Intergenerational transmission of welfare dependency: The effects of length of exposure[*]

Oscar A. Mitnik[†]

June, 2010
Preliminary and incomplete - Please do not cite

Abstract

It is well documented that a positive correlation exists between receiving welfare as a child and depending on welfare as an adult. However, previous studies have not been able to explore many aspects of this relationship. This paper uses a unique administrative dataset for California, which follows welfare recipients since their teenage years until early adulthood, to study the causal effects of different lengths of welfare exposure as a child (conditional on welfare receipt) on future welfare dependency as a young adult.

The econometric analysis relies on the estimation of a generalized propensity score (GPS) to remove the biases associated with differences in the observed characteristics of individuals. In addition, for some analyses, family-level unobserved heterogeneity is controlled for by relying on pairs of siblings subject to different lengths of exposure.

The results show that there seems to be no causal effect of length of exposure on future welfare dependency, after taking into account childbearing before age 18. Conditional on childbearing before age 18, there are very small effects of length of exposure on adult welfare dependency, but this dependency is more than twice as large for teenage mothers than for non-mothers. All the results hold when controlling for unobserved heterogeneity. In addition, the length of exposure before age 13 does not seem to affect the probability of childbearing before age 18. The results suggest that, at least regarding the intensive margin on welfare use, policies like time-limits are not likely to reduce the intergenerational correlation of welfare dependency.

JEL Classification: *I38, J62*
Keywords: *Welfare Dependency, Teenage Childbearing, Multiple Treatments*

1   Introduction

There is consensus among social scientists that there is strong intergenerational correlation in welfare participation, in particular for women. The most recently estimate of this correlation in the U.S. puts it at 0.30. This figure was obtained by Page (2004) by computing the difference between the probability that a woman participated in welfare conditional on her parents *having participated* (0.47), and the probability that a woman participated conditional on her parents *not having participated* (0.17).[1] In this paper I focus on the first of these two conditional probabilities, studying the role of *length of exposure* to welfare as a child on the probability (and intensity) of welfare dependency as an adult.[2] The main question I try to answer is whether there is a *causal* relationship between exposure and future dependency. In addition, I explore the effects of the interaction of exposure and teenage childbearing on welfare dependency.

A number of hypotheses have been advanced to explain the intergenerational transmission of welfare dependency. They can be divided into two groups, depending on whether they pose that the welfare system causes intergenerational dependency or not.

Those that hypothesize that there may be a causal effect, have suggested several channels for this effect. One of the channels through which welfare exposure can *cause* future dependency is by lowering the stigma associated to welfare participation for the children of welfare recipients (Antel, 1992; Moffitt, 1992).[3] Another channel works by lowering the transaction or information costs of future participation (Moffitt, 1992). For instance by allowing children to learn to "play the system" (Antel, 1992, Gottschalk, 1996). In addition to these "direct effects", Moffitt (1992) suggests that there could be "indirect effects" operating through the labor supply or human capital investment decisions of the family members

---

[1]Page (2004) calculates these probabilities using data on women in the Panel Study of Income Dynamics (PSID) who were between the ages of 27 and 42 at the time of the 1993 interview. Her analysis uses a relatively longer period than previous studies, which found the intergenerational correlation to be between 0.17 and 0.28. See her study for more details.

[2]I will use the term *welfare* throughout this study to refer both to the Aid to Families with Dependent Children (AFDC) program and the Temporary Assistance for Needy Families (TANF) program, which replaced AFDC after 1996.

[3]Moffitt (1983) proposes a model of welfare stigma in which welfare participation appears as an additional parameter in the utility function, either as a constant component, or as a variable component that depends on the size of benefits. He finds empirical evidence of a strong flat stigma but not of a variable stigma. That is, stigma affects the decision to go on welfare, and is not related to welfare benefits.

(including the child). A model that includes several of these mechanisms, and is in line with many of the arguments in the public debate over the welfare system, is the "welfare culture" model suggested by Mead (1986, 1992). Corcoran (1995) describes this model as emphasizing the negative effects of the welfare system in "values, attitudes and behaviors of parents and neighbors", the decrease in the stigma associated to the welfare system, and the development of "self-defeating work attitudes and poor work ethics." According to the model, these attitudes and behaviors are transmitted from the parents to the children. "In addition, parental welfare recipiency provides children with poor role models for work and marriage. Girls raised in welfare-dependent homes and communities are more likely to drop out of high school, to have illegitimate births, and to go to welfare themselves" (Corcoran, 1995, pp. 244).

As Corcoran and Adams (1995) remark, most of the discussion on the negative effects of the welfare system focuses on long-term dependency. This suggests that analyses of intergenerational dependency should find non-linear effects, where the intergenerational effects become larger when parents rely heavily on welfare. An additional argument for expecting non-linear effects is made by Beaulieu, Duclos, Fortin and Rouleau (2005), who hypothesize that the effects of reduced stigma or learning the system should occur in relatively short exposure periods. If this is correct, the marginal intergenerational effects would be stronger under low intensive use of welfare. Together with the increasing effect of long-term exposure, one would expect to find a fairly non-linear relationship between exposure and future dependency. The main focus of this study is identifying whether those hypothesized non-linear effects exist, as a way of providing evidence bearing on whether a causal relationship underlies the observed intergenerational correlation in welfare dependency.

The main argument against the causal hypothesis is that there are "correlated disadvantages" –it is not welfare exposure which causes intergenerational dependency, but other parental disadvantages (Solon, Corcoran, Gordon and Laren, 1988; Antel, 1992; Corcoran, 1995; Gottschalk, 1990, 1992, 1996). For example, children raised in low-income families might face limited educational opportunities that also affect their future employment opportunities, and which eventually make them dependent on welfare as adults. Alternatively, if the parents live in a neighborhood with poor transportation and limited work opportunities, which make them dependent on welfare aid, these factors may also affect the children as adults if they live in the same neighborhood. In all these cases one would expect children's welfare use patterns to be strongly correlated with their parent's use, but welfare itself would not have any causal effect. One could expect also a correlation between the *intensity* of exposure to welfare and future welfare use, but only because both are themselves correlated with economic and social disadvantages.

The policy implications of these two different sources of intergenerational correlation in welfare dependency are very different. If long-term welfare dependency creates a "culture of welfare", then policies like time limits could be successful in ending long-term dependency

for both the current generation and for future generations.[4] Likewise, if the effects operate through decreasing stigma or information costs, limiting access to welfare may have the effect of decreasing future welfare dependency even for short-term welfare recipients. [5]

On the other hand, if the observed correlation is just the outcome of correlated disadvantages, then the policy implications would be very different. Policies would need to address the barriers faced by both parents and children (for example, those related to educational and employment opportunities, transportation availability, etc). Then, if providing enough income to parents allows them to make the appropriate health and education investments in their children, increasing the benefits associated to welfare may be justified.[6]

To study the potential effects of *length of exposure* to welfare as a child on welfare dependency as an adult, I use an administrative dataset comprised of *all* welfare recipients in California in a 14-year period, which provides high quality information on welfare exposure. To the best of my knowledge, this is the first time that an administrative dataset has been used to analyze the intergenerational correlation of welfare in the U.S.[7]

This data has several advantages. First, because of the availability of relatively long-term information on *monthly* welfare use, it allows for a more detailed analysis of the nature of the relationship between welfare exposure and future use than in the previous literature.[8] Having relatively long observation "windows" reduces the risk of having a "window problem" (Wolfe, Haveman, Ginther and An, 1996), which arises when observing a variable for a short period of time fails to appropriately represent the circumstances of events of the entire period.[9] Second, because the data on welfare use is monthly, it allows for a much better description of both exposure and adult welfare use. All previous studies of the U.S. welfare system have relied on the National Longitudinal Survey (NLS), the National Longitudinal Survey of Youth (NLSY), or the Panel Study of Income Dynamics (PSID). In the earlier years these surveys only had annual indicators of whether the family had participated in welfare, which could potentially lead to overestimating welfare exposure and use. Although later both the NLSY and PSID included monthly measures, these suffer from "seam bias"–a disproportionate number of monthly transitions are observed at the seams of the survey period (Pepper, 2000). Third, the dataset used in this paper is much larger than the previously used datasets, and does not

---

[4]This was one of the arguments that occupied a prominent role in the public discussion leading to the welfare reform of 1996, by which TANF replaced AFDC. Eliminating the "culture of welfare" was one of the main objectives of this reform. In fact, the Personal Responsibility and Work Opportunity Reconciliation Act of 1996 (PRWORA) that implemented the reform makes reference in its findings to the fact that "children born into families receiving welfare assistance are three times more likely to be on welfare when they reach adulthood than children not born into families receiving welfare" (Government Printing Office, 1996).

[5]If this logic is followed to the extreme, as Corcoran (1995) remarks, the policy implication would even be to eliminate welfare altogether.

[6]This argument abstracts of the parent's labor-supply effects of higher benefits.

[7]Beaulieu, Duclos, Fortin and Rouleau (2005) use a similar administrative dataset for Québec, to analyze the effects of exposure in the social assistance program in Canada.

[8]The available information allows to study up to eight years of welfare use during adulthood for some cohorts.

[9]For example, Wolfe, Haveman, Ginther and An (1996) use 10-year windows (between the ages of 6 and 15) as appropriate representations of the entire childhood period.

suffer from typical problems associated to longitudinal surveys like attrition, underreporting and non-response biases.[10]

There are disadvantages associated to the administrative nature of the data used in this paper. First, it does not include many variables usually available in survey data. Second, it is not a nationally representative survey. However, given that California represented on average 20% of the national caseload in the period under analysis, we can be reasonably confident that any results obtained with this data will apply to the U.S. as a whole. Finally, the main disadvantage of this dataset is that it is restricted to welfare recipients only. Hence, it is not possible to compare the welfare dependency of those exposed to welfare as children with that of those not exposed. Nevertheless, understanding the effects of exposure *conditional* on welfare participation it still is very important for the policy debate on the welfare system.

This paper complements the existing literature on intergenerational transmission of welfare dependency, in which the effects of length of exposure have not received much attention.[11] The earlier literature on intergenerational transmission of welfare was surveyed by Moffitt (1992). He found that all studies showed consistent evidence of positive correlation between parental welfare receipt and the later behavior of daughters, but the results could not be interpreted as causal. The main shortcoming of the earlier studies is that they do not control for factors, unrelated to welfare exposure, affecting the welfare dependency of both parents and daughters.

A group of later studies have addressed the issue of causality explicitly. Some of these studies tried to account for unobserved heterogeneity directly in the estimation (Antel, 1992; Gottschalk, 1992 and 1996; An, Haveman and Wolfe, 1993; Borjas and Sueyoshi, 1997). Solon, Corcoran, Gordon and Laren (1988) studied pairs of siblings as a way of controlling for unobserved family factors, while Levine and Zimmermann (1996) resorted to instrumental variables both for mother's welfare participation and income. Pepper (2000) performed a non-parametric bounds analysis where bounds on the intergenerational effect were derived from various identifying assumptions. A few studies (Gottschalk, 1992, 1996, and An, Haveman and Wolfe, 1993) modeled also the probability of childbearing, not only welfare use by daughters. Most studies found a causal effect of mothers' welfare participation on daughters' welfare participation, although the magnitude of the effects varies widely.[12] Only a few studies considered explicitly the effects of length of welfare exposure (McLanaham, 1988;

---

[10]This point is made by Beaulieu, Duclos, Fortin and Rouleau (2005) in their study using Canadian administrative data.

[11]The literature on intergenerational dependency is part of a large literature on the effects of parental welfare receipt on a variety of children's outcomes. For a survey of the earlier studies in this literature, see Haveman and Wolfe (1995). For an analysis of the effects on children outcomes of welfare reform, see Grogger and Karoly (2005). For a recent survey of the intergenerational mobility literature see Black and Devereux (2010).

[12]Antel (1992) found positive effects as well as did Borjas and Sueyoshi (1997). An, Haveman and Wolfe (1993) found positive effects in both the probability of having a teen-out-of-wedlock birth, and on AFDC participation, conditional on childbearing, as did Gottschalk (1996). Solon, Corcoran, Gordon and Laren (1988) found no effect of parental welfare receipt, while Levine and Zimmerman (1996) found a very small or zero effect. Pepper (2000) found a variety of effects, depending on the identifying assumptions.

4

Gottschalk, 1992, 1996; Pepper 1995, 2000), but no clear pattern emerges from the results. As mentioned above, all these studies used either the NLS, NLSY or PSID, thus suffering from data limitations in measuring exposure, and from relatively small sample sizes. This greatly hindered the scope of their analyses of the effects of length of exposure.[13] Finally, the study of the Canadian social assistance program by Beaulieu, Duclos, Fortin and Rouleau (2005), employed administrative data on assistance recipients similar to the data used in this paper. Using a model that jointly estimated parental and children assistance participation, they found a U-shape non-linear relationship between parental receipt and children receipt, which implies larger effects of exposure at both low and high levels of parental receipt.

The empirical strategy of this paper is different from those usually utilized in the prior literature. Using program evaluation terminology, I treat length of exposure to welfare as a multivalued treatment for which a *causal* treatment effect (or *dose-response function*) can be estimated. A key component in this methodology if the estimation of a *generalized propensity score* (GPS). The GPS, introduced by Imbens (2000), permits, as in the binary-treatment case, removing the biases associated with differences in the observed characteristics of individuals. I estimate this score by using a flexible parametric specification, which succeeds in eliminating a large fraction of the differences in the observed characteristics of the women subject to different lengths of exposure. As in much of the previous literature, only the effects for daughters of welfare recipients are analyzed, because they are the most likely to depend on welfare as adults (around 80% of adult welfare recipients are single mothers). Additionally, to deal with potential unobserved family characteristics correlated with length of exposure as a child and with welfare use as an adult, I conduct analyses on a subsample of pairs of sisters subject to different lengths of exposure. This allows the removal of any unobserved time-invariant family-level heterogeneity.[14]

The data shows that there is a strong positive correlation between welfare exposure as a teenager and welfare use as a young adult. Even after adjusting for covariates, the probability of any welfare use by women ages 19 to 22 with welfare exposure of only three months at ages 13 to 17 is around 30%, compared to 45% for women with a length of exposure of 50 to 60 months. Nevertheless, once the causal treatment effects are estimated, this difference mostly disappears –the probability of any welfare use as adult remains around 40% regardless

---

[13]The study by Pepper (2000) is worth mentioning further because is the only prior study to explicitly consider the effects of length of exposure on welfare dependency. Even though it suffers from the sample size limitations imposed by the PSID, it systematically studies the effects of different parametric and non-parametric assumptions on deriving lower and upper bounds for the intergenerational effect of welfare receipt for women. The results of his study depend on the assumptions one is willing to make. Not making any assumptions, the bounds on the effect of an additional 3-4 years of exposure (compared to zero exposure) include zero. However, under some assumptions (ordered outcomes and exogenous local unemployment rates, or under exogenous duration of parental receipt) the effects of extra exposure become positive.

[14]The strategy of comparing pairs of siblings is also used in the study of intergenerational correlation of welfare participation by Solon et al. (1988), although its implementation and interpretation is different to the one of this paper. Levine and Zimmerman (2005), on the other hand, use a similar strategy for their study of the effects of welfare exposure on child development.

of length of exposure as a child.[15] When analyzing the effects by race/ethnicity, the story is similar – there is no effect of length of exposure. However, the *levels* of welfare dependency are much higher for blacks (and to a lesser extent for Hispanics) than for other race groups.

A finding of special interest is that teenage childbearing is associated with a large positive effect on future welfare dependency. Indeed, this appears to be the most important determinant of future welfare use. Given that childbearing occurs during the period of exposure, it can be considered an additional outcome variable, as previously done in the literature (e.g. An, Haveman and Wolfe, 1993; Gottschalk, 1996). Because of concerns regarding the endogeneity of length of exposure between ages 13 and 17 and childberaing in those years, I rely on a sample of individuals for which exposure between the ages of 9 and 12 is available. The results indicate that the probability of childbearing between ages 13 and 17 does not depend on the length of exposure between ages 9 and 12.

All together, the results of this paper indicate that, conditional on welfare participation, there does not seem to be a causal effect of the length of exposure on intergenerational welfare dependency, or on the probability of childbearing. This conclusion needs to be interpreted with caution, because it only refers to the intensive and not to the extensive margin (the participation decision) of welfare exposure. Thus, it is not possible to rule out explanations of intergenerational correlation of welfare dependency that admit formulations in which length of exposure plays no role (e.g. stigma reduction, learning the rules of the system).

In terms of welfare policy, the results indicate that policies like time-limits, at least regarding the intensive margin welfare use decision, may be not likely to reduce the intergenerational correlation of welfare dependency. In addition, policies aimed at reducing the prevalence of teenage pregnancies (for example, by providing sexual education and contraceptive services) could have an important effect in lowering welfare dependency.

This paper is organized as follows. Next section describes the data, while the third section describes the empirical strategy. The fourth section presents the estimation results, and the final section concludes.

2   Data

This paper employs an administrative longitudinal dataset for welfare recipients in the State of California, the Medi-Cal Eligibility Determination System (MEDS) dataset. It contains information for all the individuals that have ever received welfare in California (which represented around 20% of all welfare recipients in the U.S. in the period) from January 1987 to June 2001.[16] It includes demographic characteristics and monthly welfare use histories

---

[15]The story is similar if the outcome considered is the number of months of welfare use. In that case, the covariates-adjusted number of months is 8 for women with three months of exposure and around 14 for those with more than 50 months of exposure. However, after estimating the causal treatment effects, the average number of months of welfare use remains around 11, regardless of the length of exposure.

[16]As noted in the Introduction, I use "welfare" as a general way of referring to the AFDC and TANF Programs. As Klerman and Haider (2004) explain, MEDS is a monthly roster of all individuals eligible for

for assistance units which permits the construction of very detailed measures of exposure to welfare for different groups.

The data then offers big advantages with respect to the U.S.-based datasets used by previous studies on intergenerational welfare dependency (NLS, NLSY and PSID), because it has much more detailed welfare use histories. In addition, large sample sizes permit the estimation of multivalued treatment effects (dose-response functions) which would be very hard to estimate with precision with the typical sample sizes of the longitudinal surveys previously used in this literature.

However, the administrative nature of MEDS has some disadvantages and its use is subject to some caveats. First, the information in the MEDS dataset is on "assistance units", or cases, as defined by welfare offices, which for a majority of the cases will be equivalent to a family, but not necessarily. Second, similarly to other administrative datasets, important variables typically available in surveys, like income and education, are not available. Third, the information on welfare spells is extremely detailed, but it is only for California: welfare use in other states will not be captured by this dataset. This could potentially underestimate both the exposure length as a child, and the welfare dependency as an adult. As long as the inflows and outflows remain relatively constant during the analysis period, this should be a minor concern.

This study concentrates only on daughters of welfare recipients because this is the group that is more at risk of welfare dependency in adulthood (a large majority of welfare recipients are single mothers). Following the strategy of previous studies (e.g. Pepper, 2000), both the exposure to welfare as a teenager, and the welfare dependency as adult are defined over fixed-length "windows". In the main analyses the windows are from ages 13 to 17 for exposure, and from ages 19 to 22 for dependency. In additional analyses, variable welfare dependency windows from ages 19-23 to 19-26 are analyzed.

The sample is composed of all women in the MEDS dataset born between January 1974 and June 1978, and who where welfare recipients for at least three months between ages 13 to 17.[17] These "birth year" cohorts satisfy the constraint that data is available for them both for ages 13-17 and for ages 19-22. In addition, because the MEDS dataset starts in 1987, this implies that, for example, for the cohort born in 1978 exposure data is available as far back as age 9, and that the outcome windows can be extended (for the 1974 cohort) up to age 26.

A total of 219,081 women were selected that satisfied both the birth cohort and exposure criteria. As with any administrative dataset inconsistencies and lack of information in some variables implied that not all of those observations were usable. Indeed, 9,606 observations had to be dropped because they lacked valid social security number information, which is

---

Medi-Cal, California's Medicaid program. Because welfare recipients are categorically eligible for Medi-Cal and that source of eligibility is noted, MEDS provides a monthly roster of the welfare population in California.

[17]Cases in which a person received welfare for less than three months in the period are not considered because one- or two-month spells are many times just the result of lags in the administrative process by which individuals are added to and dropped from assistance units, according to analysts in the California Department of Social Services. Therefore, including those cases could potentially introduce measurement error bias.

necessary to track people over time, as well as 27,918 were dropped for lacking necessary demographic information. In addition,12,518 individuals were dropped for belonging to assistance units with an implausible structure to be considered a family.[18]

An additional group of individuals that was excluded from the analyses was that of women whose first ever appearance in the data as a welfare recipient was as teenage mothers (i.e. mothers of ages 13 to 17). Given that welfare exposure as a child is considered as being a reflection of the actions of the parents, not of the children, including this group does not seem appropriate. These women are making a participation decision, not being "exposed" to welfare. Then, it is hard to argue there for the existence of an independent effect of exposure to welfare as a teenager on welfare dependency as an adult. In contrast, all those women which enter welfare as dependent children and successively become teenage mothers are included in the analyses, because, as it will be explained in section 3.4, teenage childbearing will be considered as an additional outcome in some of the analyses.

Identifying teenage mothers, however, is challenging due to the absence of relationship codes information, and it requires the imposition of a set of rules based on the available data. The rules used are the following: a teenage girl is classified as a teenage mother if: a) her difference of age with the youngest person in the case was greater or equal to 14 years; *and* b) she continued receiving welfare after turning 19 (maximum cutoff age for eligibility of dependent children even if attending school) *or* she was the oldest female in the case, *or* she was less than 14 years younger than the oldest person in the case. Using these rules 12,193 women are identified as mothers whose first time on welfare was as a teenage mother, and therefore were dropped from the analyses, while 33,309 women are identified as teenage mothers with positive exposure prior to becoming teenage mothers, and therefore were kept in the analysis sample.[19]

In summary, the analysis sample to which I refer as "full sample" is composed of 156,846 women born between January 1974 and June 1978, and who were welfare recipients for at least three months from ages 13 to 17. In many of the analyses other three subsamples are extracted from this full sample. Two samples will be referred to as "teen mothers sample" and "non mothers sample". The first one is formed by the 33,309 women who become teenage mothers after being exposed to welfare, while the second one is the complement (123,537 observations). Finally, the "siblings sample" is composed of those cases in which two women (part of the full sample) belong to the same family (assistance unit). In the very few cases where more than two women belong to an assistance unit, only the two oldest women are selected. There are 24,546 individuals in this sample (i.e. 12,273 pairs of siblings).

County-level data was added to the data based on the county of residence (according

---

[18]Assistance units which were considered problematic were those with more than 2 adults in the case, and those for which the number of kids between the ages 0 to 18 was more than 9.

[19]Note that even those women whose first time on welfare is within the six months prior to the birth of their baby were dropped from the analyses, because welfare eligibility rules allow a women to receive benefits since the second trimester of the pregnancy.

to welfare records) or the latest recorded county for the months in which the individuals are not welfare recipients. The county-level data includes information on the share of total county expenditures in public health programs[20] and data on local economic conditions. Unemployment rates are the average for the exposure period of monthly rates by county from the Bureau of Labor Statistics (BLS) Local Area Unemployment Statistics.[21] Employment and quarterly earnings in the retail sector come from the Quarterly Census of Employment and Wages (QCEW) data produced by BLS,[22] while annual county population estimates come from the U.S. Census Bureau.

Table 1 presents descriptive statistics for the four samples considered, for the different outcome variables, as well as for the exposure variable and all the covariates used in the empirical analysis. It is clear that the teenage mothers sample is very different from the point of view of the outcome variables, with much higher welfare use. at the different outcome windows. In terms of exposure, however, it is higher than for the non mothers, but the difference between the samples is much less pronounced. Regarding outcomes and covariates, except for an over-representation of Hispanics and a slightly higher percentage of teenage mothers, the full sample and the siblings sample are very similar (of course, by definition that is not true regarding the average number of siblings), even though exposure appears as higher in the siblins sample. The teen mothers sample appears as rather different to the non mothers sample also regarding covariates, with 40% more blacks and 25% more Hispanics, and a higher percentage of English speakers. Also the 1974 cohort is over represented.

The bottom of Table 1 presents information on the average of county level variables faced by these women during ages 9 to 12, 13 to 17, and 22 to 26. These measures are based on the county of residence (according to welfare records) or the latest recorded county for the months in which the individuals are not welfare recipients. The expenditures variables will be included in the empirical analysis as proxies for the availability of family planning, contraceptive and abortion services in the counties. The local economic conditions are relevant because the literature has shown (e.g. Hoynes, 2000) that county-level unemployment, and sectorial employment and earnings affect both the rates of entry to and exit from welfare. Even though it is not apparent from the table, there is wide variation in these measures at the county level and over time.

---

[20]This share was constructed as the ratio of the total county expenditures in public health over the total county expenditures by fiscal year as provided by the California State Controller's Office. The amounts where first expressed in monthly terms by dividing by 12 and then averaged over the age year of each individual.

[21]In 2004, for the period starting in January of 1990, BLS changed the methodology of estimating Local Area Unemployment Statistics, with new models, featuring real-time benchmarking to monthly Current Population Survey (CPS) employment and unemployment totals. Unfortunately this new statistics do not go back to 1987, the first year for which welfare data is available in this paper. Therefore, by suggestions from analysts at the Labor Market Information (LMI) Division of the California Employment Development Department (EDD), a pre-1990 series of unemployment rates calculated by LMI was "bridged" to the new BLS data by applying to the new series the monthly observed changes in unemployment rates for the period 1987-1990, to generate a series for that period which is consistent in levels with the new series.

[22]The QCEW is a quarterly count of employment and wages reported by employers covering 98 percent of U.S. jobs, available at the county, MSA, state and national levels by industry.

Finally, to complete the descriptive analysis of the data, Figure 1 shows for each sample the distribution of the exposure variable. As it is apparent from the figures, the histograms are quite similar, except for the teenage mothers sample, where there is a higher proortion of women with longer exposure.

3 Empirical strategy

The nature of the problem and the richness of the data allow applying methods recently developed in the program evaluation literature to deal with continuous treatments. There has been increasing interest in the literature in both multi-valued (Imbens, 2000; Lechner, 2001) and continuous treatments (Imbens, 2000; Hirano and Imbens, 2004; Behrman, Cheng and Todd, 2004; Flores, 2004; Imai and van Dyk, 2004).

I follow the approach of Hirano and Imbens (2004), which seems the best suited to this study because it allows for estimation of the whole *dose-response function* of an (ordered) continuous treatment. In this case this implies to trace the response of the outcome of interest to every value of the treatment variable, the exposure to welfare.

The approach proposed by Hirano and Imbens (HI from now on), which extends the typical potential outcomes approach in program evaluation to continuous treatments, is presented in the next subsection. I will follow closely both their presentation and notation.

3.1 Bias removal using the GPS

Let's suppose we have a random sample of units, indexed by $i = 1, ..., N$. For each unit i there is a set of *potential outcomes*, $Y_i(t)$, for $t \in \mathcal{T}$, referred to as the unit-level dose-response function. We want to estimate the average dose-response function $\mu(t) = E[Y_i(t)]$. For each unit i we have a vector of covariates X_i, and the level of the treatment received T_i. We observe X_i, the treatment received T_i and the potential outcome corresponding to that treatment, $Y_i = Y_i(T_i)$. Imbens (1999, 2000) generalizes the unconfoundness assumption for binary treatments made by Rosenbaum and Rubin (1983) to the multivalued case, calling it *weak unconfoundness*. It implies that

$$Y(t) \perp T | X \ \textit{for all } t \in \mathcal{T}.$$

Calling $r(t, x) = f_{T|X}(t|x)$, the conditional density of the treatment given the covariates, then, Imbens (1999, 2000) define the *Generalized Propensity Score* (GPS) as

$$R = r(T, X). \tag{1}$$

This GPS has a balancing property similar to the standard propensity score in the binary case; together with the assumption of unconfoundness, this implies that the assignment to treatment is unconfounded given the GPS. It is important to keep in mind the distinction

between two different random variables once one estimates the GPS: the probability that an individual gets the treatment she actually received, $R_i = r_i(T_i, X_i)$, and the probability she receives a particular treatment t conditional on her covariates, $R_i^t = r_i(t, X_i)$. Clearly, $R_i^t = R_i$ for those units with $T_i = t$.

There are two alternatives to exploit the GPS in estimation, to remove biases associated to observables. The first one, proposed by Hirano and Imbens (2004) is to follow two steps, or a partial mean approach (Newey, 2004). First, it is necessary to estimate the conditional expectation of the outcome as a function of the treatment level T and the GPS, R:

$$\beta(t, r) = E[Y|T = t, R = r]. \tag{2}$$

Second, it is necessary to estimate the dose-response function at each particular level of the treatment. This is attained by averaging the conditional expectation function over the GPS at each particular level of the treatment,

$$\mu(t) = E[\beta(t, r(t, X))]. \tag{3}$$

As Hirano and Imbens (2004) stress, $\beta(t, r)$ does not have a causality interpretation, but $\mu(t)$ corresponds to the value of the dose-response function for treatment value t, which compared to another treatment level t' does have a causal interpretation. Note, that instead of estimating $\beta(t, r)$ one could estimate directly $\beta(t, X)$, which implies using a partial mean approach directly on covariates. In the results below I use that estimator to compare with the GPS-based estimator.

In addition to employing the GPS within a partial mean framework to estimate $\mu(t)$, the GPS can also be used to control for covariates using a weighting approach (e.g., Imbens, 2000; Cattaneo, 2009).[23] Similar to the binary treatment case, it is possible to weight the observations receiving a given treatment level t by the probability of receiving the treatment they actually received conditional on X (i.e., R_i). More specifically, in our context we can write $\mu(t)$ as (Imbens, 2000):

$$\mu(t) = E\left[\frac{Y_i \cdot 1(T_i = t)}{R_i}\right]. \tag{4}$$

The intuition behind weighting by R_i is creating a sample in which the covariates are balanced across all treatment arms (or sites), and then calculating the average outcome for those units with $T_i = t$ in that sample to estimate $\mu(t)$. In the binary treatment literature, the weights implied by (4) are usually normalized to add to one (e.g., Imbens, 2004; Busso

---

[23]See Flores et al. (2009) for a discussion of weighting-by-the-GPS estimators in a continuous treatment setting.

et al., 2009a, 2009b). Thus, the *inverse probability weighting (IPW)* estimator is given by

$$\widehat{\mu(t)}^{ipw} = \left[\sum_{i=1}^{N} \frac{Y_i \cdot 1\,(T_i = t)}{R_i} \right] \left[\sum_{i=1}^{N} \frac{1\,(T_i = t)}{R_i} \right]^{-1}. \tag{5}$$

Cattaneo (2009) analyzes the asymptotic properties of IPW estimators such as (5) when the GPS is nonparametrically estimated using a series-based estimator, and shows that under certain conditions, these estimators are asymptotically normal and efficient, in the sense of achieving the semiparametric efficiency bound. Note that, similar to the binary treatment case, $\widehat{\mu(t)}^{ipw}$ for $t = 1, \ldots, k$ equal the coefficients in a weighted linear regression of Y_i on the set of k dummy variables $1\,(T_i = j)$, with weights equal to $w_i = \sqrt{1/R_i}$.

In a parametric context, estimators combining IPW and linear regression share a "double robustness" property, which states that these estimators are consistent as long as either $E\,[Y_i|T_i, X_i]$ or the GPS is correctly specified (e.g., Robins and Rotznitzky, 1995; Scharfstein et al., 1999; Wooldridge, 2007). Thus, I implement this estimator by running the above mentioned regression, including covariates.

In summary, the method requires estimating the GPS and the $\beta()$ conditional expectation function. To estimate the GPS a flexible parametric method is followed in this paper, using a multinomial logit regression to estimate the probability of each treatment level. Using a multinomial logit has the advantage of not imposing many assumptions of the shape of the distribution function. A disadvantage of using a multinomial logit is that it is very difficult to estimate more than a certain number of categories. To avoid that problem, I discretized the exposure variable in groups of two months, leaving the last month by itself. This greatly reduced the computational burden, without affecting the final results.

The second step is to calculate the conditional expectation function of Y_i, given T_i and R_i as a flexible function of those two arguments. As in HI, a quadratic approximation is used:

$$E[Y_i|T_i, R_i] = \alpha_0 + \alpha_1 T_i + \alpha_2 T_i^2 + \alpha_3 R_i + \alpha_4 R_i^2 + \alpha_5 T_i R_i \tag{6}$$

where for each individual the observed T_i and estimated GPS $\hat{R}_i$ is used, to estimate this equation by OLS. To obtain the whole dose-response function now, for *each* treatment level t, is necessary to estimate the average potential outcome at treatment level t as

$$\widehat{E[Y(t)]} = \frac{1}{N} \sum_{i=1}^{N} \left(\hat{\alpha}_0 + \hat{\alpha}_1 t + \hat{\alpha}_2 t^2 + \hat{\alpha}_3 \hat{r}(t, X_i) + \hat{\alpha}_4 \hat{r}(t, X_i)^2 + \hat{\alpha}_5 t \hat{r}(t, X_i) \right). \tag{7}$$

Standard errors are obtained by bootstrapping to take into account the estimation of $\hat{R}$ and the $\hat{\alpha}$'s.

3.2 Controlling for unobserved heterogeneity

Given the availability of data on siblings who experienced different levels of exposure (more than 40% of the siblings experienced differences in the level of exposure of at least 6 months), one can resort to comparing siblings as a way to control for unobserved family-level heterogeneity that does not change over time. This is particularly important in this case, because the data does not have many of the variables that have been used in the literature to account for the effect of family and environmental factors. Still, because this can only control for time-unvarying factors, any variable factors (like income) will be unaccounted for. Nevertheless, the siblings comparisons should provide the most robust results.

This comparison of siblings is implemented in the estimation by using family fixed effects in the estimation of (6). This provides an estimation of the conditional expectation function of Y, which arises from the within-siblings differences in exposure levels.

3.3 The role of overlap

In the program evaluation literature, in the binary-treatment case, it is a well-known fact that it is necessary to have overlap in the covariates distribution of the treatment and control groups to obtain estimators that are unbiased (see Heckman, Ichimura and Todd, 1997; Dehejia and Wahba, 1999). In fact, one of the key reasons why the methods based on the propensity score succeed in eliminating biases is because they imply the comparison of individuals across the control and treatment that are indeed comparable.

A similar issue arises when estimating treatment effects with multi-valued treatments using the GPS; one would like to be sure that is comparing individuals that are comparable. However, it is much more complicated to introduce the concept of overlap in this framework. In this paper I deal explicitly with the issue of overlap in the covariates (or as it is also known "common support condition") by imposing a criteria for whether observations satisfy the common support condition or not.

In the binary treatment literature there is no agreement on a criteria to use. Recently, Crump, Hotz, Imbens and Mitnik (2007) have suggested a practical rule, based on efficiency, by which a lower and upper cutoff points are applied to the propensity score distribution. Applying a similar method to the case of continuous treatments is more complex because here we would like to impose a condition in which at *every treatment level* all the individuals should be comparable.

With that logic in mind, I implement the following procedure to try to guarantee overlap. Using the same notation of above, $\hat{R}_i$, the estimated GPS for every individual with a particular treatment level T_i, is compared to $\hat{r}(t = T, X_i)$, the predicted probability of t=T, for all the individuals which received a treatment level different from T_i. Then, individuals which received a treatment level T are considered to satisfy the overlap condition if their GPS is within the interval determined by the highest minimum and the lowest maximum observed across several distributions. Those distributions are the predicted probabilities of treatment

T, calculated for individuals that did not receive treatment level T. For example, let's take individuals with treatment level T=10. For each group of individuals with treatment levels *different* from T=10, we calculate the minimum and maximum of their distribution of predicted probability of receiving treatment T=10. The highest of the minimum values across all those distributions, and the lowest of the maximum values across all those distributions, form the limits of the interval within which the GPS for the individuals that received treatment T=10 has to fall to be considered within the common support condition. This is repeated for every level of treatment, until a set of individuals that are comparable across all treatment levels has been determined. Note that the intervals can also be created by using percentile cutoffs, instead of the minimum and maximum values.

Intuitively, what this condition does is to make sure that at each treatment levels, only individuals that are comparable will be used to estimate the treatment effects.

3.4   The role of teenage childbearing

As it is clear from Tables 2 and 3, and from Figures 3 and 4, the women who become teenage mothers during the period of exposure to welfare are both different in their observable characteristics from the non mothers, and present a very different pattern of welfare dependency. One approach to dealing with this is just to estimate the dose response functions separately for both groups, and analyze them as with any other categorical variable for which one would estimate separate treatment effects (like race).

However, this does not address the key issue which is that teenage childbearing is not an exogenous event. As it is shown by Wolfe, Wilson and Haveman (2001), teenage childbearing is an optimal choice driven by the comparison of its benefits and costs. Indeed, the generosity of the benefits of the welfare system has been suggested as one of the mechanisms by which welfare affects the childbearing decision. The evidence on the link of welfare benefits with the childbearing decision is very weak (Acs, 1996; Hoynes, 1997; Wolfe, Wilson and Haveman, 2001; Schaefer, Hamersma and Vander Veen, 20002), so one may be justified in simply doing separate analyses for mothers and non mothers.

Nevertheless, in this paper *exposure* is assumed to be an exogenous variable for a daughter, given by the decisions of her parents. Once a woman becomes pregnant she is eligible for welfare benefits on her own, independently of her parents decisions, which implies that any observed welfare use after childbearing cannot be considered exogenous.

One alternative would be to consider exposure only up to the point where a teenager becomes pregnant. That solution is also problematic, though, because it does not take into account that we cannot observe the counterfactual level of exposure if she had not become pregnant. In this way, trying to infer the effect of exposure from daughters who become pregnant very early on in the analysis window would overestimate this effect.

What one would like to do is to identify the pure effect of exposure to welfare on future welfare use, for the average woman. That is, the effect net of how prone a woman is to

become pregnant (or carry on a pregnancy). Black and Smith (2004) face a similar problem when estimating the wage effects of college quality. The standard approach in that literature is to include as a covariate years of schooling. But, college quality has both an effect on years of schooling and on wages. They resort to estimating the propensity score with and without that variable and find that it does makes a difference, and report the results under each situation.

In this paper I will follow a similar strategy, but instead of including teenage motherhood in the estimation of the GPS, I will allow a more flexible approach by estimating the GPS separately for both groups, as well as treatment effects for both groups. Those results would need to be considered not as causal treatment effects, but just informative of the nature of the relationship. In addition, I will perform some analyses where I only use the years of exposure prior to teenage childbearing, to assess if the results are sensitive to the exogeneity of exposure assumption.

The results of all these alternative specifications are presented in the next section.

4   Results

The first step is to estimate the GPS. For that, as described in the previous section, a multinomial logit model was estimated to predict the probability of each treatment level, for each individual, using the variables listed in Tables 4 and 5.[24] The estimated GPS for each individual, and the predicted probability of exposure at each treatment level, were then used to identify the group of observations that satisfy the common support condition. Both, min-max rules and percentile-based rules were used, selecting the ones that seemed to make the overlap better (see the discussion below).

Figure 5 shows for the four samples (full, siblings, mothers, non mothers) the distribution exposure and the 95th and 5th percentile of the estimated GPS, for individuals satisfying the common support condition.

Before analyzing the estimated treatment effects, one would like to know if the GPS is working properly in terms of eliminating the biases associated to differences in observed characteristics of the individuals. This is analyzed in the following subsection.

4.1   Testing for GPS balancing of covariates

The typical approach in a binary treatment case to assess the specification of the propensity score is to compare the means of the treated versus control units before and after matching or blocking or weighting by the propensity score. Here it is more complicated because the treatment is continuous. HI propose blocking on both the treatment variable (exposure to welfare in this case) and on the estimated GPS. After applying the common support

---

[24]As explained above, to simplify the estimation of the multinomial logit model, months of exposure were grouped by two months.

condition, I apply their idea in the following way. First, the exposure variable was divided in four groups, and within each group took the predicted GPS for the median of the exposure variable. Second, each group was divided into 20 blocks by percentiles of that predicted score evaluated at the median (using only the individuals who belong in the group). Once the cutoff points for these percentiles were calculated, inside *each* of these blocks one can calculate the difference of means of the individuals within the block compared to all other individuals who have an estimated value of the GPS such that they belong to that block, but have a exposure level different to the one being evaluated. The idea is to test if for each of these blocks the difference in means of the individuals belonging to the exposure group is statistically significant with respect to all the individuals with different exposure level, but similar GPS. This was implemented by creating dummy variables for each of the percentiles of the GPS distribution, which were interacted with a dummy for belonging to the exposure group under analysis, and running a regression of the covariate on those dummies. The weighted sum (by the size of the group) of those 20 coefficients represents the difference of means of interest. This needs to be repeated for *each* exposure group and for *each covariate*. If the GPS is balancing properly the covariates, we would expect all those differences of means to be statistically not significant.

Tables 4 and 5 present the results of performing these calculations on each of the four samples of interest. I will analyze only Table 4, panel A (full sample), the other tables show similar results. To facilitate the comparison of the differences of means, the first and second column in each table show the average and standard deviation for each covariate. The differences, though, are expressed in terms of standard deviations. Note also that the number of observations differ in each sample with respect to the numbers reported in Table 1, because the common support condition has been imposed.

The left panel of the Table ("unadjusted") shows the standardized difference of means between all the individuals in the particular exposure group compared to all other individuals. It is clear that for some variables the covariates are very unbalanced. In the right panel ("adjusted") the differences of means are calculated now by the described procedure of blocking on the GPS. For all the exposure groups, the difference of means decreased substantially. Note that many of the differences, even though they are small, are still statistically significant. This is a consequence of the large dataset, which generates very small standard errors. Still, it is clear that the GPS is working well in reducing (and eliminating) differences in observed characteristics of the individuals.

Next we analyze the estimated treatment effects.

4.2   Treatment effects

Treatment effects (dose-response functions) were estimated for the four samples, based on exposure for ages 13-17 and welfare use in ages 19-22.[25]   The estimated coefficients and

---

[25]The estimated regression functions analog to (6) are presented in Appendix Table A1.

standard errors are presented in Table 6, while a graphical representation of those same effects is presented in Figures 6 and 7. It is clear from the figures that there is no causal effect of additional length of welfare exposure on welfare use – the dose-response functions are almost completely flat. This is true both for any welfare use (top panels in each figure) and for the number of months of welfare use (bottom panels in each figure). Figure 6 also shows that it is almost the same whether one uses the full sample, or restricts the analysis to the siblings sample, although the siblings analysis shows some differences for the highest levels of exposure. As explained above, the results for the siblings sample control for time-invariant unobserved family heterogeneity by calculating (6) with family fixed effects. They suggest that, at least for the period analyzed, unobserved family level heterogeneity is not a great concern.

Figure 7 shows for comparison purposes the dose-response functions for the non mothers and teenage mothers samples. For the non mothers the effects are also flat, and at lower level than for the full sample. For the mothers there is a very small positive slope, but more importantly the extent of welfare dependency (both in terms of probability of welfare use and number of months of welfare use) is close to three times the levels for the non mothers sample (note the differences in scales in the graphs). As it was explained in the previous section, dividing the analysis in these two samples does not take into account the potential endogeneity of teenage childbearing and exposure. Still, it seems like an analysis worth performing.

Treatment effects by race/ethnicity

Figures 8 and 9 present the results of estimating dose-response functions by race/ethnicity groups.[26] Both for probability of any use and for months of use, the effects are again flat with respect to the length of exposure. However, it seems interesting to analyze the differences in levels between the groups. The highest levels of dependency is for blacks, followed by Hispanics, whites and other races (which has levels less than half of those for blacks). The size of the differentials between blacks and the other groups in months of use is striking, almost two times higher.

Treatment effects for alternative windows of exposure and use

Figure 10 present the results on the probability of welfare use of making the welfare use windows longer.[27] The effects of marginal exposure are mostly zero, consistent with the previous results.

Figures 11 and 12 show the effects of starting the exposure period at 10, 11 and 12 years instead of 13 years. Extending the window of potential exposure does not seem to have any large effect on the levels of welfare dependency. If anything the longer the potential exposure

---

[26]To save space no additional tables with treatment effects will be reported, only the figures. Those tables are available upon request.

[27]The results for months of exposure are not shown in this subsection because the patterns are essentially identical to the ones for any welfare use. The results are available upon request.

period, the lower the level of welfare use appears. When the same analysis is performed in Figure 12 for the siblings sample, the treatment effects are again flat. This could arise because in Figure 11 the previous levels of exposure (before age 13) are not controlled for. Compare with the results for exposures at ages 13 to 17, in which previous exposure was controlled for. Nevertheless, the within siblings comparison seems to solve that problem.

The role of teenage childbearing

Two additional analyses are conducted regarding the role of teenage childbearing. In Figure 13, teenage childbearing is considered an outcome itself, in line with other studies in the prior literature (e.g.. Gottschalk, 19992 and 1996; An, Haveman and Wolfe, 1993). But, to avoid the potential endogeneity of exposure in ages 13-17 with the childbearing decision, exposure only in ages 10-12 is used as a treatment. The results are in line with previous findings, there is no effect of exposure on teenage childbearing.

An alternative analysis is to study what effect has in the results when one uses the exposure only prior to the time of teenage childbearing. This is done just as a way of understanding better the potential problems caused by the definition of the exposure variable. In the top left corner of the Figure a description of the data is presented, that makes clear the issue: when considering the net exposure variable, at very high levels of exposure, the proportion of women with teenage childbearing approaches zero. Of course, this is just a mechanic effect, if a woman has 60 months of exposure, it is because she did not become a teenage mother. The right top panel shows the effects of this alternative measure of exposure on teenage childbearing itself, and the two bottom panels show the effects on probability of and months of welfare use. Although, the left top panel would have suggested otherwise, it is interesting to find that still there are no effects of exposure, even with the alternative exposure measure.

Summary of the results

The overall picture from the results is clear. First, there are no effects of differential lengths of exposure to welfare on either the probability of welfare use or on the number of months of welfare use during the ages 19 to 22. This is also true for longer exposure and use windows. Second, the differentials in welfare dependency among black women (and to a lesser extent for Hispanics), compared to the other racial/ethnic group are large. It is not clear what causes those differentials, but it seems like an issue worth of further research. Third, teenage childbearing seems to be the key element in explaining future dependency. A woman that becomes a mother has twice the probability of depending on welfare, compared to a non mother.

5 Conclusions

This paper used a unique administrative dataset and recently developed estimators in the program evaluation literature, to estimate the differential effects of length of exposure to welfare on the probability of use and number of months of use, in early adulthood. All together, the results of this paper indicate that, conditional on welfare participation, there is no causal effect of the length of exposure on intergenerational welfare dependency, or on the probability of childbearing. This conclusion needs to be interpreted with caution, because it only refers to the intensive and not to the extensive margin (the participation decision) of welfare exposure. Thus, it is not possible to rule out explanations of intergenerational correlation of welfare dependency that admit formulations in which length of exposure plays no role (e.g. stigma reduction, learning the rules of the system). However, the results are clearly not consistent with arguments based on the existence of a "culture of welfare" and are consistent with the correlated disadvantages theory.

In terms of welfare policy, the results indicate that policies like time-limits are not likely to reduce the intergenerational correlation of welfare dependency. In addition, policies aimed at reducing the prevalence of teenage pregnancies (for example, by providing sexual education and contraceptive services) could have an important effect in lowering welfare dependency.

References

[1] Acs, G. 1996. The Impact of Welfare on Young Mothers' Subsequent Childbearing Decisions. Journal of Human Resources 31(4): 898-915.

[2] An, Chong-Bum, Robert Haveman, and Barbara Wolfe. 1993. Teen Out-of-Wedlock Births and Welfare Receipt: The Role of Childhood Events and Economic Circumstances. Review of Economics and Statistics 75(2): 195-208.

[3] Antel, John. 1992. The Intergenerational Transfer of Welfare Dependency: Some Statistical Evidence. Review of Economics and Statistics 74(3): 467-473.

[4] Behrman, Jere, Yingmei Chen and Petra Todd. 2004. Evaluating Preschool Programs When Length of Exposure to the Program Varies: A Nonparametric Approach. Review of Economics and Statistics 86(1): 108-132.

[5] Black, Dan A. and Jeffrey A. Smith. 2004. How robust is the evidence on the effects of college quality? Evidence from matching. Journal of Econometrics 121: 99-124.

[6] Black, Sandra, and Paul Devereux. 2010. Recent Developments in Intergenerational Mobility in *Handbook of Labor Economics*, O. Ashenfelter and D. Card (eds.), Amsterdam: Elsevier, 2010, forthcoming.

[7] Borjas, George J. and Glenn T. Sueyoshi. 1997. Ethnicity and the Intergenerational Transmission of Welfare Dependency. Research in Labor Economics 16: 271-295.

[8] Cattaneo, Matias. 2009. Efficient Semiparametric Estimation of Multi-valued Treatment Effects under Ignorability. University of Michigan, Department of Economics (July).

[9] Corcoran, and Adams. 1995. Family and neighborhood welfare dependency and sons' labor supply. Journal of Family and Economic Issues 16(2): 239-264.

[10] Corcoran, M. 1995. Rags to Rags: Poverty and Mobility in the United States. Annual Review of Sociology 21: 237-67.

[11] Currie, Janet and Nancy Cole. 1993. Welfare and Child Health: The Link Between AFDC Participation and Birth Weight. American Economic Review 83(4): 971-985.

[12] Dehejia, Rajeev H., and Sadek Wahba. 1999. Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. Journal of the American Statistical Association 94: 1053-1062.

[13] Dehejia, Rajeev H., and Sadek Wahba. 2002. Propensity Score-Matching Methods for Nonexperimental Causal Studies. The Review of Economics and Statistics 84(1): 151-161.

[14] Flores, Carlos A.. 2004. Estimation of Dose-Response Functions and Optimal Doses with a Continuous Treatment, UC Berkeley, Ph.D. dissertation.

[15] Flores, Carlos A. and Alfonso Flores-Lagunes. 2009. Identification and Estimation of Causal Mechanisms and Net Effects of a Treatment under Unconfoundedness. University of Miami, mimeo.

[16] Flores, Carlos A., Alfonso Flores-Lagunes, Arturo Gonzalez, and Todd C. Neumann. 2009. Estimating the Effects of Length of Exposure to a Training Program: The Case of Job Corps. University of Miami, Department of Economics (March).

[17] Flores, Carlos A. and Oscar A. Mitnik. 2009. Evaluating Nonexperimental Estimators for Multiple Treatments: Evidence from Experimental Data. IZA Discussion Paper no. 4451 (September).

[18] Frangakis, Constantine and Donald B. Rubin. 2002. Principal Stratification in Causal Inference. Biometrics 58: 21-29.

[19] Gottschalk, Peter. 1996. Is the correlation in welfare participation across generations spurious?. Journal of Public Economics 63: 1-25.

[20] Government Printing Office. 1996. Personal Responsibility and Work Opportunity Reconciliation Act, 104th Congress Congressional Bills. Washington D.C.

[21] Grogger, Jeffrey, and Lynn A. Karoly. 2005. Welfare reform. Harvard University Press.

[22] Haveman, Robert, and Barbara Wolfe. 1995. The Determinants of Children's Attainments: A Review of Methods and Findings. Journal of Economic Literature 33(4): 1829-1878.

[23] Heckman, James., Hidehiko Ichimura, and Petra Todd. 1997. Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. Review of Economic Studies 64(4): 605-654.

[24] Hirano, Kei and Imbens, Guido. 2004. The Propensity Score with Continuous Treatments in *Missing Data and Bayesian Methods in Practice: Contributions by Donald Rubin's Statistical Family*, Wiley.

[25] Hoffman, Saul D., and E. Michael Foster. 2000. AFDC Benefits and Nonmarital Births to Young Women. Journal of Human Resources 35(2): 376-391.

[26] Hoynes, Hilary. 1997. Does welfare play any role in female headship decisions? Journal of Public Economics 65, no. 2:89-117.

[27] Hoynes, Hilary. 2000. Local Labor Markets and Welfare Spells: Do Demand Conditions Matter? Review of Economics and Statistics 82(3): 351-368

[28] Imai, Kosuke and David A. van Dyk. 2004. Causal Inference with General Treatment Regimes: Generalizing the Propensity Score. Journal of the American Statistical Association 99(467): 854-866.

[29] Imbens, Guido W. 1999. The Role of the Propensity Score in Estimating Dose-Response Functions. NBER Technical Working Paper Series no. 237 (April).

[30] Imbens, Guido. 2000 W. The role of the propensity score in estimating dose-response functions. Biometrika 87(3): 706-710.

[31] Klerman, Jacob A. and Steven J. Haider. 2004. A Stock-Flow Analysis of the Welfare Caseload. Journal of Human Resources 39(4), Fall: 865-886.

[32] Lechner, Michael. 2001. Identification and estimation of causal effects of multiple treatments under the conditional independence assumption, in *Econometric Evaluation of Labour Market Policies*, ed. by M. Lechner, and F. Pfeiffer, Heidelberg: Physica/Springer, 43–58.

[33] Levine, Phillip B. and David J. Zimmerman. 1996. The Intergenerational Correlation in AFDC Participation: Welfare Trap or Poverty Trap? Institute for Research on Poverty Discussion Paper no. 1109-96.

[34] Levine, Phillip B. and David J. Zimmerman. 2005. Children's welfare exposure and subsequent development. Journal of Public Economics 89: 31-56.

[35] Martin, M. A. 2003. The Role of Family Income in the Intergenerational Association of AFDC Receipt. Journal of Marriage and Family 65(2): 326-340.

[36] McLanahan, Sara S. 1988. Family Structure and Dependency: Early Transitions to Female Household Headship. Demography 25(1): 1-16.

[37] Moffitt, Robert. 1983. An Economic Model of Welfare Stigma. American Economic Review 73(5): 1023-1035.

[38] Moffitt, Robert. 1992. Incentive Effects of the U.S. Welfare System: A Review. Journal of Economic Literature, 30(1): 1-61.

[39] Moffitt, Robert. 1998. The Effect of Welfare on Marriage and Fertility: What Do We Know and What Do We Need to Know?, in *Welfare, the Family, and Reproductive Behavior: Research Perspectives*, Ed. Robert A Moffitt, 204, National Academies Press.

[40] Moffitt, Robert A. 2000. Welfare Benefits and Female Headship in U.S. Time Series. American Economic Review 90(2), Papers and Proceedings: 373-377.

[41] Page, Marianne. 2004. New Evidence on Intergenerational Correlations in Welfare Participation, in *Generational Income Mobility in North America and Europe"*, ed. by M. Corak, Cambridge University Press, 226-244.

[42] Pepper, John. 2000. The Intergenerational Transmission of Welfare Receipt: A Nonparametric Bounds Analysis. Review of Economics and Statistics, 82(3): 472-488.

[43] Rosenbaum, Paul R. and Donald B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70: 41-55.

[44] Solon, Gary, Mary Corcoran, Roger Gordon and Deborah Laren. 1988. Sibling and Intergenerational Correlation in Welfare Program Participation. Journal of Human Resources 23(3): 388-396.

[45] Wolfe, Barbara, Robert Haveman, Donna Ginther, and Chong Bum An. 1996. The "Window Problem" in Studies of Children's Attainments: A Methodological Exploration. Journal of the American Statistical Association 91(435): 970-982.

Table 1. Descriptive statistics

Variable		Full sample	Siblings sample	Non mothers sample	Teen mothers sample
Outcome variables					
Any welfare use at ages	**19-22**	0.407	0.426	0.292	0.831
		(0.491)	(0.495)	(0.455)	(0.375)
	19-23	0.440	0.455	0.327	0.848
		(0.496)	(0.498)	(0.469)	(0.359)
	19-24	0.470	0.482	0.357	0.861
		(0.499)	(0.500)	(0.479)	(0.346)
	19-25	0.499	0.503	0.386	0.873
		(0.500)	(0.500)	(0.487)	(0.333)
	19-26	0.528	0.525	0.416	0.892
		(0.499)	(0.500)	(0.493)	(0.310)
# Months of welfare use at ages	**19-22**	11.51	12.17	7.45	26.55
		(16.59)	(16.87)	(13.86)	(17.21)
	19-23	14.70	15.35	9.79	32.27
		(20.32)	(20.61)	(17.23)	(20.84)
	19-24	18.08	18.71	12.28	38.06
		(24.01)	(24.23)	(20.57)	(24.30)
	19-25	21.50	22.07	14.84	43.53
		(27.48)	(27.83)	(23.75)	(27.51)
	19-26	24.86	25.80	17.44	48.85
		(30.78)	(31.35)	(26.88)	(30.38)
Treatment variable					
# Months of welfare exposure ages 13-17		34.52	39.09	32.88	40.61
		(20.29)	(19.10)	(20.61)	(17.77)
Individual-level covariates					
Race					
White		0.315	0.273	0.339	0.225
		(0.464)	(0.446)	(0.473)	(0.417)
Hispanic		0.357	0.397	0.340	0.419
		(0.479)	(0.489)	(0.474)	(0.493)
Black		0.208	0.204	0.191	0.270
		(0.406)	(0.403)	(0.393)	(0.444)
Other		0.121	0.127	0.130	0.087
		(0.326)	(0.333)	(0.336)	(0.281)
Language: English		0.837	0.830	0.813	0.925
		(0.369)	(0.375)	(0.390)	(0.263)
Family structure					
Avg. # siblings		1.578	2.268	1.569	1.612
		(1.323)	(1.281)	(1.314)	(1.353)
Avg. # adults		0.913	0.950	0.968	0.710
		(0.527)	(0.495)	(0.526)	(0.480)
Teenage Mothers					
% Teen Mothers		0.212	0.235		1.000
		(0.409)	(0.424)		-
Mother at age 13		0.023	0.033		0.107
		(0.149)	(0.177)		(0.309)
Mother at age 14		0.027	0.037		0.129
		(0.163)	(0.188)		(0.335)
Mother at age 15		0.047	0.054		0.219
		(0.211)	(0.227)		(0.414)
Mother at age 16		0.065	0.066		0.307
		(0.247)	(0.247)		(0.461)
Mother at age 17		0.051	0.046		0.238
		(0.219)	(0.208)		(0.426)

(Continues in next page)

Table 1. Descriptive statistics (cont.)

Variable		Full sample	Siblings sample	Non mothers sample	Teen mothers sample
Birth cohort					
1974		0.206	0.215	0.200	0.228
		(0.404)	(0.411)	(0.400)	(0.420)
1975		0.228	0.235	0.225	0.238
		(0.420)	(0.424)	(0.418)	(0.426)
1976		0.244	0.246	0.244	0.243
		(0.429)	(0.431)	(0.429)	(0.429)
1977		0.254	0.244	0.258	0.238
		(0.435)	(0.430)	(0.437)	(0.426)
1978		0.069	0.060	0.074	0.053
		(0.254)	(0.237)	(0.261)	(0.224)
Months prior welfare exposure × birth cohort					
1975 (age 12)		0.90	1.02	0.91	0.86
		(5.53)	(5.90)	(5.52)	(5.60)
1976 (ages 11-12)		4.59	5.30	4.48	4.98
		(11.15)	(11.89)	(11.01)	(11.66)
1977 (ages 10-12)		3.00	3.45	2.89	3.37
		(7.49)	(7.98)	(7.37)	(7.90)
1978 (ages 9-12)		1.44	1.64	1.38	1.66
		(3.76)	(3.97)	(3.69)	(3.99)
County-level covariates					
County share of health	9-12	0.041	0.040	0.041	0.040
expenditures at ages		(0.012)	(0.012)	(0.012)	(0.012)
	13	0.043	0.042	0.043	0.042
		(0.013)	(0.012)	(0.013)	(0.012)
	14	0.046	0.045	0.046	0.045
		(0.014)	(0.013)	(0.014)	(0.013)
	15	0.049	0.048	0.049	0.048
		(0.015)	(0.014)	(0.015)	(0.014)
	16	0.052	0.052	0.053	0.052
		(0.016)	(0.015)	(0.016)	(0.015)
	17	0.055	0.054	0.055	0.055
		(0.016)	(0.016)	(0.016)	(0.016)
Local economic conditions by age					
Unemployment Rate (%)	9-12	6.4	6.5	6.4	6.5
		(2.8)	(2.8)	(2.8)	(2.8)
	13-17	8.5	8.6	8.5	8.6
		(3.3)	(3.4)	(3.3)	(3.3)
	22-26	7.8	7.9	7.7	8.1
		(3.6)	(3.6)	(3.6)	(3.5)
Employment/population	9-12	0.179	0.179	0.179	0.178
retail sector		(0.03)	(0.03)	(0.03)	(0.03)
	13-17	0.176	0.176	0.176	0.175
		(0.03)	(0.02)	(0.03)	(0.02)
	22-26	0.174	0.173	0.174	0.173
		(0.02)	(0.02)	(0.02)	(0.02)
Average earnings in	9-12	4.94	4.93	4.94	4.94
retail sector ($1,000/qtr)		(0.52)	(0.52)	(0.52)	(0.52)
	13-17	4.85	4.84	4.85	4.85
		(0.48)	(0.47)	(0.48)	(0.47)
	22-26	4.83	4.81	4.84	4.81
		(0.52)	(0.52)	(0.53)	(0.51)
Number of observations		156,846	24,546	123,537	33,309

Note: Standard deviations in parentheses

Table 2. Balance in covariates before and after GPS adjustment
    Full sample

	Difference of means one treatment group vs. all other treatments pooled								Means for each treatment group									
Variable	Unadjusted (before overlap)				Adjusted by GPS blocking				Unadjusted (before overlap)					Inverse probability weighting				
	[3,15]	[16,35]	[36,59]	[60]	[3,15]	[16,35]	[36,59]	[60]	[3,15]	[16,35]	[36,59]	[60]	p-val	[3,15]	[16,35]	[36,59]	[60]	p-val
Teenage mother	-0.35*	-0.03*	0.32*	0.00	-0.24*	0.02*	0.16*	-0.21*	-0.26	-0.02	0.22	0.00	0.00	-0.22	0.00	0.08	-0.19	0.00
White	0.25*	0.13*	-0.09*	-0.33*	0.00	0.05*	0.05*	-0.07*	0.18	0.09	-0.07	-0.28	0.00	-0.01	0.03	0.04	-0.04	0.00
Hispanic	0.02*	0.03*	0.00	-0.07*	-0.04*	0.00	0.02*	0.01	0.02	0.02	0.00	-0.06	0.00	0.01	0.03	0.03	0.02	0.15
Black	-0.17*	-0.10*	0.11*	0.18*	0.01	-0.05*	-0.02*	0.03*	-0.13	-0.07	0.08	0.15	0.00	0.03	-0.01	0.00	0.05	0.00
English	0.03*	0.00	0.08*	-0.15*	-0.02*	-0.01*	0.05*	0.01	0.02	0.00	0.06	-0.12	0.00	0.00	0.02	0.07	0.06	0.00
Average # siblings	-0.29*	-0.13*	0.16*	0.29*	0.00	-0.03*	0.04*	0.00	-0.22	-0.09	0.11	0.24	0.00	-0.02	-0.03	0.01	-0.03	0.00
Average # adults	-0.06*	0.02*	-0.01	0.05*	0.03*	0.02*	-0.01	-0.02*	-0.05	0.01	-0.01	0.04	0.00	0.04	0.03	0.01	-0.02	0.00
Prioor welfare exposure × birth cohort																		
Age 12 × 1975	-0.09*	-0.05*	0.05*	0.10*	0.00	-0.01	0.00	0.01	-0.11	-0.08	-0.01	0.04	0.00	-0.05	-0.05	-0.05	-0.04	0.28
Ages 11-12 × 1976	-0.32*	-0.18*	0.15*	0.41*	0.01	-0.01*	0.00	0.02*	-0.23	-0.12	0.11	0.34	0.00	0.01	-0.02	-0.02	0.00	0.04
Ages 10-12 × 1977	-0.33*	-0.18*	0.16*	0.39*	0.00	-0.01	0.01*	0.01	-0.24	-0.12	0.12	0.33	0.00	0.00	-0.01	-0.01	0.00	0.60
Ages 9-12 × 1978	-0.34*	-0.16*	0.16*	0.38*	-0.03*	0.01	0.02*	0.00	-0.24	-0.11	0.12	0.32	0.00	-0.03	0.00	0.00	-0.01	0.02
County share of health expenditures																		
Ages 9-12	-0.02*	-0.02*	-0.02*	0.08*	0.03*	-0.01*	-0.05*	0.05*	-0.02	-0.01	-0.02	0.06	0.00	0.05	0.01	-0.02	0.04	0.00
Age 13	-0.03*	-0.02*	-0.03*	0.11*	0.03*	-0.01*	-0.06*	0.03*	-0.04	-0.04	-0.04	0.07	0.00	0.03	-0.02	-0.05	0.00	0.00
Age 14	-0.03*	-0.03*	-0.04*	0.13*	0.03*	0.00	-0.06*	-0.01	-0.04	-0.04	-0.04	0.09	0.00	0.02	-0.02	-0.05	-0.05	0.00
Age 15	-0.03*	-0.03*	-0.05*	0.14*	0.03*	0.00	-0.07*	-0.01	-0.04	-0.04	-0.05	0.10	0.00	0.02	-0.02	-0.06	-0.05	0.00
Age 16	-0.04*	-0.03*	-0.05*	0.16*	0.03*	0.00	-0.07*	0.00	-0.04	-0.03	-0.04	0.12	0.00	0.02	-0.01	-0.06	-0.03	0.00
Age 17	-0.06*	-0.03*	-0.04*	0.17*	0.04*	-0.01	-0.08*	0.06*	-0.03	-0.02	-0.02	0.15	0.00	0.05	0.01	-0.04	0.04	0.00
Number of observations	36,617	38,742	48,934	27,998	33,639	37,088	45,116	23,563	36,617	38,742	48,934	27,998		33,639	37,088	45,116	23,563	

Note: All variables have been standardized to mean 0 and standard deviation 1.
    * indicates difference of means is statistically significant at the 5% significance level. "p-val" refers to the p-value of the joint test of all means are equal.

Table 3. Balance in covariates before and after GPS adjustment
    Siblings sample

Variable	Difference of means one treatment group vs. all other treatments pooled								Means for each treatment group									
	Unadjusted (before overlap)				Adjusted by GPS blocking				Unadjusted (before overlap)					Inverse probability weighting				
	[3,15]	[16,35]	[36,59]	[60]	[3,15]	[16,35]	[36,59]	[60]	[3,15]	[16,35]	[36,59]	[60]	p-val	[3,15]	[16,35]	[36,59]	[60]	p-val
Teenage mother	-0.49*	-0.15*	0.36*	0.03	-0.36*	-0.02	0.26*	-0.18*	-0.41	-0.11	0.23	0.03	0.00	-0.36	-0.04	0.12	-0.17	0.00
White	0.27*	0.19*	-0.06*	-0.33*	-0.03	0.03	0.04*	-0.08*	0.22	0.14	-0.04	-0.26	0.00	0.00	0.03	0.04	-0.04	0.00
Hispanic	0.03	0.01	0.01	-0.05*	-0.09*	-0.02	0.00	0.05*	0.03	0.01	0.00	-0.04	0.01	-0.03	0.02	0.02	0.06	0.04
Black	-0.17*	-0.12*	0.10*	0.12*	0.06*	-0.03	0.01	-0.02	-0.14	-0.09	0.07	0.10	0.00	0.03	0.00	0.03	0.01	0.35
English	-0.02	0.00	0.14*	-0.18*	-0.05*	-0.03	0.09*	-0.07*	-0.01	0.01	0.09	-0.13	0.00	-0.02	0.03	0.10	0.02	0.00
Average # siblings	-0.17*	-0.16*	0.03*	0.26*	0.04	-0.04*	-0.03	0.08*	-0.14	-0.12	0.02	0.21	0.00	-0.02	-0.06	-0.06	0.02	0.00
Average # adults	0.12*	0.06*	-0.09*	-0.04*	0.04	0.01	-0.03*	0.02	0.10	0.04	-0.06	-0.03	0.00	0.01	-0.02	-0.05	-0.02	0.06
Prioor welfare exposure × birth cohort																		
Age 12 × 1975	-0.07*	-0.06*	0.03*	0.07*	-0.02	-0.02	0.02	0.00	-0.10	-0.08	-0.02	0.02	0.00	-0.07	-0.06	-0.03	-0.05	0.17
Ages 11-12 × 1976	-0.22*	-0.18*	0.06*	0.28*	0.08*	-0.01	0.01	0.00	-0.18	-0.13	0.05	0.23	0.00	0.05	-0.02	-0.01	-0.01	0.24
Ages 10-12 × 1977	-0.29*	-0.16*	0.08*	0.29*	0.01	0.02	0.01	0.00	-0.23	-0.11	0.06	0.23	0.00	-0.03	0.01	-0.02	-0.01	0.66
Ages 9-12 × 1978	-0.37*	-0.20*	0.13*	0.33*	-0.13*	0.00	0.06*	-0.01	-0.31	-0.14	0.09	0.27	0.00	-0.16	-0.02	0.00	-0.04	0.00
County share of health expenditures																		
Ages 9-12	-0.01	-0.02	-0.05*	0.10*	0.04	-0.01	-0.07*	0.07*	-0.01	-0.02	-0.03	0.07	0.00	0.04	0.01	-0.02	0.05	0.00
Age 13	-0.02	-0.03	-0.06*	0.13*	0.06*	-0.01	-0.08*	0.06*	-0.04	-0.03	-0.05	0.08	0.00	0.05	0.00	-0.05	0.02	0.00
Age 14	-0.01	-0.04*	-0.07*	0.15*	0.08*	-0.01	-0.08*	0.03	-0.02	-0.04	-0.06	0.10	0.00	0.04	-0.02	-0.06	-0.01	0.00
Age 15	-0.01	-0.04*	-0.08*	0.16*	0.08*	-0.01	-0.08*	0.04	-0.02	-0.04	-0.06	0.11	0.00	0.04	-0.03	-0.07	-0.01	0.00
Age 16	-0.04*	-0.03	-0.08*	0.17*	0.07*	0.00	-0.08*	0.03	-0.04	-0.03	-0.06	0.12	0.00	0.03	-0.01	-0.06	-0.01	0.00
Age 17	-0.07*	-0.04*	-0.06*	0.18*	0.07*	0.00	-0.08*	0.07*	-0.05	-0.02	-0.03	0.15	0.00	0.05	0.01	-0.04	0.05	0.00
Number of observations	3,710	5,821	9,215	5,241	3,383	5,455	8,447	4,436	3,710	5,821	9,215	5,241		3,383	5,455	8,447	4,436	

Note: All variables have been standardized to mean 0 and standard deviation 1.
    * indicates difference of means is statistically significant at the 5% significance level. "p-val" refers to the p-value of the joint test of all means are equal.

Table 4. Balance in covariates before and after GPS adjustment
  Non mothers sample

Variable	Difference of means one treatment group vs. all other treatments pooled								Means for each treatment group									
	Unadjusted (before overlap)				Adjusted by GPS blocking				Unadjusted (before overlap)					Inverse probability weighting				
	[3,15]	[16,35]	[36,59]	[60]	[3,15]	[16,35]	[36,59]	[60]	[3,15]	[16,35]	[36,59]	[60]	p-val	[3,15]	[16,35]	[36,59]	[60]	p-val
White	0.21*	0.11*	-0.05*	-0.34*	0.00	0.04*	0.04*	-0.04*	0.15	0.08	-0.04	-0.28	0.00	-0.04	0.02	0.03	-0.06	0.00
Hispanic	0.05*	0.02*	-0.02*	-0.07*	0.00	0.00	0.00	0.02	0.04	0.01	-0.01	-0.06	0.00	0.04	0.02	0.02	0.03	0.16
Black	-0.14*	-0.07*	0.09*	0.16*	0.00	-0.04*	-0.01	0.01	-0.10	-0.05	0.07	0.13	0.00	0.06	0.02	0.02	0.08	0.00
English	0.08*	0.00	0.05*	-0.17*	0.00	-0.02*	0.04*	0.06*	0.06	0.00	0.04	-0.13	0.00	0.02	0.02	0.06	0.09	0.00
Average # siblings	-0.25*	-0.07*	0.12*	0.25*	-0.01	-0.01	0.03*	-0.04*	-0.18	-0.05	0.09	0.21	0.00	0.02	0.01	0.02	-0.01	0.01
Average # adults	-0.06*	0.08*	-0.01*	-0.01	0.00	0.06*	-0.01*	-0.08*	-0.04	0.06	-0.01	-0.01	0.00	0.01	0.05	0.00	-0.07	0.00
Prioor welfare exposure × birth cohort																		
Age 12 × 1975	-0.09*	-0.04*	0.05*	0.10*	-0.01	-0.01*	0.01	0.00	-0.11	-0.07	-0.01	0.04	0.00	-0.04	-0.05	-0.05	-0.03	0.26
Ages 11-12 × 1976	-0.32*	-0.15*	0.14*	0.42*	-0.01	-0.02*	0.01	0.00	-0.22	-0.10	0.11	0.35	0.00	0.03	0.01	0.00	0.02	0.00
Ages 10-12 × 1977	-0.32*	-0.14*	0.15*	0.40*	-0.01	-0.01	0.02*	-0.01	-0.22	-0.09	0.12	0.34	0.00	0.03	0.01	0.00	0.02	0.05
Ages 9-12 × 1978	-0.33*	-0.13*	0.14*	0.41*	-0.03*	0.01*	0.01*	-0.01	-0.23	-0.08	0.11	0.35	0.00	-0.01	0.02	-0.01	0.00	0.03
County share of health expenditures																		
Ages 9-12	-0.04*	-0.02*	-0.01	0.09*	0.02*	-0.01*	-0.03*	0.02*	-0.03	-0.02	-0.01	0.07	0.00	0.04	0.01	-0.01	0.04	0.00
Age 13	-0.04*	-0.03*	-0.02*	0.12*	0.02*	-0.02*	-0.05*	0.01	-0.05	-0.04	-0.03	0.07	0.00	0.02	-0.01	-0.04	0.00	0.00
Age 14	-0.05*	-0.03*	-0.02*	0.13*	0.02*	-0.01	-0.05*	-0.04*	-0.05	-0.04	-0.03	0.09	0.00	0.01	-0.02	-0.05	-0.05	0.00
Age 15	-0.05*	-0.03*	-0.03*	0.14*	0.02*	0.00	-0.06*	-0.03*	-0.05	-0.04	-0.04	0.10	0.00	0.01	-0.02	-0.06	-0.05	0.00
Age 16	-0.06*	-0.03*	-0.03*	0.16*	0.02*	0.00	-0.06*	-0.02*	-0.05	-0.03	-0.03	0.12	0.00	0.01	-0.01	-0.06	-0.03	0.00
Age 17	-0.07*	-0.04*	-0.03*	0.18*	0.03*	-0.01	-0.06*	0.04*	-0.04	-0.02	-0.01	0.15	0.00	0.05	0.01	-0.03	0.05	0.00
Number of observations	32,802	30,760	34,066	22,010	29,797	29,281	32,324	19,183	32,802	30,760	34,066	22,010		29,797	29,281	32,324	19,183	

Note: All variables have been standardized to mean 0 and standard deviation 1.
   * indicates difference of means is statistically significant at the 5% significance level. "p-val" refers to the p-value of the joint test of all means are equal.

Table 5. Balance in covariates before and after GPS adjustment
Teenage mothers sample

	Difference of means one treatment group vs. all other treatments pooled								Means for each treatment group									
Variable	Unadjusted (before overlap)				Adjusted by GPS blocking				Unadjusted (before overlap)					Inverse probability weighting				
	[3,15]	[16,35]	[36,59]	[60]	[3,15]	[16,35]	[36,59]	[60]	[3,15]	[16,35]	[36,59]	[60]	p-val	[3,15]	[16,35]	[36,59]	[60]	p-val
White	0.30*	0.20*	-0.08*	-0.32*	-0.03	0.04*	0.08*	-0.08*	0.26	0.15	-0.05	-0.26	0.00	-0.06	0.01	0.01	-0.07	0.00
Hispanic	0.03	0.06*	-0.02*	-0.06*	-0.07*	0.00	0.00	0.02	0.02	0.05	-0.01	-0.05	0.00	-0.01	0.01	0.03	0.03	0.52
Black	-0.22*	-0.18*	0.08*	0.25*	0.06*	-0.02	-0.06*	0.03*	-0.19	-0.13	0.05	0.21	0.00	0.06	0.01	-0.03	0.04	0.00
English	0.01	0.01	0.02*	-0.06*	0.00	0.01	0.03*	-0.03	0.02	0.01	0.02	-0.04	0.00	-0.01	0.03	0.03	0.00	0.56
Average # siblings	-0.64*	-0.35*	0.28*	0.42*	0.04	-0.01	0.01	0.02	-0.57	-0.27	0.15	0.34	0.00	0.02	0.02	0.06	0.01	0.01
Average # adults	-0.76*	-0.26*	0.31*	0.33*	0.05*	0.08*	0.01	-0.05*	-0.67	-0.20	0.17	0.26	0.00	0.11	0.14	0.11	-0.01	0.00
Prioor welfare exposure × birth cohort																		
Age 12 × 1975	-0.11*	-0.09*	0.06*	0.09*	-0.01	-0.02	0.02*	-0.01	-0.13	-0.10	0.01	0.05	0.00	-0.01	-0.06	-0.02	-0.05	0.06
Ages 11-12 × 1976	-0.39*	-0.31*	0.17*	0.37*	-0.10*	-0.04*	0.04*	0.01	-0.34	-0.22	0.10	0.31	0.00	-0.13	-0.04	0.02	-0.05	0.00
Ages 10-12 × 1977	-0.40*	-0.29*	0.16*	0.36*	-0.08*	-0.02	0.02	0.04*	-0.34	-0.21	0.10	0.30	0.00	-0.05	0.00	0.04	-0.01	0.01
Ages 9-12 × 1978	-0.38*	-0.27*	0.19*	0.27*	-0.08*	-0.01	0.04*	0.00	-0.33	-0.19	0.11	0.23	0.00	-0.11	0.04	0.07	-0.01	0.00
County share of health expenditures																		
Ages 9-12	0.06*	0.01	-0.05*	0.04*	0.08*	-0.02	-0.06*	0.03	0.05	0.00	-0.03	0.03	0.00	0.06	-0.01	-0.03	0.01	0.04
Age 13	0.04*	-0.02	-0.06*	0.09*	0.09*	-0.02	-0.06*	0.02	0.02	-0.03	-0.05	0.06	0.00	0.06	-0.03	-0.05	-0.03	0.03
Age 14	0.02	-0.03*	-0.06*	0.13*	0.07*	-0.02	-0.07*	-0.01	0.01	-0.03	-0.05	0.09	0.00	0.04	-0.03	-0.05	-0.07	0.04
Age 15	0.03	-0.03*	-0.08*	0.14*	0.07*	-0.02	-0.09*	0.01	0.02	-0.03	-0.05	0.11	0.00	0.03	-0.04	-0.06	-0.06	0.05
Age 16	0.03	-0.03*	-0.08*	0.16*	0.08*	-0.02	-0.09*	0.02	0.02	-0.03	-0.05	0.12	0.00	0.04	-0.03	-0.06	-0.05	0.02
Age 17	0.01	-0.03*	-0.08*	0.16*	0.09*	0.00	-0.10*	0.05*	0.02	-0.01	-0.04	0.14	0.00	0.11	0.00	-0.06	0.00	0.00
Number of observations	3,815	7,982	14,868	5,988	2,548	6,773	13,775	5,183	3,815	7,982	14,868	5,988		2,548	6,773	13,775	5,183	

Note: All variables have been standardized to mean 0 and standard deviation 1.

  * indicates difference of means is statistically significant at the 5% significance level. "p-val" refers to the p-value of the joint test of all means are equal.

Table 6. Differences in months of welfare exposure ages 13-17 among sibling-pairs

Difference (montths)	# sibling-pairs	Percentage
No difference	3,031	24.7
1-12 months	5,197	42.3
13-24 months	2,668	21.7
25-36 months	1,028	8.4
37-48 months	282	2.3
49-60 months	67	0.6
Total	12,273	100.0

Table 7. Exposure ages 9-12 and percentage of teenage mothers
       Women born in 1978

Exposure ages 9-12	# observartions	Percentage	Teenage Mother
No Exposure	5,067	46.62	0.147
1-11 months	1,665	15.32	0.132
12-23 months	1,426	13.12	0.156
24-35 months	916	8.43	0.187
36-47 months	1,002	9.22	0.221
48 months	793	7.3	0.237
Total	10,869	100.0	0.163

Figure 1
Months of exposure to welfare ages 13–17
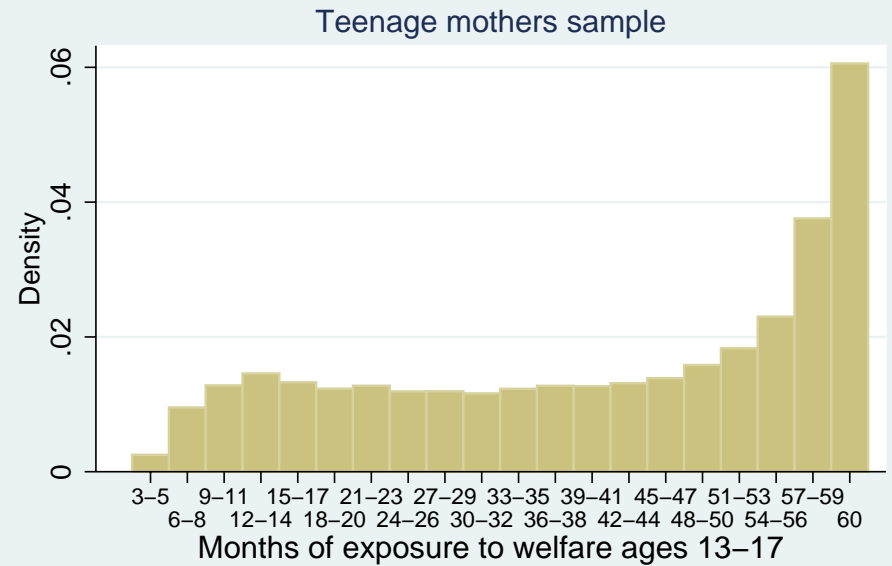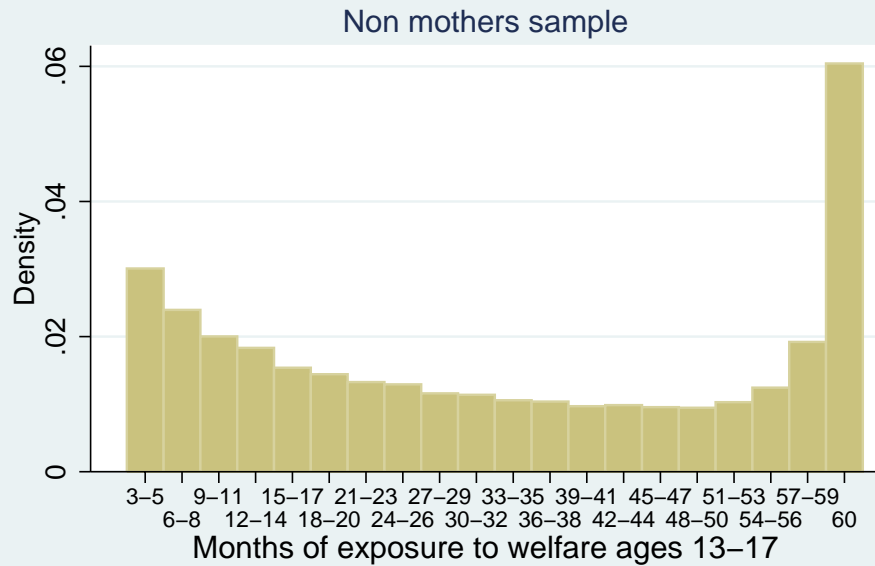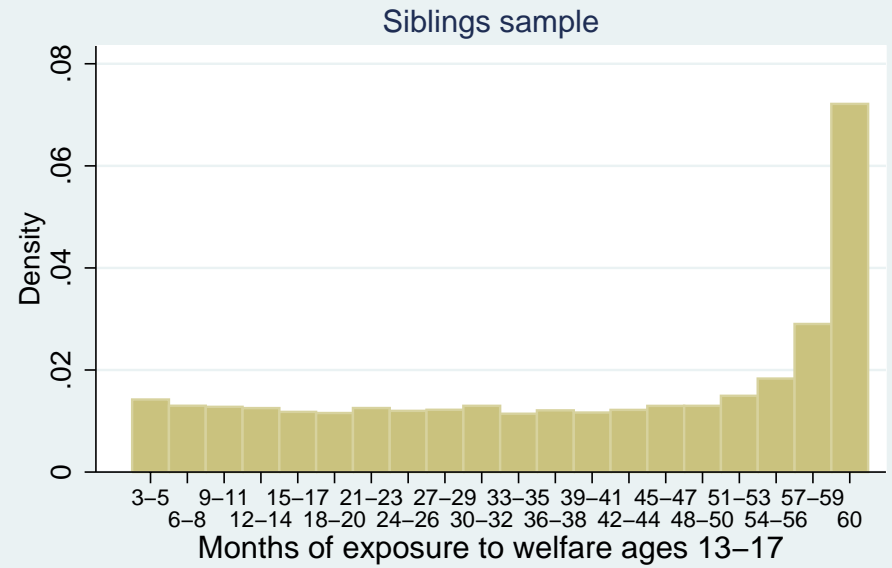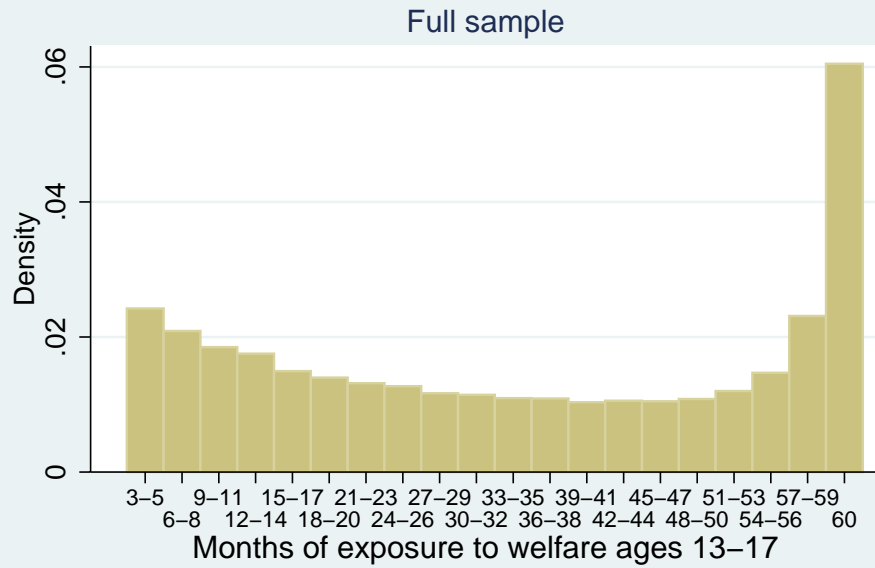
Figure 2
Kernel densities of estimated GPS
Full sample

A. Before imposing overlap
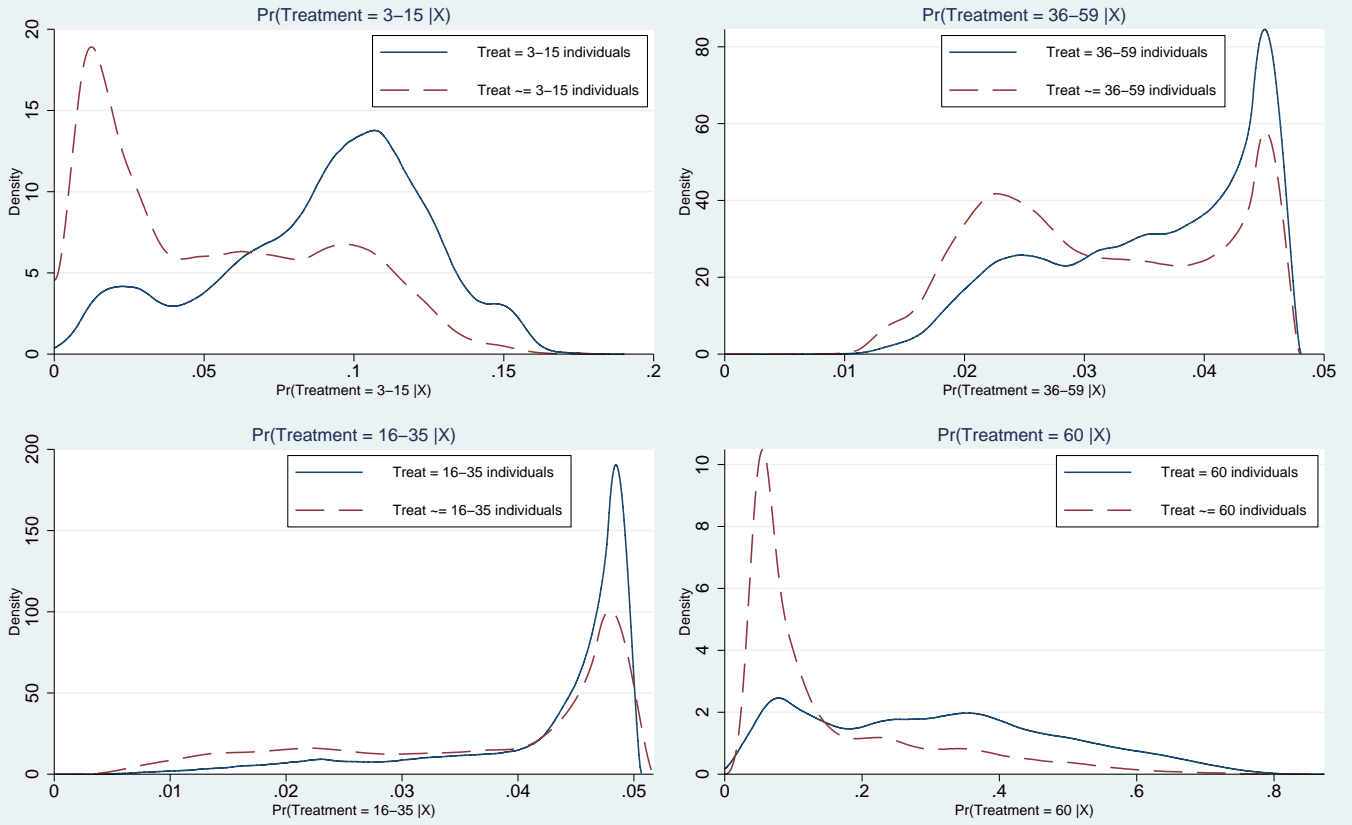
B. After imposing overlap
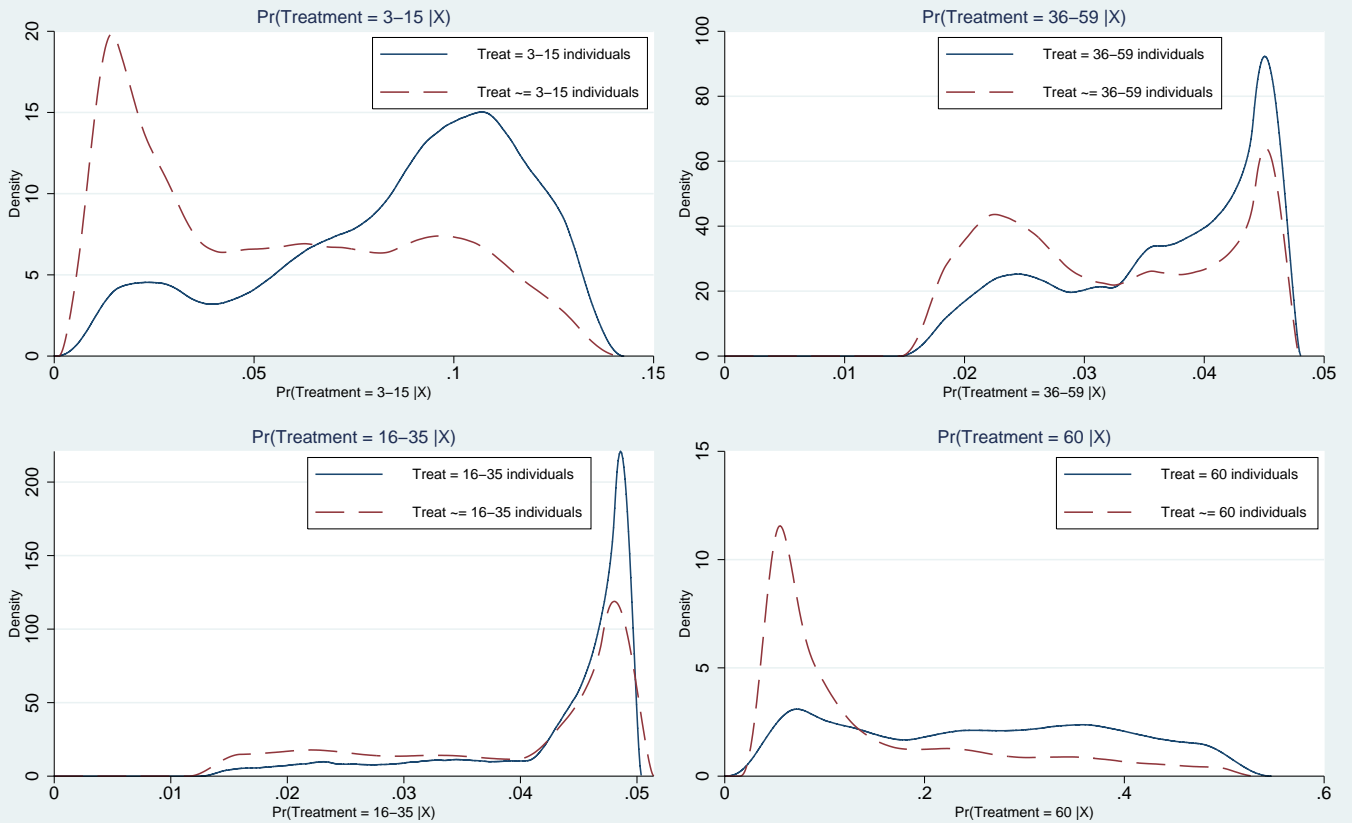
Figure 3
Treatment effects on any welfare use ages 19–22
Full sample
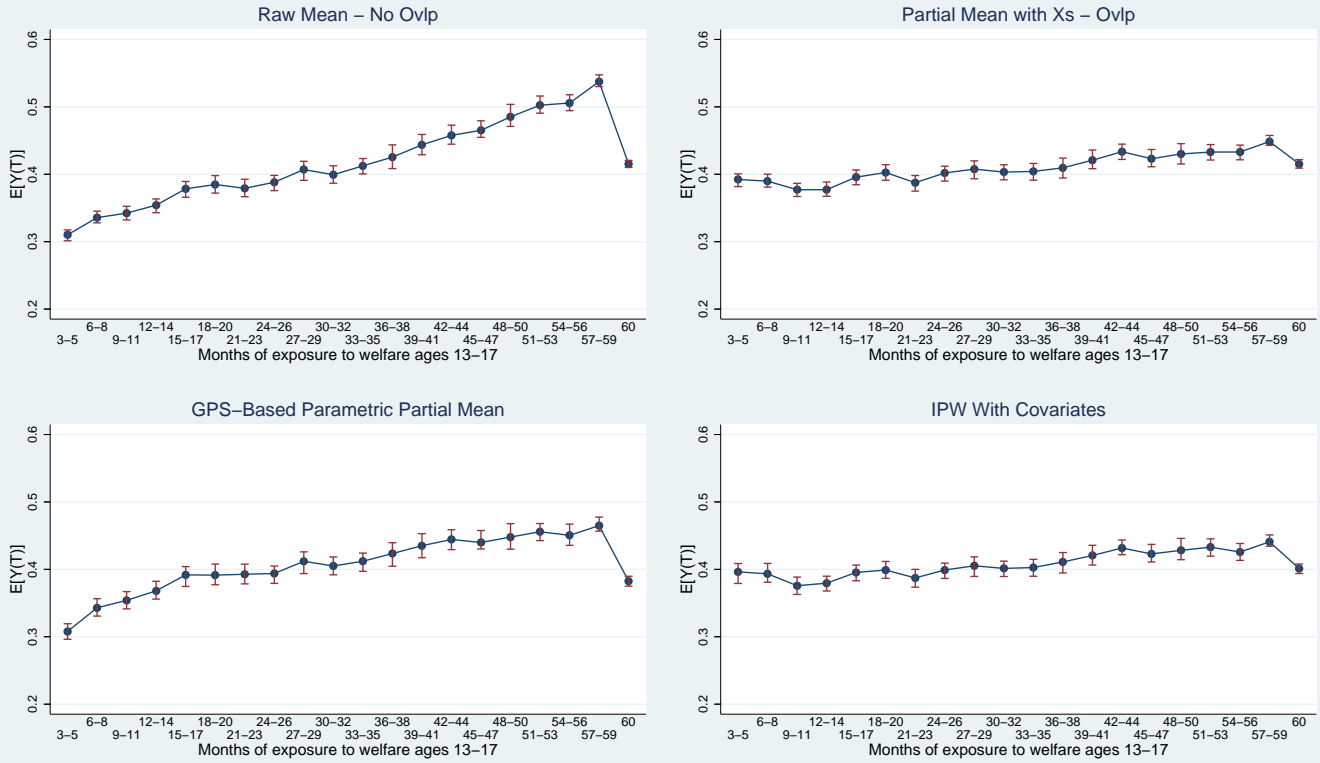
Figure 4
Treatment effects on months of welfare use ages 19–22
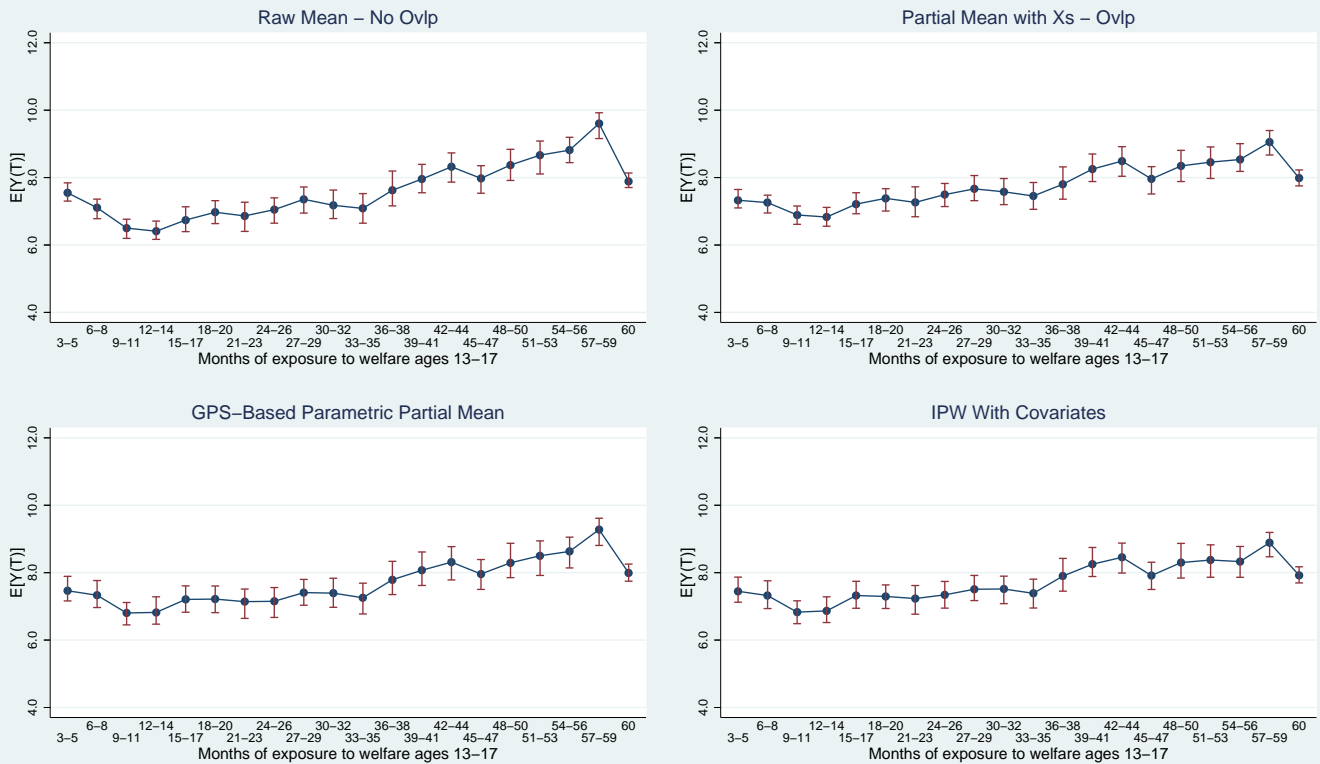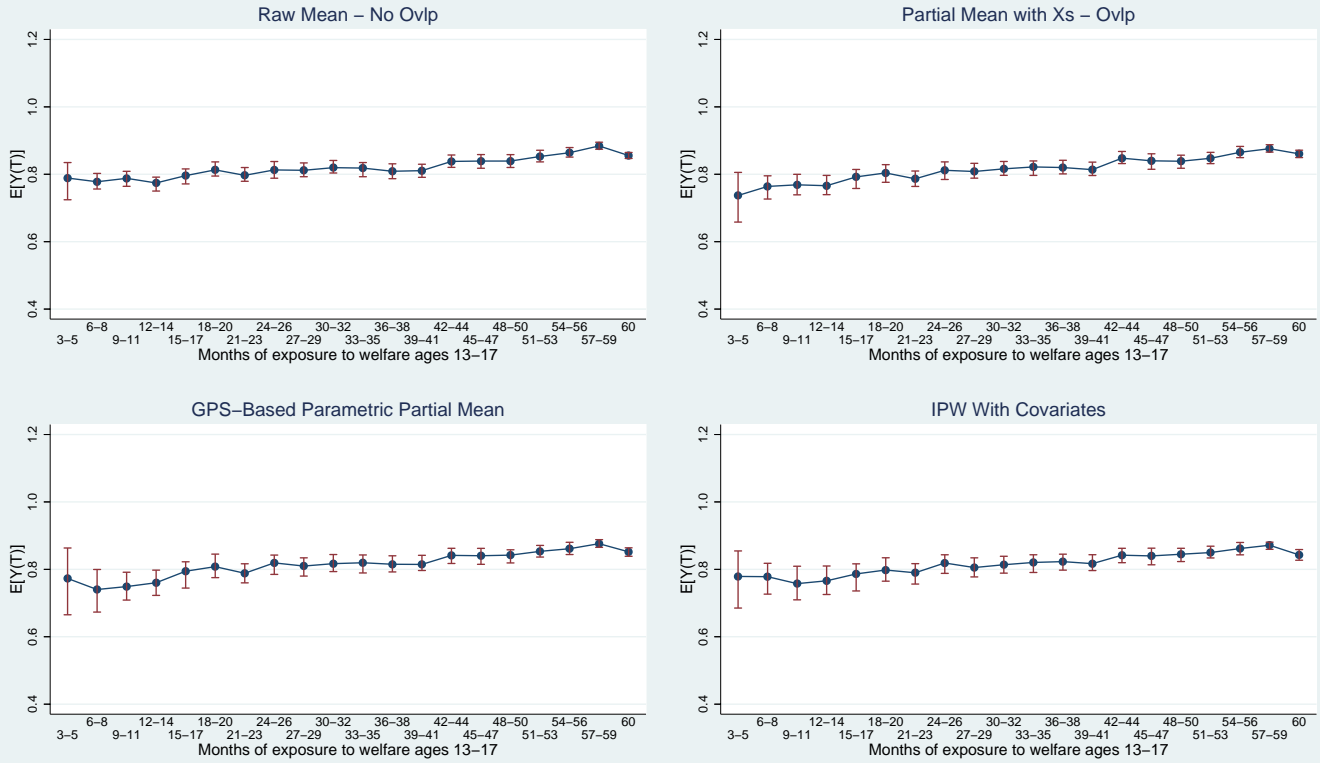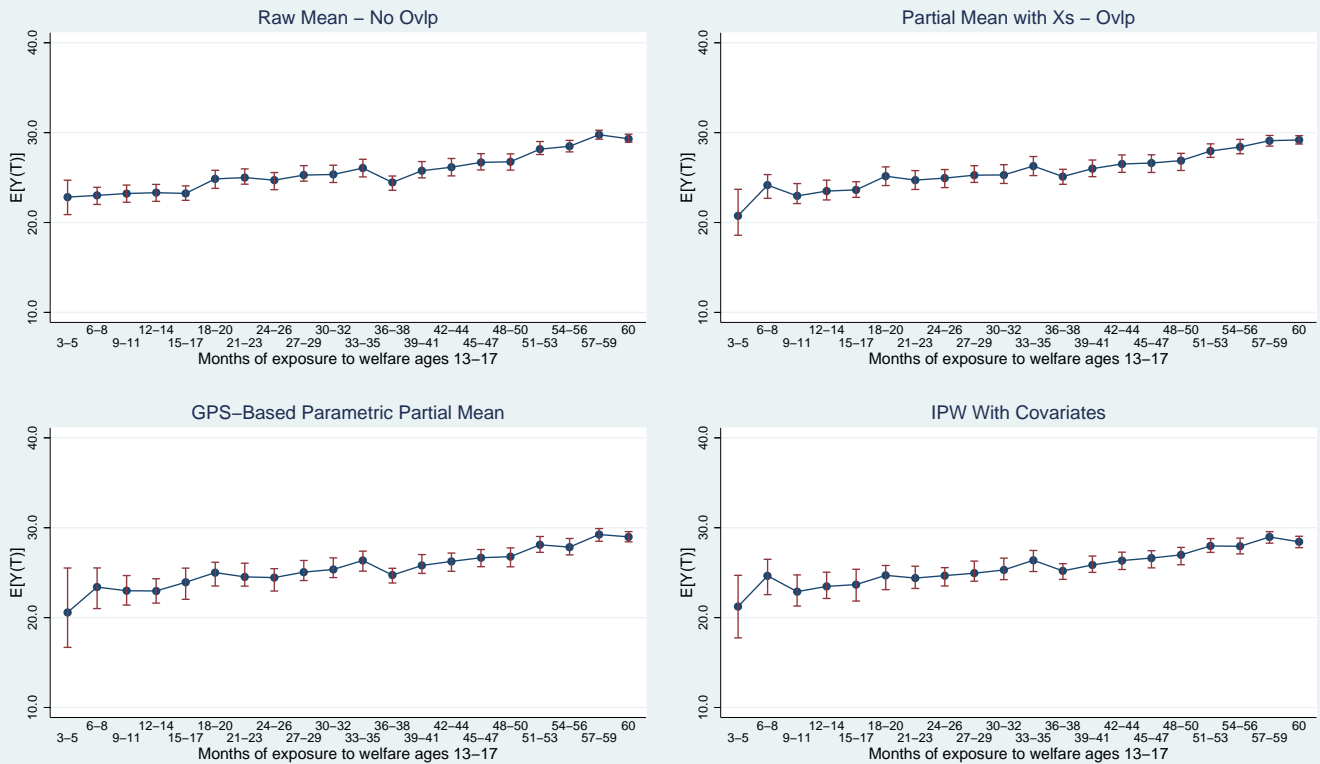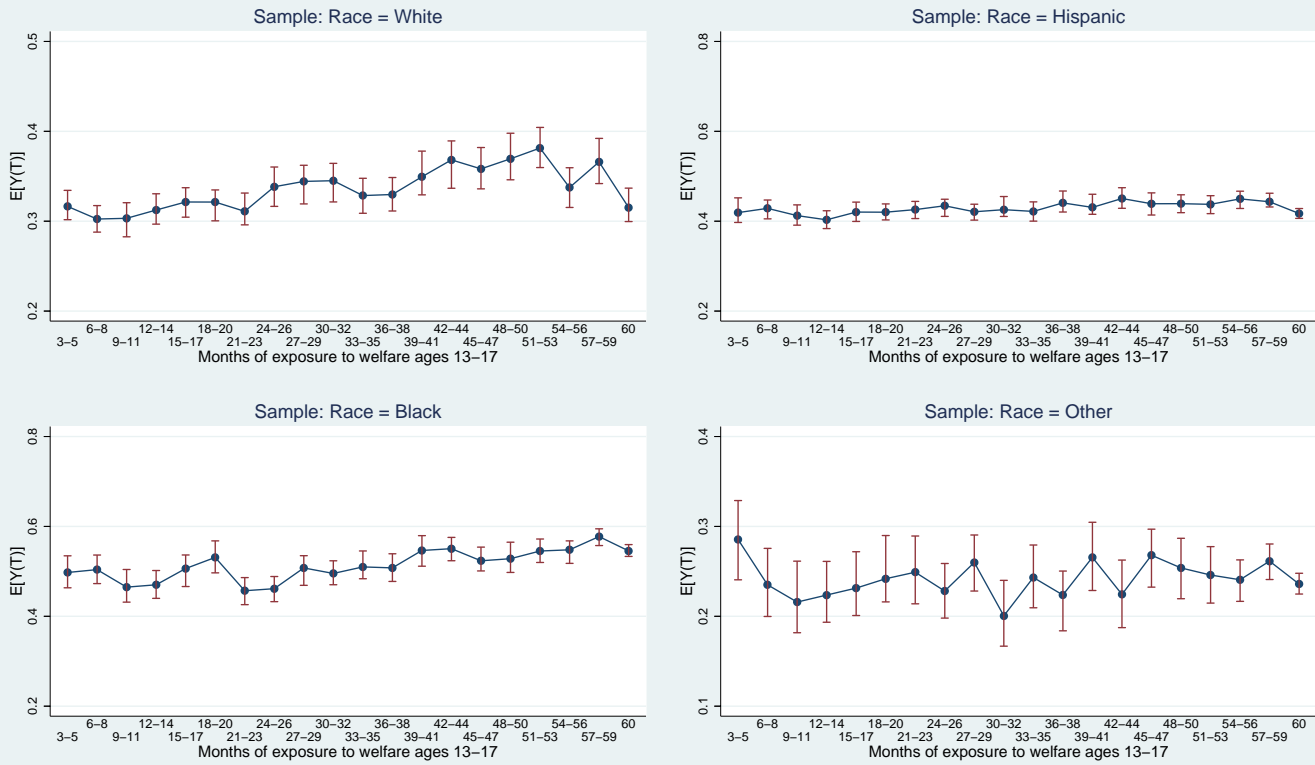### Full sample
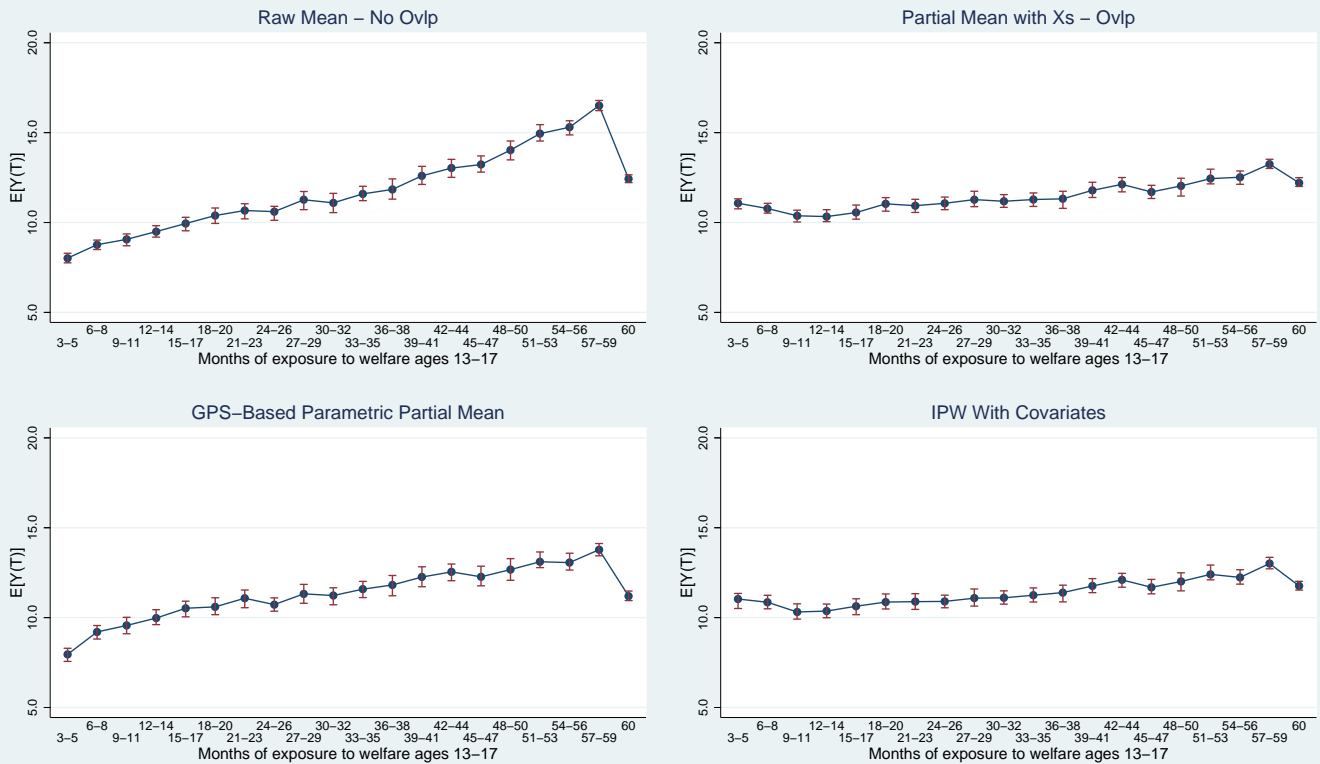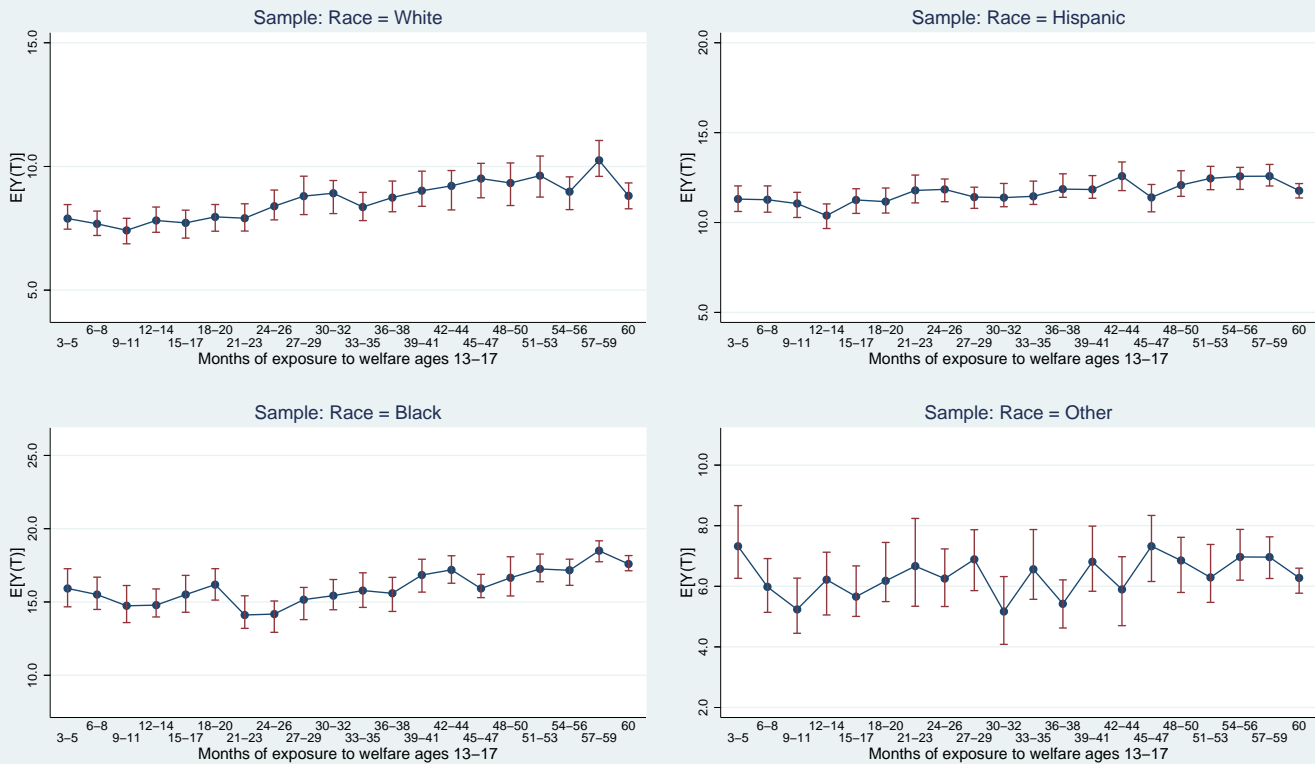
Figure 5
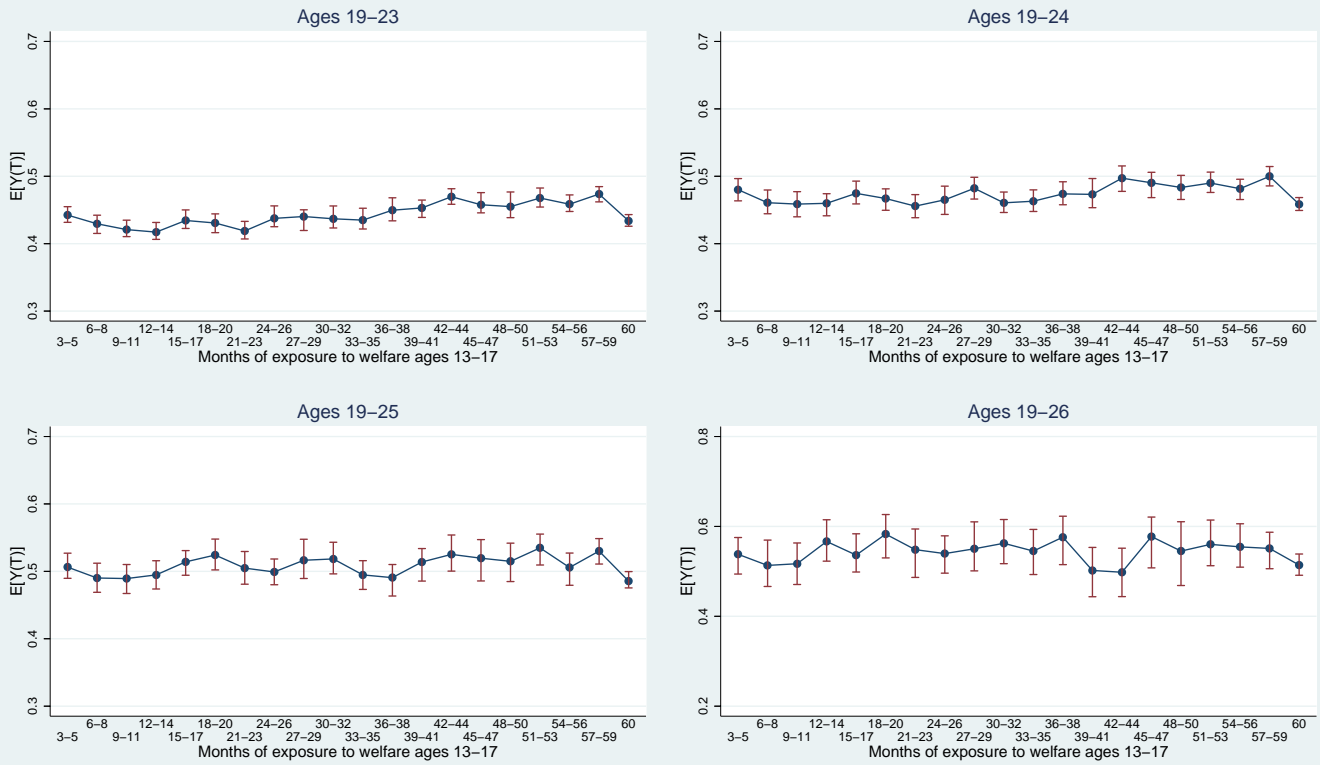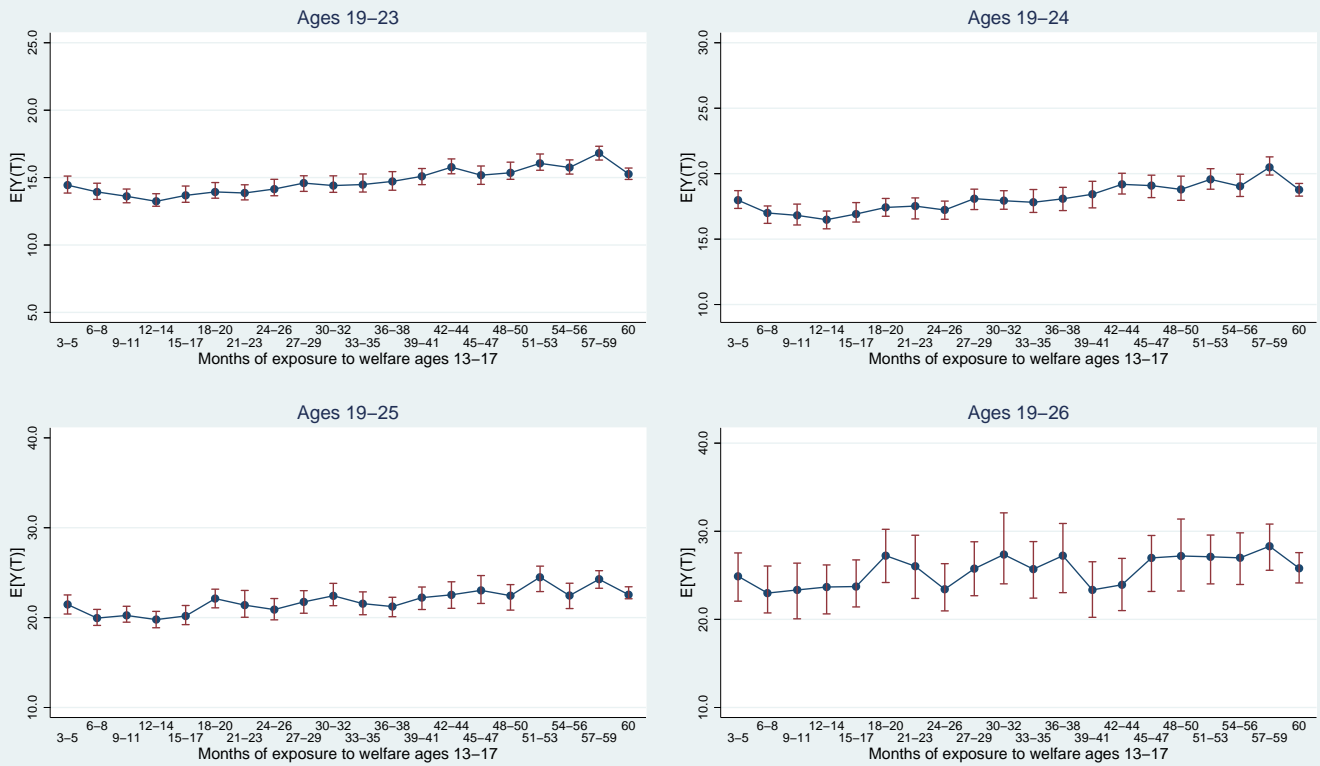### Treatment effects on any welfare use ages 19–22
Full sample – Not controlling for teen mom indicator
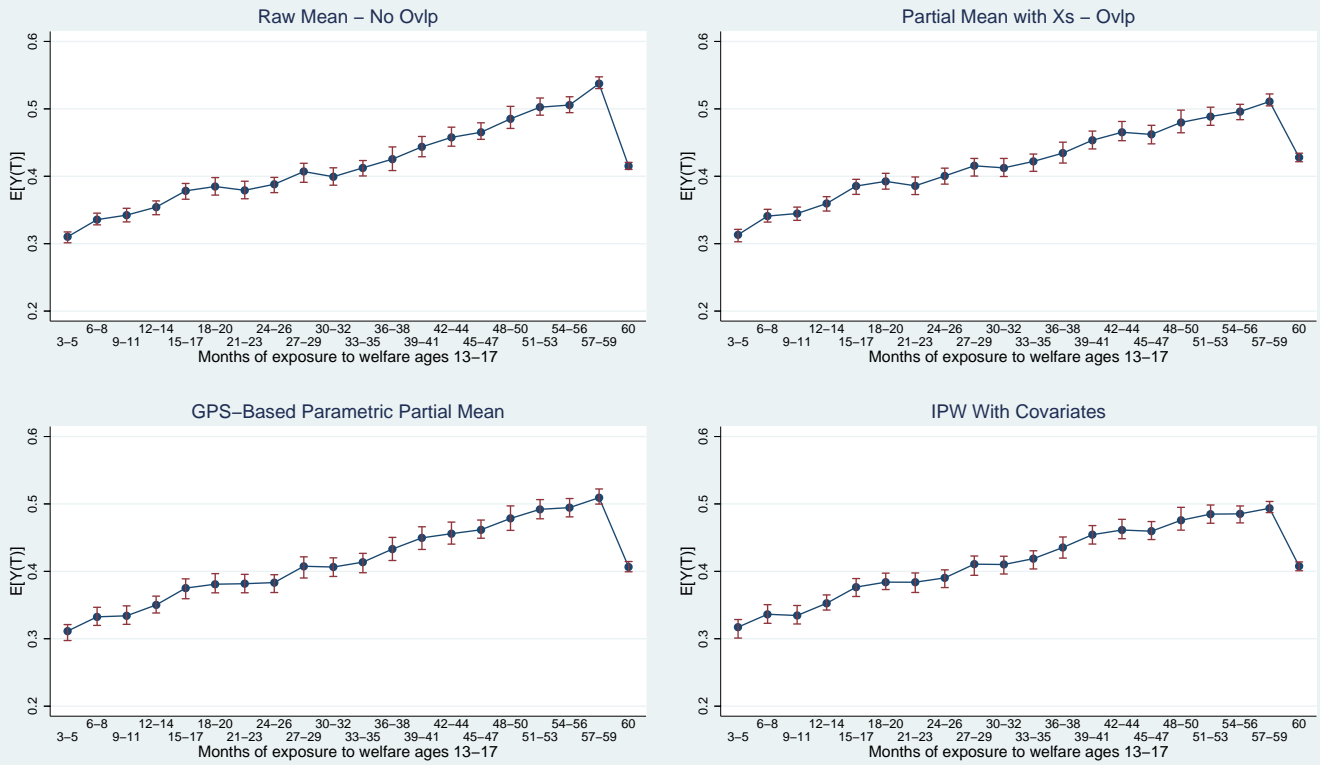
Figure 6
Treatment effects on months of welfare use ages 19–22
Full sample – Not controlling for teen mom indicator
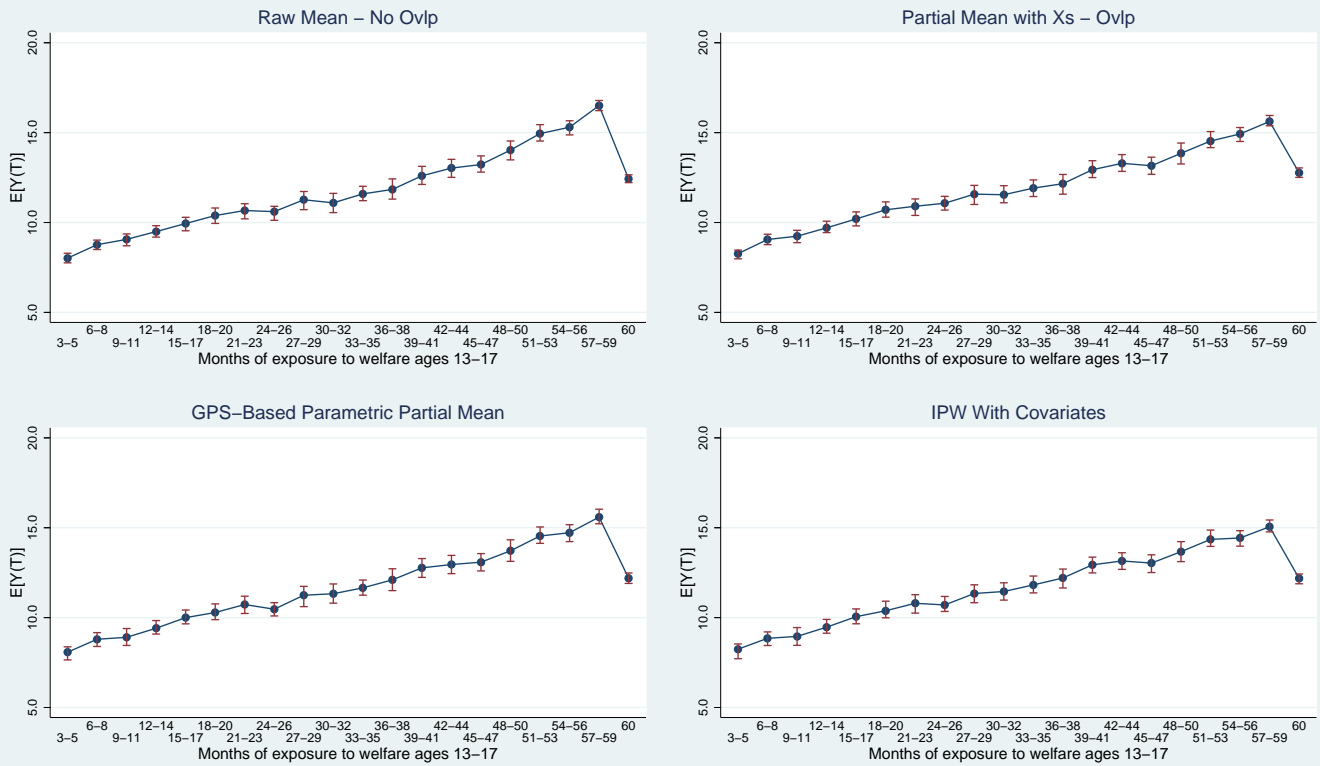
Figure 7
Treatment effects on any welfare use ages 19–22
Non mothers sample

Raw Mean – No Ovlp   Partial Mean with Xs – Ovlp   GPS–Based Parametric Partial Mean   IPW With Covariates
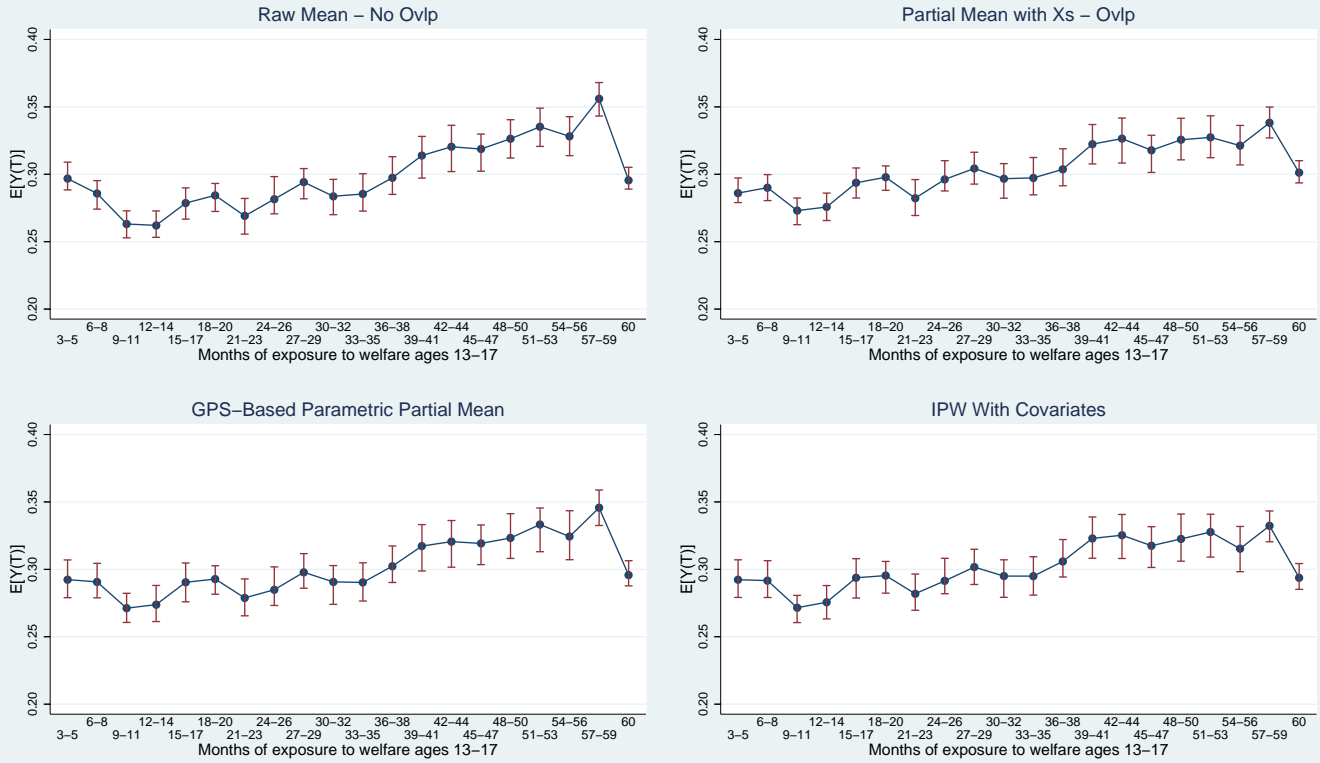
Figure 8
Treatment effects on months of welfare use ages 19–22
Non mothers sample

Raw Mean – No Ovlp   Partial Mean with Xs – Ovlp   GPS–Based Parametric Partial Mean   IPW With Covariates

Figure 9
Treatment effects on any welfare use ages 19–22
Teenage mothers sample

Figure 10
Treatment effects on months of welfare use ages 19–22
Teenage mothers sample

Figure 11
Treatment effects on any welfare use ages 19–22
Full sample by race

Sample: Race = White

Sample: Race = Hispanic

Sample: Race = Black

Sample: Race = Other

Figure 12
Treatment effects on months of welfare use ages 19–22
Full sample by race

Sample: Race = White

Sample: Race = Hispanic

Sample: Race = Black

Sample: Race = Other

Figure 13
Treatment effects on any welfare use ages 19–23 to 19–26
Full sample

Figure 14
Treatment effects on months of welfare use agess 19–23 to 19–26
Full sample

Figure 15
Treatment effects on any welfare use ages 19–22
Siblings sample – With no family fixed effects
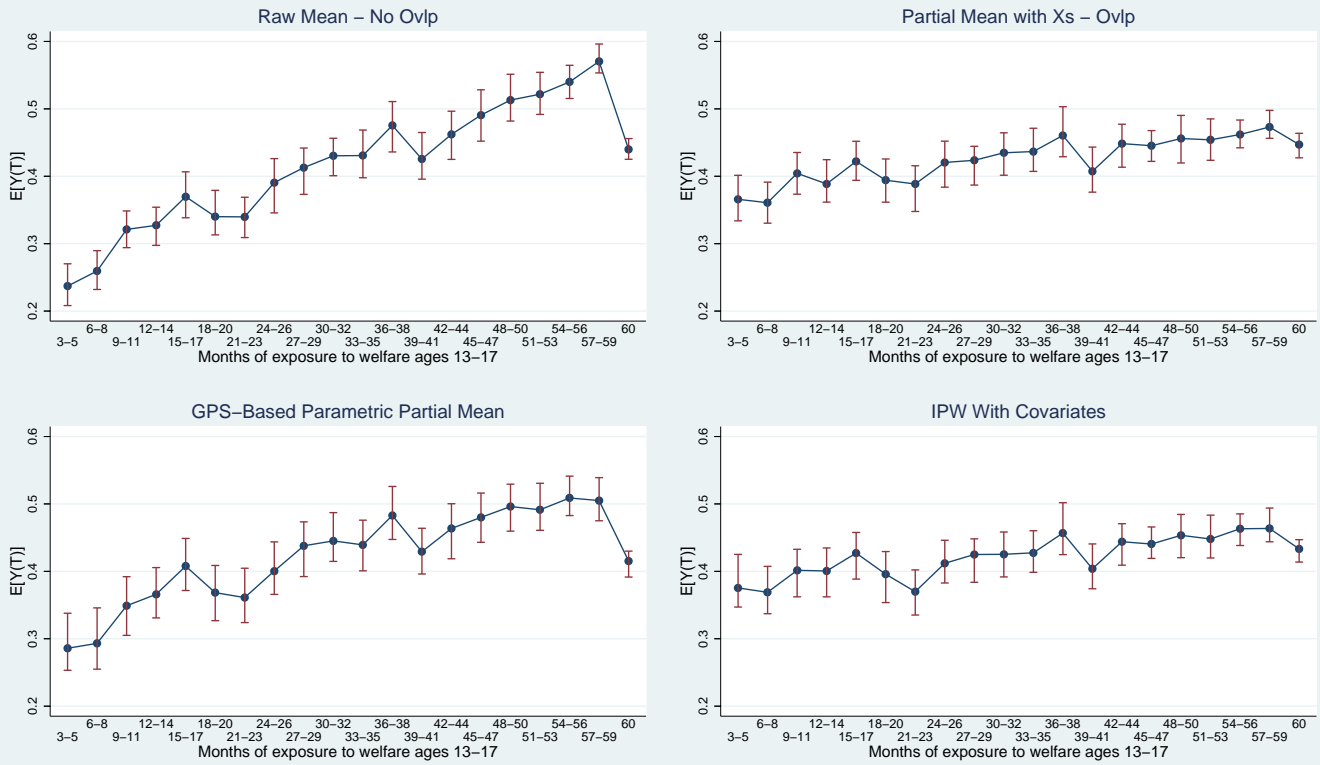
Figure 16
Treatment effects on any welfare use ages 19–22
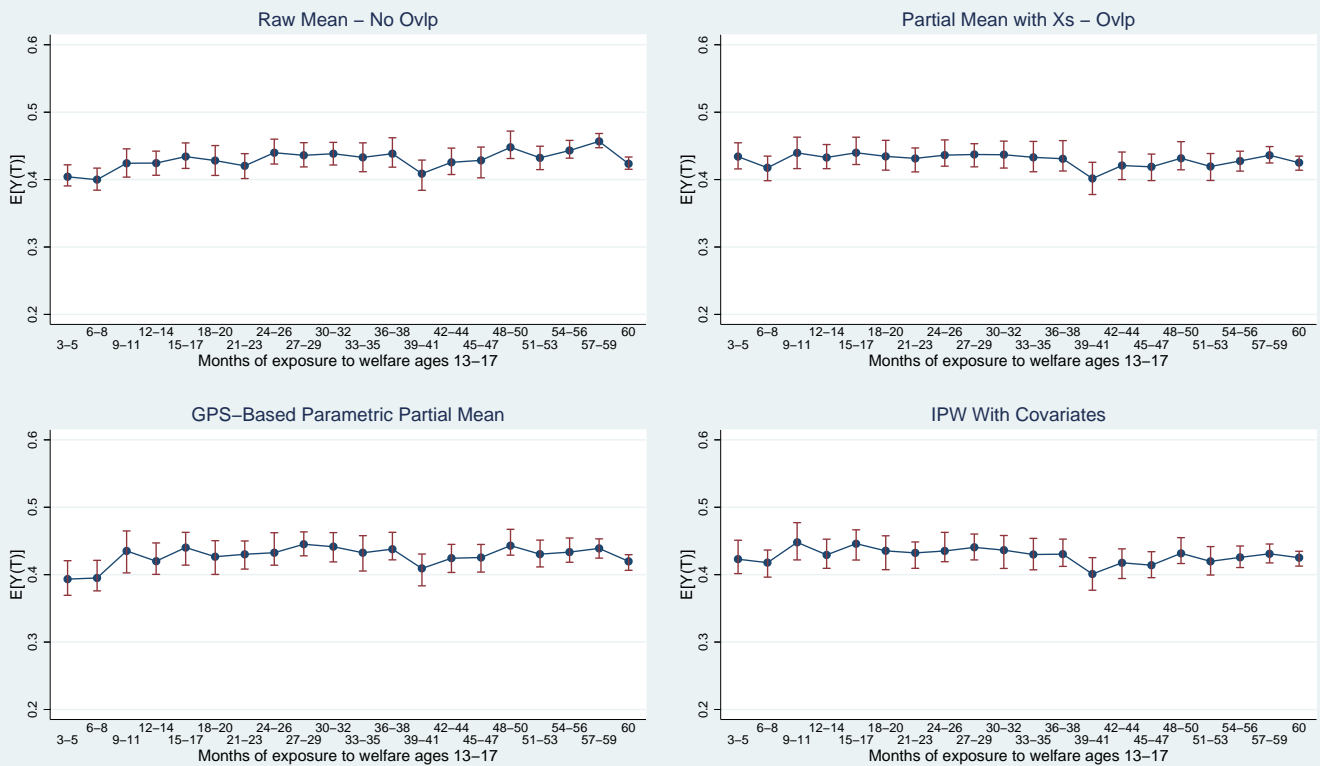Siblings sample – With family fixed effects

Figure 17
Treatment effects on months of welfare use ages 19–22
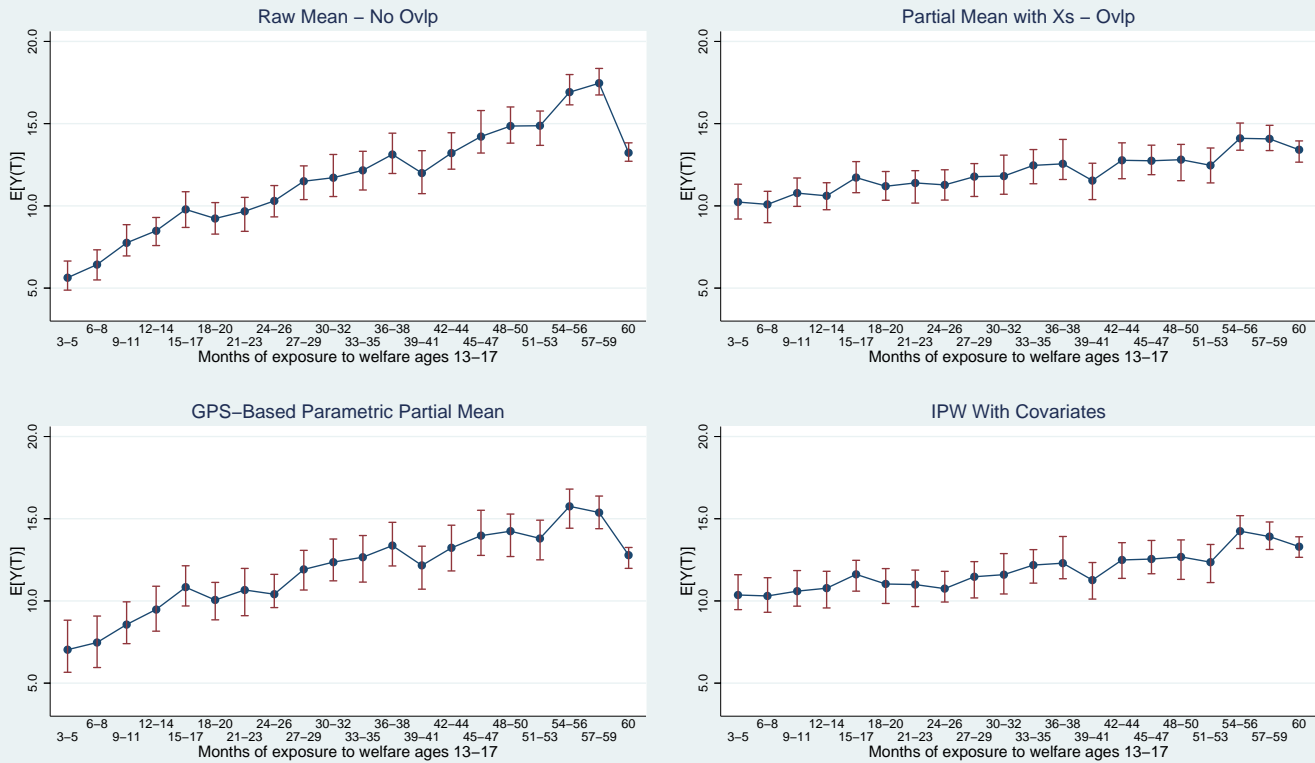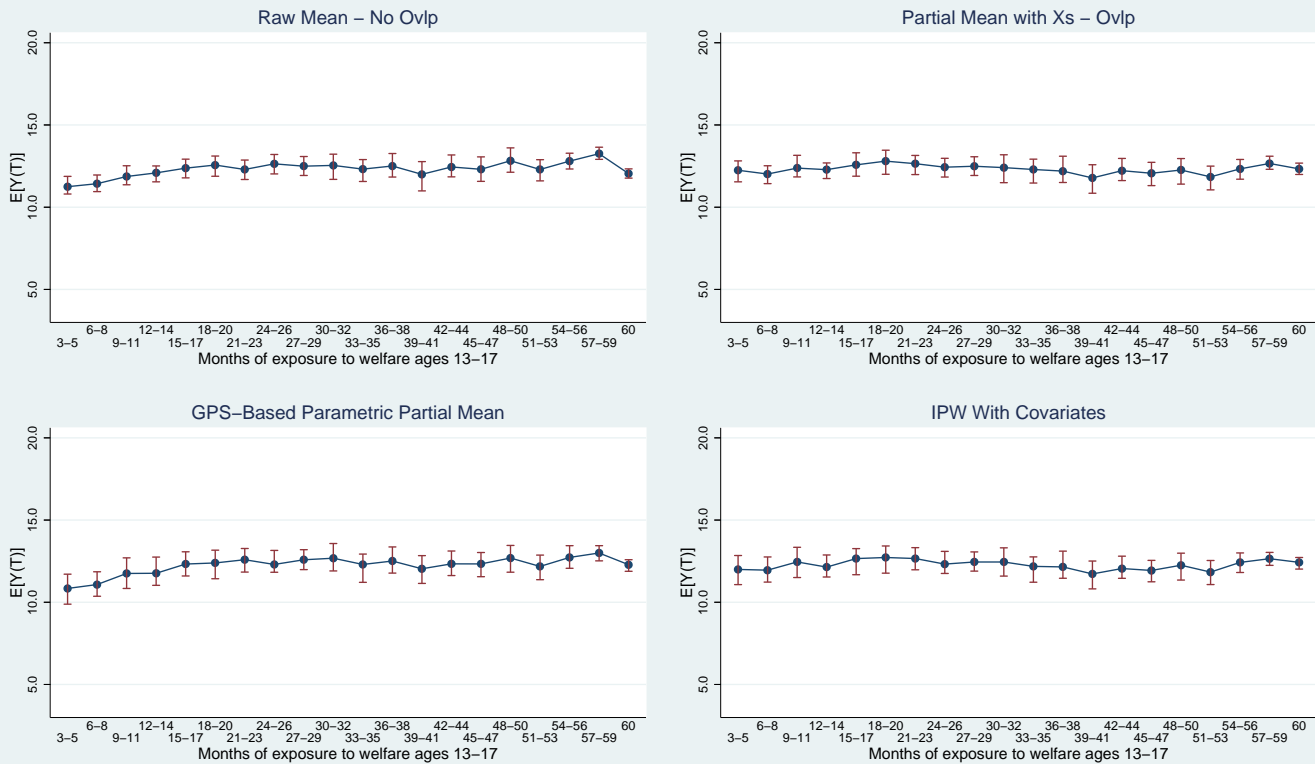## Siblings sample – With no family fixed effects

Raw Mean – No Ovlp

Partial Mean with Xs – Ovlp

GPS–Based Parametric Partial Mean

IPW With Covariates

Figure 18
Treatment effects on months of welfare use ages 19–22
Siblings sample – With family fixed effects

Raw Mean – No Ovlp

Partial Mean with Xs – Ovlp

GPS–Based Parametric Partial Mean

IPW With Covariates

Figure 19
Treatment effects on any welfare use ages 19–22
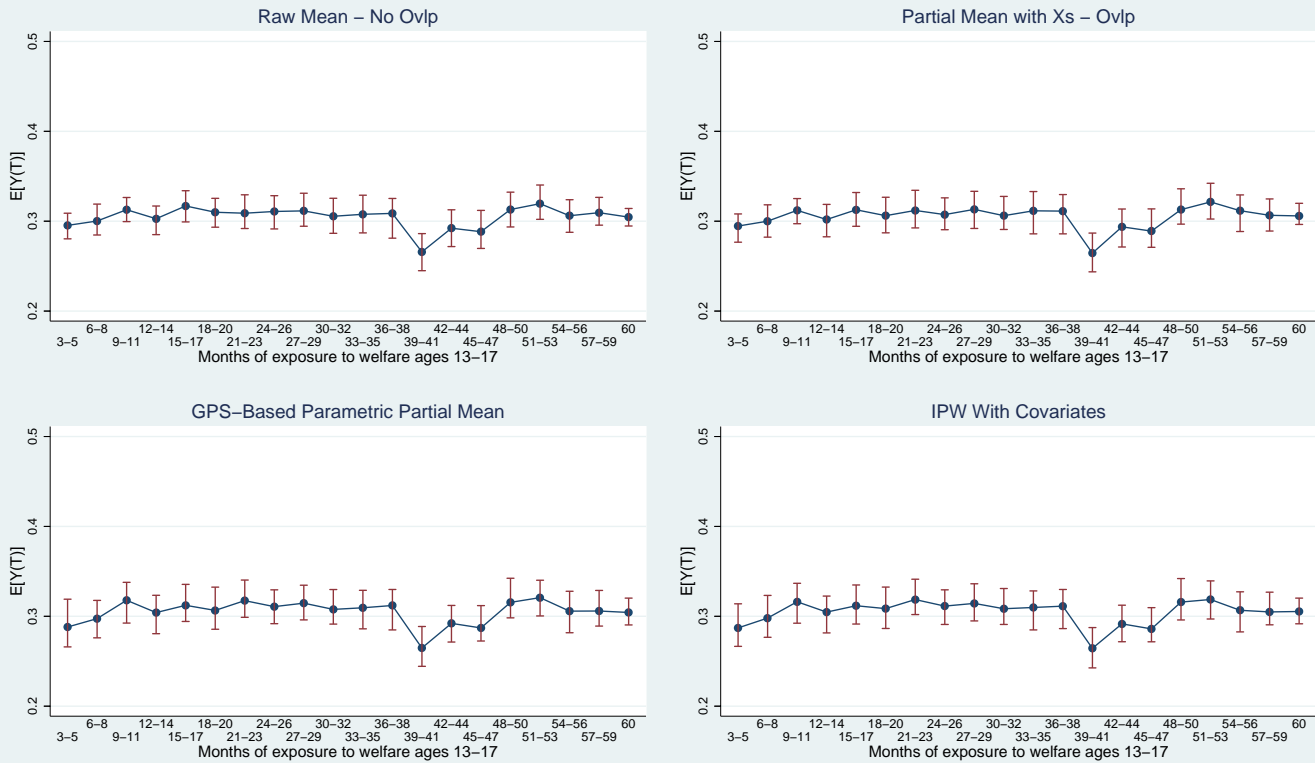## Siblings sample: non mothers – With family fixed effects

Figure 20
Treatment effects on months of welfare use ages 19–22
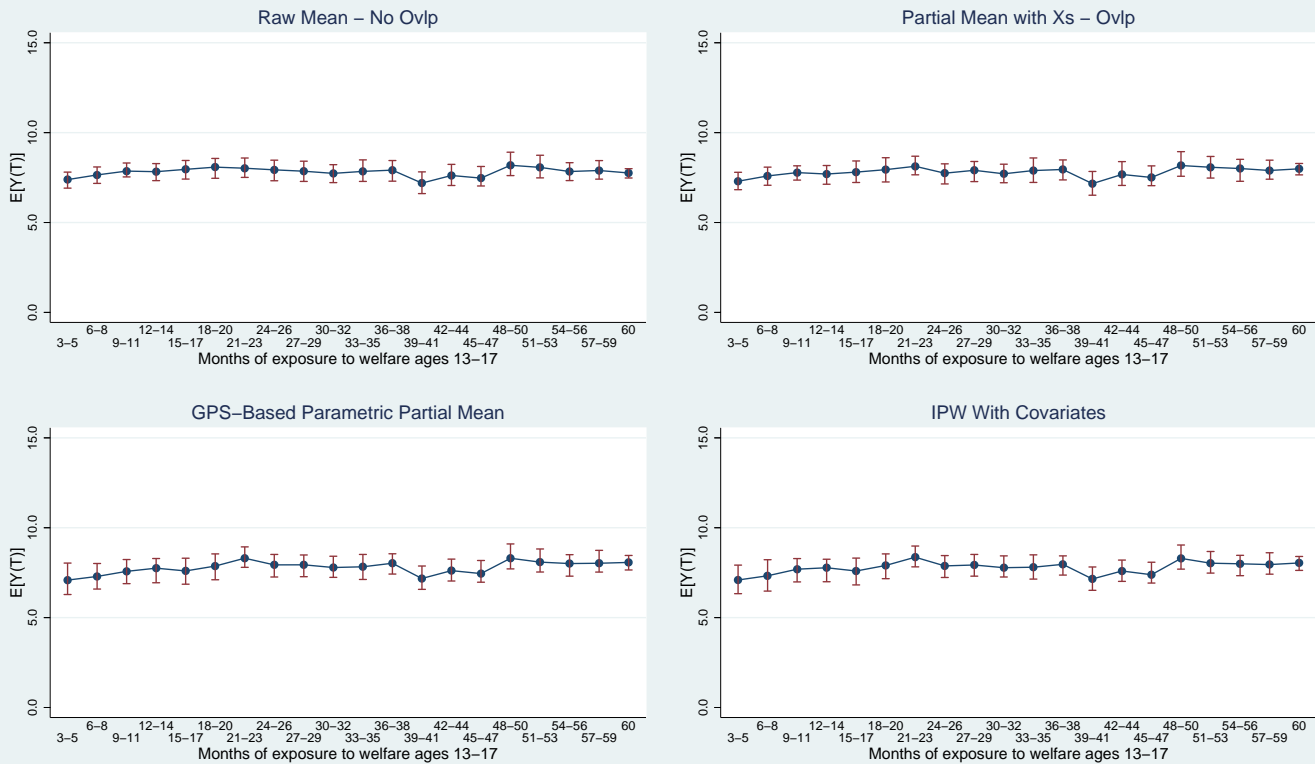## Siblings sample: non mothers – With family fixed effects

Figure 21
Treatment effects on probability of becoming a teenage mother
Sample: women born in 1978