation 1a with the employment, earnings and welfare benefit outcomes defined earlier).

We estimate subgroup impacts within each state by adding interactions between treatment status and dummy variables corresponding to a particular characteristic, X', as shown in Equation 1b:

(1b) $$f_m^* = \beta X_1 + \sum b_nTX'_n + \varepsilon,$$

where n indexes the different subgroup dummies for X' and b_n is the impact estimate for each subgroup. A global F test on the equivalence of the set of b_n's for each characteristic indicates whether subgroup differences are statistically significant.

We use a linear probability model to estimate these equations, as ordinary least squares (OLS) regression software facilitated weighting and computation of a large number of estimates in percentage-point terms. Results from logistic regression analyses conducted as checks were highly consistent with the OLS estimates.

Impact estimates in each state capture the total effect of welfare reforms only to the degree that the control group environment remains constant under the traditional AFDC rules. To create the clearest possible distinctions, each state assigned specially-trained caseworkers to handle each group. Furthermore, important rule changes were hard-wired to random assignment status, preventing accidental application of sanctions, time limits, family cap, and other provisions to control group members. It was not possible, however, to shield control group members from wider public discourse about welfare reform or from general changes in welfare office environments. Surveys have documented varying levels of incorrect reporting by both treatment and control group members about the rules that applied to them (Camasso et al., 2003). To the extent that such errors reflect actual misperceptions (and not simply measurement error), they imply that experimental estimates will underestimate welfare reforms' full impacts on economic and demographic behaviors.

We note also that the Delaware experiment ended in the second year, when ABC became the state's TANF program and former control group members became subject to welfare reform rules and services. Employment and earnings impacts quickly faded as the control group became subject to the ABC rules. Negative welfare benefit impacts continued to grow, and, by the time of the follow-up survey, were larger than those in any of the other three states.[10] We thus might expect any demographic impacts in ABC to derive mainly from benefit effects and be somewhat less representative of full program impacts than in the other states. For reasons explained below, this shortcoming of the ABC experiment, and any tendency of the experiments more generally to undercount wider cultural impacts, are not necessarily liabilities in the non-experimental portion of our analysis and indeed may improve the conditions for the technique we use.

---

[10] This continued growth reflected the progressive and permanent nature of ABC's financial penalties, higher sanctioning rates during than after the experimental period, and the persistence of a substantial difference in assistance time clocks. See Fein et al. (2000) for further discussion.

The non-experimental analyses address the question of whether there are causal pathways running from welfare reforms' impacts on economic outcomes to each of the four demographic behaviors of interest. We use a non-experimental approach that exploits variation in experimental treatment effects on employment, earnings, and welfare benefits. Our model for each family formation outcome is:

(2) $$f_m^* = E\,\gamma + X_1\,\delta + \mu,$$

where f_m^* is each woman's latent propensity to experience family formation outcome m and γ is a vector of coefficients capturing the effects of economic variables E (employment, earnings, and welfare benefits) on family formation. The vector X_1 is a set of control variables—specifically dummy variables representing the baseline characteristics listed in Table 1—with coefficients δ, and μ is an error term. Consistency of the estimated coefficients γ requires (a) that omitted right-hand-side variables not be correlated with demographic outcomes and E and (b) that causality run only from E to f_m^* and not in the reverse direction. The potential for endogeneity bias in Equation 2 is non-trivial for all there economic outcomes. Welfare benefit amounts are conditioned on family size and composition and employment decisions are shaped by family obligations and resources available from spouses and partners. The potential for endogeneity is somewhat less in analyses of marriage and births than for cohabitation and doubling up, because marriage and births are measured as changes from baseline. There also is some protection against reverse causation in the way our economic variables are lagged (restricted to the first two years of the follow-up period)—but only to a degree, as some temporal overlap remains.[11]

To address endogeneity more fully, we estimate Equation 2 using instruments for E. The first-stage IV equations are:

(3) $$E = X_1\,\Pi_1 + X_2\,\Pi_2 + \upsilon,$$

where X_1 is the same set of control variables as in Equation 2, X_2 is a vector of independent variables assumed to be correlated with E and uncorrelated with f_m^* except through their associations with E (i.e., the excluded instruments); Π_1 and Π_2 are coefficients for X_1 and X_2, respectively; and υ is a random

[11] The surveys measured demographic outcomes only as of the survey date, thus capturing behavior over the entire follow-up period.

error term. If the assumptions are met, solving Equation 3 and using predicted values of E in Equation 2 provides consistent estimates for γ. Because there are three economic predictors (employment, earnings, and welfare benefits), we need at least three instruments in order for the equations to be identified.

Our excluded instruments, X_2 , are interactions between the treatment variable, T, and a series of fixed baseline characteristics. The random nature of the treatment variable provides excellent insurance against correlation with omitted variables in Equation 2 (Angrist and Krueger, 2001). Indications of substantial heterogeneity in welfare reforms' economic impacts (Michalopolous & Schultz, 2001; Bitler et al., 2004c) led us to hope that these instruments might differentiate among the three endogenous regressors in Equation 2 (Gennetian et al., 2002).

We use a backwards elimination strategy to select interactions whose contributions to explained variance in Equation 2 were statistically significant. For each economic outcome (employment, earnings, and welfare benefits) we started with a fully-saturated OLS model and then successively eliminated all interactions involving T that were not statistically significant (at $p<.10$). We included in the final model for Equation 3 terms that were significant for one or more economic variable, as well as all baseline characteristics X_1. The resulting instruments, X_2, include the treatment dummy, T, its interactions with: three state dummies (Minnesota as the omitted category), education, new/ongoing applicant, new/ongoing* Delaware, and new/ongoing*Indiana.

Partial F statistics indicate strong correlations of these instruments with employment (12.0) and welfare benefits (12.8) and a somewhat weaker correlation with earnings (3.5). The instruments may not differentiate adequately among the economic variables if the latter are inter-correlated, however, leading to "instrument irrelevance" (Baum et al., 2003). One way to assess instrument irrelevance is to compare standard partial R^2 statistics for each economic regressor with a partial R^2 statistic that adjusts for inter-correlation.[12] We discuss these statistics in the section on results.

[12] The difference between the unadjusted and adjusted partial R^2s cannot be tested formally since the Shea (adjusted) statistic's distribution has not been derived (Baum et al., 2003). Stock and Yogo (2004) provide critical values for testing the global contribution of a set of excluded instruments when there are multiple endogenous regressors based on the Cragg-Donaldson statistic, but this test does not discern whether some regressors are not identified.

In estimating γ, we assume a probit function relating observed family formation outcomes (marriage, cohabitation, doubling up, and births) to the independent variables. For endogeneity and overidentification tests, we use a robust ordinary least squares routine, as these tests have not been developed for the probit model. Estimated coefficients and standard errors from OLS are very similar to those from probit, suggesting that OLS diagnostic tests provide useful guides.

The overidentification tests are relevant to the second critical IV assumption—statistical independence between the instruments and second-stage error term. In our application, this assumption requires that any impacts of welfare reform on family formation captured in the experiments occur only through policy effects on employment, earnings, and welfare benefits and not from policy effects through other channels.

In our literature review, we suggest that this assumption may be justified. We identified two potential additional pathways by which welfare reforms might influence demographic behavior. First, family decisions might turn on perceived financial pressures in anticipation (and independent) of any actual financial impacts. Second, demographic behaviors might respond to non-economic influences such as new messages about family responsibilities and any expanded access to reproductive health services. However, we expect any such influences to be weak relative to direct economic effects. Furthermore, estimates from the current analysis are vulnerable only to the degree that the experimental designs measured these kinds of indirect effects. In general, the experiments were unable to restrict study participants' exposure to any wider signals reforms sent into the community at large concerning financial responsibilities or demographic behaviors. Such signals may have led some control group members to believe incorrectly that they were subject to time limits or other welfare reform rules (Camasso et al., 2003). By contrast, when it came to the way clients were handled in the welfare offices, the designs generally did create strong contrasts in procedures applying to employment, earnings, and welfare benefits—partly because these were the direct focus of the projects and partly because safeguards against inappropriate handling were hard-wired into automated systems.

As a general test of potential bias from any omitted mediators, we report overidentification tests—in the form of Hansen's J statistic (Baum et al., 2003)—that detect correlations between the excluded instruments, X_2, and the error terms in Equation 2. As discussed under results, we find no statistically significant correlations of this type.

Experimental Impact Findings

We examine first pure experimental impacts for the four demographic outcomes in each state (Equations 1a and 1b). Findings indicate, for each outcome, a statistically significant impact in just one state—a different state for each outcome (Table 2). Welfare reform raises the percent married in Delaware by 3.5 percentage points ($p<.05$), increases cohabitation in Florida by 2.8 points ($p<.10$), reduces doubling up in Minnesota by 3.5 points ($p<.10$), and increases births in Indiana by 2.9 points ($p<.10$). The number of statistically significant impacts (four) modestly exceeds the number expected by chance (two) in a test of 16 estimates at the 10-percent level. [13] The estimates are consistent with existing findings from the four evaluations, allowing for differences due in follow-up and sample definitions.[14]

Compared with the demographic outcomes, impacts on average economic outcomes (measured over the first two follow-up years) are larger and more variable across states (bottom panel, Table 2, $p<.01$ for all three F-tests of state differences). We see sizeable proportionate reductions in average benefits in Delaware (19 percent impact) and Indiana (10 percent impact), zero impact in Florida, and a substantial proportionate *increase* in average benefits in Minnesota (14 percent impact). For employment, impacts are positive in all four states but substantially larger in Minnesota (26 percent) than elsewhere (7-12 percent). For average earnings, impacts are fairly similar in Minnesota (17 percent), Florida (15 percent), and Indiana (17 percent) and much smaller and statistically insignificant in Delaware (7 percent).

This variation in economic impacts reflects policy differences and that fact that Delaware maintained experimental distinctions for only a year and a half. At the extremes, Minnesota's generous earnings

[13] Differences in impacts across states were not statistically significant for any of the four outcomes in global F tests ($p>.10$ for all F-tests; see Table 2). More powerful tests based on comparisons with pooled estimates for states with non-significant point estimates found a significant difference between impacts on marriage in Delaware compared to all other states ($p<.10$) and nearly significant differences ($p<.15$) for impacts on cohabitation in Florida and doubling up in Minnesota.
[14] Fein (2001) finds a smaller, but still statistically significant and positive, impact on marriage in an earlier (one-year) follow-up survey in Delaware. Gennetian & Knox (2003) report a slightly smaller positive cohabitation impact (.024) for a sample that includes some men. The Indiana evaluation (Beecroft et al., 2003) did not report impacts for births in the full adult survey sample, and the Minnesota study (Miller et al., 2000) did not report impacts for doubling up.

disregards produced large positive benefit impacts, whereas Delaware's tough sanctions and time limits led to relatively large and long-lasting negative benefit impacts. Bloom et al. (2000) report that benefit impacts were small in Florida until the third follow-up year, when clients began to reach the 24-month time limit. Impacts on employment and earnings are smaller in Delaware than in the other states due to the extension of work requirements to control group members in the second follow-up year.

A comparison of economic and demographic impacts across states hints at a potential linkage between the degree of financial pressure from benefit reductions and adjustments in living arrangements and fertility. Delaware, Florida, and Indiana, which featured time limits and sanctions for a wide range of personal behaviors, increased marriage and cohabitation and decreased fertility, whereas Minnesota, with generous financial incentives, reduced doubling up.[15] Non-experimental analyses in the next section provide more direct estimates of the relationships between economic impacts and demographic behavior.

As discussed in the methods section, our non-experimental approach exploits variation in experimental impacts across subgroups to instrument for economic outcomes in equations for each family formation outcome. To assess variability in demographic and economic impacts, we tested interactions between treatment status and the seven personal and family characteristics in Table 1 for each state and outcome. For the demographic outcomes, we find that only 11 sets of subgroup contrasts were statistically significant at the 10-percent level out of a total of 112 interactions tested (7 characteristics * 4 states * 4 outcomes). Because this is exactly the number we would expect by chance, we must regard the results—summarized for the 11 significant interactions in Table 3—with some caution.

Subgroup patterns for the four demographic outcomes exhibit the kinds of inconsistencies reported in previous research (e.g., Fein et al., 2002; Gennetian & Knox, 2003). We do find some common patterns, such as in Delaware and Florida, where reforms are associated with higher cohabitation among women with only one child and with lower or similar levels of cohabitation among women with two or more children. Such findings fit with the notion that children may constrain partnering responses, but it is unclear why differences would occur for cohabitation and not for marriage. Other findings are less

[15] As noted above, financial pressure in Florida increased in Year 3 as treatment group members began reaching time limits, and a significant treatment-control welfare benefit gap emerged (Bloom et al., 2000).

consistent with theory. For instance, reforms reduce cohabitation (Delaware) and marriage (Florida) among whites but have the opposite effects for blacks.

By contrast, there is a great deal more subgroup variation in the estimated economic impacts. Of a total of 84 contrasts tested (7 characteristics * 4 states * 3 outcomes), 64 were statistically significant at the 10-percent level. We do not report details here, as the number of results is large and similar analyses have been reported previously (e.g., Michalopolous & Schwartz, 2001).[16] Rather, having established that there is important subgroup variation in economic impacts, we now discuss results from an instrumental variables strategy that exploits this variation to assess whether economic impacts bring demographic changes.

Instrumental Variables Estimates of Demographic Responses to Economic Impacts

As discussed earlier, we used a backwards elimination strategy to identify treatment * subgroup interactions that were correlated with economic outcomes and could serve as the excluded instruments. The instruments included: the main treatment term and its interactions with state, new applicant status, education, and state * new applicant status. The two main requirements for consistent IV estimates are that these instruments be sufficiently correlated with employment, earnings, and welfare benefits and otherwise uncorrelated with marriage, cohabitation, doubling up, and births.

Partial correlation statistics for Equation 3 indicate fairly strong associations of the instruments (X_2) with employment and benefits, but not with earnings (see Table 4).[17] Although the conventional partial F-statistic for instruments in the earnings equation is statistically significant ($p<.01$), it is not nearly as large as those for employment and welfare benefits. More importantly for purposes of multivariate regression, comparison of the regular and adjusted partial R^2 statistics (rows 2 and 3) indicates that the instruments cannot distinguish variation in employment from variation in earnings. As a consequence, when all instruments are included in a multivariate model they may not identify the effects of either employment or (particularly) earnings very well. The reason is that employment impacts appear to drive earnings impacts to a degree that produces a high correlation between the two impacts across subgroups.

[16] This analysis is available on request from the first author.
[17] In addition to the excluded instruments, first-stage models (see Equation 3 above) also included state dummies and main effects for all seven baseline characteristics.

When we remove earnings from the equations, partial R^2s in the bottom row of Table 4 show satisfactory discrimination between employment and welfare benefits. Estimates for employment now capture earnings effects as well as any non-economic effects of work, however, a point to remember when interpreting findings from models excluding earnings.

Key results appear in Table 5, expressed as changes in the probability of each demographic outcome associated with a 10-percentage point increase in average quarterly employment rates, and with $100 increases in average quarterly earnings and welfare benefits measured over the first two follow-up years.[18] These effects are estimated while setting other characteristics in the model to their mean values. Models 1 and 3 include earnings among the dependent variables, and Models 2 and 4 exclude earnings for reasons discussed above.

Whereas standard probit estimates indicate a small negative effect for employment on marriage, the IV models suggest a relatively large positive effect. The difference between the standard and IV estimates (in models 2 and 4) is statistically significant, indicating potentially serious biases in the standard estimates. However, the IV estimates do not differ significantly from zero (due to large standard errors).[19] Overall, the indications do not strongly indicate a relationship between employment and marriage. Nor is there any evident relationship between employment and cohabitation (second panel, first row).

There is a comparatively strong indication that increased employment leads to less doubling up with adults other than a spouse or partner. Probit estimates are negative across standard and IV models and statistically significant in three models. The IV estimate of the effect of a 10-percentage point increase in employment—a two point reduction in the fraction doubling up (model 4)—is much larger than the corresponding standard probit estimate.

[18] A one-standard deviation increase is .366 for the average quarterly employment rate (with a mean of .487); $1,251 for the average quarterly earnings amount (with a mean of $975); and $720 for the average quarterly welfare benefits amount (with a mean of $729).

[19] These tests are based on statistics computed using an OLS IV program (Stata's ivreg2) that provided effect estimates nearly identical to those from the probit analysis (Stata's ivprobit, which does not provide Hausman tests). T-statistic = 1.76 for the difference in employment effects on marriage in Models 2 and 4. For the remaining 19 comparisons between standard and IV estimates in Table 5, Hausman tests found no statistically significant differences.

Standard probit estimates show a small, statistically significant decrease in births for increased employment. Here, the IV estimates, which differ in sign from the standard probit estimates, may not merit much weight given problems instrumenting earnings in Model 3 and the fact that Hausman tests do not indicate statistically significant differences between the un-instrumented and instrumented estimates.

In assessing the effects of earnings we are restricted to the regular probit estimates (in Model 1), as first-stage regression diagnostics found that the instruments could not discern earnings from the better instrumented effects of employment. Without valid instruments for comparison, there is no basis for testing for endogeneity. Limited to Model 1, we see small negative coefficients for earnings on all four outcomes in Model 1, none of which is statistically significant.

Finally, we find consistent evidence that lower welfare benefits are associated with higher rates of marriage and doubling up and lower rates of childbearing. A $100 decline in average benefits is associated with a statistically significant one-percentage point increase in the proportion married in Models 1 and 2 and fairly similar changes in the corresponding IV estimates. The effect on doubling up is smaller (a .4 to .5 percentage point increase) but also statistically significant (in Models 1 and 2) and consistent with the IV estimates. Finally, a $100 decline in welfare benefits brings only a small (.2 percentage point) reduction in the probability of a birth. Though statistically significant only in Model 2, this effect is consistent with the direction of the corresponding IV estimates. Neither benefits, nor either of the other two economic outcomes, have any apparent effects on cohabitation.

Overall, diagnostic statistics give some basis for confidence in both the IV and standard probit estimates. The first-stage regressions are sufficiently powerful to avoid bias from weak instruments, at least for employment and welfare benefits.[20] Although unbiased, the IV estimates nevertheless are not very precise. We also computed Hansen's J statistic, which tests statistical independence between a set of excluded regressors and the main outcome, conditional on the predicted endogenous regressors and other second-stage covariates (Baum et al., 2003). Examination of these statistics found no violations of

[20] A common rule of thumb in models with only one endogenous regressors is that F statistics greater than 10 provide reasonable insurance against such bias (Bound et al., 1995). Other than the Shea statistic, analogous standards have not been developed to assess each of a set of multiple endogenous regressors Baum et al. (2003), although Stock and Yogo (2001) provide critical levels for the global contribution of such a set of regressors.

the independence assumption for any of the four family formation outcomes in either the three- or two-endogenous regressor models in Table 5.[21] Finally, as discussed above, Hausman tests revealed only one instance in which the standard and IV estimates differed statistically. We thus have some confidence that statistically significant standard estimates are meaningful, particularly where they carry the same sign as the IV point estimates.

Sensitivity Analyses

We performed several additional analyses to assess results' robustness to the choice of excluded instruments, to time frames covered by economic measures, and to the inclusion of each of the four states in the sample. First, we assessed the effects of dropping several interactions between characteristics and treatment status from first-stage IV models that did not entail an excessive loss of statistical power. The results are not very sensitive to the selection of excluded instruments.

Next, we examined the effects of defining economic outcomes over the entire follow-up period, rather than just the first two follow-up years.[22] Restricting economic measures to the first two follow-up years injects useful temporal separation between independent variables and family formation outcomes (measured at four years on the average), but misses any important effects of economic impacts occurring later in the follow-up period. Although coefficient magnitudes and significance levels change somewhat, key qualitative findings are not affected. Specifically, employment still has little effect on marriage or cohabitation but is related negatively to doubling up and births; earnings effects on all outcomes remain weakly negative (statistically significantly so for doubling up and births); and welfare benefits are still correlated negatively with marriage and doubling up, positively with births, and not at all with cohabitation.[23]

---

[21] The statistics failed to reject independence at $p<.30$ for marriage, doubling up, and births and at $p<.20$ for cohabitation in both the three- and two-economic variable models.
[22] The re-analysis included redoing the backwards selection process identifying interactions between treatment status and covariates to be included in the first-stage IV equations. The resulting all-year models were fairly similar to models for the two-year versions of the economic models.
[23] Tests of overidentification restrictions indicate potential instrument invalidity for the marriage model ($p<.12$ for Hansen's J statistic in Model 4). The lack of a similar problem in models based on two-year economic outcomes (e.g., Table 6) suggests that the lagged specifications may help to minimize feedback effects from unmeasured mediators over the longer term.

Finally, we test the effects of excluding each state to assess the possible effects of idiosyncratic time patterns in economic impacts occurring after the first two follow-up years, particularly in Delaware, Florida, and Indiana.[24] We start by re-estimating Model 1 from Table 5 four times, each time systematically excluding a different state. Model 1 allows us to estimate earnings effects and provides good precision, apparently without serious endogeneity bias. We opt for examining estimates for three states after excluding a fourth rather than making estimates for individual states, since our finding of minimal endogeneity is based on our previous four-state analysis. These analyses show little sensitivity to the exclusion of Delaware, Florida, or Indiana. Although significance levels fluctuate from sample to sample, the magnitudes and signs of most estimates are the same as in the all-state sample. The finding again does not suggest that defining economic regressors over the first two follow-up years is problematic.

We do find that excluding Minnesota markedly sharpens many of the effects we noted in the all-state sample. Given that time patterns for impacts in Minnesota were relatively stable after the first two years, it is possible that policy differences between the three TANF-like programs and MFIP are responsible. To take a closer look, Table 6 provides estimates for Models 1-4 for the sample limited Delaware, Florida, and Indiana.[25]

For welfare benefits, the three-state effects for marriage (negative) and births (positive) are larger than those in the all-state sample. In both instances, the un-instrumented and instrumented estimates have the same sign, and the latter are substantially larger than the former. For marriage, Hausman tests reject the equivalence of estimates in Models 1 and 3, suggesting that the un-instrumented probit estimates may represent a lower bound to the true effect. For doubling up, un-instrumented benefit effects are smaller than in Table 5 and are no longer statistically significant. The corresponding IV estimates are larger

[24] In Delaware, the end of random assignment in the middle of the second follow-up year led to the disappearance of employment and earnings impacts, whereas welfare benefit impacts actually increased due to the built-in treatment-control differential in time clocks and continuing effects of permanent sanctions (Fein et al., 2001). In Florida, benefit reductions grew after the second year, as participants began to reach time limits (Bloom et al., 2000). In Indiana, the size of both positive earnings and negative benefit impacts grew over time (Beecroft et al., 2003). By contrast, there was little change in impacts in the third, compared with earlier, years for the MFIP demonstration (Miller et al., 2000).

[25] Again, we redid the backwards elimination process to identify the best excluded instruments for first-stage regressions. Indications from the standard and Shea partial R^2 statistics showed the models performed similarly to, if with somewhat less power than, the four-state sample.

than in the four-state sample, but Hausman tests do not indicate statistically significant differences in the standard and IV estimates.

Indications of employment effects on marriage and births continue to be inconsistent in the three-state sample, and earnings continue to show at most weak relations to the outcomes. The three-state estimates also continue to suggest that increased work leads to decreased doubling up, with magnitudes similar to estimates from the all-state model.

Although further analysis would be needed to establish the reasons for these state differences, there certainly were marked policy differences between Minnesota and the other three states that might be responsible. Minnesota stands out particularly for its generous financial work incentives (leading to positive impacts on welfare benefits) and the fact that during the demonstration there were no time limits, much less use of sanctions, and none of the parental responsibility incentives. Reforms in Delaware, Florida and Indiana did feature the latter policies but made comparatively little use of earned income disregards. All four states adopted mandatory work requirements, perhaps explaining why employment and earnings effects are not affected by excluding Minnesota from the sample.

Discussion

Purely experimental results from this analysis accord with existing research suggesting that welfare reforms of the 1990s had at most small effects on demographic behavior in some states and subgroups. At the state level, we find that welfare reform has small, statistically significant positive impacts on marriage in Delaware and cohabitation in Florida, and negative impacts on births in Indiana and doubling up in Minnesota. Point estimates are very similar for those outcomes previously reported from these studies, if slightly sharpened by our restriction to women who were unmarried at the point of random assignment. Also as in past studies, we find a mixed assortment of differences in family formation impacts across subgroups within states. The findings contribute to the existing universe of point estimates; namely, by adding results for two states (Delaware and Indiana) and two outcomes (doubling up and births) not previously analyzed in detail.

Nearly all previous empirical analysis has focused on discerning the overall effects of welfare reform policy bundles, and of key components of these bundles (e.g., family cap, work requirements, time limits), on demographic outcomes. In this paper we suggest that one reason results have been so inconclusive is that analyses have not specified and measured the intervening mechanisms. Since the direct objective of welfare reform was to move single mothers from welfare to work, a good place to start is by examining the effects on demographic behavior of policy-induced changes in employment, earnings, and welfare benefits.

Toward this end, we have exploited experimental variation in economic impacts across states and subgroups to instrument for hypothesized economic mediators. A widely-recognized vulnerability of the welfare reform experiments is their inability to insulate control groups completely from possible wider cultural reverberations of new signals emanating from welfare offices. This vulnerability works to our advantage, making the experiments better candidates for identifying direct economic effects using IV methods given reduced bias from other mediators not captured in the designs. Although these conditions help us to identify the effects of economic impacts on women who received welfare during the early years of welfare reform, the analysis does not tell us whether and how the "culture of welfare" may have changed or how family formation behavior may be influenced over the wider population in the long-term.

Diagnostic statistics suggest that the approach has some promise: the instruments were significantly associated with the first stage outcomes and over-identification tests did not suggest the presence of additional, unmeasured mediators of welfare reform's impacts on demographic behavior. As usual in IV analyses, the benefits of statistical consistency came at the expense of a large loss in statistical precision. We thus found the IV estimates to be useful guides to interpretation, but not completely satisfactory replacements for the un-instrumented estimates.

Substantively, we find little consistent evidence that impacts on employment and earnings have had much effect on the marriage or cohabitation. The analysis provides strong indications, however, that going to work reduces the likelihood of sharing living quarters with other adults (possibly by a substantial amount) and of having additional children (by a small amount). The findings tend to reinforce a view of employment as bringing countervailing influences on marriage—reducing need for a spouse's income while increasing women's financial viability as partners, and increasing exposure to potential partners in the workplace while depleting time and energy for forming and sustaining intimate relationships. By comparison, both expectations and findings are clearer for employment's effects on doubling up and births. Statistically significant negative effects for doubling up affirm the view that financial necessity is a predominant motive for sharing living quarters with other adults. Negative effects for births support the hypothesis that employment increases the psychic and financial costs of additional childbearing.

In contrast to employment and earnings, we find that welfare benefits are negatively related to marriage. Benefits reduce the need for a spouse's income, and financial disincentives continue to face single parent recipients contemplating marriage to a man with income. The implications of benefit reductions thus are clearer on net than those of earnings increases. Our results indicate little consistent welfare benefit effect on cohabitation, which we expected might increase also as a response to increased financial need accompanying benefit reductions. The fact that welfare rules always have taken a more liberal stance towards cohabitants' income may explain why the presence of benefits has little inhibiting effect on cohabitation. More generally, the analyses provide little evidence that cohabitation is sensitive to policy impacts on any of the three economic outcomes. Previous research shows that cohabitation is a relatively fluid arrangement, with less income pooling and lower bars to entry and exit, compared with marriage. For this reason, welfare mothers may not look to informal unions for economic support when financial strains arise.

The findings also support the hypothesis that welfare benefits are negatively related to doubling up. This effect is concentrated in Minnesota, where policies led to higher average benefits (implying less doubling up).

Finally, the results support expectations that reductions in welfare benefits lead to lower fertility. As is the case for marriage, this effect also is concentrated in the three states whose programs most closely resembled TANF. This convergence may be due to the fact that these three programs produced reductions in average benefits, whereas the Minnesota program provided increased benefits. Our analysis does not indicate why fertility did not increase with higher benefits in Minnesota. Recent work in behavioral economics indicates that people respond differently to financial losses and gains for essentially non-economic reasons (e.g., emotions), but such speculation takes us well beyond the limits of the present analysis.

In sum, the findings reinforce our initial premise that the demographic impacts of welfare reforms reflect the net effects of divergent forces and that there is likely to be more structure to the "mixed" findings produced to date than readily meet the eye. The expected effects of women's employment and earnings are more ambiguous for marriage and cohabitation than for doubling up and births. Nevertheless, because welfare reforms typically had opposing impacts on employment-earnings (positive) and welfare benefits (negative) we recommend greater attention to distinguishing these influences in future work.

Differences in economic effects on marriage, cohabitation, and doubling up—and in net impacts on these outcomes across states—suggest that these living arrangements offer alternative responses to changes in economic status that will suit some circumstances better than others. We have speculated on the conditions favoring each of these responses, but a more definitive understanding also requires closer analysis.

There are several additional ways the analyses might be refined and extended. Adding data from more experiments would help in assessing the generality of findings across a wider range of policies, populations, and socioeconomic environments. Such an expansion also might introduce additional cross-site and cross-subgroup variation in experimental impacts, thereby improving the power of the IV estimators. Where experiments obtained measures of perceived financial pressures and of family values, further analysis to rule out or in potential mediating roles for these influences would be useful.

References

Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives, 15*(4), 69-85.

Baum, C. F., Schaffer, M. E., & Stillman, S. (2003). *Instrumental variables and GMM: Estimation and testing* (Tech. Rep. No. Working Paper No. 545). Boston College, Department of Economics.

Becker, G. (1991). *A Treatise on the Family (enlarged edition).* Cambridge, MA: Harvard University Press.

Beecroft, E., Lee, W., Long, D., Holcomb, P. A., Thompson, T. S., Pindus, N., et al. (2003). *The Indiana welfare reform evaluation: Five-year impacts, implementation, costs and benefits*. Bethesda, MD: Abt Associates Inc.

Bitler, M. P., Gelbach, J. B., Hoynes, H. W., & Zavodny, M. (2004a). The impact of welfare reform on marriage and divorce. *Demography, 41*(2), 213-236.

Bitler, M. P., Gelbach, J. B., & Hoynes, H. W. (2004b). *Welfare reform and children's living arrangements.*, Manuscript.

Bitler, M. P., Gelbach, J. B., & Hoynes, H. W. (2004c). *What mean impacts miss: Distributional effects of welfare reform experiments.*, Manuscript.

Bloom, D., Farrell, M., & Fink, B. (2002). *Welfare time limits: State policies, implementation, and effects on families*. New York, NY: MDRC.

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.* New York: Manpower Demonstration Research Corporation.

Bloom, D., & Michalopoulos, C. (2001). *How welfare and work policies affect employment and income: A synthesis of research*. New York, NY: MDRC.

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

Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association, 90*(430), 433-450.

Bramlett, M., & Mosher, W. (2002). Cohabitation, marriage, divorce, and remarriage in the United States. *National Center for Health Statistics. Vital Health Statistics, 23*(22).

Burlingame, P., Hutson, R., & Levin-Epstein, J. (1999). *Making the link: Pregnancy prevention and the new welfare era*. Washington, D.C.: Center for Law and Social Policy.

Camasso, M., Harvey, C., Jagannathan, R., & Killingsworth, M. (1998). *New Jersey's Family Development Program: Results on program impacts, experimental control group analysis*. Trenton, N.J.: New Jersey Department of Family Services.

Camasso, M. J., Jagannathan, R., Harvey, C., & Killingsworth, M. (2003). The use of client surveys to gauge the threat of contamination in welfare reform experiments. *Journal of Policy Analysis and Management, 22*(2), 207-223.

Clarkberg, M. (1999). The price of partnering: The role of economic well-being in young adults' first union experiences. *Social Forces, 77,* 945-968.

Conger, R., Ebert-Wallace, L., Sun, Y., Simons, R., McLoyd, V., & Brody, G. (2002). Economic pressure in African-American families: A replication and extension of the Family Stress Model. *Developmental Psychology, 38*(2), 179-193.

Conger, R. D., Rueter, M. A., & Elder, G. H., Jr. (1999). Couple resilience to economic pressure. *Journal of Personality and Social Psychology, 76*(1), 54-71.

Dyer, W. T., & Fairlie, R. W. (2004). Do family caps reduce out-of-wedlock births? Evidence from Arkansas, Georgia, Indiana, New Jersey, and Virginia. *Population Research and Policy Review, 23,* 441-473.

Edin, K. (2000a). Few good men: Why poor mothers don't marry or remarry. *The American Prospect,* 26-31.

Edin, K. (2000b). What do low-income single mothers say about marriage? *Social Problems, 47,* 112-133.

Ellwood, D. T. (2000). The impact of the earned income tax credit and social policy reforms on work, marriage, and living arrangements. *National Tax Journal, 53*(4), 1063-1106.

Ellwood, D. T., & Jencks, C. (2001). *The growing differences in family structure: What do we know? Where do we look for answers?*. Unpublished manuscript, Harvard University.

Fairlie, R. W., & London, R. A. (1997). The effect of incremental benefit levels on births to AFDC recipients. *Journal of Policy Analysis and Management, 16*(4), 575-596.

Fein, D., Burstein, N., Fein, G., & Lindberg, L. (2003). *The Determinants of Marriage and Cohabitation among Disadvantaged Americans: A Literature Review*. Bethesda, MD: Abt Associates Inc.

Fein, D., Duberstein Lindberg, L., London, R., & Mauldon, J. (2002). *Welfare reform and family formation: Assessing the effects.* Research Brief No. 1, The Welfare Reform and Family Formation Project, Bethesda, MD: Abt Associates Inc.

Fein, D., Long, D., Behrens, J., & Lee, W. (2000). *Turning the corner: Delaware's A Better Chance welfare reform program at four years*. Bethesda, MD: Abt Associates Inc.

Fein, D. J. (2001). Will welfare reform influence marriage and fertility? Early evidence from the ABC demonstration. *Evaluation and Program Planning, 24,* 427-444.

Fitzgerald, J. M., & Ribar, D. C. (2004). Welfare reform and female headship. *Demography, 41*(2), 189-212.

Fox, G. L., & Chancey, D. (1998). Sources of economic distress: Individual and family outcomes. *Journal of Family Issues, 19*(6), 725-749.

Fraker, T. M., Ross, C. M., Stapulonis, R. A., Olsen, R. B., Kovac, M. D., Dion, M. R., et al. (2002). *The Evaluation of Welfare Reform in Iowa: Final Impact Report*. Washington, D.C.: Mathematica Policy Research Inc.

Gais, T., Nathan, R., Lurie, I., & Kaplan, T. (2001). Implementation of the Personal Responsibility Act of 1996. In R. Blank & R. Haskins (Eds.), *The New World of Welfare* (pp. 35-69). Washington D.C.: The Brookings Institution.

Gennetian, L., & Knox, V. (2003). *Staying single: The effects of welfare reform policies on marriage and cohabitation.* (The Next Generation Project, NY: Manpower Demonstration Research Corporationn, Working Paper No. 13).

Gennetian, L. A., Bos, J. M., & Morris, P. A. (2002). *Using instrumental variables to learn more from social policy experiments*. New York, NY: MDRC.

Grogger, J., Karoly, L., & Klerman, J. (2002). *Consequences of welfare reform: A research synthesis* (Tech. Rep. No. DRU-2676). Los Angeles, CA: RAND.

Hamilton, G., Freedman, S., Gennetian, L., Michalopoulos, C., Walter, J., Adams-Ciardullo, D., et al. (2001). *National evaluation of welfare-to-work strategies: How effective are different welfare-to-work approaches? Five-year adult and child impacts for eleven programs*. New York, NY: MDRC.

Harknett, K., & Gennetian, L. A. (2003). The effect of an earnings supplement on union formation. *Demography, 40*(3), 451-478.
Hutson, R. Q., & Levin-Epstein, J. (200). *Linking family planning with other social services: The perspectives of state family planning administrators*. Washington, D.C.: Center for Law and Social Policy.
Joyce, T., Kaestner, R., Korenman, S., & Henshaw, S. (2004). Family cap provisions and changes in births and abortions. *Population Research and Policy Review, 23,* 475-511.
Kaestner, R., Korenman, S., & O'Neil, J. (2003). Has welfare reform changed teenage behaviors? *Journal of Policy Analysis and Management, 22*(2), 225-48.
Kearney, M. S. (2004). Is there an effect of incremental welfare benefits on fertility behavior? A look at the family cap. *The Journal of Human Resources, 39*(2), 295-325.
Levine, P. (2002). *The impact of social policy and economic activity throughout the fertility decision tree.* (. Cambridge, MA: National Bureau of Economic Research.
London, R. (2000). The interaction between single mothers' living arrangements and welfare participation. *Journal of Policyi Analysis and Management, 19*(1), 93-117.
Manning, W. D., & Smock, P. J. (1995). Why marry? Race and the transition to marriage among cohabitors. *Demography, 32*(4), 509-520.
Mauldon, J. G., London, R. A., Fein, D. J., Patterson, R., & Sommer, H. (2004). Attitudes of welfare recipients toward marriage and childbearing. *Population Research and Policy Review, 23,* 595-640.
McLaughlin, D. K., & Lichter, D. T. (1997). Poverty and the marital behavior of young women. *Journal of Marriage and the Family, 59,* 582-594.
Mellor, J. M. (1998). The effect of family planning programs on the fertility of welfare recipients. *Journal of Human Resources, 33*(4), 866-895.
Michalopoulos, C., & Schwartz, C. (2001). *What Works Best for Whom: Impacts of 20 Welfare-to-Work Programs by Subgroup*. New York, NY: MDRC.
Miller, C., Knox, V., Gennetian, L., Dodoo, M., Hunter J.A., & Redcross, C. (2000). *Reforming Welfare and Rewarding Work: Final Report on the Minnesota Family Investment Program*. New York: Manpower Demonstration Research Corporation.
Miller, C., Knox, V., Gennetian, L., Dodoo, M., Hunter, J., & Redcross, C. (2000). *Reforming Welfare and Rewarding Work: Final Report on the Minnesota Family Investment Program* (Volume 1: Effects on Adults). New York, NY: MDRC.
Moffitt, R. (2002). *The Temporary Assistance to Needy Families Program.* NBER Working Paper No. 8749. Cambridge, MA: National Bureau of Economic Research.
Moffitt, R. (1998). The effect of welfare on marriage and fertility. In R. Moffitt (Ed.), *Welfare, the family, and reproductive behavior: Research perspectives* (pp. 50-97). Washington, DC: National Academy Press.
Moffitt, R., Reville, R., & Winkler, A. E. (1998). Beyond single mothers: Cohabitation and marriage in the AFDC program. *Demography, 35*(3), 259-278.
Mosher, W. et al. (2004). *Use of Contraception and Use of Family Planning Services in the United States, 1982-2002* (Advance Data from Vital and Health Statistics No. 350). Washington, D.C.: Centers for Disease Control.
Murray, C. (2001). Family formation. In R. Blank & R. Haskins (Eds.), *The New World of Welfare* (pp. 137-168). Washington, D.C.: Brookings Institution.
Rosenbaum, D. (2000). *Taxes, the Earned Income Tax Credit, and marital status.* (Joint Center for Poverty Research, Chicago, IL, Joint Center for Poverty Research Working Paper No. 177).

Rossi, P. (2000). New Jersey's Family Development Program: An overview and critique of the Rutgers evaluation. In D. Besharov & P. Germanis (Eds.), *Preventing subsequent births to welfare mothers* (pp. 29-46). College Park, MD: University of Maryland.

Schoeni, R. F., & Blank, R. M. (2000, February). *What has welfare reform accomplished? Impacts on welfare participation, employment, income, poverty, and family structure.* Unpublished manuscript.

Scott, E., Edin, K., London, A., & Mazelis, J. (1999, September). *My children come first: Welfare-reliant women's post-TANF views of work-family trade-offs and marriage.* Presented at the For Better or Worse: State Welfare Reform and the Well-Being of Low-Income Families and Children Research Conference, Washington, DC: Northwestern University & University of Chicago Joint Center for Poverty Research.

Stock, J. H., & Yogo, M. (2004). *Testing for weak instruments in linear IV regression.*, Department of Economics, Harvard University, Manuscript.

Sweeney, M. M. (2002). Two decades of family change: The shifting economic foundations of marriage. *American Sociological Review, 67,* 132-147.

Turturro, C., Benda, B., & Turney, H. (1997). *Arkansas welfare waiver demonstration project: Final report*. Little Rock, AR: University of Arkansas at Little Rock.

U.S. House of Representatives. (1996). *Personal Responsibility and Work Opportunity Reconciliation Act of 1996* (104th Congress, 2nd Session No. 104-725). Washington, D.C.: U.S. Government Printing Office.

Vinokur, A. D., Price, R. H., & Caplan, R. D. (1996). Hard times and hurtful partners: How financial strain affects depression and relationship satisfaction of unemployed persons and their spouses. *Journal of Personality and Social Psychology, 71*(1), 166-179.

Wilson, W. (1987). *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy.* Chicago: University of Chicago Press.

Figure 1. Main Pathways of Expected Influence of Welfare Reforms on Family Formation

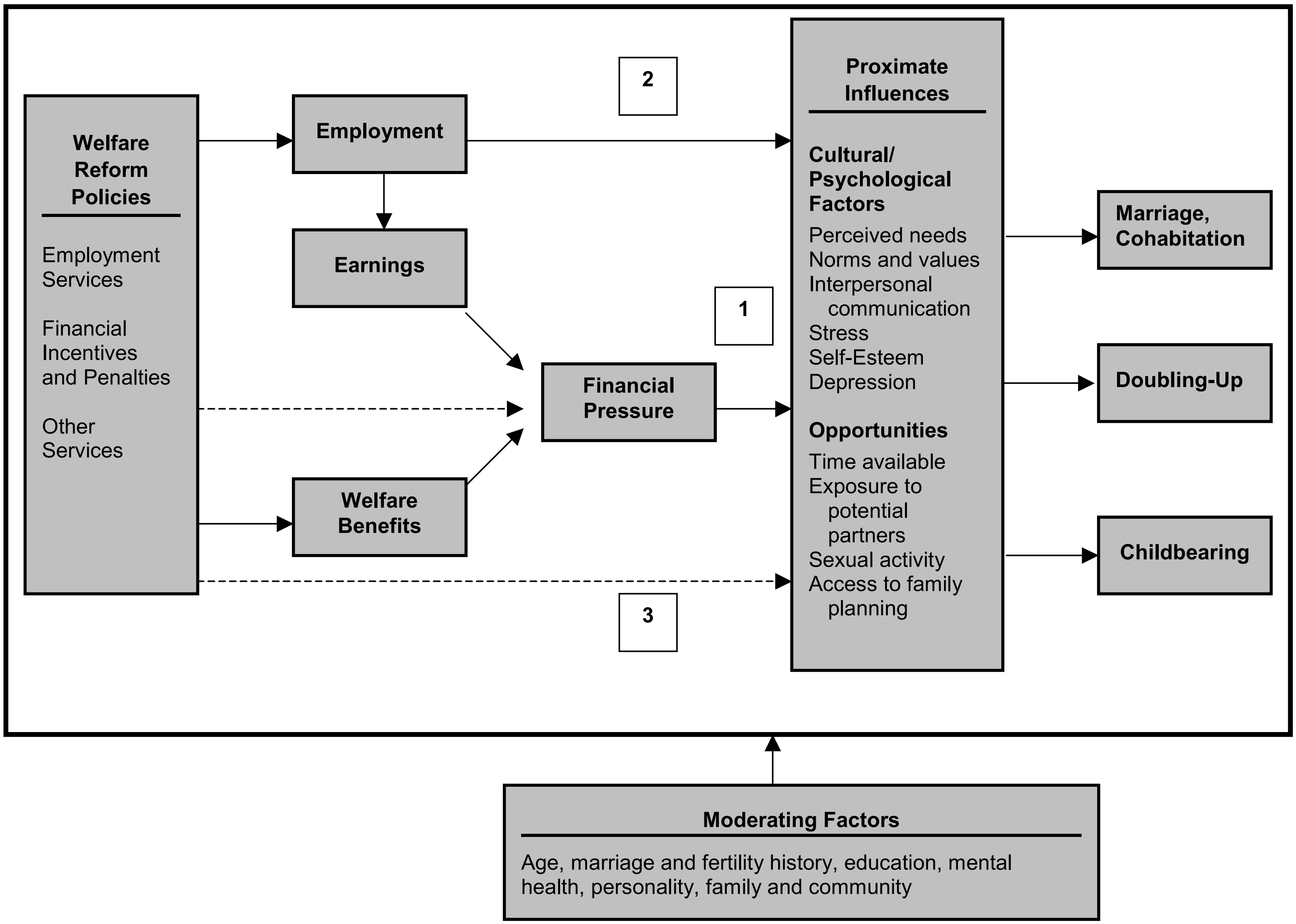

Table 1. Descriptive Statistics for Key Analysis Variables

Measure		State				
		Delaware	Florida	Indiana	Minnesota	All States
Outcomes for Control Group						
Lives with Spouse at Survey (%) ***		11.5	18.7	20.6	12.0	16.0
Lives with Partner at Survey (%) ***		9.2	9.4	12.1	15.4	11.9
Lives with Others at Survey (%) **		19.2	18.5	14.4	18.1	17.3
Had Birth since Baseline (%) ***		31.7	27.2	34.0	26.5	30.4
Over First Two Follow-up Years:						
Percent Employed in Average Quarter ** (Standard Deviation)		49.5 (33.1)	44.4 (36.9)	44.6 (37.0)	43.3 (37.7)	45.1 (36.5)
Average Total Quarterly Earnings*** (Standard Deviation)		$971.81 (1114.62)	$786.12 (1039.81)	$813.48 (1093.39)	$979.70 (1428.42)	$885.97 (1195.63)
Average Total Quarterly Welfare Benefits*** (Standard Deviation)		$483.36 (347.37)	$439.12 (369.30)	$458.40 (322.69)	$1,368.75 (869.52)	$721.87 (687.49)
Sample Size		706	809	1,087	1,055	3,658
Baseline Characteristics for All (Percentage Distributions)						
Age***						
	<25	38.2	33.2	38.9	35.8	36.6
	25-34	42.7	44.8	41.3	41.0	42.2
	35+	19.1	22.0	19.9	23.2	21.2
Number of Children***						
	1	34.4	41.0	48.1	49.0	44.2
	2	33.3	29.7	29.0	27.9	29.7
	3+	32.3	29.3	22.9	23.1	26.2
Age of Youngest Child**						
	<3	49.6	43.5	47.2	46.1	46.5
	3-5	23.8	26.9	23.8	23.2	24.3
	6+	26.6	29.6	29.0	30.6	29.2
Ever Married***						
	No	65.9	53.5	63.8	63.8	61.9
	Yes	34.1	46.5	36.2	36.2	38.1
Race***						
	White	35.7	42.7	54.0	54.1	48.0
	Black	58.5	55.5	40.7	34.1	45.5
	Other	5.7	1.8	5.3	11.7	6.5
Years of School Completed***						
	<12	45.2	45.5	40.4	35.5	41.0
	12	41.3	53.0	41.2	46.8	45.5
	>12	13.5	1.5	18.4	17.7	13.5
New Applicant***						
	No	57.0	55.0	66.0	60.6	60.2
	Yes	43.0	45.0	34.0	39.4	39.8
Average Months from Baseline to Survey Interview***		46.0	51.4	60.2	40.5	49.7
Sample Size		1,410	1,622	2,111	2,167	7,310

Note: Samples restricted to single mothers who were unmarried at baseline (random assignment).
*** State differences statistically significant at the 99-percent confidence level; ** at the 95-percent level; * at the 90-percent level.

Table 2. Experimental Impact Estimates for Demographic and Economic Outcomes by State

	Delaware			Florida			Indiana			Minnesota			
Outcome	(1) Control Mean	(2) Impact	(3) % Impact (2)/(1)=	(4) Control Mean	(5) Impact	(6) % Impact (5)/(4)=	(7) Control Mean	(8) Impact	(9) % Impact (8)/(7)=	(10) Control Mean	(11) Impact	(12) % Impact (11)/(10)=	(13) Prob. no State Difference
Probabilities for Demographic Outcomes at the Time of the Follow-up Survey													
Lives with Spouse	0.115	0.035**	0.304	0.187	-0.019	-0.102	0.206	-0.001	-0.005	0.120	0.001	0.008	0.297
Lives with Partner	0.092	-0.002	-0.022	0.094	0.028*	0.298	0.121	0.002	0.017	0.154	0.007	0.045	0.453
Lives with Other Adult	0.192	0.008	0.042	0.185	-0.009	-0.049	0.144	-0.005	-0.035	0.181	0.035*	0.193	0.466
Had Birth Since Baseline	0.317	-0.011	-0.035	0.272	0.002	0.007	0.340	-0.029*	-0.085	0.265	0.004	0.015	0.559
Average Quarterly Economic Outcomes over the First Two Follow-up Years													
Employment (probability)	0.495	0.033*	0.067	0.444	0.053***	0.119	0.446	0.034**	0.076	0.433	0.114***	0.263	0.000
Earnings ($000s)	0.972	0.066	0.068	0.786	0.116**	0.148	0.813	0.136***	0.167	0.980	0.164**	0.167	0.001
Welfare Benefits ($000s)	0.483	-0.091***	-0.188	0.439	-0.015	-0.034	0.458	-0.047***	-0.103	1.369	0.189***	0.138	0.000

Note: Statistics for demographic outcomes represents proportions of control group members with each outcome as of the follow-up survey, and the impact on this outcome (i.e., difference between treatment and control group means). Employment statistics represent average quarterly employment rates for control group, and the impact on this outcome, over the first two follow-up years. Statistics for earnings and welfare benefits in $000s, as indicated. Column 13 provides probabilities that differences in impacts across states exceed those that would arise by chance.

*** Impact estimate differs from zero at the 99-percent level; ** at the 95-percent level; * at the 90-percent level.

Table 3. Experimental Impact Estimates for Subgroups for Which Differences Are Statistically Significant (p<.10 for F-statistic)

Characteristic		State and Outcome (Mean Probability for Control Group and Impact)					
		MN: Live w/Others		**IN: Live w/Spouse**			
Age		**Cntl. Mean**	**Impact**	**Cntl. Mean**	**Impact**		
	<25	0.243	-0.032	0.243	-0.025		
	25-34	0.184	-0.079***	0.228	-0.027		
	35+	0.082	0.035	0.088	.092**		
		MN: Live w/Others		**DE: Live wPartner**		**FL: Live w/Partner**	
Number of Children		**Cntrl. Mean**	**Impact**	**Cntl. Mean**	**Impact**	**Cntl. Mean**	**Impact**
	1	0.200	-0.007	0.092	0.033	0.097	.070***
	2	0.166	-0.044	0.127	-.059**	0.105	-0.013
	3+	0.156	-.086**	0.051	0.017	0.081	0.009
		MN: Live w/Others		**MN: Live w/Partner**			
Age of Youngest Child		**Cntl. Mean**	**Impact**	**Cntl. Mean**	Impact		
	<3	0.229	-0.037	0.183	-0.038		
	3-5	0.165	-.081***	0.172	-0.001		
	6+	0.123	0.000	0.094	.080**		
		FL: Live with Others					
Ever Married		**Cntl. Mean**	**Impact**				
	No	0.216	-.052**				
	Yes	0.149	0.032				
		FL: Live w/Spouse		**DE: Live w/Partner**			
Race		**Cntl. Mean**	**Impact**	**Cntl. Mean**	**Impact**		
	White	0.293	-.059*	0.164	-.063**		
	Black	0.099	0.022	0.043	.032**		
	Other	0.538	-.366**	0.102	0.003		
		DE: Live w/Spouse					
New Applicant		**Cntl. Mean**	**Impact**				
	No	0.090	.048**				
	Yes	0.147	0.014				

Note: The analysis involved testing for whether treatment effects on marriage, cohabitation, doubling up, and births differed across each of seven characteristics in each of the four states—a total of 118 sets of interactions. The seven characteristics included education (<12, 12, >12 years) in addition to the six shown in the table. F-tests of global differences in impacts across subgroups were not statistically significant (p>.10) for the remaining 107 contrasts.

*** Point estimate differs from zero at the 99-percent confidence level; ** at the 95-percent level; * at the 90-percent level.

Table 4. Summary Statistics for First-Stage IV Models for Each Economic Regressor

Statistic	Employment	Earnings	Welfare Benefits
F	12.0	3.5	12.8
Partial R^2	.015	.004	.016
Shea Partial R^2			
Including earnings	.003	.001	.010
Excluding earnings	.013	---	.014

Table 5. Percentage Point Effects on Demographic Outcomes of Increases in Employment, Earnings, and Welfare Benefits (Standard Errors in Parentheses): Full Sample

	Standard Probit Estimates		IV Probit Estimates	
Outcome and Policy Change	**(1)**	**(2)**	**(3)**	**(4)**
Living with Spouse				
10-Point Increase in	-0.14	-0.26**	4.36	1.65
Quarterly Employment	(0.17)	(0.12)	(3.82)	(1.30)
$100 Increase in Average	-0.05		-1.43	
Quarterly Earnings	(0.05)		(1.87)	
$100 Increase in Average	-0.91***	-0.89***	-1.11	-0.51
Welfare Benefits	(0.09)	(0.09)	(1.08)	(0.72)
Living with Unmarried Partner				
10-Point Increase in	0.18	0.08	-2.41	-0.37
Quarterly Employment	(0.15)	(0.11)	(3.31)	(1.15)
$100 Increase in Average	-0.05		1.10	
Quarterly Earnings	(0.05)		(1.65)	
$100 Increase in Average	-0.11	-0.09	0.87	0.43
Welfare Benefits	(0.08)	(.08)	(0.93)	(0.63)
Living with Other Adult				
10-Point Increase in	-0.37**	-0.49***	-2.72	-2.28*
Quarterly Employment	(0.18)	(0.13)	(3.68)	(1.29)
$100 Increase in Average	-0.06		0.23	
Quarterly Earnings	(0.06)		(1.85)	
$100 Increase in Average	-0.21**	-0.18*	-0.56	-0.66
Welfare Benefits	(0.10)	(0.09)	(1.01)	(0.73)
Birth since Random Assignment				
10-Point Increase in	-0.51**	-0.76***	3.02	0.17
Quarterly Employment	(0.24)	(0.17)	(4.71)	(1.82)
$100 Increase in Average	-0.12		-1.53	
Quarterly Earnings	(0.08)		(2.33)	
$100 Increase in Average	0.17	0.23*	0.19	0.92
Welfare Benefits	(0.13)	(0.13)	(1.55)	(1.03)

Note: Results indicate the change in probabilities for each outcome associated with the specified increases in average employment, earnings, and benefits measured over the first two follow-up years. Excluded instruments are treatment status and interactions between treatment status and: state, new/ongoing applicant, Delaware*new/ongoing applicant, Indiana*new/ongoing applicant, and education. Models also control for: age, number of children, age of youngest child, ever married, years of school completed, new/ongoing applicant, and number of follow-up months from random assignment to survey interview. Probit estimates from Stata's dprobit and divprobit programs. Hausman tests of the corresponding robust estimates from Stata ivreg2 fail to reject the null hypothesis of no difference in estimates between Models 1 and 3 and between Models 2 and 4, except in for employment and marriage (t statistic=1.76 for Models 2 and 4). Hansens J statistics fail to reject independence between excluded instruments and errors in the second-stage family formation equations in all cases.

*** Statistically significant at the 99-percent level; ** at the 95-percent level; * at the 90-percent level.

Table 6. Percentage Point Effects on Demographic Outcomes of Increases in Employment, Earnings, and Welfare Benefits (Standard Errors in Parentheses): Sample Restricted to Delaware, Florida, and Indiana

Outcome and Policy Change	Standard Probit Estimates		IV Probit Estimates	
	(1)	(2)	(3)	(4)
Living with Spouse				
10-Point Increase in Quarterly Employment	-0.56**	-0.31**	5.72	-1.72
	(0.21)	(0.15)	(5.75)	(2.04)
$100 Increase in Average Quarterly Earnings	0.11*		-4.48	
	(0.07)		(3.03)	
$100 Increase in Average Welfare Benefits	-1.42***	-1.47***	-6.87**	-3.88*
	(0.19)	(0.19)	(3.37)	(2.09)
Living with Unmarried Partner				
10-Point Increase in Quarterly Employment	0.24	0.04	-2.32	0.24
	(0.18)	(0.13)	(3.51)	(1.62)
$100 Increase in Average Quarterly Earnings	-0.10		1.52	
	(0.06)		(1.82)	
$100 Increase in Average Welfare Benefits	-0.23	-0.19	0.81	-0.13
	(0.15)	(0.15)	(2.11)	(1.67)
Living with Other Adult				
10-Point Increase in Quarterly Employment	-0.49**	-0.60***	-1.97	-2.68
	(0.22)	(0.16)	(4.26)	(2.01)
$100 Increase in Average Quarterly Earnings	-0.05		-0.43	
	(0.07)		(2.24)	
$100 Increase in Average Welfare Benefits	-0.05	-0.03	-2.03	-1.81
	(0.19)	(0.19)	(2.39)	(2.06)
Birth since Random Assignment				
10-Point Increase in Quarterly Employment	-0.47	-0.67***	4.37	1.03
	(0.29)	(0.21)	(5.63)	(2.66)
$100 Increase in Average Quarterly Earnings	-0.10		-1.99	
	(0.10)		(2.94)	
$100 Increase in Average Welfare Benefits	0.79***	0.84***	2.86	4.01
	(0.25)	(0.24)	(3.20)	(2.61)

Note: Results indicate the change in probabilities for each outcome associated with the specified increases in average employment, earnings, and benefits measured over the first two follow-up years. Excluded instruments are treatment status and interactions between treatment status and state, new/ongoing applicant, and applicant status by Delaware and Florida. Models also control for: age, number of children, age of youngest child, ever married, years of school completed, new/ongoing applicant, and number of follow-up months from random assignment to survey interview. Probit estimates from Stata's dprobit and divprobit programs. Hausman tests of the corresponding robust estimates from Stata ivreg2 fail to reject the null hypothesis of no difference in estimates between Models 1 and 3 and between Models 2 and 4, except for earnings and marriage (in Models 1 and 3, t=1.89) and welfare benefits and marriage (in Models 1 and 3, t=1.92). The Hansen's J statistics fail to reject independence between excluded instruments and errors in the second-stage family formation equations in all cases.

*** Statistically significant at the 99-percent level; ** at the 95-percent level; * at the 90-percent level.