The Effects of Early Childhood Exposure to Welfare Reform on Adult Health and Employment*

Mehrdad Baratian
Tulane University

October 2026

Abstract

In models of child development, a child's later outcomes depend partly on family income and parental time. Welfare rules that require work trade one for the other, so theory does not say whether children gain or lose. I study the 1990s welfare reforms that imposed such rules, following young children into adulthood in the Panel Study of Income Dynamics. I compare children born in the same year across states whose rules differed in strictness, by their families' likelihood of receiving welfare. Among children the rules likely reached, those with more exposure to stricter rules before age five had better general and mental health as adults. But their adult employment did not change, and the estimates for their hours and earnings are imprecise. On net, my findings suggest that these children gained as adults from the trade-off between family income and parental time that the rules imposed.

JEL codes: I38, I14, J13, H53

1 Introduction

In economic models of child development, a young child's later outcomes depend partly on family income and parental time (Del Boca et al., 2014; Bernal and Keane, 2011; Mullins, 2026). Parental employment raises income and costs time, so theory does not say whether a parent's work leaves a young child better or worse off (Del Boca et al., 2014). When a welfare rule requires work, every parent it covers must make that choice. A parent who complies gains earnings and loses time at home, and one who does not keeps the time and

*Latest version: `https://m-baratian.github.io/Baratian_JMP.pdf`

loses the aid. The welfare reform of the 1990s tied cash aid for low-income families to work. This paper measures the net effect of the income gained and the time lost on the young children in those families. It follows those children into adulthood and asks whether more years under stricter rules changed their health and their employment as adults.

The evidence so far points one way for income and both ways for time. More family income raises children's test scores and their health as adults (Agostinelli and Sorrenti, 2021; Aizer et al., 2016; Hoynes et al., 2016). In contrast, less time with a parent can help or harm a child, because the effect depends on the care that replaces that time (Bernal and Keane, 2011; Bailey et al., 2021; Baker et al., 2019; Duncan et al., 2022). Most of these findings come from programs that gave families more resources, such as food stamps, cash transfers and Head Start (Agostinelli and Sorrenti, 2021; Hoynes et al., 2016; Aizer et al., 2016; Bailey et al., 2021; Baker et al., 2019; Duncan et al., 2022). But the welfare reform of the 1990s was a contraction that ended the cash entitlement and tied what remained to work. Most studies of what the reform did to children measure its effects in childhood, and they find harm (Leonard and Mas, 2008; Herbst, 2017; Kalil et al., 2023; Dave et al., 2021; Paxson and Waldfogel, 2003). By contrast, few studies follow the reform's children into adulthood, and their findings are mixed (Hartley et al., 2022; Corman et al., 2022; Dore et al., 2025). Of these studies, Dore et al. (2025) follows children exposed before age five in the Panel Study of Income Dynamics (PSID). Among those children it finds more psychological distress and chronic conditions in adulthood. However, the studies that follow the children into adulthood measure a child's exposure only by time under reform (Hartley et al., 2022; Corman et al., 2022; Dore et al., 2025), so they treat a state that only raised its earnings disregard the same as a state that imposed full-family sanctions. It is therefore not known whether a child's adult outcomes depend on how strict a state's work rules were, and not only on how long they applied.

In this paper, I estimate how the strictness of a state's work rules before a child's fifth birthday changed her health and her employment as an adult. In every state and year, I score five rules that govern work and combine the scores into one strictness index. Each rule's weight in the index is its effect on parental employment in earlier experimental and quasi-experimental studies. To measure a child's exposure, I add up the years she spent under reform before age five and weight each year by the index in her birth state. Because the rules reached a child only through her family, I multiply each child's exposure by the predicted probability that her family was on welfare. Then I compare children born in the same year in states that reformed in the same year. These children spent the same years under reform, so the comparison is between stricter and less strict states. The children in this comparison are those born between 1975 and 1999 whom the PSID follows into adulthood.

For these children, I measure health from age 18 and employment from age 24. I find that, where the rules were stricter, children whose families were likely to be on welfare had better general and mental health as adults. However, I find no effect of the stricter rules on the employment of these children as adults.

Work rules like those studied here are now the standard form of welfare in the United States. The reform of the 1990s is where they began. Between 1993 and 1998 every state replaced Aid to Families with Dependent Children (AFDC), an entitlement to cash for poor families with children, with aid that a parent had to work for and that ran out after a fixed number of months. States moved first under waivers from the federal program and then under Temporary Assistance for Needy Families (TANF), which Congress created in 1996. As the rules changed, welfare participation among single mothers fell by 23 percentage points between 1993 and 2002 and their employment rose by 11.3 percentage points (Fang and Keane, 2004). Studies that estimate the effect of the reform itself find the same move into work among less-educated single mothers and a rise in their earnings and income (Kaushal and Kaestner, 2001; Bollinger et al., 2009). In my tabulations of the Current Population Survey (CPS), the employment of less-educated single mothers rose from 52 to 67 percent between 1993 and 1999, and their annual hours of paid work rose from 989 to 1,268 (Figure 1). They were the group most likely to receive welfare, while only 1 percent of higher-educated married mothers had welfare income. For comparison, the employment of these married mothers rose from 73 to 75 percent over the same years, and their annual hours of paid work from 1,306 to 1,355. Because a mother's day has a fixed number of hours, each hour that moves into paid work is an hour taken from time at home, including time caring for her child. On the children's side, nearly half of the children on welfare when the rules changed were under six (Herbst, 2017). For them the reform came in the years when a family's cash matters most for later health (Hoynes et al., 2016) and when a parent's time matters most for a child's development (Del Boca et al., 2014).

Three problems make the effect of the reform on children hard to estimate. First, no state kept the old welfare rules to serve as a comparison. Second, a reform date does not show how strict a state's rules were. Third, the rules reached only children whose families were on welfare. In the preferred specification, I address the first problem by comparing children born in the same year in states that reformed in the same year, so the comparison needs no state that kept the old rules. These children spent the same years under reform, so the preferred estimates compare stricter with less strict rules, not reform with no reform. To address the second problem, I score each rule that governs work by how hard it pushed parents toward work and combine the scores into one state-year strictness index. Each rule's score runs year by year from 0 at the pre-reform program to 1 at the strictest setting any state

used. Each rule's weight in the index is its effect on parental employment in experimental and quasi-experimental studies (see Appendix A; Miller et al., 2000; Smedslund et al., 2006; Grogger, 2003; Fang and Keane, 2004; Looney, 2005). To address the third problem, I compare children with a higher and a lower predicted probability of being on welfare and ask whether the gap in their adult health and employment was larger in states with stricter rules. Throughout, a child's exposure counts the years she spent under reform from conception to age five, because both the literature (Hoynes et al., 2016) and my own analysis (Section 6) show that window to be the critical period for adult health. Each year a child spent under reform before age five is weighted by the value of the index in her birth state in that year. The weighted total is the child's strictness-adjusted years of exposure. The preferred estimates quoted below are per strictness-adjusted year. One such year is one year with an index of 1 or two years with an index of 0.5.

I multiply each child's exposure by the predicted probability that her family was on welfare. To predict this probability, I fit a gradient-boosted trees model to families in the March CPS before reform and apply it to the characteristics each PSID child's family had before she turned five. From age 18, the PSID children born between 1975 and 1999 contribute about 60,000 person-waves. I measure their adult outcomes with three indices, for general health, mental health and employment, and on each index a higher value is better.

I estimate four difference-in-differences specifications. In all four, the first difference is between children with a higher and a lower predicted probability of being on welfare, and the second is between children with more and less exposure to reform, as each specification measures it. In the first specification, exposure is the years a child spent under any reform, no matter how strict the rules were, and I call this raw exposure. The second specification measures exposure instead in strictness-adjusted years. The preferred specification is the third of the four. It keeps strictness-adjusted exposure and adds fixed effects for each birth year in each reform wave, so each child is compared only with children born in the same year in states that reformed in the same year. The fourth specification is the same as the third, except that it breaks the strictness index into its five rules. Each rule comes with its own exposure.

In the preferred specification, stricter rules left children with a high predicted probability of being on welfare healthier as adults. For a child whose family was certain to be on welfare, each strictness-adjusted year of exposure raises the general health index by 0.104 standard deviations and the mental health index by 0.158. Both health estimates are significant at the 10 percent level. Across the specifications, the general health estimate is 0.061 with raw exposure in the first, 0.090 with strictness-adjusted exposure in the second and 0.104 in the preferred third. In contrast, the employment index shows no effect for these children. Of its

three parts, the share of adults employed does not move, and the estimates for hours and earnings are too imprecise to say whether they move. However, in the fourth specification (Appendix B), a rule letting working families keep more of their grant raised the employment index, while a rule requiring work from parents of younger children reduced it. These two rule effects offset each other when the five rules are combined in the strictness index, and this offset is part of why the employment index shows no effect.

These results are relevant today because work requirements are being extended to programs that did not have them and to parents who were previously excused. The 2025 reconciliation law attaches one to Medicaid (Congressional Research Service, 2025b) and widens the one in the Supplemental Nutrition Assistance Program (Congressional Research Service, 2025a). The results show that the welfare reform rules that govern work did not harm the children they reached and improved their adult health.

The paper makes two contributions to existing work. The first is to measure the net outcome of the trade-off between the earnings a parent's work adds and the time it takes from a young child, using the welfare reform of the 1990s. On the income side of that trade-off, earlier work finds long-run gains from expansions that gave poor families food or health insurance (Hoynes et al., 2016; Bailey et al., 2024; Boudreaux et al., 2016; Goodman-Bacon, 2021b; Miller and Wherry, 2019). On the same side, Dustmann et al. (2024) find lower schooling and earnings in adulthood among the children of refugee families in Denmark after a cut of about 30 percent in their transfers. On the time side, earlier work finds that the care replacing a parent's time decides the effect on the child, with poor substitutes lowering children's outcomes and good ones more than making up for the lost time (Bernal and Keane, 2011; Baker et al., 2019; Bailey et al., 2021; Duncan et al., 2022). The reform studied here moved both sides of the trade-off at once, because its work rules raised earnings and took a parent's time. The second contribution is a measure of how strict each state's rules were, whereas the studies of this reform's children in adulthood measure only when the rules began (Corman et al., 2022; Dore et al., 2025; Hartley et al., 2022). The measure dates a state's reform from the first year a work rule was in force statewide. Beyond the date, the measure scores how strict each rule that governs work was, so two states that reformed in the same year can receive different scores and therefore different treatment. The five scores form one state-year strictness index, and that index can be attached to any child-level data that record a birth state and a birth year.

2 Policy Background

From 1935 to 1996, Aid to Families with Dependent Children (AFDC) paid a monthly cash grant to low-income families with children, and each state set the size of that grant. In 1994 the maximum grant for a family of three with no other income was $366 a month in the median state, $120 in Mississippi and $923 in Alaska.[1] That year AFDC paid a grant to 5.0 million families in an average month,[2] and in 1996 one single mother in four was in a family on AFDC (Fang and Keane, 2004). In contrast, federal law set who qualified and what a parent had to do to keep the grant. A family qualified if it had a child under 18 whose parent was absent, incapacitated, deceased or unemployed, and if its income was below the state's limit and its assets at most $1,000 (U.S. House Committee on Ways and Means, 1996, pp. 384–394). From 1988 the federal Job Opportunities and Basic Skills (JOBS) program offered mostly education and training to build parents' skills, and a parent had to take part in JOBS if the state called her (U.S. House Committee on Ways and Means, 1996, p. 411). The state could not call a parent whose youngest child was under 3, or a parent in a few smaller exempt groups,[3] and in 1994 the parents of a child under 3 alone headed 40.5 percent of all the families on AFDC (U.S. House Committee on Ways and Means, 1996, p. 411). A state had to call the other parents only if it had a place for them in JOBS (U.S. House Committee on Ways and Means, 1996, p. 411), and the number of places was limited because federal funds for JOBS were capped at about $1 billion a year (U.S. House Committee on Ways and Means, 1996). In addition, the law set a minimum participation rate, and in 1994 a state had to have at least 15 percent of the adults who were not exempt in an activity (U.S. House Committee on Ways and Means, 1996, p. 412). The exempt groups together were about half of all the adults on AFDC, so that minimum came to about 7 percent of all the adults on AFDC (U.S. House Committee on Ways and Means, 1996, p. 422). A state that missed the minimum could lose part of its federal matching funds, but the law let the penalty be waived, and by 1996 the penalty had been waived every time (U.S. House Committee on Ways and Means, 1996, p. 412). In practice, states reported that in 1994, 20.4 percent of the families on AFDC had a member in a JOBS activity (Blank, 2002). An activity could be schooling or job search rather than work (U.S. House Committee on Ways and Means,

[1] University of Kentucky Center for Poverty Research National Welfare Data, in 1994 dollars; the median is over the fifty states and the District of Columbia.

[2] Author's tabulation of the University of Kentucky Center for Poverty Research National Welfare Data, summed over the fifty states and the District of Columbia.

[3] The other exempt groups were parents who were ill, incapacitated or of advanced age, those needed at home to care for an ill family member, those already working 30 hours a week, pregnant women from the second trimester, children under 16 or in school full time, and recipients living where the state ran no program (U.S. House Committee on Ways and Means, 1996, p. 411).

1996, p. 411). So the work requirement bound few of the parents on AFDC, both in law and in practice.

Those federal rules were the same in every state until states began to obtain waivers. Under Section 1115 of the Social Security Act, a state could apply to the Department of Health and Human Services for a waiver, a permission to set aside parts of the AFDC rules (Blank, 2002). The Clinton administration encouraged waivers that applied statewide (Blank, 2002). By August 1996, 27 states had such a waiver in place (Blank, 2002). Most of these waivers made the work requirement stricter (Blank, 2002). Instead, some let a working family keep more of its earnings before the grant was reduced (Bitler et al., 2005). In August 1996 the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) replaced AFDC with Temporary Assistance for Needy Families (TANF). Twenty-four states started their TANF programs by the end of 1996, and 25 more states and the District of Columbia started theirs during 1997.[4] California started its program last, in January 1998. TANF ended the entitlement to aid, so a state could choose which families it supported (Blank, 2002). In addition, the federal government paid each state a fixed block grant in place of the matching funds it had paid under AFDC (Blank, 2002). The block grant did not fall when the number of families on aid fell, from 5.0 million in 1994 to 2.2 million in 2000.[5] With fewer families to pay, the states needed less of the block grant for cash. The states spent the savings on services instead. Cash payments took 75 percent of spending on family welfare under AFDC, and 40 percent under TANF in fiscal year 2001 (Burke, 2003). Child care was the largest of those services (Burke, 2003), so part of what the reform spent went to the care of children whose parents it had moved into work. Besides the block grant, TANF had two federal rules, and every other rule was the state's to set. No family could receive federally funded aid for more than 60 months in a lifetime, although a state could exempt up to a fifth of its caseload from this limit or set a shorter one (Blank, 2002). A further rule required a state to have half of its adult recipients in work activities for about 30 hours a week (Blank, 2002; Burke, 2003).[6] A state that fell short of that half lost part of its block grant (Burke, 2003). Schooling could not count toward the first 20 of the 30 hours (Blank, 2002).

The strictness index of the introduction is built from the work requirement, the exemption age, the sanction, the earnings disregard and the time limit. Each of these five rules had

[4]The waiver and TANF dates, and the rule settings below, are from the summaries of the Office of the Assistant Secretary for Planning and Evaluation in the Department of Health and Human Services and from the Urban Institute's Welfare Rules Database, described in Section 3.

[5]Average monthly caseload, University of Kentucky Center for Poverty Research National Welfare Data, summed over the fifty states and the District of Columbia.

[6]The statutory rate was 50 percent by 2002. A state's required rate was reduced by the fall in its caseload since 1995, so the effective rate was lower (Burke, 2003).

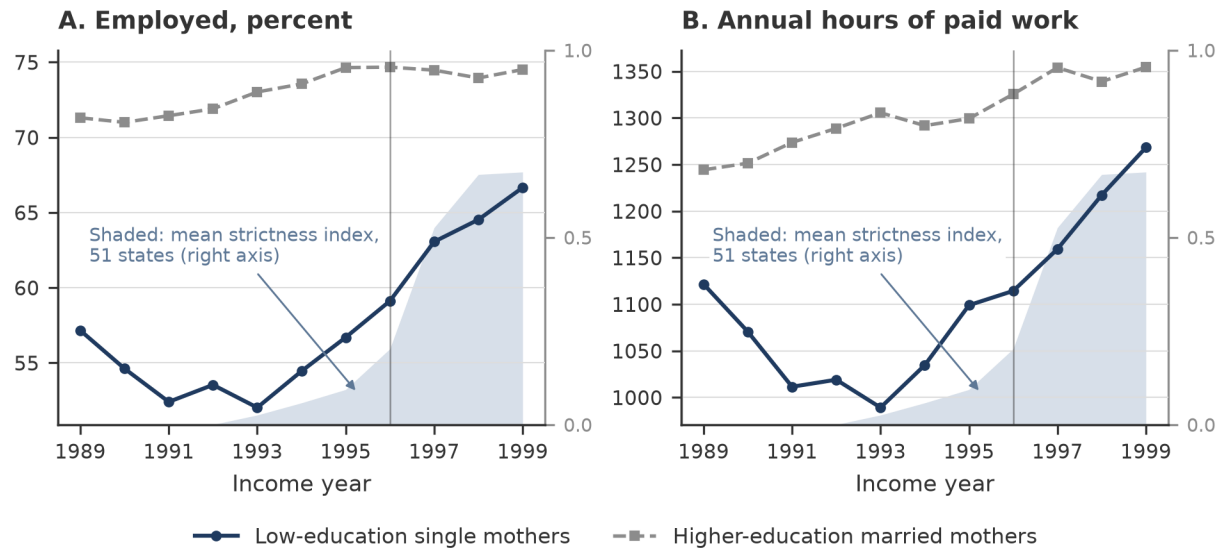


Figure 1: Employment and annual hours of mothers by likelihood of welfare receipt, 1989 to 1999

Notes: The data are the March Current Population Survey, income years 1989 to 1999, weighted by the survey's sample weights. Panel A is the share employed, measured as status in the survey week and plotted against the preceding income year. Panel B is annual hours, usual weekly hours times weeks worked in the income year, and zero for a mother who did not work. Mothers are women aged 18 to 54 with at least one child. The solid line is low-education single mothers, those with a high school diploma or less who are not married with a spouse present, 29.3 percent of whom had cash welfare income over income years 1989 to 1999. The dashed line is higher-education married mothers, 0.8 percent of whom did. The shaded area, on the right axis, is the mean strictness index across the 51 jurisdictions, and the vertical line marks 1996, the year PRWORA passed.

one setting under AFDC and a range the states could choose after reform. Table 1 lists each rule's AFDC setting, the change the paper counts as a move toward work, and the states' settings in 1998. The work requirement under AFDC was JOBS, and a parent had to take part in it if the state called her (U.S. House Committee on Ways and Means, 1996, p. 411). Under TANF every state required recipients to work or to look for work as a condition of the grant (Blank, 2002). Under AFDC a parent was exempt from JOBS until the youngest child turned three, so the state could not call her before then (U.S. House Committee on Ways and Means, 1996, p. 411). This exemption age was the state's to set under TANF, and a state could require work from the birth of the youngest child (Herbst, 2017). The sanction for a parent who did not comply removed the adult's share of the grant under AFDC and left the children's share in place (U.S. House Committee on Ways and Means, 1996, p. 524). A TANF state could remove the whole family's grant (Blank, 2002). The earnings disregard is the part of a family's earnings that did not reduce the grant. Under AFDC, in a family's first four months of work the disregard was $120 a month plus one third of the remaining

earnings. For the next eight months it was $120 a month (Bitler et al., 2006). After those twelve months only $90 was disregarded, so each dollar earned above $90 cut the grant by a dollar (Bitler et al., 2006). Under a waiver or TANF a state could disregard far more, and from 1996 Connecticut disregarded all earnings up to the federal poverty line (Bitler et al., 2006). AFDC set no time limit, so a family could receive aid until its youngest child turned 18 (U.S. House Committee on Ways and Means, 1996, p. 385). Under TANF the federal limit was 60 months, and a state could set a shorter one (Blank, 2002). Where a state moved a rule away from its AFDC setting, the move pushed parents toward work, by making four of the rules stricter and the disregard more generous (Bitler et al., 2005).

I give each state one reform year, dated from when its waiver or its TANF program took effect. A state's reform year is the first year in which one of the five rules was in force across the whole state at a setting different from the one AFDC had.[7] The setting must also have moved in the direction that pushed toward work. That direction is stricter for the work requirement, the exemption age, the sanction and the time limit, and more generous for the disregard, because a larger disregard raised what a working family kept and so made work pay (Bitler et al., 2005). A rule that took effect after June counts from the following year. For example, California's waiver raised the disregard in December 1992, so California is dated 1993. Because a state is dated only from the five rules, a waiver that only added a family cap does not start its reform year. A family cap denied a family an added grant for a child born while the family was on aid, so it changed the grant after a birth and not what the parent had to do to keep it. Georgia's waiver of 1994 was such a waiver, so Georgia is dated from its TANF program of January 1997. Nor does a waiver that ran in only a few counties start a state's reform year, because it did not reach all of the state's children. A state with such a waiver is dated from the year a rule applied statewide. The same rule dates every one of the states. Of the fifty states and the District of Columbia, three began in 1993, three in 1994 and three in 1995. Eight more states began in 1996, twenty-four states began in 1997, and the last ten began in 1998. Figure 2 plots these counts beside the years in which the waivers and the TANF programs took effect, and Figure D1 in Appendix D maps the reform year of each state. In these counts, every state dated before 1997 was dated from a waiver, because no TANF program could start before PRWORA passed in August 1996. The ten states dated 1998 started their TANF programs after June 1997.

By 1998 every state had reformed, so the rules in force in that year show how far each state had moved from the AFDC rules. All 51 states required recipients to work or look for work. On the exemption age, 27 states set it at 12 months and 19 states set it lower. Five of those 19 states set it at zero, so a parent was expected to work from the month the child was

[7]The dates are the implementation dates in the ASPE summaries, not the approval dates.

Table 1: The five rules under AFDC and the states' settings in 1998

Rule	Setting under AFDC	Move that counts as toward work	Settings across the 51 states in 1998
Work requirement	JOBS, mostly education and training. A parent had to take part if the state called her. The state could not call a parent of a child under 3, and it could call other parents only as far as its capped federal JOBS funds allowed. In 1994, 20.4 percent of families were in a JOBS activity	A statewide requirement to work or look for work as a condition of the grant	All 51 required work or job search
Exemption age	36 months	A lower age	12 months in 27 states; below 12 in 19 (zero in five); 18 months in two, 24 in one, 36 in New Hampshire, 48 in Texas
Sanction	Adult's share of the grant removed; children's share kept	The whole family's grant removed	Whole family's grant in 35 states; adult's share in 16
Earnings disregard	\$120 plus one third of the rest for four months, \$120 for eight more, \$90 thereafter	A larger disregard (earnings kept before the grant is reduced)	Larger than AFDC's in 31 states, of which 11 disregarded earnings up to the poverty line; AFDC's rule in 20
Time limit	None	A limit shorter than 60 months	Shorter than 60 months in 20 states (18 to 48 months); the federal 60 months alone in 31

Notes: The AFDC settings are the federal rules of 1994 to 1996 (U.S. House Committee on Ways and Means, 1996; Bitler et al., 2006). The share of families in a JOBS activity in 1994 is from Blank (2002). The 1998 settings are the author's tabulation of the state-by-year policy panel described in Section 3, built from the summaries of the Office of the Assistant Secretary for Planning and Evaluation and the Urban Institute's Welfare Rules Database, over the fifty states and the District of Columbia. A move in the direction shown starts a state's reform year and raises its score on the strictness index; the exemption-age settings above 36 months (Texas) do not count as a move. Twenty-two states also had a family cap in 1998, which is not one of the five rules and does not enter the index.

born.[8] The remaining five states set it above 12 months. Two states set it at 18 months and

[8]The 1998 counts in this paragraph are the author's tabulation of the state-by-year policy panel described in Section 3.

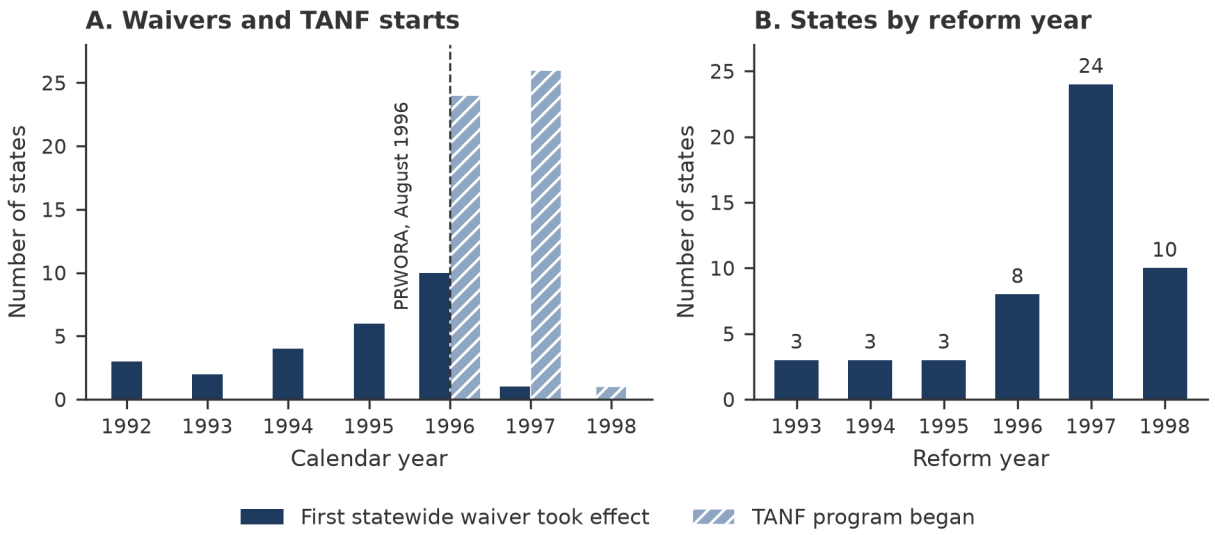


Figure 2: Waivers, TANF starts and reform years, 1992 to 1998

Notes: Panel A counts, by calendar year, the states whose first statewide waiver took effect and the states whose TANF program began; the dashed line marks August 1996, when PRWORA passed. Panel B counts states by the reform year this paper assigns, the first year in which one of the five rules was in force statewide at a setting that moved toward work, with a change after June dated to the following year. Both panels cover the fifty states and the District of Columbia. Dates are the implementation dates in the summaries of the Office of the Assistant Secretary for Planning and Evaluation.

one state at 24, New Hampshire kept the AFDC age of 36 months, and Texas raised it to 48. Neither New Hampshire's setting nor Texas's is a move toward work, so both states are dated from the first of their other rules that moved toward work. On the sanction, 35 states took the whole family's grant when a parent did not meet the work requirement, and 16 kept the AFDC rule of removing only the adult's share. On the disregard, 31 states disregarded more earnings than the $90 that AFDC allowed after twelve months of work, and in 11 of them a family of three kept part of its grant until its earnings reached the federal poverty line. The other 20 states kept the AFDC disregard rule. On the time limit, 20 states set a limit shorter than the federal 60 months, from Delaware's 48 down to Tennessee's 18, and the other 31 states applied the federal limit alone. So two states with the same reform year could differ in which rules had moved and in how far each rule had moved, because a state's reform year is the year in which its first rule moved toward work.

Apart from the cash grant, the reform left most of a family's other aid in place. PRWORA required states to keep Medicaid for any family that still met the AFDC eligibility rules of 1996, so losing the grant did not by itself end a family's eligibility for Medicaid (Bitler et al., 2005). Food stamp rules for families with children did not change (Bitler and Hoynes, 2016). So for a family on welfare the reform changed the rules that tied its grant to work, and little else. States changed those rules in different years, and how far each of the five rules had

moved from the AFDC setting differed across states and from year to year. In Section 3 I score each rule from 0 at the AFDC setting to 1 at the setting that pushed hardest toward work, and combine the five scores into the strictness index. In Section 4 I compare children who were born in the same year and whose states reformed in the same year, so that what differs between them is the strictness index of their birth states in the years from conception to age five.

3 Data

3.1 The PSID and the sample

The comparison in Section 4 requires three things for each child: the state and month of birth, the characteristics of the child's family during the first five years of the child's life, and the child's health and employment as an adult. The Panel Study of Income Dynamics (PSID) records all three things for the same person. It began in 1968 with about 5,000 families, and about 1,900 of those families were drawn from a low-income sample. Since then it has interviewed the members of those families and their descendants in 43 waves through 2023. The low-income sample is an oversample, so the estimates use the PSID weights described in Section 4. A child's birth state is the state where the family lived when the child was born.[9] I measure a child's exposure to reform from the birth state, as Hoynes et al. (2016) do from the county of birth.

The sample is the 8,408 children born between 1975 and 1999 who descend from the PSID's original 1968 families and were observed both in childhood and at age 18 or older. The families the PSID added after 1968 to represent immigrants and Latinos are left out, because PRWORA cut the eligibility of many non-citizens for aid. That cut is a different change from the work rules this paper studies. A child must have been seen in an interviewed household before age 18, because the birth state and the family characteristics come from that childhood record.[10] The cohorts stop in 1999 because a child born later is not yet 24

[9]For the 5,956 children who were first seen in an interviewed household by age 1, the birth state is the state where the family lived at that interview. Of the other children, 945 take the state of birth a parent reports in the PSID's childbirth history. For the 1,409 children with both sources, the parent's report agrees with the state at the first interview for 94 percent. Another 1,504 children were first seen in an interviewed household after age 1 and have no parent's report, so they take the state where their original PSID family lived in 1969. Without these 1,504 children, the main estimates for the general and mental health indices in Section 5 move from 0.104 to 0.116 and from 0.158 to 0.149, and both remain significant at the same levels.

[10]The PSID has 25,942 members born between 1975 and 1999, and 19,614 of them descend from the original 1968 families. Of these members, 19,582 were born in the 50 states or the District of Columbia, and 14,020 of those were seen in an interviewed household before age 18. I keep the 8,408 of these children who are still in the panel at age 18 or older. To check whether the attrition between childhood and age 18 depends on the reform, I regress staying in the panel to age 18 on exposure to reform in early childhood

in the 2023 wave, and 24 is the age from which employment is measured. The cohorts born before 1988 turned five before the first state reformed in 1993, and I keep them because the birth-state effects, the survey-wave effects and the controls are more precisely estimated with them. The 8,408 children contribute 59,730 person-waves at age 18 or older. Of these person-waves, 36,315 are at age 24 or older. Early childhood here means the years from conception to age five. I call a child exposed if some part of early childhood fell after the birth state's reform. Exposed children contribute 12,612 of the 59,730 person-waves and 5,256 of the 36,315 at age 24 or older. Children whose early childhood fell entirely after the reform contribute 3,201 of the 12,612 exposed person-waves.

3.2 The strictness index

For each state and year, the strictness index measures how far the five rules described in Section 2 had moved from their AFDC settings toward work, on a scale from 0 to 1. It has one value for each of the 51 states and each year from 1992 to 2003. It is 0 in every state in 1992, because 1992 is the last year in which every state still ran the AFDC rules. A state's index stays at 0 until its reform year. From that year on, the index is a weighted average of the scores of the state's five rules in that year. The rules through 1998 come from the summaries of each state's waiver and TANF program published by the Office of the Assistant Secretary for Planning and Evaluation (ASPE). By 1998 every state had started its TANF program, and few states changed the five rules after that. From 1999 each state keeps its 1998 index value unless the Urban Institute's Welfare Rules Database records a change in one of the five rules. Between 1998 and 2003 the index rose by 0.03 in the average state, from 0.67 to 0.70, and 37 of the 51 states kept their 1998 value.

Each of the five rules is scored from 0 at its AFDC setting to 1 at the setting that pushed hardest toward work. The work requirement and the sanction are switches and score 0 or 1. The work requirement scores 1 in a year in which a statewide requirement to work or look for work was in force, and 0 under AFDC. For the sanction, the score is 1 where a parent who did not comply lost the whole family's grant and 0 where she lost only the adult's share. The other three rules are quantities, and each score rises in a straight line from its 0 setting to its 1 setting. The time limit scores 0 both with no limit and at the federal 60 months, because that limit applied in every state and no family could reach it before late 2001, five years after TANF began. Then the score rises in a straight line from 0 at 60 months to 1 at the 18 months Tennessee chose. The exemption age scores 0 at the 36 months AFDC allowed and 1 where a parent had to work whatever the age of the youngest child. The

among the 14,020 children, with birth-year and birth-state fixed effects. The coefficient on exposure is not statistically significant, so the attrition is not related to exposure.

Table 2: Consensus employment effects of the five rules and the weights they give

Rule	Employment effect (percentage points)	Weight	Studies
Work requirement	4.0	0.419	Fang and Keane (2004); Smedslund et al. (2006); Miller et al. (2000); Scrivener et al. (2002)
Earnings disregard	2.5	0.262	Miller et al. (2000); Matsudaira and Blank (2014)
Time limit	1.5	0.157	Grogger (2003); Fang and Keane (2004); Grogger and Michalopoulos (2003)
Sanction	1.2	0.126	Looney (2005); Fang and Keane (2004)
Exemption age	0.35	0.037	Pingle (2003); Looney (2005)
Sum	9.55	1.000	

Notes: Each effect is a consensus value, in percentage points, for the effect of the rule on the employment of parents on welfare, set from the experimental and quasi-experimental studies listed. Each weight is the effect divided by the sum of the five effects, 9.55, and rounded to three decimals. Appendix A gives each study's estimate and how the consensus value was set.

earnings disregard is scored by the breakeven point, the monthly earnings at which a family of three loses its grant (Bitler et al., 2006). Its score is 0 at the breakeven point under the state's own AFDC rules and 1 where the breakeven point reaches the federal poverty line. For every rule, a higher score means a stronger push toward work. The other four rules push harder when they become stricter, and the disregard pushes harder when it becomes more generous, because a larger disregard made work pay. For the other four rules, a setting more lenient than AFDC's scores 0, because the score measures only moves toward work. So the 48-month exemption age Texas set scores 0. Appendix A gives each formula with a worked example.

I combine the five scores into the strictness index with fixed weights that sum to one. For each rule, Table 2 lists the experimental and quasi-experimental studies of its effect on parental employment and a consensus effect that I set from those studies, as Appendix A explains. A rule's weight is its consensus effect divided by the sum of the five consensus effects. The effects are 4.0 percentage points for the work requirement, 2.5 for the earnings disregard, 1.5 for the time limit, 1.2 for the sanction and 0.35 for the exemption age. So the weights are 0.419, 0.262, 0.157, 0.126 and 0.037, in the same order. To estimate the effect of each rule on its own, I also enter the five rule scores separately in the model rather than combined into the index. Appendix Table B1 reports the estimate for each rule.

Figure 3 plots the index for each state and each year from 1992 to 1998, with the states ordered by reform year. Figure D2 in Appendix D maps each state's average of the same

index from its reform year through 1998. By 1998 every state had a work requirement, so the differences between states in 1998 come from the other four rules. In that year the index averaged 0.67 across the 51 states, with a standard deviation of 0.13, and it ran from 0.44 in Alaska to 0.98 in Connecticut. States that reformed in the same year differed almost as much, with a standard deviation of 0.12 within a reform year. Among the 24 states that reformed in 1997, the index ran from 0.44 to 0.88. That spread is the variation the preferred specification uses, because it compares children who were born in the same year and whose states reformed in the same year.

3.3 Exposure

For each child I take the 5.75 years from conception to the fifth birthday, with conception set at nine months before birth. Exposure is how many of those years fell on or after the birth state's reform year. The whole reform year counts, because the reform is dated by year and the index takes one value per year. Strictness-adjusted exposure is the sum of the years after the reform, each weighted by the index of the birth state in that year. In these units, one strictness-adjusted year is one year under an index of 1, or two years under an index of 0.5. Every estimate that uses the index is per strictness-adjusted year of exposure. On average, an exposed child spent 3.8 years under reform, and because the index in those years averaged 0.62, those years count as 2.4 strictness-adjusted years. A child conceived in or after the reform year spent all 5.75 years under reform, and those years count on average as 3.8 strictness-adjusted years.

3.4 Predicted welfare receipt

The rules reached a child only through a family that received AFDC, so I need to know how likely each child's family was to receive AFDC. Whether the family actually received AFDC cannot be used for this. Receipt is a choice, so families that received AFDC differ from those that did not in ways the data do not show (Hoynes and Schanzenbach, 2009, pp. 121–122). In addition, after reform a family's receipt was itself changed by the rules, so receipt measured then is an outcome of the treatment. To capture how likely each child was to be treated, Hoynes et al. (2016) multiply the food stamp treatment by the participation rate of the group the child's family belonged to, out of twelve groups defined by the parent's education, race and marital status. I follow the same approach, but I estimate a probability of AFDC receipt for each family rather than a rate for each group. This probability, P, is the chance that a family with the characteristics of the child's family received AFDC before any state reformed. In the estimating equation I interact each measure of a child's exposure

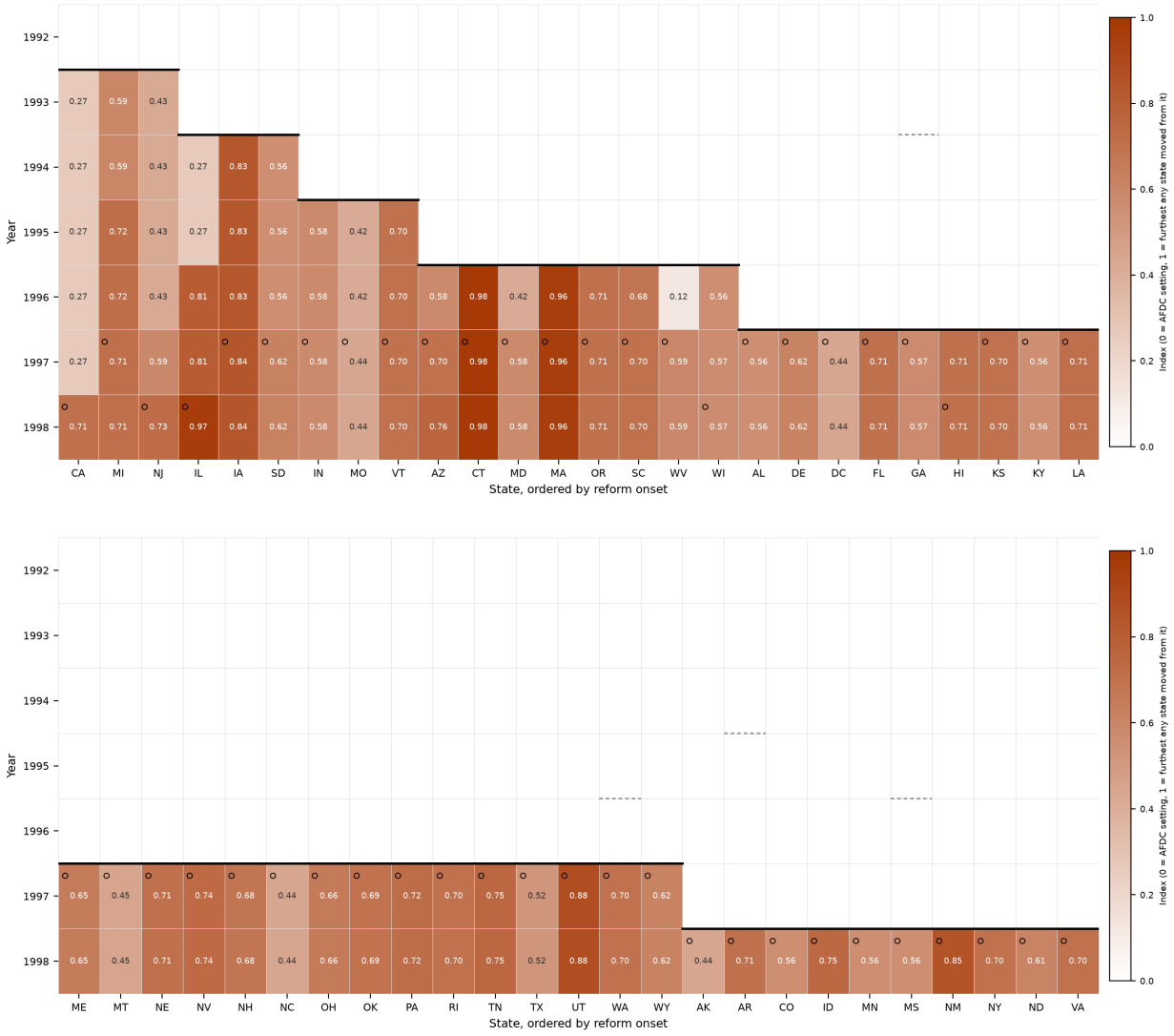


Figure 3: Strictness index by state and year, 1992 to 1998

Notes: The upper panel holds the 26 states that reformed earliest and the lower panel the remaining 25, each ordered by the year of reform. Rows are years and columns are states. The index combines five work rules with fixed weights, and is 0 at the AFDC setting and 1 at the setting that pushed hardest toward work; blank cells are 0. A black edge marks the first year in which one of the five rules had moved away from its AFDC setting across the whole state, a circle marks the first TANF year, and a grey dashed edge marks a state that left the AFDC rules earlier on a waiver that changed no work-related rule. The rules are from the summaries of the Office of the Assistant Secretary for Planning and Evaluation, for the fifty states and the District of Columbia.

with P.

P comes from eight characteristics of the family: whether the head finished 12 years of school, the head's race, age and marital status, the size of the family, the number of its children under 18, the number of other adults in it, and the state. Each is averaged over the interviews in the child's ages 0 to 5, so a category such as marital status enters as a

share of those interviews. Income, employment and hours are left out, because they are what the rules were meant to move. I fit a gradient-boosted trees model of AFDC receipt on the eight characteristics. The model is fitted to 102,884 families with a child under 6 in the March Current Population Survey (CPS) of 1984 to 1993. Those surveys cover income years 1983 to 1992, before any state reformed. I use the CPS because it has many more families observed before reform than the PSID. In addition, fitting the model on the CPS leaves the PSID families out of the fit, so they can be used to test it. To test the model, I score the PSID families interviewed between 1990 and 1993 with a child under 6 and compare each score with whether the family reported AFDC. The area under the receiver operating characteristic curve (AUC) for this test is 0.903. An AUC of 0.903 means that a family that received AFDC is scored above one that did not 90 percent of the time. On the scale of Hosmer et al. (2013, ch. 5), an AUC from 0.7 to 0.8 is acceptable, from 0.8 to 0.9 is excellent, and 0.9 or above, as here, is outstanding. Then I apply the model to every child in the sample, exposed or not. In the sample, P is 0.008 at the 10th percentile and 0.460 at the 90th. Through the interaction with P, the effect of exposure changes with P across this range, from families very unlikely to families fairly likely to be on welfare. Appendix E gives the details of the model.

3.5 Outcomes

I measure adult health from age 18 and employment from age 24. Hoynes et al. (2016) start their economic outcomes at 25 so that education is complete. By contrast, I use 24, because the last cohort in the sample was born in 1999 and reaches 24 in the 2023 wave. For general health the outcomes are self-rated health on the survey's five-point scale and not being obese, meaning a body-mass index below 30 from reported height and weight. The two mental health outcomes are the absence of a psychiatric diagnosis by a doctor and the six-item Kessler distress scale. The employment outcomes are being employed, annual hours of work and own labor earnings. Earnings enter in the inverse hyperbolic sine rather than the logarithm, because the logarithm is not defined at zero earnings. For earnings well above zero the two are close, so a coefficient on the inverse hyperbolic sine reads as an approximate percent change, as one on the logarithm does. Each outcome is oriented so that a higher value is better. Table 3 gives the survey question behind each outcome and how it is coded. General health, mental health and employment each form an index. Each index is the average of its components' z-scores, standardized on the children with no exposure (Kling et al., 2007). The main interview asks these health and employment questions only of the head of the household and the spouse, so for adult children aged 18 to 28 who still

Table 3: Survey items and coverage of the adult outcomes

Outcome	PSID item	Coding (higher is better)	Ages	First wave	Index
Self-rated health	General health status, excellent to poor (also asked in the TAS)	1 = poor to 5 = excellent	18+	1993	General health
Not obese	Height and weight, reported by the respondent (also in the TAS)	1 if body-mass index below 30	18+	1999	General health
No psychiatric diagnosis	Whether a doctor has ever said the respondent has an emotional, nervous or psychiatric problem	1 if not	18+	1999	Mental health
Kessler distress scale	Six items on the past 30 days: sadness, nervousness, restlessness, hopelessness, everything an effort, worthlessness; each from all of the time to none of the time	One minus the summed score (0 to 24) divided by 24	18+	2001	Mental health
Employed	Employment status, first mention (also in the TAS)	1 if working now	24+	1999	Employment
Annual hours	Hours worked in the prior year	Hours	24+	1999	Employment
Own labor earnings	Labor income in the prior year, wages and other labor income (also in the TAS from 2017)	Inverse hyperbolic sine	24+	1999	Employment

Notes: Items from the PSID main interview, asked of the household head and the spouse, and from the Transition into Adulthood Supplement (TAS), from 2005, for adult children aged 18 to 28 who still live with a parent. First wave is the first survey wave in which the outcome is observed at the stated ages for the 8,408 children of Section 3. Each index is the average of its components' z-scores, computed with the mean and standard deviation of the children with no exposure; where a component is missing, the index is the average of the components observed. Every person-wave, TAS answers included, carries the PSID individual longitudinal weight from the main file.

live with a parent the answers come from the Transition into Adulthood Supplement (TAS), fielded by the PSID since 2005.

3.6 Summary statistics

Table 4 compares the exposed children with the children with no exposure, where a child counts as exposed when part of the 5.75 years from conception to the fifth birthday fell on or after the birth state's reform year. The exposed were born twelve years later on average, because the first state reformed in 1993 and a child had to turn five in 1993 or later to be exposed. Their predicted probability of welfare receipt, P, averages 0.228, against 0.167 for the unexposed. On average the exposed spent 54 percent of their early childhood with a married mother, against 69 percent for the unexposed. Their adult outcomes are close to those of the unexposed, with self-rated health of 3.65 against 3.69 and employment at 24 or older of 75 against 78 percent. But the exposed are twelve years younger and were interviewed in later waves, so these comparisons mix the reform with age and calendar time. For that reason I compare children who were born in the same year and whose states reformed in the same year. Children in such a group spent the same years under reform, and they differ in P and in how strict their states' rules were.

4 Empirical Strategy

4.1 The estimate, built in four steps

I build the estimate in four steps, and each step adds one thing to the specification before it. The first measures reform only by the years a child spent under it. It compares the adult outcomes of children who spent more of their early childhood under reform with those who spent less, and asks whether that comparison looks different for children whose families were more likely to be on welfare. This makes each step a difference-in-differences, with the first difference between children with a high and a low predicted probability of being on welfare and the second between children with more and less exposure to the rules. For child i born in state s in year b and observed in survey wave t,

$$y_{isbt} = \beta_1 E_i + \beta_2 E_i P_i + \gamma P_i + \alpha_s + \lambda_b + \phi_{d(s)b} + \psi_t + X'_{it}\delta + \varepsilon_{isbt}. \tag{1}$$

Here y is an outcome from the list in Section 3, oriented so that a higher value is better. E_i is how many years between conception and the fifth birthday fell under any reform in the child's birth state, so it runs from 0 for a child who turned five before her state reformed to 5.75 for a child conceived after it. P_i is the predicted probability that a family like the child's received AFDC before any state reformed. β_1 is the effect of one more year of E_i on a child whose family had no predicted chance of receipt, and β_2 is how much larger or smaller

Table 4: Summary statistics by exposure to reform

	(1) Unexposed	(2) Exposed	(3) All
Panel A. Characteristics (one row per child)			
Birth year	1982.6	1994.7	1986.7
Predicted welfare receipt P	0.167	0.228	0.187
Parent with less than 12 years of school	0.237	0.265	0.247
Nonwhite	0.454	0.514	0.474
Male	0.502	0.517	0.507
Share of ages 0 to 5 with mother married	0.693	0.539	0.642
Family income to needs, early childhood	2.66	2.86	2.73
Years of exposure	0.00	3.80	1.26
Strictness-adjusted years of exposure	0.00	2.36	0.79
Children	5,609	2,799	8,408
Panel B. Health outcomes, age 18 or older			
Self-rated health (5 = excellent)	3.69	3.65	3.68
Not obese	0.694	0.722	0.700
Distress scale (1 = no distress)	0.824	0.783	0.817
No psychiatric diagnosis	0.877	0.849	0.873
Person-waves	47,118	12,612	59,730
Panel C. Employment outcomes, age 24 or older			
Employed	0.780	0.753	0.776
Annual hours	1,726	1,695	1,722
Own labor earnings, inverse hyperbolic sine	9.60	9.31	9.56
Person-waves	31,059	5,256	36,315

Notes: Means, unweighted. The sample is the 8,408 PSID children born 1975 to 1999 described in Section 3. A child is exposed when part of the 5.75 years from conception to the fifth birthday falls on or after the birth state's reform year. Years of exposure is the number of those years that fall on or after the reform year. Strictness-adjusted years weight each of those years by the strictness index in force. The index is 0 at the AFDC rules and 1 at the strictest setting any state used. Predicted welfare receipt P is the modelled probability that a family with the child's early-childhood characteristics received AFDC before any state reformed. Panel A has one row per child; family characteristics are measured over the child's ages 0 to 5, and Nonwhite and Male refer to the child. Panels B and C report person-waves rather than children, pooling every survey wave in which the person is observed at the stated age, from the core interview or the Transition into Adulthood Supplement, so a person contributes as many rows as the waves in which she is observed. Every outcome is oriented so that a higher value is better. The distress scale is one minus the six-item Kessler score divided by 24. Earnings enter in the inverse hyperbolic sine, the form used in Section 5.

that effect is for a child whose family was certain to be on welfare. β_2 is the coefficient of interest, and it is an intention-to-treat effect, since P is a probability and not the family's actual receipt. A family with a high P that never took up aid counts as exposed all the same, so the estimate is smaller in size than the effect on the families the rules actually reached.

The regression includes birth-state fixed effects α_s, birth-year fixed effects λ_b, census-division by birth-year fixed effects $\phi_{d(s)b}$ and survey-wave fixed effects ψ_t, where $d(s)$ is the census division of state s. The birth-state effects adjust for each state's level, and the census-division by birth-year effects account for differences in cohort trends across census divisions, so a child is compared only with the other children of her division born in the same year. Besides these fixed effects, the regression includes the main effect of P_i, and its coefficient γ is the difference in adult outcomes between children of likely-recipient families and the rest when exposure is zero. In addition, the controls X_{it} are the child's sex, race, age and age squared, and the log of the family's income-to-needs ratio over the same years from conception to the fifth birthday. These controls have coefficients δ, and ε_{isbt} is the error term.

Within a division and a birth year, β_2 can use every way in which exposure and the chance of receipt differ between children. Exposure differs between the states of a division because they reformed in different years, and within a state it differs across cohorts because a later cohort spent more of its window under the rules. The chance of receipt differs in a third way, between families within a state and cohort. β_2 draws on all three at once, through how the gap in adult outcomes between likely-recipient children and the rest moves with exposure across the states and cohorts of a division.

What this first specification cannot do is tell one reform from another, because E_i counts a year the same whether the state required a parent to work or only raised the earnings disregard. Corman et al. (2022) and Dore et al. (2025) measure exposure this way, from the date a state's reform began, and their estimates are therefore of reform on average, across whatever the states chose to do.

The second step puts what the rules required into the exposure measure and changes nothing else. The exposure measure is now strictness-adjusted exposure, in which the exposure is weighted by the strictness index of the birth state in each year, and the index is zero in AFDC years by definition. W_i is this weighted exposure measured in years, so it runs from 0 for a child never under reform to 5.75 for a child under the strictest rules for the whole window. W_i replaces E_i in equation (1), so a unit of exposure is now one year under the strictest rules rather than one year under any reform. As in the first step, a child is compared with the other children of her division born in the same year.

That comparison has a cost when effects build over time, because a state that reformed

in 1993 is a comparison for a state in its division that reformed in 1998. With staggered adoption, a pooled difference-in-differences estimate is a weighted average of all such comparisons and can be biased when treatment effects change over time (Goodman-Bacon, 2021a). The estimators of Callaway and Sant'Anna (2021) and de Chaisemartin and D'Haultfœuille (2020) avoid this bias by comparing treated states only with states not yet treated, but this sample has none for the cohorts born from 1993 on. Every state reformed between 1993 and 1998, so every state reformed during the early childhood of these cohorts. These cohorts hold 77 percent of the exposed children. Therefore, the third step removes the comparisons between early and late reformers rather than correcting them. It replaces the birth-year effects λ_b with reform-wave by birth-year effects $\lambda_{w(s)b}$, where $w(s)$ is the year in which state s reformed. From here on, a child is compared only with children born in the same year in states that reformed in the same year as hers. For the same child i, the third step is

$$y_{isbt} = \beta_1 W_i + \beta_2 W_i P_i + \gamma P_i + \alpha_s + \lambda_{w(s)b} + \phi_{d(s)b} + \psi_t + X'_{it}\delta + \varepsilon_{isbt}. \tag{2}$$

Within a wave, states share the year of reform, so an early reformer is never a comparison for a late one. Goodman-Bacon (2021b) uses Medicaid-year-by-cohort fixed effects for the same purpose in his study of Medicaid. I take this third step as the preferred specification, because the year a state reformed can no longer enter the comparison, and every number the paper quotes comes from it.

What remains to identify β_2 is the difference in P across families and the difference in strictness-adjusted exposure W across children. Inside a wave, the states reformed in the same year, so for any cohort their children spent the same years under reform, and W differs between them only because their rules differed. A stricter state has a higher W for the same years. β_1 is the effect of W on a child whose family had no predicted chance of receipt, and β_2 is how much larger that effect is for a child whose family was certain to be on welfare. β_2 measures how much more a stricter state moved the adult outcomes of children of likely-recipient families, among children born in the same year in states that reformed in the same year. So a shock that coincided with a state's reform and reached all its children alike enters β_1 but not β_2. By contrast, the only shock that can reach β_2 is one that moved likely-recipient families relative to the rest, in step with the strictness of the state's rules. The identifying assumption is that no shock of that kind occurred. That is, among states that reformed in the same year, the gap in adult outcomes between children of likely-recipient families and other children born in the same year would have moved across cohorts in the same way in every state had their rules not differed. It would fail if something else that varied across states with the strictness of reform reached likely-recipient families alone, such as a state

EITC adopted in the same years.

The reform years also fix which cohorts carry exposure. It begins with the 1988 cohort, the first that can carry any, because the first state reformed in 1993 and the window reaches age five. The sample reaches back to the 1975 cohort, so the thirteen cohorts born before 1988 carry no exposure and do not identify β_2, and I keep them because the fixed effects and the controls are more precisely estimated with them than without.

The fourth step keeps everything in equation (2) and replaces W_i with the five rules that make up the index, entered together, each with its own main effect and its own interaction with P_i. With W_i^k the years of the window weighted by rule k's own 0 to 1 score, in the same units as W_i, the fourth step is

$$y_{isbt} = \sum_{k=1}^{5} \left(\beta_1^k W_i^k + \beta_2^k W_i^k P_i\right) + \gamma P_i + \alpha_s + \lambda_{w(s)b} + \phi_{d(s)b} + \psi_t + X_{it}'\delta + \varepsilon_{isbt}. \qquad (3)$$

β_2^k is the effect of one year under rule k at its strictest setting for a child with $P = 1$ relative to a child with $P = 0$, with the other four rules held fixed. Nothing about the comparison changes at this step. Appendix B reports the estimates of β_2^k to show which of the five rules the result comes from.

4.2 What the tables report

Tables 5 to 7 present the results for general health, mental health and employment, with one column for each of the first three steps. Each column reports β_2, the effect of one year of exposure for a child with $P = 1$ relative to a child with $P = 0$. In column 1 the year is a year under any reform, and in columns 2 and 3 it is a year under the strictest rules.

4.3 Inference

Standard errors are clustered on the birth state, because the treatment varies by state and cohort and errors can be correlated within a state across its cohorts and survey waves. That leaves only 51 clusters, the 50 states and the District of Columbia, so cluster-robust p-values can overstate precision, and for β_2 in the first three steps I therefore report p-values from a restricted wild-cluster bootstrap with 999 Rademacher draws, in which the null is imposed (Cameron et al., 2008). The fourth step carries cluster-robust p-values alone, because the bootstrap is run one coefficient at a time and that step has five interactions. Every regression also carries the PSID's individual longitudinal weight. The weight corrects both the survey's oversample of low-income households and the families it loses as the panel runs on.

4.4 Validity of the design

The threat the identifying assumption leaves standing is a shock to a state and cohort that reached likely-recipient families alone and coincided with the state's reform. The census-division by birth-year effects take out whatever part of such a shock was common to a division's states. What survives is a shock to one state's likely-recipient families that moved with that state's reform and not with its division's. The state EITC named in Section 4.1 is a shock of this kind. I test for it in Section 6 by dropping the children whose window from conception to age five could include a year with a refundable state EITC that started in 1993 or later. If the estimates hold without those children, a state EITC did not produce them.

The same regression gives a second check in the coefficient on exposure alone, β_1. This coefficient is the effect of exposure on children whose families had no predicted chance of receipt. The welfare rules paid those families nothing, so if reform reached their children in no other way, such as through the local labour market, β_1 should be zero. If β_1 is zero, stricter rules did not coincide with a shock that reached all families in the state, whether or not they could receive welfare. But β_1 cannot detect a shock that reached only likely-recipient families, because it measures the effect on the children of the other families. The state EITC test described above is the check on a shock of that kind. Hoynes et al. (2016) include the same kind of coefficient in their own model. There it is the effect of food stamps for a group whose participation rate is zero, and they expect it to be zero. Section 6 reports β_1 for the ten outcomes as a placebo test of the design.

5 Results

Tables 5 to 7 report the estimates for the general health, mental health and employment indices and their components. Each table has one block for each outcome and one column for each of the first three specifications of Section 4. Within a block the first row is β_2, the coefficient on exposure interacted with P, with its clustered standard error and its wild-cluster bootstrap p-value. The second row is β_1, the coefficient on exposure alone, and Section 6 uses it as a placebo test of the design. In column 1 exposure is counted in years under any reform, and columns 2 and 3 use strictness-adjusted years instead. In each column β_2 is the effect of one more year of that column's exposure for a child with $P = 1$ against one with $P = 0$. Column 3 is the preferred specification, and in it children are compared only with children of the same birth year in states that reformed in the same year. Those children have the same years under reform, so the estimate comes from how strict their states' rules

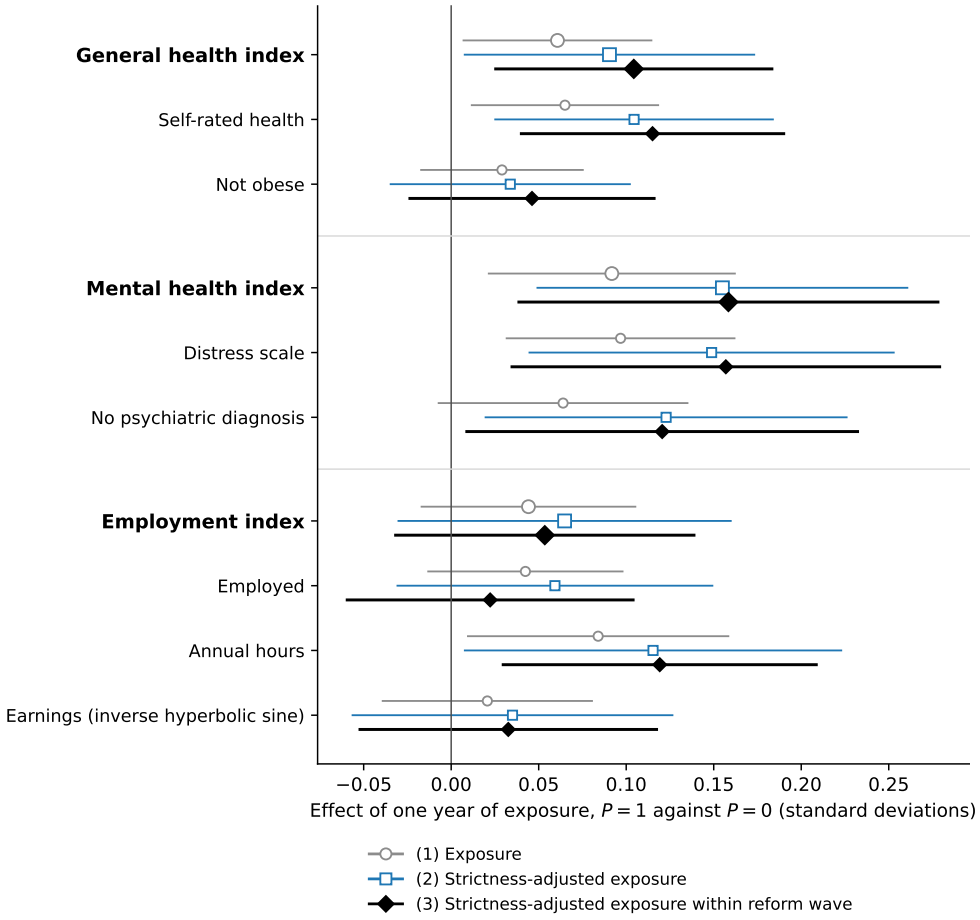


Figure 4: Effects of exposure on every outcome under the first three specifications

Notes: Each marker is β_2 from Tables 5 to 7, the effect of one more year of exposure for a child with $P = 1$ against one with $P = 0$. The year is a year under any reform in the first specification and a strictness-adjusted year in the second and third. Each component is divided by its standard deviation among unexposed children. Bars are 90 percent confidence intervals from standard errors clustered on the birth state. The data are the PSID cohorts born 1975 to 1999, and the sample and controls are those of Tables 5 to 7.

were and from their families' P. Every number in the abstract and the introduction comes from that column. Appendix Table B1 reports a fourth specification that gives each of the five rules its own exposure. Figure 4 plots the estimates of the first three specifications for the three indices and their seven components, each with its 90 percent confidence interval. The confidence intervals in this section's figures are built from the clustered standard errors, so an interval can exclude zero when the bootstrap p-value is above 0.10. In Figure 4 the components are in standard deviations of the unexposed children, so every outcome is on the scale of the indices.

Table 5 presents the estimates for the general health index and its two components, self-rated health and obesity. Under the preferred specification in column 3, each strictness-

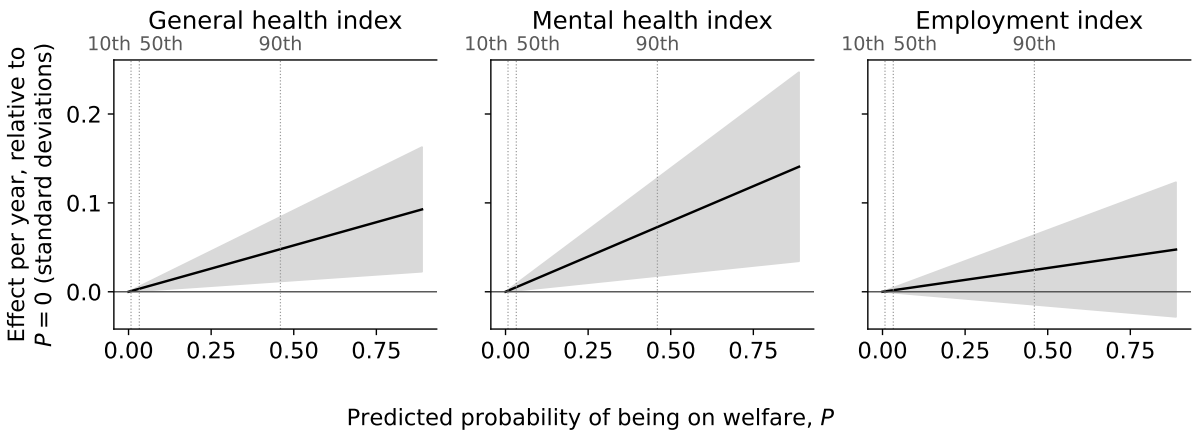


Figure 5: Effect of exposure across the predicted probability of welfare receipt

Notes: The line is $\beta_2 P$ from column 3 of Tables 5 to 7, for each index, and so the effect of one strictness-adjusted year of exposure for a child at that value of P against a child with $P = 0$. The band is the 90 percent confidence interval from standard errors clustered on the birth state. The dotted lines are the 10th, 50th and 90th percentiles of P among children, weighted by the PSID weight at first observation. The data are the PSID cohorts born 1975 to 1999, and the sample and controls are those of Tables 5 to 7.

adjusted year of exposure raises the general health index of a child with $P = 1$ by 0.104 standard deviations relative to a child with $P = 0$, and the estimate is significant at the 10 percent level. Figure 5 plots this estimate across P, as the effect of one strictness-adjusted year of exposure for a child at each value of P relative to a child at $P = 0$, with its 90 percent confidence interval. The dotted lines mark the 10th, 50th and 90th percentiles of P, 0.008, 0.033 and 0.46. Of the two components, only self-rated health is significant at the 10 percent level, rising 0.111 points on the five-point scale per strictness-adjusted year. That is 3.0 percent of its mean of 3.70. Not being obese moves the same way and is imprecise, rising 2.1 percentage points per year with a 90 percent confidence interval from -1.1 to 5.3 points. A child conceived in or after the reform year spent all 5.75 years from conception to age five under reform, and those years count on average as 3.8 strictness-adjusted years (Section 3). Since each strictness-adjusted year raises the general health index by 0.104 standard deviations, the 3.8 years together raise it by 0.40 standard deviations for a child whose family was certain to be on welfare.

Columns 1 and 2 of Table 5 show the two steps before the preferred estimate. With the years under any reform unweighted, the index coefficient is 0.061, and scoring the same years by strictness raises it to 0.090, so most of the rise across the three columns comes from measuring what the rules required. The reform-wave by birth-year effects of the third step then raise it by a further 0.014, to 0.104. Self-rated health follows the same path, 0.063, 0.101 and 0.111. Not obese stays between 0.013 and 0.021 and is not significant in any

Table 5: Exposure to welfare reform and adult general health

	(1) Years under any reform	(2) Strictness-adjusted years	(3) Column 2 with reform-wave × birth-year effects
General health index			
Exposure $\times P$	0.061	0.090	0.104*
	(0.033)	(0.050)	(0.048)
	[0.11]	[0.12]	[0.08]
Exposure	0.000	0.010	-0.057
	(0.025)	(0.039)	(0.070)
Mean of outcome	0.057	0.057	0.057
Person-waves	40,731	40,731	40,731
Self-rated health (5 = excellent)			
Exposure $\times P$	0.063*	0.101*	0.111*
	(0.031)	(0.046)	(0.044)
	[0.10]	[0.08]	[0.08]
Exposure	0.004	0.031	-0.020
	(0.026)	(0.040)	(0.063)
Mean of outcome	3.704	3.704	3.704
Person-waves	40,713	40,713	40,713
Not obese			
Exposure $\times P$	0.013	0.016	0.021
	(0.013)	(0.019)	(0.020)
	[0.34]	[0.46]	[0.32]
Exposure	-0.002	-0.007	-0.033
	(0.008)	(0.013)	(0.028)
Mean of outcome	0.728	0.728	0.728
Person-waves	39,607	39,607	39,607

Notes: Each block is a separate estimate of the first three steps of Section 4, one per column, for the outcome named, on the 1975 to 1999 cohorts described in Section 3, weighted by the PSID longitudinal weight. Exposure is measured in years of the 5.75-year window from conception to the fifth birthday, years under any reform in column 1 and strictness-adjusted years in columns 2 and 3, so the interaction is the effect of one year for a child with $P = 1$ against one with $P = 0$. Every column carries birth-state, census-division by birth-year and survey-wave effects and the controls of Section 4. Columns 1 and 2 also carry birth-year effects, and column 3 replaces them with reform-wave by birth-year effects and is the preferred estimate. Standard errors clustered on the birth state are in parentheses and wild-cluster bootstrap p-values (999 Rademacher draws, null imposed) in brackets; stars mark significance at the 10, 5 and 1 percent levels (*, **, ***), by the bootstrap p-value for the interaction and the cluster-robust p-value for the main effect. The index is in standard deviations of the unexposed children; the components are in their own units. Outcomes are measured from age 18. The bootstrap p-value for self-rated health in column 1 is 0.098, so it carries a star at two decimals.

column. The index coefficient is significant at the 10 percent level only in column 3. The second row of each block is β_1, the coefficient on exposure alone, and Section 6 takes it up as the design's placebo.

Table 6 turns to the mental health index, the distress scale and the share with no psychiatric diagnosis. In column 3 each strictness-adjusted year raises the index by 0.158 standard deviations, and the estimate is significant at the 10 percent level. The same 3.8 strictness-

Table 6: Exposure to welfare reform and adult mental health

	(1) Years under any reform	(2) Strictness-adjusted years	(3) Column 2 with reform-wave × birth-year effects
Mental health index			
Exposure $\times P$	0.092*	0.155**	0.158*
	(0.043)	(0.064)	(0.073)
	[0.08]	[0.04]	[0.05]
Exposure	0.011	0.015	-0.067
	(0.030)	(0.043)	(0.082)
Mean of outcome	-0.084	-0.084	-0.084
Person-waves	25,136	25,136	25,136
Distress scale (1 = no distress)			
Exposure $\times P$	0.017*	0.026*	0.028*
	(0.007)	(0.011)	(0.013)
	[0.05]	[0.07]	[0.09]
Exposure	-0.003	-0.003	-0.012
	(0.005)	(0.007)	(0.013)
Mean of outcome	0.818	0.818	0.818
Person-waves	23,030	23,030	23,030
No psychiatric diagnosis			
Exposure $\times P$	0.021	0.040*	0.040*
	(0.014)	(0.021)	(0.022)
	[0.18]	[0.07]	[0.07]
Exposure	0.009	0.011	-0.022
	(0.009)	(0.013)	(0.027)
Mean of outcome	0.843	0.843	0.843
Person-waves	25,103	25,103	25,103

Notes: Each block is a separate estimate of the first three steps of Section 4, one per column, for the outcome named, on the 1975 to 1999 cohorts described in Section 3, weighted by the PSID longitudinal weight. Exposure is measured in years of the 5.75-year window from conception to the fifth birthday, years under any reform in column 1 and strictness-adjusted years in columns 2 and 3, so the interaction is the effect of one year for a child with $P = 1$ against one with $P = 0$. Every column carries birth-state, census-division by birth-year and survey-wave effects and the controls of Section 4. Columns 1 and 2 also carry birth-year effects, and column 3 replaces them with reform-wave by birth-year effects and is the preferred estimate. Standard errors clustered on the birth state are in parentheses and wild-cluster bootstrap p-values (999 Rademacher draws, null imposed) in brackets; stars mark significance at the 10, 5 and 1 percent levels (*, **, ***), by the bootstrap p-value for the interaction and the cluster-robust p-value for the main effect. The index is in standard deviations of the unexposed children; the components are in their own units. Outcomes are measured from age 18.

adjusted years of a child conceived in or after the reform year raise it by 0.60 standard deviations for a child whose family was certain to be on welfare. Both components improve and are significant at the 10 percent level. On the distress scale, which runs from 0 to 1 with 1 for no distress, the gain is 0.028 per year, 3.4 percent of its mean, and the share with no psychiatric diagnosis rises by 4.0 percentage points per year. Across the columns the index rises from 0.092 to 0.155 to 0.158, so here almost the whole rise is the strictness step.

Table 7 turns to the employment index and its three components, employment, annual hours and earnings, measured from age 24. In column 3 the effect of one strictness-adjusted year on the employment index is 0.053 standard deviations and is not significant at the 10 percent level. Of the three components, only annual hours is significant at the 10 percent level, rising by 113 hours per strictness-adjusted year. That is 6.4 percent of its mean of 1,764 hours. Annual hours count a year without work as zero, so the rise in hours can come from two sources, more people working during the year or longer hours among those already working. Employment shows whether more people work, but only on the day of the interview, whereas annual hours cover the whole previous year. The effect on employment is 1.0 percentage point per year and is not significant at the 10 percent level. This estimate is too imprecise for me to tell the two sources apart. For earnings the point estimate is about 12 percent per year but imprecisely estimated. This percentage comes from reading the coefficient on the inverse hyperbolic sine of earnings as a log change. The estimates in Table 7 also rest on fewer exposed children than the health estimates, because employment is measured from age 24. The exposed children were born from 1988 to 1999, so the PSID observes them at age 24 or older only in its last six waves, from 2013 to 2023. In the sample of Table 4, exposed children make up 14 percent of the person-waves observed at age 24 or older, against 21 percent of the person-waves observed at age 18 or older, the ages the health estimates use. Most of these children's adult employment is still to be observed.

The fourth step of Section 4, deferred above, is Appendix Table B1. It reports the five rules entered together in the preferred specification, for the three indices. Each rule enters with its own exposure, built the way W is built but from that rule's score alone. A rule's score is 0 at the AFDC setting and 1 at the setting of that same rule that pushed a parent hardest toward work. A child's exposure to the rule is the sum of the years of her window spent under it, each weighted by that score. A cell in Table B1 is therefore the effect of one year under that rule at its score-1 setting rather than under the AFDC setting, for a child with $P = 1$ against one with $P = 0$ and with the other four rules held fixed. Figure 6 plots these cells, one panel per index, each with its 90 percent confidence interval. For the earnings disregard the score-1 setting is the most generous one, under which a family kept part of its grant until its earnings reached the poverty line, and Appendix A gives the anchor for each of the other four. On the two health indices the only cell significant at the 10 percent level is the time limit on general health. Each year under the time limit raises the general health index by 0.085 standard deviations, with a cluster-robust p-value of 0.10. The other four rules on that index have p-values above 0.39. On mental health every p-value is above 0.28. For the employment index two rules are large and opposed, 0.221 standard deviations per year for the disregard with a p-value below 0.01 and -0.229 for the exemption age with a

Table 7: Exposure to welfare reform and adult employment

	(1) Years under any reform	(2) Strictness-adjusted years	(3) Column 2 with reform-wave × birth-year effects
Employment index			
Exposure $\times P$	0.044	0.065	0.053
	(0.037)	(0.058)	(0.052)
	[0.27]	[0.33]	[0.36]
Exposure	0.026	0.007	-0.027
	(0.016)	(0.034)	(0.062)
Mean of outcome	0.127	0.127	0.127
Person-waves	29,849	29,849	29,849
Employed			
Exposure $\times P$	0.018	0.026	0.010
	(0.015)	(0.024)	(0.021)
	[0.26]	[0.35]	[0.67]
Exposure	0.010	-0.001	-0.006
	(0.007)	(0.015)	(0.022)
Mean of outcome	0.801	0.801	0.801
Person-waves	29,788	29,788	29,788
Annual hours			
Exposure $\times P$	80	110	113
	(43)	(62)	(52)
	[0.16]	[0.19]	[0.10]
Exposure	21	30	84
	(18)	(33)	(55)
Mean of outcome	1,764	1,764	1,764
Person-waves	28,350	28,350	28,350
Own labor earnings, inverse hyperbolic sine			
Exposure $\times P$	0.077	0.130	0.121
	(0.135)	(0.206)	(0.191)
	[0.59]	[0.56]	[0.56]
Exposure	0.084	0.007	-0.204
	(0.067)	(0.125)	(0.242)
Mean of outcome	9.777	9.777	9.777
Person-waves	29,307	29,307	29,307

Notes: Each block is a separate estimate of the first three steps of Section 4, one per column, for the outcome named, on the 1975 to 1999 cohorts described in Section 3, weighted by the PSID longitudinal weight. Exposure is measured in years of the 5.75-year window from conception to the fifth birthday, years under any reform in column 1 and strictness-adjusted years in columns 2 and 3, so the interaction is the effect of one year for a child with $P = 1$ against one with $P = 0$. Every column carries birth-state, census-division by birth-year and survey-wave effects and the controls of Section 4. Columns 1 and 2 also carry birth-year effects, and column 3 replaces them with reform-wave by birth-year effects and is the preferred estimate. Standard errors clustered on the birth state are in parentheses and wild-cluster bootstrap p-values (999 Rademacher draws, null imposed) in brackets; stars mark significance at the 10, 5 and 1 percent levels (*, **, ***), by the bootstrap p-value for the interaction and the cluster-robust p-value for the main effect. The index is in standard deviations of the unexposed children; the components are in their own units. Outcomes are measured from age 24. The bootstrap p-value for annual hours in column 3 is 0.103, so it carries no star at two decimals.

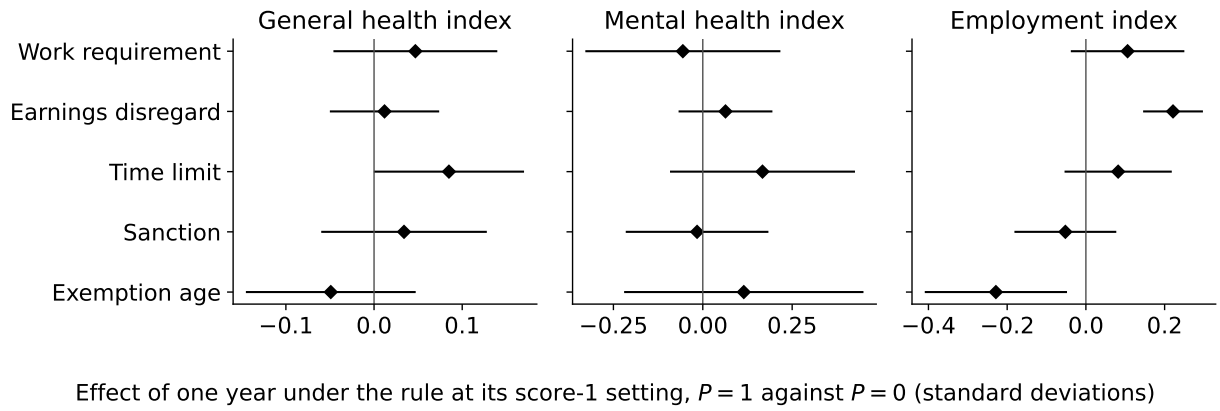


Figure 6: Effects of the five rules entered together, by index

Notes: Each marker is a cell of Appendix Table B1, the effect of one year under the rule named on its row at its score-1 setting rather than under the AFDC setting, for a child with $P = 1$ against one with $P = 0$, with the other four rules held fixed. Bars are 90 percent confidence intervals from standard errors clustered on the birth state. The data are the PSID cohorts born 1975 to 1999, and the sample and controls are those of Tables 5 to 7.

p-value of 0.035, while the index itself in Table 7 is 0.053.

The earnings disregard let a working family keep more of its grant, so each dollar the parent earned added more to the family's money. The exemption age set how young a child could be when the parent had to work, and at its strictest setting the parent had to work from the month the child was born. Unlike the disregard, this rule changed the parent's time at home and not the family's money. In Table B1 the disregard raised the children's employment as adults, and the exemption age lowered it. Table 7 combines the five rules into one exposure, the strictness-adjusted years. In that estimate the gain from the disregard and the loss from the exemption age offset each other, and the employment index shows no significant effect. The gain from the disregard is consistent with the long-run gains from family income in the studies Section 1 cites. Those studies also find that children are harmed when low-quality child care replaces a parent's time, and the loss from the exemption age is consistent with that finding. On the two health indices neither of these rules is significant. Of the other three rules, only the time limit is significant, on general health. It ended the grant and pushed the parent into work at the same time, so its effect cannot be split between money and time. The work requirement and the sanction are not significant on any index. So Table B1 does not show how much of the health gains in Tables 5 and 6 came from the family's money and how much from the parent's time.

6 Robustness

The regression enters exposure twice, alone with coefficient β_1 and interacted with P with coefficient β_2. The estimate of β_2 is wrong only if something besides the rules changed across a state's birth cohorts in step with how strict its rules were. That something would also have to reach the children of likely-recipient families more than other children. A shock to a state's economy, labour market or other policies would usually reach families in the state whether or not they could receive welfare. The coefficient on exposure alone, β_1, is the effect of exposure on children whose families had no predicted chance of receipt. The rules paid those families nothing, so the rules cannot move β_1, but a shock that reached them as it reached other families would move it. This reading assumes that the reform itself reached their children in no other way, such as through the local labour market. Therefore, a β_1 of zero says that the children the rules could not reach did no better or worse when they were born under stricter rules. In that case, stricter rules did not come with a shock that reached every family in the state. So what could still bias β_2 is a shock that moved with strictness and reached only likely-recipient families. Such a shock leaves β_1 at zero, so β_1 cannot rule it out. The main shock of that kind is a state EITC, and I test it later in this section. Both coefficients appear in Table 8 for the ten outcomes of Section 5. As in Section 4, β_1 carries a cluster-robust p-value and β_2 a wild-cluster bootstrap p-value. I report the two coefficients in the preferred specification and in the same regression without the census-division by birth-year effects, to show what those effects remove.

In the preferred specification β_1 is negative on nine of the ten outcomes and is not significant at the 10 percent level on any of them. On the general health index it is -0.057 with a standard error of 0.070 and a p-value of 0.41, against a β_2 of 0.104. On the mental health index it is -0.067 with a standard error of 0.082 and a p-value of 0.41, against a β_2 of 0.158. On the employment index it is -0.027 with a p-value of 0.66. Annual hours is the one outcome where β_1 is positive, 84 hours with a p-value of 0.12, against a β_2 of 113.

Without the census-division by birth-year effects, β_1 is significant at the 10 percent level on six of the ten outcomes. None of the six is in the employment family. On the two health indices it is negative and significant, -0.113 with a p-value of 0.04 on general health and -0.176 with a p-value of 0.006 on mental health. Every outcome is oriented so that a higher value is better, so a negative β_1 of that size says that children whose families had no predicted chance of receipt did worse, as their exposure rose across cohorts, in the states whose rules were stricter. That is the shape a state-cohort shock would leave if it hit all children alike. The division effects take such a shock out wherever it moved all the children of a division's cohorts together, and with them in place the placebo passes on every reported outcome.

On the general health index β_2 is 0.105 standard deviations per year without the effects and 0.104 with them. On the mental health index it rises from 0.096 to 0.158, and its bootstrap p-value falls from 0.20 to 0.05. So the mental health estimate is significant at the 10 percent level only with the effects in place. With the effects, β_2 compares only children born in the same year in the same census division. The mental health estimate rises when the comparisons across divisions are removed, so those comparisons had been pulling that estimate down.

The main candidate for a shock that reached likely-recipient families more than other families is a refundable state EITC adopted in the reform years. A refundable state EITC pays cash to working families with low earnings, and the families the rules pushed into work are among them. If the states that adopted such a credit in the reform years differed in strictness from the other states that reformed in the same year, the credit's effect on the children of those families would be counted as an effect of the rules. A nonrefundable credit would not matter, because it pays nothing to a family that owes no income tax. In the two tests that follow, I drop the children whose birth state started a new refundable credit before they turned five, so I need not know which way those states differed.

By the state EITC panel of the University of Kentucky Center for Poverty Research, seven places started a refundable EITC in the reform years or after them, 1993 or later: New York in 1994, Massachusetts in 1997, Colorado in 1999, the District of Columbia in 2000, New Jersey in 2001, Oklahoma in 2002 and Rhode Island in 2003. The first cut drops every child born in one of these seven places in or after the fifth year before its EITC started, because that child's window from conception to age five could include a year with the EITC. For New York, whose credit started in 1994, that means children born in 1989 or later. Those seven places keep their children born earlier, so no birth state leaves the sample. That cut removes 181 of the 6,225 children in the general health sample. On the rest, with everything else in the preferred specification kept, β_2 on the general health index rises from 0.104 to 0.116 and has a bootstrap p-value of 0.059 (Appendix Table C1). On the mental health index it moves less, from 0.158 to 0.151, with a bootstrap p-value of 0.064. On the employment index it falls from 0.053 to 0.023, with a bootstrap p-value of 0.68, so the index was a null before the cut and stays one after it. In a stricter cut I treat any rise in a state's refundable credit, measured as a share of the federal one, in the same way as a start. This cut counts starts and rises from 1989, the first year the EITC panel covers, while the first cut counted only starts from 1993 on. Under this rule, the stricter cut also drops the children of Minnesota and Vermont born in or after the fifth year before their credit started or rose. Minnesota started its credit in 1991, and Vermont raised its credit in 2000. The stricter cut removes 220 children from the general health sample in all, 39 more than the

first. Every Minnesota child in the sample was born in or after 1986, the fifth year before its credit started, so Minnesota itself leaves the sample under this cut and 50 birth states remain. The two health estimates are then 0.117 on general health and 0.144 on mental health, with bootstrap p-values of 0.067 and 0.077. The employment index stays a null at 0.028, with a bootstrap p-value of 0.62. Under both cuts the two health estimates keep their sign and remain significant at the 10 percent level, and neither moves by more than 0.3 of its standard error from the preferred estimates in Tables 5 and 6. Therefore, a state EITC that started or rose in the reform years does not produce the two health estimates.

In the preferred specification, exposure runs from conception to age five, because early childhood is a critical period of development for adult health (Hoynes et al., 2016). But a family could receive welfare only while it had a child under 18, so the rules could reach a child from conception to age 17. I test the choice of age five by measuring exposure in the same way from conception to age 17, in place of exposure to age five. With exposure to age 17, β_2 is 0.035 per strictness-adjusted year on the general health index and 0.046 on the mental health index, against 0.104 and 0.158 with exposure to age five (Appendix Table F1). Each coefficient is the effect of one strictness-adjusted year, and the window from conception to age 17 is 18.75 years long against 5.75 for the window from conception to age five. Each coefficient times the length of its window is the effect of a whole window spent under the strictest rules, for a child with $P = 1$ against one with $P = 0$. On general health that effect is 0.60 standard deviations with exposure to age five and 0.66 with exposure to age 17. On mental health it is 0.91 with exposure to age five and 0.85 with exposure to age 17. The window to age 17 contains the window to age five, so a whole window to age 17 would be worth more than a whole window to age five if the years after age five mattered. Since the two whole-window effects are nearly equal, a strictness-adjusted year after age five adds almost nothing to either index. Therefore, the effect of the rules on adult health comes almost entirely from exposure before age five, as Hoynes et al. (2016) find for food stamps.

In the preferred specification the placebo passes, and under both EITC cuts the two health estimates keep their sign and stay significant at the 10 percent level. Without the division effects, the mental health estimate is not significant.

7 Conclusion

This paper presents new evidence that exposure to the welfare reform of the 1990s in early childhood improved adult health among children whose families were likely to be on welfare, and more so where the rules were stricter. Theory does not say whether such children should gain or lose, because the reform could raise a family's income and at the same time take a

Table 8: Exposure among children whose families had no predicted chance of receipt

	Without division effects		Preferred column	
	β_1	β_2	β_1	β_2
	(1)	(2)	(3)	(4)
General health index	-0.113**	0.105*	-0.057	0.104*
	(0.054)	(0.048)	(0.070)	(0.048)
	[0.04]	[0.07]	[0.41]	[0.08]
Self-rated health	-0.091**	0.111*	-0.020	0.111*
	(0.045)	(0.042)	(0.063)	(0.044)
	[0.04]	[0.05]	[0.75]	[0.08]
Not obese	-0.039*	0.021	-0.033	0.021
	(0.022)	(0.021)	(0.028)	(0.020)
	[0.08]	[0.38]	[0.23]	[0.32]
Mental health index	-0.176***	0.096	-0.067	0.158*
	(0.064)	(0.072)	(0.082)	(0.073)
	[0.01]	[0.20]	[0.41]	[0.05]
Distress scale (higher = less distress)	-0.031***	0.018	-0.012	0.028*
	(0.011)	(0.013)	(0.013)	(0.013)
	[0.00]	[0.24]	[0.38]	[0.09]
No psychiatric diagnosis	-0.049**	0.025	-0.022	0.040*
	(0.020)	(0.022)	(0.027)	(0.022)
	[0.01]	[0.29]	[0.42]	[0.07]
Employment index	-0.011	0.040	-0.027	0.053
	(0.041)	(0.053)	(0.062)	(0.052)
	[0.79]	[0.47]	[0.66]	[0.36]
Employed	-0.004	0.008	-0.006	0.010
	(0.016)	(0.021)	(0.022)	(0.021)
	[0.80]	[0.72]	[0.78]	[0.67]
Annual hours	2	88	84	113
	(52)	(57)	(55)	(52)
	[0.96]	[0.19]	[0.12]	[0.10]
Earnings, inverse hyperbolic sine	0.052	0.090	-0.204	0.121
	(0.150)	(0.193)	(0.242)	(0.191)
	[0.73]	[0.66]	[0.40]	[0.56]

Notes: The second pair of columns is one estimate of equation (2), the preferred specification, on the children of Section 3, and is column 3 of Tables 5 to 7. The first pair is the same regression without the census-division by birth-year effects. β_1 is the coefficient on exposure, the effect of one strictness-adjusted year on a child whose family had no predicted chance of receipt ($P = 0$), and β_2 is the coefficient on exposure $\times P$, the additional effect for a certain recipient. Standard errors clustered on the birth state are in parentheses. In brackets, β_1 carries its cluster-robust p-value and β_2 its wild-cluster bootstrap p-value from 999 draws, and stars mark significance at the 10, 5 and 1 percent levels (*, **, ***) by the same p-value; the two pairs of columns come from separate runs of the bootstrap. Each row uses the sample of its outcome in Tables 5 to 7, with the person-waves reported there.

parent's time away from the child. Between 1993 and 1998 every state replaced the cash entitlement of Aid to Families with Dependent Children (AFDC) with aid that a parent had to work for, that could be cut if the parent did not, and that ran out after a fixed number of months. A parent who complied went to work and kept the grant, so the family had more money and the child had less of the parent's time. Each state wrote its own rules, so two states that reformed in the same year could ask very different things of a parent. For that reason, I measure the strictness of each state's rules year by year with an index that runs from 0 under the AFDC rules to 1 at the strictest setting any state used. I weight each of the five rules that govern work by the size of its effect on parents' employment in the experimental literature. The five rules are the work requirement, the earnings disregard, the time limit, the sanction and the exemption age. A child's exposure is then the years from conception to age five spent under those rules, each weighted by the index in force, so that a year at an index of 0.5 counts as half a strictness-adjusted year of exposure. The rules could reach a child only through the family's welfare status, so I compare that exposure between children whose families were likely to be on welfare and children whose families were not. I make that comparison only among children born in the same year, in the same census division and in states that reformed in the same year. Each child's predicted probability of being on welfare comes from a model fitted on families observed before any state reformed, applied to the family's characteristics over the child's first five years. The sample is the 8,408 children born between 1975 and 1999 in the Panel Study of Income Dynamics (PSID).

For a child whose family was certain to be on welfare, each strictness-adjusted year of exposure before age five raises the general health index by 0.104 standard deviations and the mental health index by 0.158 standard deviations in adulthood. Both indices are oriented so that a higher value is better health. Both estimates are significant at the 10 percent level under a wild-cluster bootstrap over the 51 birth states. Within the general health index, self-rated health carries most of the gain. Both parts of the mental health index improve, the distress scale and the indicator for having no psychiatric diagnosis. Beyond the number of years, the strictness of the rules matters, because the gains are larger when a child's years under reform are weighted by strictness than when the same years are counted without that weight. The gains come from the years before age five, because a whole window from conception to age 17 is worth about as much as a whole window to age five. Hoynes et al. (2016) find the same timing for food stamps. As a placebo, the same regression also estimates the effect of exposure on children whose families had no predicted chance of being on welfare. That placebo effect is not significant for any outcome in the paper. Finally, both health estimates hold when I drop the children whose window from conception to age five could include a year of a refundable state earned income tax credit, a second policy that

paid working families in those years.

On the employment index the estimate is 0.053 standard deviations per strictness-adjusted year of exposure, with a standard error of 0.052, so I cannot reject the null. The estimate is imprecise in part because the children exposed to the reform before age five have had little time in the labor market. The employment outcomes are measured from age 24 onward. About a third of these children were born in states that reformed in 1997, more than in any other reform year. A child born in such a state in 1997 is 26 in the 2023 wave, so the PSID observes that child in only two interviews, in 2021 and 2023. As a result, the children exposed before age five make up only 14 percent of all employment observations. By contrast, health is measured from age 18, so the same children make up 21 percent of all health observations. In addition, two of the five rules move the children's adult employment in opposite directions. A second regression gives each of the five rules its own exposure, built from that rule's own score in place of the combined index. In that regression, letting a working family keep more of its grant (the earnings disregard) raised the children's employment as adults. Requiring a parent to work before the child turned three (the exemption age) lowered it. In the employment index these two effects offset each other. As these cohorts age, later waves of the PSID will show whether the reform changed their employment as well.

Studies that measured children while their families were under the reform mostly found harm or mixed effects (Leonard and Mas, 2008; Herbst, 2017; Kalil et al., 2023; Dave et al., 2021; Baltagi and Yen, 2016), whereas the outcomes here are measured years later, from age 18 and from age 24. Dustmann et al. (2024) find harm in adulthood, but from a cut in transfers to refugee families in Denmark with no work requirement attached, unlike the reform studied here. Dore et al. (2025) study the same reform in the PSID and find more psychological distress among adults exposed in early childhood, but three features of their design leave that estimate open to bias. They count the years after a state's reform date, whatever its rules required. With only state and interview-year fixed effects, they compare states that had already reformed with states that reformed later. Under staggered adoption that comparison can bias the estimate, even in sign, when the effect grows with time under reform. They also assign children to the group likely to be on welfare using childhood circumstances that the adult recalls in the same interview that measures the distress, so an adult in poor mental health may recall a harder childhood and be counted as treated. This paper avoids the first two by scoring each state's rules and by comparing only states that reformed in the same year. It also avoids the third, because the family circumstances it uses were recorded by the parent while the children were young. Unlike Dore et al.'s estimate, the gain here has the same sign as the effect of food stamps in early childhood (Hoynes et al., 2016). Food stamps improved adult health by adding to a family's resources, whereas

the reform took away the cash entitlement. If the gain here also works through resources, the families who met the work requirement probably did not lose resources on net, once the parent's earnings are counted.

The health estimate combines the family's added earnings and the parent's lost time at home, and my design does not separate them. Mullins (2026) separates them in a model of maternal work and investment in children, estimated on three random-assignment welfare-to-work experiments. In that model, whether a child gains depends on the family's income and on the kind of care the child receives while the parent works. The reform itself paid for part of the care that working parents needed. The 1996 welfare law also created the Child Care and Development Fund, a child care subsidy program for low-income working families (Herbst, 2017). Under that law's Temporary Assistance for Needy Families (TANF) block grant, the states moved their welfare spending from cash to services year by year. Cash benefits fell from 75 percent of federal and state spending on welfare for families before TANF to 40 percent by 2001 (Burke, 2003). Over the same years child care grew to become the largest service paid for with TANF funds (Burke, 2003). So the reform these children were exposed to was more than a work requirement, because it also let a working family keep part of its grant and paid for child care through TANF. The PSID does not record the care these children received, so my design cannot say how much of the gain came through child care.

These estimates matter now because work requirements are being extended to programs that did not have them and to parents who were previously excused. The 2025 reconciliation law attaches a work requirement to Medicaid (Congressional Research Service, 2025b) and widens the one in the Supplemental Nutrition Assistance Program (Congressional Research Service, 2025a). Rules that push a parent toward work take time the parent would otherwise spend with the child. But the rules studied here came with a cash grant that a working family could partly keep, and with child care paid for through TANF. Under those rules, the children whose families were likely to be on welfare were healthier as adults. So the effect of a new work requirement on the children in the families it covers may depend on what comes with it, such as a benefit that a working family can partly keep and help with child care.

References

Agostinelli, F. and Sorrenti, G. (2021). Money vs. time: Family income, maternal labor supply, and child development. Working Paper 273, University of Zurich, Department of Economics. Revised version, February 2021. SSRN 3102271.

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

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

Bailey, M. J., Sun, S., and Timpe, B. (2021). Prep school for poor kids: The long-run impacts of head start on human capital and economic self-sufficiency. *American Economic Review*, 111(12):3963–4001.

Baker, M., Gruber, J., and Milligan, K. (2019). The long-run impacts of a universal child care program. *American Economic Journal: Economic Policy*, 11(3):1–26.

Baltagi, B. H. and Yen, Y.-F. (2016). Welfare reform and children's health. *Health Economics*, 25(3):277–291.

Bernal, R. and Keane, M. P. (2011). Child care choices and children's cognitive achievement: The case of single mothers. *Journal of Labor Economics*, 29(3):459–512.

Bitler, M. and Hoynes, H. (2016). The more things change, the more they stay the same? The safety net and poverty in the Great Recession. *Journal of Labor Economics*, 34(S1):S403–S444.

Bitler, M. P., Gelbach, J. B., and Hoynes, H. W. (2005). Welfare reform and health. *Journal of Human Resources*, 40(2):309–334.

Bitler, M. P., Gelbach, J. B., and Hoynes, H. W. (2006). What mean impacts miss: Distributional effects of welfare reform experiments. *American Economic Review*, 96(4):988–1012.

Blank, R. M. (2002). Evaluating welfare reform in the united states. *Journal of Economic Literature*, 40(4):1105–1166.

Bollinger, C., Gonzalez, L., and Ziliak, J. P. (2009). Welfare reform and the level and composition of income. In Ziliak, J. P., editor, *Welfare Reform and its Long-Term Consequences for America's Poor*, pages 59–103. Cambridge University Press, Cambridge.

Boudreaux, M. H., Golberstein, E., and McAlpine, D. D. (2016). The long-term impacts of Medicaid exposure in early childhood: Evidence from the program's origin. *Journal of Health Economics*, 45:161–175.

Burke, V. (2003). Welfare reform: TANF trends and data. CRS Report 98-369 EPW, Congressional Research Service. Updated March 12, 2003.

Callaway, B. and Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90(3):414–427.

Congressional Research Service (2025a). Supplemental nutrition assistance program (snap) and related nutrition programs in p.l. 119-21: An overview. CRS Report R48552, Congressional Research Service.

Congressional Research Service (2025b). Work requirements: Comparison of medicaid and supplemental nutrition assistance program (snap) after p.l. 119-21. CRS Report R48755, Congressional Research Service.

Corman, H., Dave, D. M., Schwartz-Soicher, O., and Reichman, N. E. (2022). Effects of welfare reform on household food insecurity across generations. *Economics and Human Biology*, 45:101101.

Dave, D., Corman, H., Kalil, A., Schwartz-Soicher, O., and Reichman, N. E. (2021). Intergenerational effects of welfare reform: Adolescent delinquent and risky behaviors. *Economic Inquiry*, 59(1):199–216.

de Chaisemartin, C. and D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996.

Del Boca, D., Flinn, C., and Wiswall, M. (2014). Household choices and child development. *Review of Economic Studies*, 81(1):137–185.

Dore, E. C., Hamad, R., Komro, K. A., and Livingston, M. D. (2025). The long-term health effects of welfare reform. *Social Science & Medicine*, 371:117878.

Duncan, G., Kalil, A., Mogstad, M., and Rege, M. (2022). Investing in early childhood development in preschool and at home. BFI Working Paper 2022-58, Becker Friedman Institute, University of Chicago. Prepared for the Handbook of the Economics of Education, Volume 6.

Dustmann, C., Landersø, R., and Andersen, L. H. (2024). Unintended consequences of welfare cuts on children and adolescents. *American Economic Journal: Applied Economics*, 16(4):161–185.

Fang, H. and Keane, M. P. (2004). Assessing the impact of welfare reform on single mothers. *Brookings Papers on Economic Activity*, 2004(1):1–116.

Goodman-Bacon, A. (2021a). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

Goodman-Bacon, A. (2021b). The long-run effects of childhood insurance coverage: Medicaid implementation, adult health, and labor market outcomes. *American Economic Review*, 111(8):2550–2593.

Grogger, J. (2003). The effects of time limits, the EITC, and other policy changes on welfare use, work, and income among female-headed families. *Review of Economics and Statistics*, 85(2):394–408.

Grogger, J. and Michalopoulos, C. (2003). Welfare dynamics under time limits. *Journal of Political Economy*, 111(3):530–554.

Hartley, R. P., Lamarche, C., and Ziliak, J. P. (2022). Welfare reform and the intergenerational transmission of dependence. *Journal of Political Economy*, 130(3):523–565.

Herbst, C. M. (2017). Are parental welfare work requirements good for disadvantaged children? Evidence from age-of-youngest-child exemptions. *Journal of Policy Analysis and Management*, 36(2):327–357.

Hosmer, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied Logistic Regression*. Wiley, Hoboken, NJ, 3 edition. Chapter 5, Assessing the Fit of the Model, pp. 143–202.

Hoynes, H., Schanzenbach, D. W., and Almond, D. (2016). Long-run impacts of childhood access to the safety net. *American Economic Review*, 106(4):903–934.

Hoynes, H. W. and Schanzenbach, D. W. (2009). Consumption responses to in-kind transfers: Evidence from the introduction of the food stamp program. *American Economic Journal: Applied Economics*, 1(4):109–139.

Kalil, A., Corman, H., Dave, D., Schwarz-Soicher, O., and Reichman, N. E. (2023). Welfare reform and the quality of young children's home environments. *Demography*, 60(6):1791–1813.

Kaushal, N. and Kaestner, R. (2001). From welfare to work: Has welfare reform worked? *Journal of Policy Analysis and Management*, 20(4):699–719.

Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, 75(1):83–119.

Leonard, J. and Mas, A. (2008). Welfare reform, time limits, and infant health. *Journal of Health Economics*, 27(6):1551–1566.

Looney, A. (2005). The effects of welfare reform and related policies on single mothers' welfare use and employment. Technical report, Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series 2005-45, Washington, DC.

Matsudaira, J. D. and Blank, R. M. (2014). The impact of earnings disregards on the behavior of low-income families. *Journal of Policy Analysis and Management*, 33(1):7–35.

Miller, C., Knox, V., Gennetian, L. A., Dodoo, M., Hunter, J. A., and Redcross, C. (2000). Reforming welfare and rewarding work: Final report on the Minnesota Family Investment Program, volume 1: Effects on adults. Technical report, Manpower Demonstration Research Corporation, New York.

Miller, S. and Wherry, L. R. (2019). The long-term effects of early life Medicaid coverage. *Journal of Human Resources*, 54(3):785–824.

Mullins, J. (2026). A structural meta-analysis of welfare reform experiments and their impacts on children. *Journal of Political Economy*, 134(1):435–477.

Paxson, C. and Waldfogel, J. (2003). Welfare reforms, family resources, and child maltreatment. *Journal of Policy Analysis and Management*, 22(1):85–113.

Pingle, J. F. (2003). What if welfare had no work requirements? the age of youngest child exemption and the rise in employment of single mothers. Technical report, Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series 2003-57, Washington, DC.

Scrivener, S., Hamilton, G., Farrell, M., Freedman, S., Friedlander, D., Mitchell, M., Nudelman, J., and Schwartz, C. (2002). WRP: Final report on Vermont's welfare restructuring project. Technical report, Manpower Demonstration Research Corporation, New York.

Smedslund, G., Dalsbø, T. K., Steiro, A., Winsvold, A., and Clench-Aas, J. (2006). Work programmes for welfare recipients. *Campbell Systematic Reviews*, 2(1).

U.S. House Committee on Ways and Means (1996). 1996 green book: Background material and data on programs within the jurisdiction of the committee on ways and means. Committee Print WMCP 104-14, U.S. Government Printing Office, Washington, DC. Section 8, Aid to Families with Dependent Children and Related Programs.

A Construction of the strictness index

This appendix gives what is needed to recompute the index from the policy panel. Work is the largest part of what PRWORA changed, and the index scores every rule that governs it. The statute and the waivers changed other things besides, among them the family cap, two-parent eligibility, the immigrant bar, and school and immunization requirements for a recipient's children. I leave those out because they reach a family through some channel other than the parent's job, or because no study measures what they did to employment.

The index for state s in year t is

$$S_{st} = \sum_{c=1}^{5} w_c \, I_{cst}, \tag{4}$$

where I_{cst} is the intensity of rule c in that state and year, a number between 0 and 1, and w_c is a fixed weight. The weights sum to one, so the index also runs from 0 to 1. Every intensity is 0 at the AFDC setting, so a state under AFDC on every rule scores 0. No state-year reaches 1, because no state chose the strictest setting on every rule at once. The highest value in the panel is Connecticut's 0.977 from 1996.

A.1 Scoring each rule

Two of the five rules are switches and take the value 0 or 1. The work requirement is 1 in a year in which a statewide requirement to work or look for work was in force and 0 otherwise. Under AFDC a parent had to take part in JOBS only if the state called her (Section 2), so no state had a statewide requirement and every AFDC year scores 0. The sanction is 1 when a family that did not meet the work requirement lost its whole grant and 0 when it lost only the adult's share, which was the AFDC rule.

The other three rules are quantities. Each is scored as the state's setting minus the AFDC setting, divided by the strict anchor minus the AFDC setting, and kept between 0 and 1. The time limit is measured in months of assistance a family may receive. The AFDC setting is no limit. The federal 60-month limit of 1996 is scored the same, because the federal clock did not bind for any family in the exposure window before 2001. The strict anchor is the 18-month limit Tennessee adopted, the shortest in any state. The score is

$$I_{\text{time}} = \max\left\{0, \min\left[1, \frac{60 - \text{months}}{60 - 18}\right]\right\}, \tag{5}$$

so a 24-month limit scores 0.857, 36 months scores 0.571 and 48 months scores 0.286. The

exemption is the age of the youngest child, in months, below which a parent was excused from work. Under AFDC that age was 36 months. The strict anchor is 0, immediate work, which five states adopted. The score is

$$I_{\text{exempt}} = \max\left\{0, \min\left[1, \frac{36 - \text{months}}{36}\right]\right\}, \tag{6}$$

so a 12-month exemption, the setting of 27 jurisdictions in 1998, scores 0.667. A state that set the age above 36 months is more lenient than AFDC and scores 0, since the scale does not represent leniency below the baseline.

A.2 The earnings disregard and its exit point

The disregard is scored by the exit point, the monthly earnings at which a family of three loses its grant. Under AFDC a working family kept 30 dollars plus one third of its earnings for four months and 30 dollars for the next eight. After those twelve months only the 90-dollar work-expense deduction remained. Every dollar earned above that deduction then reduced the grant by a dollar, so the sustained exit point under AFDC was

$$\text{exit}_{\text{AFDC}} = 90 + B_s, \tag{7}$$

where B_s is the state's maximum monthly grant for a family of three. A state that disregarded a flat amount F and a share p of the remainder had a grant that reached zero when the counted earnings equalled the grant. The exit point is therefore

$$\text{exit}_{\text{TANF}} = F + \frac{B_s}{1 - p}. \tag{8}$$

The strict anchor is the federal poverty line for a family of three, FPL_t, and the score is

$$I_{\text{disregard}} = \max\left\{0, \min\left[1, \frac{\text{exit}_{\text{TANF}} - \text{exit}_{\text{AFDC}}}{\text{FPL}_t - \text{exit}_{\text{AFDC}}}\right]\right\}. \tag{9}$$

Because the AFDC exit point contains the state's own grant, the same disregard formula scores differently in different states. A 400-dollar grant with a half disregard moves the exit point from 490 to 800 dollars, and a 700-dollar grant with the same formula moves it from 790 to 1,400, which is past the poverty line and scores 1. Eleven states carried the exit point to or past the poverty line and score 1. The ceiling is deliberate, since almost no family in the caseload earned above the poverty line while receiving assistance, so a disregard beyond that point changed no one's return to work. A disregard that applied only for an introductory

period and then reverted to the AFDC rule scores 0, because the sustained exit point is unchanged. Three states had such rules. The score is computed once, in the year the state adopted its own disregard rule, with that year's grant and poverty line, and held fixed until the state changed its sustained formula again. Computing it every year would have let the score drift downward as the poverty line rose against a fixed nominal grant. In this panel 121 state-years would then have registered a fall in strictness with no policy behind it.

A.3 The weights

Each weight starts from the consensus effect of the rule on parental employment, in percentage points, divided by the sum of the effects. The consensus effects were assembled from the experimental and quasi-experimental studies of each rule. Table A1 lists them with the studies behind them. For the work requirement the national estimates sit low and the experimental ones high. Fang and Keane (2004) attribute about 1.9 percentage points of the 1993 to 2002 rise in single mothers' employment to work requirements. Smedslund et al. (2006) put the pooled effect of 46 randomized welfare-to-work programs near 3 points. The three-arm Minnesota experiment isolates the mandate from the financial incentives and finds it added 6 to 10 points to employment among long-term recipients (Miller et al., 2000). Vermont's three-arm experiment found that its incentive-only arm moved employment little while the mandate arm did (Scrivener et al., 2002). The consensus of 4 points sits between the national and the experimental estimates. For the disregard the Minnesota incentive-only arm raised employment by 7 points in the first year and 3.6 later (Miller et al., 2000), while Matsudaira and Blank (2014) find that the statutory disregard expansions alone did little, because few recipients used them. The consensus of 2.5 points sits between them, and it is the least settled of the large weights. For the time limit, Grogger (2003) attributes about 0.7 points of the employment rise to time limits and Fang and Keane (2004) about 1.1, and the structural estimates of Grogger and Michalopoulos (2003) are larger, so the consensus is 1.5. For the sanction, Looney (2005) finds a clear effect on welfare use and an imprecise one on employment. Fang and Keane (2004) find that moving from lenient to strict sanction and exemption rules changes work participation by half a point, so the consensus of 1.2 points is inferred from the effect on welfare use rather than measured on employment directly. For the exemption age, Pingle (2003) finds no employment gain among mothers newly subject to work requirements and Looney (2005) about 0.7 points, so the consensus is their midpoint, 0.35.

Table A1: Study estimates behind the consensus employment effects and the weights

Rule	Study	Design	Estimate (pp)	Consensus (pp)	Weight
Work requirement	Fang and Keane (2004)	Structural model of work and welfare, CPS 1980–2002	1.9	4.0	0.419
	Smedslund et al. (2006)	Meta-analysis of 46 randomized welfare-to-work programs	3		
	Miller et al. (2000)	MFIP experiment, full program against incentives only	6.3 to 10.2		
	Scrivener et al. (2002)	WRP experiment, full program against incentives only	–[a]		
Earnings disregard	Miller et al. (2000)	MFIP experiment, incentives only against AFDC	7.0, then 3.6	2.5	0.262
	Matsudaira and Blank (2014)	Changes in state disregard rules after 1996	0		
Time limit	Grogger (2003)	State time limits, by age of the youngest child	0.7	1.5	0.157
	Fang and Keane (2004)	Same structural model	1.1		
	Grogger and Michalopoulos (2003)	FTP experiment, time limit separated by a model	–[b]		
Sanction	Looney (2005)	Changes in state sanction rules over time	–[c]	1.2	0.126
	Fang and Keane (2004)	Same model, sanction and exemption rules made strict	0.5		
Exemption age	Pingle (2003)	Newly subject mothers against still-exempt mothers	0	0.35	0.037
	Looney (2005)	Changes in state exemption ages over time	0.7		
Sum				9.55	1.000

Notes: Each estimate is the study's effect of the rule on the employment of single mothers or parents on welfare, in percentage points (pp). Every study except Smedslund et al. (2006) estimates the effect of the welfare reform of the 1990s, under a state waiver before 1996 or under TANF after it. Smedslund et al. (2006) pool randomized welfare-to-work programs in general, so their estimate is a reference point for the size of a work-requirement effect and not a measure of this reform. Where a study reports the share of the 1993 to 2002 rise in single mothers' employment that it attributes to the rule, the estimate is that share of the rise. MFIP, WRP and FTP are state waiver experiments that assigned welfare families at random to the new rules or to AFDC, and an incentives-only arm received the new financial incentives without the work mandate. MFIP is the Minnesota Family Investment Program. Its estimates are for urban long-term recipients, over follow-up years one to three for the work requirement, and in year one and then in years two and three for the disregard. WRP is Vermont's Welfare Restructuring Project and FTP is Florida's Family Transition Program. The structural model of Fang and Keane (2004) estimates how single mothers choose between work and welfare and computes the effect of each rule from those choices. CPS is the March Current Population Survey. An estimate of 0 is a study that finds no effect. [a] No single estimate. Employment rose in the arm with the mandate and not in the arm with incentives only. [b] The estimate is on welfare receipt, a fall of 16 percent, and not on employment. [c] The estimate is on welfare use, 9 percent of its decline, and the employment estimate is imprecise. The consensus effect is the value the index uses for each rule, set from the estimates in its rows as the text explains. Each weight is the consensus effect divided by the sum of the five, 9.55, rounded to three decimals.

Because these consensus effects are uncertain, I recompute the index under 5,000 random draws of weights. In every draw the work requirement gets the largest weight and the exemption age the smallest, with the disregard, the time limit and the sanction between them in that order, as in Table A1. For each draw I compare the ranking of the 51 states with their ranking under the weights of Table A1, in 1998, when every state had reformed. The rank correlation between the two rankings has a median of 0.94 across the draws, and it is above 0.82 in 95 percent of them. The preferred specification in Section 4 compares only states that reformed in the same year. So I also compute the rank correlation in 1998 within each group of states that reformed in the same year and average it over the groups, weighting each group by its number of states. This average has a median of 0.88 across the draws. Equal weights on the five rules ignore this order entirely, but the ranking they give still has a rank correlation of 0.88 with the ranking under the weights of Table A1, across the 51 states in 1998. So largely the same states come out as strict whether the five rules are weighted in the order the studies give or all count equally.

A.4 Onset and a worked example

A state's index is 0 in every year before its reform. Reform is dated from the first year in which a work-related rule was in force statewide, so a waiver that changed only the family cap, or that ran in a few counties, does not start a state's index. The index is free to fall within a state when a rule was loosened, since forcing it never to fall would fabricate policy. It does so in three states.

Florida in 1997, its first full year under TANF, shows the arithmetic. The work requirement was in force, so $I = 1$. The disregard was 100 percent for three months and the AFDC rule thereafter, so the sustained exit point did not move and $I = 0$. The time limit was 24 months in any 60, so $I = (60 - 24)/42 = 0.857$. The sanction removed the whole grant, so $I = 1$. The exemption ended when the youngest child was 3 months old, so $I = (36 - 3)/36 = 0.917$. The index is

$$S = 0.419 \times 1 + 0.262 \times 0 + 0.157 \times 0.857 + 0.126 \times 1 + 0.037 \times 0.917 = 0.71.$$

B The rules entered together

Table B1 reports the fourth specification of Section 4, the five rules entered together in the preferred specification. Section 5 reads it.

Table B1: The rules entered together

	(1) General health index	(2) Mental health index	(3) Employment index
Work requirement $\times P$	0.047	-0.056	0.106
	(0.056)	(0.164)	(0.086)
	[0.40]	[0.73]	[0.22]
Earnings disregard $\times P$	0.012	0.064	0.221***
	(0.037)	(0.078)	(0.045)
	[0.75]	[0.42]	[0.00]
Time limit $\times P$	0.085*	0.167	0.082
	(0.051)	(0.155)	(0.081)
	[0.10]	[0.28]	[0.31]
Sanction $\times P$	0.034	-0.016	-0.052
	(0.056)	(0.120)	(0.077)
	[0.55]	[0.89]	[0.50]
Exemption age $\times P$	-0.049	0.115	-0.229**
	(0.058)	(0.202)	(0.108)
	[0.40]	[0.57]	[0.03]
Person-waves	40,731	25,136	29,849

Notes: Each column is one regression with the specification of column 3 of Tables 5 to 7. In that specification the strictness-adjusted exposure is replaced by five exposures, one for each rule. A rule's exposure is the years of the 5.75-year window from conception to the fifth birthday, each weighted by the score from 0 to 1 that the rule had in the birth state in that year. Each of the five exposures enters alone and interacted with P. The cells are the interactions. Standard errors clustered on the birth state are in parentheses and cluster-robust p-values in brackets; stars mark significance at the 10, 5 and 1 percent levels (*, **, ***). A cell is the effect of one year under a rule at the setting scored 1, with the other four rules held fixed. Each cell compares a child with $P = 1$ with a child with $P = 0$. The data are the PSID cohorts born 1975 to 1999, and the sample is that of Tables 5 to 7.

C The state EITC check

Table C1: Dropping children exposed to a new or enlarged refundable state EITC

	(1) Preferred column, Tables 5 to 7	(2) Without children exposed to a refundable state EITC started in 1993 or later	(3) Column 2, also treating a rise in a credit after 1989 as a start
General health index			
Exposure $\times P$	0.104*	0.116*	0.117*
	(0.048)	(0.048)	(0.049)
	[0.08]	[0.06]	[0.07]
Exposure	-0.057	-0.061	-0.059
	(0.070)	(0.073)	(0.074)
Person-waves	40,731	40,017	39,840
Mental health index			
Exposure $\times P$	0.158*	0.151*	0.144*
	(0.073)	(0.078)	(0.078)
	[0.05]	[0.06]	[0.08]
Exposure	-0.067	-0.034	-0.056
	(0.082)	(0.084)	(0.084)
Person-waves	25,136	24,886	24,805
Employment index			
Exposure $\times P$	0.053	0.023	0.028
	(0.052)	(0.056)	(0.055)
	[0.36]	[0.68]	[0.62]
Exposure	-0.027	-0.005	0.000
	(0.062)	(0.067)	(0.069)
Person-waves	29,849	29,537	29,444

Notes: Each column of a block is one estimate of column 3 of Tables 5 to 7 for the index named. Column 2 drops every child born in New York, Massachusetts, Colorado, the District of Columbia, New Jersey, Oklahoma or Rhode Island in or after the fifth year before the state's refundable EITC started (1994, 1997, 1999, 2000, 2001, 2002 and 2003). Column 3 treats any rise in a refundable state EITC after 1989 as a start, and so it also drops the matching children in Minnesota (1991) and Vermont (2000). State EITC rules are from the University of Kentucky Center for Poverty Research National Welfare Data. Standard errors clustered on the birth state are in parentheses. The interaction carries its wild-cluster bootstrap p-value from 999 draws in brackets; column 1 takes it from Tables 5 to 7 and columns 2 and 3 from a separate run. Stars mark significance at the 10, 5 and 1 percent levels (*, **, ***), by the bootstrap p-value for the interaction and the cluster-robust p-value for the main effect.

D Where reform happened

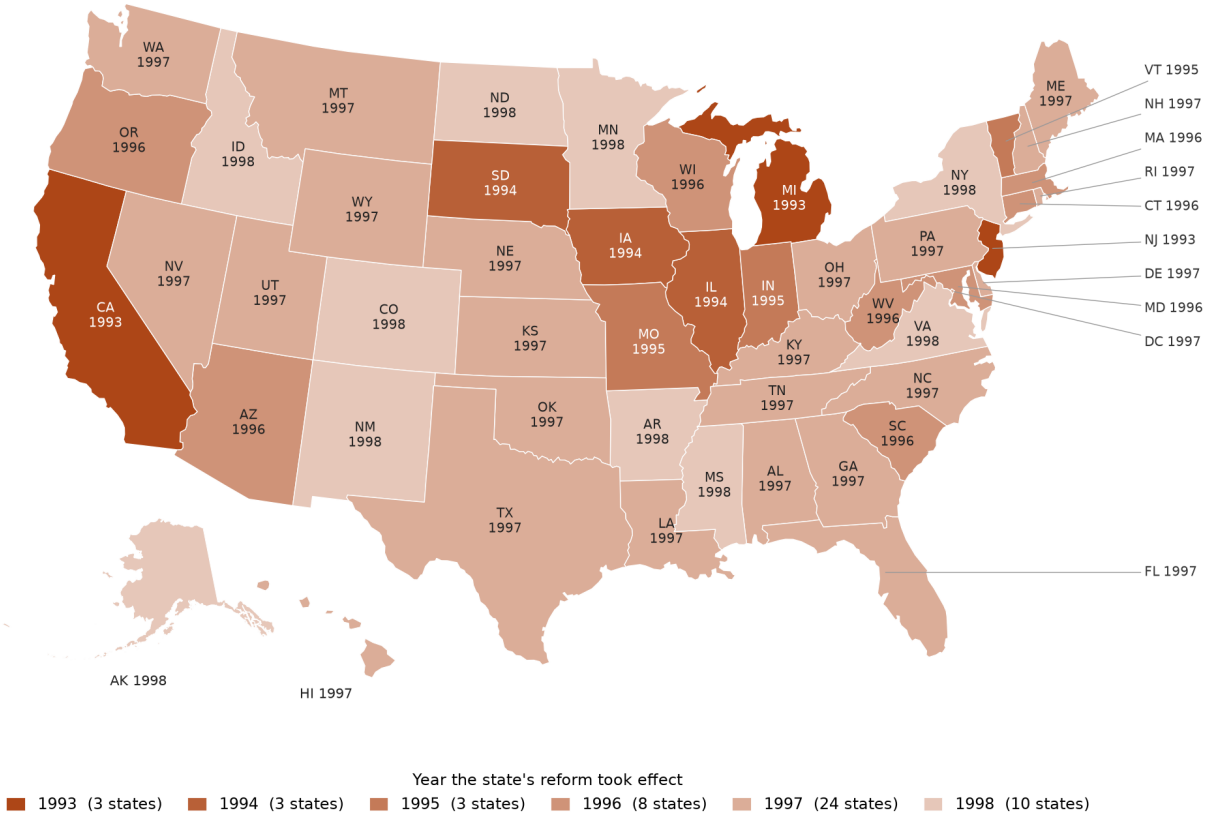


Figure D1: Year each state's reform took effect, 1993 to 1998

Notes: Reform years follow the dating rule of Section 3.3, from the summaries of the Office of the Assistant Secretary for Planning and Evaluation, for the fifty states and the District of Columbia. The legend gives the number of states in each year.

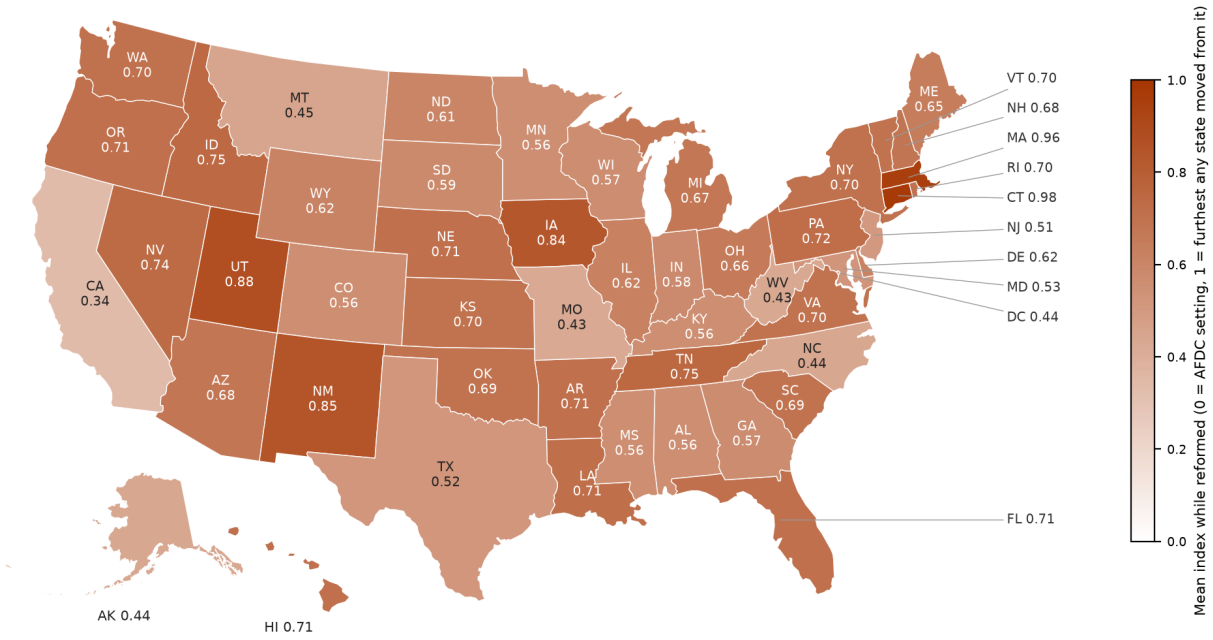


Figure D2: Mean strictness index from each state's first year above zero through 1998

Notes: States first score above zero between 1993 and 1998, so the average covers between one and six years depending on the state. The index is 0 at the AFDC setting and 1 at the furthest any state moved away from it. The rules are from the summaries of the Office of the Assistant Secretary for Planning and Evaluation, for the fifty states and the District of Columbia.

E The predicted probability of AFDC receipt

P is the probability of AFDC receipt that a model predicts from the eight characteristics of the family listed in Section 3. I define these characteristics and AFDC receipt in the same way in the CPS, where the model is fitted, and in the PSID, where it is applied. The unit is the family rather than the household, because AFDC was paid to a family. A family here is a parent with her spouse or partner and her children. AFDC receipt counts AFDC alone and not other cash assistance. A family counts as receiving AFDC in the CPS if it reports welfare income whose source is AFDC, and in the PSID if it reports AFDC income for the year. The PSID values of the eight characteristics are averaged over the interviews in the child's ages 0 to 5, because each child needs one value of each characteristic. Three of the eight characteristics rarely change across those interviews, so their averages are close to their values at any one interview. The head's education is the same at every interview for 87 percent of children, race for 99 percent and marital status for 77 percent.

For the model of P I use gradient-boosted trees, because they predict AFDC receipt better than the alternatives on CPS families held out from the fit. On those families the AUC is 0.904 for the trees, against 0.897 for a logit on the same characteristics and 0.868 for a table of receipt rates within eight groups defined by the head's education, race and marital status. I chose the depth of the trees, the learning rate and the other settings by five-fold cross-validation on the CPS alone, and the 27 combinations I tried differ by at most 1.2 percent in cross-validated log-likelihood. The chosen trees are shallow, and the model stops adding trees once its fit on held-out CPS families stops improving. The PSID played no part in fitting the model or in choosing its settings, so the test in Figure E1 uses families the model never saw.

The model scores only families in which the mother is the head, the head's wife or the head's partner. For mothers who lived in a household headed by someone else, most often their own parents, P comes from a separate model fitted on the 1,507 CPS families that lived in such a household, and a child whose mother lived this way for part of ages 0 to 5 gets a blend of the two models.

I test the gradient-boosted trees on the 7,975 PSID family-years of 1990 to 1993 with a child under 6 and a mother who is the head, the head's wife or the head's partner, and Figure E1 shows the result. For every cutoff on the predicted probability, the figure plots the share of recipients scored above that cutoff against the share of non-recipients scored above the same cutoff. A model with no information would lie on the diagonal, flagging recipients and non-recipients at the same rate. The curve of the trees rises well above the diagonal.

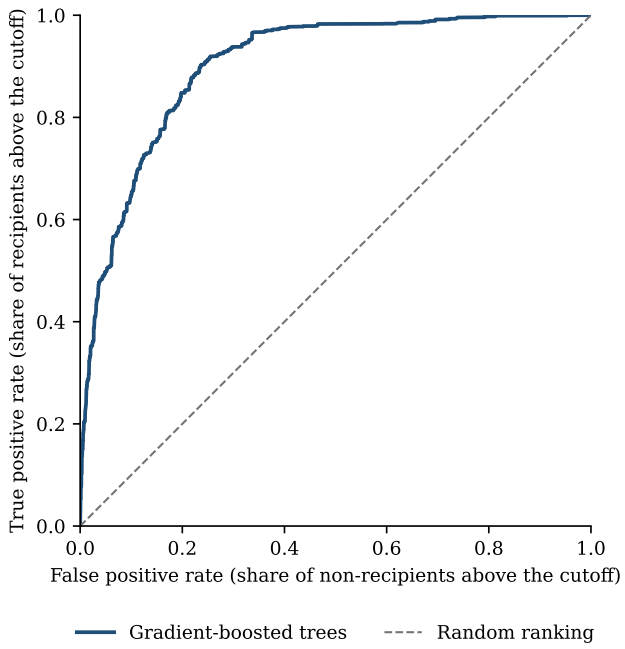


Figure E1: Receiver operating characteristic curve of the model of P

Notes: For every cutoff on the predicted probability of AFDC receipt from the gradient-boosted trees, the curve plots the share of families that received AFDC whose prediction lies above the cutoff against the share of families that did not. The dashed line is the curve of a model with no information. The area under the curve is 0.903. Data are from the PSID, 7,975 family-years of 1990 to 1993 with a child under 6 and a mother who is the head, the head's wife or the head's partner; receipt is AFDC income reported for the year. The model is fitted on the March CPS of 1984 to 1993, and the curve weights each family-year by the PSID family weight.

F The exposure window

Table F1: Exposure measured to age five and to age 17

	(1) Conception to age five	(2) Conception to age 17	(3) Person-waves
General health			
General health index	0.104*	0.035*	40,731
	(0.048)	(0.015)	
	[0.08]	[0.05]	
Self-rated health	0.111*	0.038**	40,713
	(0.044)	(0.012)	
	[0.08]	[0.02]	
Not obese	0.021	0.007	39,607
	(0.020)	(0.008)	
	[0.32]	[0.39]	
Mental health			
Mental health index	0.158*	0.046	25,136
	(0.073)	(0.025)	
	[0.05]	[0.11]	
Distress scale	0.028*	0.012**	23,030
	(0.013)	(0.004)	
	[0.09]	[0.03]	
No psychiatric diagnosis	0.040*	0.006	25,103
	(0.022)	(0.008)	
	[0.07]	[0.50]	
Employment			
Employment index	0.053	0.022	29,849
	(0.052)	(0.020)	
	[0.36]	[0.34]	
Employed	0.010	0.006	29,788
	(0.021)	(0.007)	
	[0.67]	[0.45]	
Annual hours	113.3	21.0	28,350
	(51.7)	(20.2)	
	[0.10]	[0.34]	
Own labor earnings	0.121	0.074	29,307
	(0.191)	(0.078)	
	[0.56]	[0.36]	

Notes: Each entry is β_2, the coefficient on exposure interacted with P, from the preferred specification, column 3 of Tables 5 to 7, estimated on the PSID children born 1975 to 1999. Column 1 measures exposure from conception to age five, as in those tables, and repeats their column 3. Column 2 measures it in the same way from conception to age 17. Exposure is in strictness-adjusted years in both columns, so each coefficient is the effect of one strictness-adjusted year for a child with $P = 1$ against a child with $P = 0$. The indices are in standard deviations of the unexposed children, self-rated health is on a five-point scale, the distress scale runs from 0 to 1 with 1 for no distress, annual hours are hours, earnings are the inverse hyperbolic sine of own labor earnings, and the other outcomes are shares. Every regression carries the fixed effects and controls of column 3, and both windows use the same person-waves, given in column 3. Standard errors clustered on the birth state are in parentheses, and the wild-cluster bootstrap p-value from 999 draws is in brackets. Stars mark significance at the 10, 5 and 1 percent levels (*, **, ***) by the bootstrap p-value.