Best of Both Worlds? Estimating the Treatment Effect of Teen Childbearing on Education Using Propensity Score Matching in Sibling Clusters*

Frank W. Heiland

Marxe School of Public and International Affairs & CUNY Institute for Demographic Research, Baruch College; The Graduate Center (Economics Program), CUNY,

and

Sanders Korenman

Marxe School of Public and International Affairs & CUNY Institute for Demographic Research, Baruch College, CUNY;

Revised: March 24, 2019

(For PAA 2019. Preliminary and incomplete. Do not share without authors' consent.)

Corresponding author: Sanders Korenman. School of Public and International Affairs, Baruch College, CUNY. 1 Bernard Baruch Way, D-901, New York, NY 10010. E-mail: sanders.korenman@baruch.cuny.edu.

Data Acknowledgement:
This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

---

* We thank Rachel Smith for support with the Add Health data. We are responsible for all errors.

Best of Both Worlds? Estimating the Treatment Effect of Teen Childbearing on Education Using Propensity Score Matching in Sibling Clusters *

Abstract

Sibling difference (or family Fixed Effects, "FE") methods are a well-known strategy for addressing selectivity bias due to omitted family-level variables. However, they face concerns over efficiency, generalizability and within-family selectivity. Recent advances in Propensity Score Matching (PSM) by Arkhangelsky and Imbens (2018) provide an alternative approach to estimating treatment effects in clustered data that may address some of these concerns by utilizing family-average treatment information. Using "Add Health" and NLSY79 data, we illustrate this approach in family/sibling samples and compare cluster PSM treatment effects of teenage childbearing on years-of-schooling to family FE and conventional PSM estimates. Preliminary results indicate that the PSM cluster estimates are smaller than conventional PSM estimates, and more similar to the (nearer-zero) family FE estimates. We discuss the findings in the context of recent work on method choice and heterogeneous effects in the literature on the educational consequences of teenage childbearing.

I. Introduction

Despite rapid and sustained declines in teenage fertility in the United States since the early 1990s (Martin et al. 2018), concern over its consequences continues unabated, because U.S. rates remain high in international comparisons (Kearney and Levine 2012) and due to recent threats to funding of pregnancy prevention programs (e.g., Hellman 2018). Evidence on the consequences of teenage births is mixed and contested (Kearney 2010; Geronimus and Korenman 1992; Hoffman, Foster and Furstenberg 1993; Geronimus and Korenman 1993; Ashcraft and Lang 2006; Fletcher and Wolfe 2009; Hotz et al. 2005; Levine and Painter 2003; Kane et al. 2013; Diaz and Fiel 2016; Duncan et al. 2018; Heiland, Korenman and Smith 2019; and Schulkind and Sandler 2019).

These studies have used one or more methods to support casual inference, including propensity-score matching, sibling differences, instrumental variables estimation, semi-parametric maximum likelihood and various combinations of these. All have strengths and weaknesses. With instrumental variables, potential shortcomings include implausible identifying exclusion restrictions and weak instruments. Propensity score methods can typically adjust for measured confounders only, leaving open the possibility of confounding by unobservables. And while sibling fixed-effects eliminate bias from unmeasured confounders common to siblings, limitations include small sample sizes (low power), within-family selectivity/endogeneity, limited generalizability, and possible contamination (e.g., a teen birth may affect a sibling's outcome).

Another recent strand of this literature studies whether effects are heterogeneous, explaining variation in estimates across methods and samples by differences in effects across the populations that identify the effect of a teen birth (e.g., Yakusheva 2011, Diaz and Fiel 2016, Geroniums 2003). Diaz and Fiel (2016) used smoothing-differencing and inverse probability weighting to allow heterogeneity and in NLSY79 data found that effects were more adverse among women who are *less* likely to have teen births. Similarly, in the High School and Beyond data, Yakusheva (2011) found that effects of a school-aged birth were not significant for teenagers at high-risk of a birth; a few significant effects were found among teens at *low* risk of a teen birth (e.g., Yakusheva 2011, Table 6).

Heiland, Korenman and Smith (2019) argued that sibling FE methods, despite weaknesses, have particular advantages in the presence of bias from unmeasured family background characteristics and heterogeneous effects. Sibling FE methods both reduce bias from unmeasured family characteristics common to siblings and arguably estimate the effect of interest for understanding the adverse outcomes actually observed among teenage mothers: the Average Treatment Effect on the Treated (ATET). This is the effect of a teenage birth on teen mothers, since the counterfactual outcome for a teenage mother is, essentially, her sister's outcome. The distinction between the ATET and the average effect in the population of teenagers grows in importance as teenage childbearing becomes more selected in the population, creating a greater difference between the average teen mother and the average teenager. As we argue below, as teenager fertility rates have plummeted, increasing selectivity has indeed occurred, at least judging by observable family background characteristics.

Recent advances in Propensity Score Matching (PSM) with clustered data (Arkhangelsky and Imbens 2018) provide an alternative approach to estimating treatment effects from sibling data that may overcome some of the concerns with sibling difference/FE estimates and with conventional PSM estimators. The cluster PSM approach differs from conventional matching approaches ("matching in non-clustered data") in that it exploits similarities between observations in different clusters using cluster-average treatment information. To our knowledge this strategy has not been applied in the context of sibling data which offer individuals nested within family clusters. Using recent data and large samples from the National Longitudinal Study of Adolescent to Adult Health (Add Health) and the National Longitudinal Study of Youth - 1979 Cohort (NLSY79), we apply Arkhangelsky and Imbens' approach with family cluster data, compare cluster PSM estimates of the treatment effect of teenage childbearing on educational attainment with conventional PSM estimates and family FE estimates, and discuss the relative strengths and weaknesses of matching approaches using sibling cluster data. We explore the possibility of heterogenous effects by using the cluster PSM approach to estimate both Average Treatment Effects in the population (ATE) and Average Treatment Effects on the Treated (ATET).

The remainder of the paper is organized as follows. In the next section, we introduce the Add Health and NLSY79 data and describe the samples and measures used in the analysis. Section III reports conventional and family cluster PSM estimates, estimates from sibling fixed effects methods, and conventional OLS estimates. Section VI discusses the implications of choice of method for policy in the presence of heterogeneous treatment effects.

II. Data, Samples, and Measures

We select samples and measures to compare results across estimation strategies. We take advantage of the relatively large sibling samples in Add Health and NLSY79 in order to maximize sample sizes available for sibling FE and PSM family cluster estimates.

Data and Samples

The data for this study come from the National Longitudinal Study of Adolescent Health ("Add Health") contractual dataset (Harris 2009) and the National Longitudinal Survey of Youth – 1979 Cohort. Add Health is a longitudinal, nationally representative sample of over 20,000 U.S. adolescents in grades seven through twelve who attended 80 high schools and 52 middle schools in 1994-95. The first wave was collected through student in-school questionnaires (90,118), and a subsample was interviewed in-home (20,745) and included an in-home parent questionnaire (17,670). At Wave I the Add Health "provides a nationally representative sample of … adolescents in grades 7 to 12" (Chen and Chantala, 2014, p.4); this sample (12,105) is referred to as the "core in-home sample." Three further waves of data were collected from the original in-home adolescent respondents in 1996 (II), 2001-02 (III), and 2007-08 (IV). The data for our analyses come from Wave I and Wave IV. Wave IV included 15,701 respondents who were then aged 24-34.

Supplemental samples were also drawn at Wave I, including oversamples of those of Cuban, Puerto Rican and Chinese ethnicity, black adolescents with highly-educated parents, adopted children, those with disabilities, as well as based on genetic relatedness (Chen and Chantala, 2014, p. 4). Add Health identified adolescent pairs (or multiples) related to varying degrees, including twins, full sibling, half siblings, unrelated adolescents living in the same household, and siblings of twins. Co-resident adolescent pair members were identified through reports on the in-school questionnaire or during the in-home interview. Members of pairs were both in grades seven through twelve at Wave I (see Harris et al. 2009; Harris et al. 2013). In the construction of our samples, we attempted to maximize the number of (female) siblings. The largest sample includes all (8,345) female respondents to both the Wave I and Wave IV questionnaires, including co-residing female members of the core in-home sample as well as the supplemental samples, most notably the supplemental samples of siblings.

The NLSY1979 is a longitudinally, nationally representative sample of 12,686 men and women between the ages of 14 and 21 as of January 1979. Information on personal and family characteristics was collected at the baseline interview in 1979. Information on educational attainment and fertility was collected in annual (biennially since 1994) follow-up interviews. We use data on all (6,283) female respondents from the core sample and the minority and low-income white over-samples. The poor white over-sample was dropped after 1990. We combined data from the baseline interview with education and birth history data from 1990. Siblings in NLSY79 are any co-residing female sample members that are identified as sisters on the baseline household roster.

The large number of (female) siblings in a longitudinal study with detailed information on family background and educational and fertility trajectories are recognized strengths of Add Health and NLSY79. The cohorts studied in the two surveys are almost two decades apart and have been used in numerous studies on the educational consequences of teen childbearing. This allows us to analyze the role of method choice on the treatment effect of teenage birth on education in two different generations and relate to a wide range of previous estimates.

In each survey, we implement the cluster PSM approach in three samples: (I) A sample of all women (or "all families") as mentioned above (Add Health: N=8,345; NLSY79: N=5,305). (II) A sample of all siblings consisting of female respondents who co-resided at baseline with another female survey member (Add Health: N=1,361; NLSY79: N=1,558). This sample of all siblings is of course a subsample of the all families sample. A detailed breakdown of this sample for Add Health in terms of the distribution by treatment status/family type and the corresponding means of outcome and characteristics is shown in Table 2. (III) A "discordant sibling" subsample, consisting of female siblings in families where (at least) one sibling is a teen mom and one is not (Add Health: N=289; NLSY79: N=383).

Measures

Add Health respondents' educational attainment at Wave IV is a categorical variable (h4ed2) ranging from 1-13 (1 = 8^{th} grade or less; 11 = completed a doctoral degree; 13 = completed post baccalaureate professional education). We applied a recoding suggested by Jason Fletcher (personal communication). For this "Fletcher" version, we coded "completed a doctoral degree" as 21 years of schooling. This version of the educational outcome variable has a mean of 14.4 years and a range of 8 to 21 years.[1] In the NLSY79, we used highest grade completed based on the 1990 wave (variable HGREV90).

We defined a teen birth as a dichotomous variable indicating a young woman had a live birth before exact age 19 (henceforth: "Teen Birth (by age 19)"). (We plan to examine the sensitivity of results to varying the cut-off age, using either age 18 or 20, and to restricting the sample to mothers only.)

Additional covariates used in the analysis below are respondent age and academic ability test score. In Add Health we use the Add Health Picture Vocabulary Test age-standardized percentile score from Wave I (henceforth: "PVT"), a measure of verbal ability. In NLSY79 we use the age-standardized Armed Forces Qualification Test score (henceforth: "AFQT"), a measure of verbal and math ability based on the Armed Services Vocational Aptitude Battery

[1] Stata code for our the "Fletcher" version is as follows: rename h4ed2 educationw4; recode educationw4 96=. 98=. 13=19 12=18 11=21 10=18 9=18 8=17 7=16 6=14 5=14 4=13 3=12 2=11 1=8. The h4ed2 education variable is described here: www.cpc.unc.edu/projects/addhealth/documentation/ace/tool/variable?VariableId=6896.

administered in 1980 (recoded in 2006; variable: AFQT_3). (Respondents with missing PVT or AFQT scores are flagged and a dummy for missing is included in models with PVT/AFQT score but not reported in the results tables shown below.)

Key measures constructed to implement the cluster PSM approach following Arkhangelsky and Imbens (2018) are the proportion of siblings in the family who had a teen birth ("Mean # Teen Moms"), the average age of the siblings ("Mean Age", and the average PVT/AFQT score of the siblings in the family (and the proportion missing) ("Mean PVT/AFQT" and "Mean PVT/AFQT missing").

Table 1 shows (unweighted) means for a number of individual and household-level variables including our education outcome measure, teen birth indicator, age at baseline, PVT/AFQT score, race/ethnicity, foreign-born status, number of (female) "siblings" in the household and income-to-needs ratio for the three samples in Add Health and NLSY79. Both surveys are racially/ethnically diverse and the proportions foreign-born are comparable. Educational attainment is greater among the women in the Add Health, which follows a more recent cohort. NLSY79 respondents are in household of lower socio-economic status and are more likely to have a teen birth.

III. Results

Descriptive Evidence

Table 2 shows means for all young women in Add Health who co-resided with at least one other young woman also in the sample. Of the 1,361 young women in this sample, 1,018 came from families in which no female sample member had a teen birth (before age 19), 289 came from "mixed" teen/non-teen families, i.e. families in which at least one sample member had and one did not have a teen birth, and 54 came from families in which all members had a teen birth.

Three things are apparent from this table. First, families in which no female sample member had a teen birth are more socially advantaged than those where at least one did. Their parents have more education (13.2 versus about 12.3 years), higher incomes (income/needs of 2.9 versus 1.7), they resided in wealthier areas (census tract per capita income of 13 thousand versus 10.5 thousand), they were less likely to be racially identified as non-Hispanic black (21% versus roughly 40%) and more likely to have two biological parents present at baseline (52% versus 25 to 35%). They also attained more education by Wave IV; adolescents from families where no sample members had a teen birth completed 1.5 to 1.8 more years of education than teen mothers. But they also have substantially higher PVT scores (99 vs. about 92) at Wave I. Note that adolescents from families with no teen births also completed about 1.2 years more education than *non-teen* mothers from "mixed" families (where a sibling had a teen birth), underlining the importance of controlling for family background in estimating effects of teen births on education.

Second, not surprisingly, siblings in mixed (teen/non-teen) families were very well-matched on background characteristics; they have virtually identical parental education, race/ethnicity, etc. (compare the second and third columns). The teen mothers scored slightly below their sisters who were not teen mothers (including non-mothers) on the PVT (91.9 versus 92.9) and also completed about 0.3 fewer years of education. As we will see in the next section, this unadjusted differential among matched (teen/non-teen) siblings of 0.3 years is far smaller than the corresponding unadjusted or OLS regression-adjusted differentials in the sample of all women/families. Also notable is the five-year difference in average age at first birth between matched siblings who did and did not have a teen birth (conditional on having a birth). A delay of five years, from age 17 to age 22 on average, is not trivial, and takes place across ages with considerable (but by no means universal) educational enrollment. In other words, the within mixed family teen/non-teen difference in education is a modest 0.3 years despite a delay of first birth of at least a five years over ages with high enrollment, and despite the fact that many of the non-teen mothers had not had a first birth as of the Wave IV follow-up.

Third, families in which all sample members had teen births do not appear particularly distinct or more disadvantaged than other families in which a female sample member had a teen birth (other than that they are somewhat more likely to be non-Hispanic white). This similarity in observed characteristics suggests that teens (and families) that identify sibling-difference in the Add Health estimates are not highly distinct from the families of other teen mothers.

OLS and Sibling Estimates

Table 3 shows OLS and sibling Fixed Effects (FE) results from Add Health. Table 4 reports corresponding results from NLSY79. Columns 1-2 show cross-section (pooled) OLS results ("XSEC OLS") for the sample with all families (Add Health: N=8,345; NLSY: N=5,305), columns 3-4 show OLS results for the sample of families with 2 or more siblings (Add Health: N=1,361; NLSY: N=1,558), and columns 5-6 show results from models with sibling FEs for those same samples (as required).

The educational differential by teen birth status in Add Health ranges from -1.464 and -1.362 based on unadjusted OLS estimate from the sample of all women and the sibling sample, respectively, to -0.237 for the unadjusted sibling FE estimate. The OLS estimates are significantly smaller when adjusted for age and PVT score (-1.138 and -0.960, respectively, in the two samples). Table 3 also shows that the FE point estimate declines to -0.221 when age and PVT are controlled for, which does not represent a statistically significant decline. Although this estimate is imprecise (SE=0.18), the 95% CI [-0.575, 0.134] includes neither of the two OLS estimates.

As shown in Table 4, the NLSY79 estimates are larger (in absolute terms) than the Add Health estimates. The unadjusted OLS results for the overall and sibling samples range from -2.135 to -1.946, respectively, about 0.6 of a year of schooling greater than the corresponding Add Health results. Adjusting for age and AFQT, the educational differential by teenage birth falls to -1.205 in the sample of all women and -1.008 in the sibling sample, values slightly larger than the corresponding Add Health estimates. The within-family estimates range from -0.656 unadjusted to -0.619 adjusted, which is about 0.4 years greater than the FE estimates from Add Health.

The overall impression left by the sibling fixed-effects estimates is one of small to moderate adverse effects of a teen birth relative to a sibling. Although the sibling estimates are less precise (larger SEs) than the OLS estimates, the confidence intervals are sufficiently narrow that they do not contain large adverse effects.

Propensity Score Matching Estimates

Tables 4 and 5 present our main propensity score matching results for Add Health and NLSY, respectively. All models are estimated using the command suite “teffects ipwra” in Stata 14 and share the same basic set of controls in the treatment logit: age and test score (PVT/AFQT). We report estimates of both the average treatment effect (ATE) and the average treatment effect on the treated (ATET). We implement Arkhangelsky and Imbens’ (2018) approach by exploiting similarities between observations in different families using family-average treatment information, namely the variation in the proportion of teen moms in the family.

The top panel of each table shows the estimated treatment effects based on a conventional PSM strategy using only the covariates mentioned above. Results with family-level average variables included in the treatment equation are shown in the bottom panel of the table. Specifically, all PSM cluster results are based on treatment equations that control for the proportion of teen moms in the family (in addition to age, test score and a constant term). The difference between the results in columns (1), (3) and (5) and columns (2), (4) and (6) is that, in the latter, the treatment models also control for mean age and mean test score. Both strategies are applied to the same three samples discussed above: All Families, Siblings, and Families with Discordant Siblings. The complete PSM estimates for the ATE models are shown in Appendix Tables A1-A4.

Looking at the conventional PSM results, in Add Health, the estimated average treatment effect of teen childbearing on years of schooling is -1.254 (SE=0.061) years in the full sample and -1.100 (SE=0.146) years in the (full) sibling sample. The corresponding ATEs for NLSY are -1.406 (0.078) and -1.327 (0.131). The ATET are smaller in absolute terms, especially in the NLSY sample. Overall the magnitudes of the conventional PSM estimates are similar to the adjusted OLS estimates in both samples and across both surveys. In fact, the ATET are almost identical to the corresponding adjusted OLS point estimates, while the ATE are greater in absolute terms.

The conventional PSM treatment effects are estimated with good precision and—taken at face value—they reject the small effects in Add Health and the moderate effects in NLSY79 implied by the sibling FE estimates. In addition, as shown in the Propensity Score Overlay Plots in Appendix Figures A1 and A2, the models have good overlap of propensity scores across treatment and comparison groups, i.e. there are no concerns about lack of common support. In terms of covariates in the treatment equation, both age and PVT/AFQT scores are found to be important predictors of treatment status in the matching models as shown in Appendix Tables A1 and A2 (for ATE; ATET results are available upon request).

The bottom of Tables 4 and 5 show the results from family cluster matching. Here, all treatment selection models include the proportion of teen moms in the family; additional family-level average covariates are included in models (2), (4) and (6). (Model 6 could not be estimated in NLSY due to insufficient overlap.) In the Add Health, the average treatment effect (ATE) of teen childbearing on years of schooling ranges from -1.176 (0.082) to -1.145 (0.090) years in the all women sample and from -0.915 (0.169) to -0.877 (0.176) years in the (full) sibling sample. The ATE cluster PSM estimates are about 0.1 to 0.2 years smaller than the conventional PSM ATE estimates. Similar order of magnitude reductions in the educational differential of teen birth are observed for the ATE results in NLSY79.

Looking at the ATET results, we find major differences between the conventional and the cluster PSM estimates. In the Add Health sample of all women, the average treatment effect on the treated (ATET) declines from -1.165 (0.060) based on the conventional matching strategy to -0.148 (0.135) in Model 1 and -0.190 (0.144) in Model 2. In the Add Health (full) sibling sample, the ATET declines from -0.980 (0.139) in the conventional PSM to -0.268 (0.186) in cluster PSM Model 3 and -0.315 (0.192) in cluster PSM Model 4. The corresponding cluster PSM ATET results range from -1.042 (0.143) to -1.039 (0.145) in the full sample and from -0.749 (0.177) to -0.727 (0.176) in the (full) sibling sample.

The cluster PSM strategy yields treatment effects that are most similar to the family FE estimates. In fact, the cluster PSM ATET estimates from Add Health are almost identical to the (near-zero) FE estimates and for the cluster ATET and FE estimates in the NLSY and the 95% Confidence Intervals overlap. We note that the cluster PSM effects are somewhat less precisely estimated than the conventional PSM estimates. Also, there is evidence that the application of the cluster matching strategy faces a common support challenge. As shown in Appendix Figures A3 and A4, the models with family mean treatment have high concentrations at extreme values of support. In terms of covariates in the treatment equation, as shown in Appendix Tables A1 and A2 (for ATE), the proportion of women with a teen birth in the family (a mean #Teen Moms) emerges as the key predictor of treatment (individual PVT/AFQT and age are no longer statistically significant predictors).

Inspecting the PSM estimates based on the sample of discordant siblings, we find that both matching strategies yield much smaller estimates than in the full sibling sample (see columns 5 and 6 in Tables 4 and 5). For example, in the discordant sample in Add Health, the average treatment effect of a teen birth on education (ATE) ranges from -0.192 (0.205) to -0.186 (0.203) based on the cluster PSM estimate. The corresponding conventional PSM point estimate is -0.233 (0.205), which is slightly larger (in absolute terms) than the cluster estimate, a pattern consistent with results for the larger samples. The effect magnitudes are very similar to the family FE results.

The similarity of the estimates for the discordant sibling samples is not surprising since in these samples a greater proportion of the total variation is within-family, i.e. the type of variation that identifies the FE estimates. In addition, families are more similar in this sample from an average treatment perspective, which means the cluster approach may add little. Results from a one-way ANOVA of the predicted propensity scores confirm that, in these samples, PSM tends to compare the treated more to their sibling (who, by construction of the samples, is not treated). Consistent with this, common support is not a concern for these estimates as shown by the overlap plots in the Appendix (see Panels V and VI in Figure A3 and Panel V in Figure A4).

In addition, we note that an interesting pattern emerges from the different outcome models for the "treated" (teen mothers) and "untreated" groups, particularly in models estimated on the discordant sibling subsample. First, the estimated effect of age on education is greater for the treated (teen mothers) than their untreated siblings. Specifically, in Add Health (ATE models): 0.118 (SE= 0.072) vs. -0.070 (SE=0.075); Table A3, column 6. Second, the impact of adolescent test scores on educational attainment is about twice as large for the untreated than their treated sisters: 0.046 (SE=0.010) vs. 0.020 (SE=0.009) for teen mothers.

These results are consistent with Geronimus' (2003) description of teenage mothers following an alternative life-course, and continuing to gain education as they age into their late twenties. While educational attainment does not literally decrease with age for the untreated, the negative (though not significant) coefficient could reflect cohort differences in attainment, since attainment has increased across these cohorts. The weaker relationship between test scores and attainment among teenage mothers is consistent with either teenage births reducing academic progress among more academically talented teenage mothers, or with births boosting the motivation of less-academically-oriented teenage mothers to stay in or return to school.

Discussion and Conclusion

Sibling difference methods are a well-known strategy for addressing selectivity bias due to omitted family-level variables. However, they face concerns over efficiency, generalizability and within-family selectivity. Recent advances in Propensity Score Matching (PSM) by Arkhangelsky and Imbens (2018) provide an alternative approach to estimating treatment effects in clustered data that may address some of these concerns by utilizing family-average treatment information. Using large family/sibling samples based on nationally representative data from "Add Health" and NLSY79, this paper illustrates this approach and compares cluster PSM treatment effects of teenage childbearing on years-of-schooling to family FE and conventional PSM estimates.

To our knowledge, this is the first application of Arkhangelsky and Imbens' approach to family/sibling data. Preliminary results indicate that the cluster approach yields estimates of the effect of teen birth on education that are smaller than conventional PSM estimates, which are of comparable magnitude to adjusted OLS estimates, and more similar to the (nearer-zero) family FE estimates which address bias from family-level unobservables. This indicates that the cluster PSM approach has the potential to act as a quasi-fixed effects approach in family data while delivering an additional advantages of "double robustness" to misspecifications.

The fact that the average treatment effects on the treated (ATET) in the cluster strategy were consistently more similar than the ATEs to the FE estimates is noteworthy in the context of the ongoing debate on method choice and effect heterogeneity (Heiland, Korenman and Smith 2019; Diaz and Fiel 2016; Kane et al. 2013). It is consistent with heterogeneous treatment effects in teenage childbearing. It also supports the notion that sibling FE estimates provide policy-relevant estimates of the impact of teen childbearing as those estimates generalize to the educational consequences faced by actual teenage mothers (see discussion in Heiland, Korenman and Smith 2019).

While the cluster PSM approach by Arkhangelsky and Imbens (2018) is promising in the context of family/sibling data, the implementation may face challenges. In our application, we observed that the cluster PSM effects were slightly less precise than conventional PSM estimates and we encountered lack of common support in the cluster-matched propensity scores. These issues will require further investigation.

References

Allison, P.D. (2009). *Fixed effects regression models*. Newbury Park, CA: Sage Publications.

Almanza, C.H., and Sahn, D.E. (2018). Early Childbearing, School Attainment, and Cognitive Skills: Evidence from Madagascar. *Demography* 55, 643–668.

Angrist, J., and Pischke, J.-S. (2015). *Mastering `metrics: The path from cause to effect*. Princeton NJ: Princeton University Press.

Angrist, J., and Pischke, J.-S. (2009). *Mostly harmless econometri*cs. Princeton, NJ: Princeton University Press.

Arkhangelsky, D., and Imbens, G. (2018). The Role of the Propensity Score in Fixed Effect Models (NBER Working Paper No. 24814). Cambridge, MA: National Bureau of Economic Research.

Ashcraft, A., Fernández-Val, I., and Lang, K. (2013). The consequences of teenage childbearing: Consistent estimates when abortion makes miscarriage non-random. *The Economic Journal*, 123, 875–905.

Ashcraft, A., and Lang, K. (2006). The consequences of teenage childbearing (NBER Working Paper No.12485). Cambridge, MA: National Bureau of Economic Research.

Banerjee, A., Chassang, S., and Snowberg, E. (2016). Decision Theoretic Approaches to Experiment Design and External Validity (NBER Working Paper No. 22167). Cambridge, MA: National Bureau of Economic Research.

Ben-Porath, Y. (1980). The F-Connection: Families, Friends, and Firms and the Organization of Exchange, *Population and Development Review*, 6(1), 1-30.

Boardman, J., and Fletcher, J. (2015). To cause or not to cause? That is the question, but identical twins might not have all of the answers. *Social Science and Medicine*, 127, Comments, 198–200.

Bound, J., Jaeger, D., and Baker, R. (1995). Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Variables Is Weak. *Journal of the American Statistical Association* 90, 443–450.

Bound, J., and Solon, G. (1999). Double trouble: on the value of twins-based estimation of the return to schooling. *Economics of Education Review* 18*,* 169–182.

Bound, J., and Turner, S.E. (2007). Cohort crowding: How resources affect collegiate attainment. *Journal of Public Economics* 91(5-6), 877–899.

Campbell, J. (1998). Review of Dunn, L. M., and Dunn, L. M. (1997). *Peabody Picture Vocabulary Test, Third Edition.* Circle Pines, MN: American Guidance Service. *Journal of Psychoeducational Assessment* 16, 334–338.

Chen, P., and Chantala, K. (2014). *Guidelines for Analyzing Add Health Data*. Carolina Population Center. Updated March 2014.

Cutler, D.M., and Katz, L.F. (1992). Rising Inequality? Changes in the Distribution of Income and Consumption in the 1980's. *The American Economic Review*, 82(2), Papers and Proceedings of the Hundred and Fourth Annual Meeting of the American Economic Association, 546–551.

Diaz, C.J., and Fiel, J.E. (2016). The Effect(s) of Teen Pregnancy: Reconciling Theory, Methods, and Findings, *Demography* 53, 85–116.

Domingue, B., Belsky, D., Conley, D., Harris, K., and Boardman, J. (2015). Polygenic influence on educational attainment: New evidence from the National Longitudinal Survey of Adolescent to Adult Health. *AERA Open* 1(3), 1–13.

Duncan, G. J., Lee, K. T. H., Rosales-Rueda, M., and Kalil, A. (2018). Maternal Age and Child Development. *Demography* 55, 22–29.

Fletcher, J. M., and Wolfe, B.L. (2009). Education and labor market consequences of teenage childbearing. *Journal of Human Resources* 44, 303–325.

Geronimus, A.T. (1987). On Teenage Childbearing and Neonatal Mortality in the UnitedStates. *Population and Development Review* 13(2), 245–279.

Geronimus, A.T. (2003). Damned if You Do: Culture, Identity, Privilege and Teenage Childbearing in the United States. *Social Science and Medicine* 57, 881–893.

Geronimus, A.T., and Korenman, S. (1992). The socioeconomic consequences of teen childbearing reconsidered. *Quarterly Journal of Economics* 107, 1187–1214.

Geronimus, A. T., and Korenman, S. (1993). The socioeconomic costs of teenage childbearing: Evidence and interpretation. *Demography* 30, 281–290.

Harris, K.M. (2009). The National Longitudinal Study of Adolescent to Adult Health (Add Health), Waves I & II, 1994–1996; Wave III, 2001–2002; Wave IV, 2007-2009 [machine-readable data file and documentation]. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill. doi: 10.3886/ICPSR27021.v9.

Harris, K.M. (2013). The Add Health Study: Design and Accomplishments. Carolina Population Center, University of North Carolina at Chapel Hill.

Harris, K.M., Halpern, C.T., Whitsel, E., Hussey, J., Tabor, J., Entzel, P., and Udry, J.R. (2009). The National Longitudinal Study of Adolescent to Adult Health: Research Design [WWW document]. URL: http://www.cpc.unc.edu/projects/addhealth/design.

Harris, K.M., Halpen, C., Habertsick, B., and Smolen, A. (2013). The National Longitudinal Study of Adolescent Health (Add Health) sibling pairs data. *Twin Research and Human Genetics* 16(1), 391–398.

Heiland, F., S. Korenman and R. Smith. 2019. Estimating the Educational Consequences of Teenage Childbearing: Identification, Heterogeneous Effects and the Value of Biological Relationship Data. *Economics & Human Biology* 33, 15–28.

Hellman, J. (2018). Judge rules against Trump administration in teen pregnancy prevention case. *The Hill*. April 19. https://thehill.com/policy/healthcare/383996-judge-rules-trump-administrations-cut-to-teen-pregnancy-prevention-program. Accessed March 22, 2019.

Hoffman, S.D., Foster, E.M., and Furstenberg, F.F., Jr. (1993). Reevaluating the costs of teenage childbearing. *Demography* 30, 1–13.

Hotz, V.J., McElroy, S.W., and Sanders, S.G. (2005). Teenage childbearing and its life cycle consequences exploiting a natural experiment. *Journal of Human Resources* 40, 683–715.

Imbens, G.W., and Rubin, D.B. (2015). *Causal inference for statistics, social, and biomedical sciences: An introduction*. New York: Cambridge University Press.

Kane, J.B., Morgan, S.P., Harris, K.M., and Guilkey, D.K. (2013). The Educational Consequences of Teen Childbearing. *Demography* 50, 2129–2150.

Kearney, M.S., and Levine, P.B. (2012). Why is the Teen Birth Rate in the United States so High and Why Does it Matter? *Journal of Economic Perspectives* 26(2), 141–166.

Kearney, M.S. (2010). Teen Pregnancy Prevention. In *Targeting Investments in Children: Fighting Poverty When Resources are Limited.* P.B. Levine & D.J. Zimmerman, eds. Chicago: University of Chicago Press.

King, G., Honaker, J. Joseph, A., and Scheve, K. (2001). Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation. *American Political Science Review* 95(1), 49–69.

Langa, E and Nystedt, P. (2018). Two by two, inch by inch: Height as an indicator of environmental conditions during childhood and its influence on earnings over the life cycle among twins. *Economics and Human Biology* 28, 53-66.

Leamer, E. (1983). Let's Take the Con out of Econometrics. *American Economic Review* 73, 31–44.

Lee, D. (2010). The early socioeconomic effects of teenage childbearing: A propensity score matching approach. *Demographic Research*, 23(Article 25), 697–736. doi:10.4054/DemRes.2010.23.25

Levine, D. I., and Painter, G. (2003). The schooling costs of teenage out-of-wedlock childbearing: Analysis with a within-school propensity-score-matching estimator. *The Review of Economics and Statistics* 85, 884–900.

Lundborg, P., Nystedt, P., and Rooth, D.O. (2014). Body size, skills, and income: evidence from 150,000 teenage siblings. *Demography* 51(5), 1573–1596.

Maddala, G.S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Magnuson, K., Duncan, G.J., Lee, K.T.H., and Metzger, M.W. (2016). Early school adjustment and educational attainment. *American Educational Research Journal* 53(4), 1198–1228.

Martin, J. A., Hamilton, B. E., Osterman, M. J. K., Driscoll, A. K., and Drake, P. (2018). Births: Final Data for 2017. National Vital Statistics Reports Volume 67, Number 8 (November 7, 2018).

Moulton, B. (1986). Random Group Effects and the Precision of Regression Estimates, *Journal of Econometrics* 32, 385–397.

Moulton, B. (1990). An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units. *Review of Economics and Statistics* 72(2), 334–338.

Nielsen, J.S., Bech, M., Christensen, K., Kiil, A., and Hvidt, N.C. (2017). Risk aversion and religious behaviour: Analysis using a sample of Danish twins, *Economic and Human Biology* 26, 21–29.

Schulkind, L., and Sandler, D. H. (2019). The Timing of Teenage Births: Estimating the Effect on High School Graduation and Later-Life Outcomes. *Demography* 56(1), 345–365.

Stock, J., Wright, J.H., and Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics* 20, 518–529.

Wall-Wieler, E., Roos, L.L., and Nickel, N.C. (2016). Teenage pregnancy: the impact of maternal adolescent childbearing and older sister's teenage pregnancy on a younger sister. BMC Pregnancy Childbirth 16, doi: 10.1186/s12884-016-0911-2.

Wehby, G. (2014). Breastfeeding and child disability: A comparison of siblings from the United States. *Economics and Human Biology* 15, 13-22.

Yakusheva, O. (2011). In high school and pregnant: the importance of educational and fertility expectations for subsequent outcomes. *Economic Inquiry* 49(3), 810–837.

Yakusheva, O., and Fletcher, J. (2015). Learning from Teen Childbearing Experiences of Close Friends: Evidence using Miscarriages as a Natural Experiment. *The Review of Economics and Statistics* 97(1), 29–43.

Table 1: Means by Type of Sample and Survey (unweighted), Add Health and NLSY79.

	Add Health			NLSY79		
	All Women	Fam. w/ ≥ 2 Sibs.	Discordant Sibs.	All Women	Fam. w/ ≥ 2 Sibs.	Discordant Sibs.
Number of Obs.	8,345	1,361	289	5,305	1,558	383
<u>Outcome</u>						
Years of Completed Education[1]	14.5	14.2	13.3	12.9	13.0	11.8
<u>Characteristics</u>						
Teen Birth	11.6 %	14.5 %	49.5 %	19.9 %	16.9 %	48.6 %
Two-Parent Fam.[2]	52.1 %	45.8 %	27.7 %	67.1 %	70.4 %	55.4 %
Parents' Education[3]	13.2	13.0	12.3	10.7	10.5	9.7
Income-To-Needs Ratio (% missing)[4]	3.1 (25.3)	2.8 (23.4)	1.8 (18.3)	2.3 (20.9)	2.2 (17.5)	1.5 (12.3)
# of "Siblings"[5]	1.18	2.08	2.14	0.67	2.28	2.38
Race/Ethnicity						
NH Black	23.2 %	26.5 %	43.3 %			
Hispanics	15.7 %	14.0 %	19.7 %			
NH White	53.9 %	54.2 %	35.6 %			
NH Other	7.2 %	5.4 %	1.4 %			
Black				25.5 %	30.0 %	45.2 %
Hispanic				16.4 %	16.4 %	13.8 %
Not Black or Hisp.				58.2 %	53.7 %	41.0 %
Foreign-born	6.4 %	5.0 %	3.5 %	6.3 %	5.5 %	3.7 %
PVT[6]	99.8	97.6	92.4			
AFQT[7]				41.3	40.3	25.5
Age[8]	15.5	15.5	15.2	17.7	17.3	17.0

Notes:

1. Add Health measure recoded following Fletcher, Max = 21 years.
2. Living with both parents in Wave I in Add Health and at age 14 in NLSY79.
3. Years of education completed. Add Health measure is recoded from a categorical variable on the parent survey. NLSY79 measure is father's education or mother's education if father data is missing.
4. Household income data are missing for a large proportion of the sample.
5. "Siblings" in Add Health are any co-residing female Wave I sample members. "Siblings" in NLSY79 are any co-residing female sample members identified as sisters on the baseline household roster.
6. PVT score is the age-standardized Wave I Add Health Picture Vocabulary Test (normed against a distribution with a mean of 100 and standard deviation of 15).
7. AFQT is the age-standardized Armed Forces Qualification Test score in the NLSY79 (a percentile score based on Armed Services Vocational Aptitude Battery administered in 1980 and renormed in 2006).
8. Age is measured age at baseline (Wave I in Add Health, baseline interview in 1979 in NLSY79).

Table 2: Sample Means by Family Type, Breakdown of Sibling Sample (N=1,361), Add Health. [1, 2, 3]

Variable	Fam. w/ No Sibs Teen Moms	Discordant Siblings Families, i.e. Sibs in Mixed Teen/Non-teen Families (N=289)		Fam. w/ All Sibs Teen Moms
		Not Teen Moms	Teen Moms	
Number of women (outcome variable sample size)	1,018	146	143	54
Outcome				
Years of Completed Education (Wave IV) (recoded following Fletcher, Max = 21 years)	14.6	13.4	13.1	12.8
Characteristics				
Two-Parent Family	51.6	28.6	26.8	35.4
Parent's Education (yrs. recoded from categorical)	13.2	12.3	12.3	12.4
Income-to-Needs Ratio	3.10	1.93	2.00	2.17
Race/Ethnicity				
NH Black	21.1	43.8	42.7	37.0
Hispanic	12.5	19.9	19.6	11.1
NH White	59.8	34.9	36.4	46.3
NH Other	6.6	1.4	1.4	5.6
Foreign-born	5.4	3.8	3.2	6.4
PVT score	99.4	92.9	91.9	91.5
Age (Wave IV)	28.4	28.2	27.8	28.1
Per Capita Income (census tract)	13.0	10.3	10.5	10.8
Age at first birth (conditional)	23.1	21.9	16.9	17.3
Number of siblings in sibling sample	2.1	2.1	2.1	2.1
Number of teen moms among respondents in family	0.0	1.0	1.1	2.1
Number of moms among respondents in family	1.1	1.8	1.8	2.1

Notes:

1. Teen mother is defined having a live birth before exact age 19. Non-mothers are included in the non-teen mother category.
2. Siblings are defined as co-resident female sample members at Wave I with educational outcomes at Wave IV.
3. Total number of observations (based on outcome variable) is 1,361 (=sum of obs. across four types of families).

Table 3: XSEC OLS and Sibling Fixed Effect Estimates of Teen Birth on Years of Education, Add Health.[1, 2, 3]

	Coefficients (SEs), Number of Observations (N)					
	All Families (N=8,345)		Families w/ ≥ 2 Siblings (N=1,361)			
	XSEC OLS		XSEC OLS		Sib Fixed Effects	
Variable	No Controls	Age & PVT	No Controls	Age & PVT	No Controls	Age & PVT
Teen Birth	**-1.464**	**-1.138**	**-1.362**	**-0.960**	**-0.237**	**-0.221**
	(0.061)	(0.060)	(0.148)	(0.144)	(0.170)	(0.181)
95% CI	[-1.58,-1.35]	[-1.26, -1.02]	[-1.65,-1.07]	[-1.24,-0.68]	[-0.57,0.10]	[-0.58,0.13]
Age		**0.005**		**0.060**		**0.045**
		(0.011)		(0.031)		(0.039)
PVT		**0.053**		**0.057**		**0.030**
		(0.002)		(0.004)		(0.006)
Constant	**14.655**	**14.449**	**14.418**	**12.767**	**14.255**	**12.830**
	(0.025)	(0.343)	(0.074)	(0.916)	(0.045)	(1.191)

Notes:

1. "Siblings": young women co-resident at Wave I interview.
2. Robust SEs: for pooled sample, clustered on family ID, for fixed-effects: robust SEs, not clustered.
3. Age is measured age at Wave IV. PVT score is the Wave I Add Health Picture Vocabulary Test with missing observations imputed as a value of 50. A separate dummy for PVT non-missing is included.

Table 4: XSEC OLS and Sibling Fixed Effect Estimates of Teen Birth on Years of Education, NLSY79.[1, 2, 3]

	Coefficients (SEs), Number of Observations (N)					
	All Families (N=5,305)		Families w/ ≥ 2 Siblings (N=1,558)			
	XSEC OLS		XSEC OLS		Sib Fixed Effects	
Variable	No Controls	Age & AFQT	No Controls	Age & AFQT	No Controls	Age & AFQT
Teen Birth	**-2.135**	**-1.205**	**-1.946**	**-1.008**	**-0.656**	**-0.619**
	(0.074)	(0.066)	(0.140)	(0.122)	(0.160)	(0.151)
95% CI	[-2.28,-1.99]	[-1.34, -1.07]	[-2.22,-1.67]	[-1.25,-0.77]	[-0.97,-0.34]	[-0.92,-0.32]
Age		**0.066**		**0.129**		**0.031**
		(0.012)		(0.022)		(0.027)
AFQT		**0.046**		**0.043**		**0.035**
		(0.001)		(0.002)		(0.003)
Constant	**13.291**	**11.197**	**13.321**	**10.178**	**13.102**	**11.701**
	(0.039)	(0.271)	(0.077)	(0.485)	(0.046)	(0.782)

Notes:

1. "Siblings": young women co-resident and identified as sisters in baseline interview (1979).
2. Robust SEs: for pooled sample, clustered on family ID, for fixed-effects: robust SEs, not clustered.
3. Age is measured at baseline (1979). AFQT is the Armed Forces Qualification Test score (based on Armed Services Vocational Aptitude Battery administered in 1980; recoded in 2006) with missing observations imputed as a value of 50. A separate dummy for AFQT non-missing is included.

Table 5: PSM Estimates of Effect of Teen Birth on Years of Education, Various Samples, Add Health.[1, 2]

	Treatment Effect of Teen Birth (SEs)					
	All Families (N=8,345)		Families w/ 2 Sibs. (N=1,361)		Fam. w/ Discordant Sibs. (N=289)[3]	
	(1)	(2)	(3)	(4)	(5)	(6)
			Conventional PSM			
ATE	**-1.254**		**-1.100**		**-0.233**	
	(0.061)		(0.146)		(0.205)	
95% CI	[-1.37, -1.14]		[-1.39, -0.81]		[-0.64, 0.17]	
ATET	**-1.165**		**-0.980**		**-0.252**	
	(0.060)		(0.139)		(0.212)	
95% CI	[-1.28, -1.05]		[-1.25, -0.71]		[-0.67, 0.16]	
Treatment Model (Logit) Specification						
Age	✓		✓		✓	
PVT	✓		✓		✓	
Constant	✓		✓		✓	
			Cluster PSM			
ATE	**-1.176**	**-1.145**	**-0.915**	**-0.877**	**-0.192**	**-0.186**
	(0.082)	(0.090)	(0.169)	(0.176)	(0.205)	(0.203)
95% CI	[-1.34, -1.02]	[-1.32, -0.97]	[-1.25, -0.58]	[-1.22, -0.53]	[-0.59, 0.20]	[-0.58, 0.21]
ATET	**-0.148**	**-0.190**	**-0.268**	**-0.315**	**-0.228**	**-0.267**
	(0.135)	(0.144)	(0.186)	(0.192)	(0.207)	(0.211)
95% CI	[-0.41, 0.12]	[-0.47, 0.09]	[-0.63, 0.10]	[-0.69, 0.06]	[-0.63, 0.18]	[-0.68, 0.14]
Treatment Model (Logit) Specification						
Age	✓	✓	✓	✓	✓	✓
PVT	✓	✓	✓	✓	✓	✓
Mean #Teen Moms	✓	✓	✓	✓	✓	✓
Mean Age		✓		✓		✓
Mean PVT		✓		✓		✓
Constant	✓	✓	✓	✓	✓	✓

Notes:

1. Models are estimated using the teffects command in Stata 14: “teffects ipwra (…) (…), aeq” with option “ate” or “atet” to estimate the Average Treatment Effect (ATE) or the Average Treatment Effect on the Treated (ATET).
2. Age is measured age at Wave IV. PVT score is the Wave I Add Health Picture Vocabulary Test with missing observations imputed as a value of 50. A separate dummy for PVT non-missing is included.
3. Discordant siblings are siblings from families in which at least one woman had a teenage birth and one did not. Teenage is defined as younger than exact age 19.

Table 6: PSM Estimates of Effect of Teen Birth on Years of Education, Various Samples, NLSY79.[1, 2]

	Treatment Effect of Teen Birth (SEs)					
	All Families (N=5,305)		Families w/ 2 Sibs. (N=1,558)		Fam. w/ Discordant Sibs. (N=383)[3]	
	(1)	(2)	(3)	(4)	(5)	(6)
			Conventional PSM			
ATE	**-1.406**		**-1.327**		**-0.621**	
	(0.078)		(0.131)		(0.163)	
95% CI	[-1.56, -1.25]		[-1.57, -1.06]		[-0.94, -0.30]	
ATET	**-1.165**		**-0.989**		**-0.645**	
	(0.069)		(0.120)		(0.171)	
95% CI	[-1.30, -1.03]		[-1.22, -0.75]		[-0.98, -0.31]	
Treatment Model (Logit) Specification						
Age	✓		✓		✓	
AFQT	✓		✓		✓	
Constant	✓		✓		✓	
			Cluster PSM			
ATE	**-1.276**	**-1.272**	**-1.246**	**-1.229**	**-0.642**	(*)[4]
	(0.086)	(0.085)	(0.156)	(0.156)	(0.12)	
95% CI	[-1.44,-1.11]	[-1.44,-1.10]	[-1.55,-0.94]	[-1.53,-0.92]	[-0.98,-0.30]	
ATET	**-1.042**	**-1.039**	**-0.749**	**-0.810**	**-0.680**	(*)[4]
	(0.143)	(0.145)	(0.177)	(0.172)	(0.172)	
95% CI	[-1.32, -0.76]	[-1.32, -0.76]	[-1.09 -0.40]	[-1.15, -0.47]	[-1.02, -0.34]	
Treatment Model (Logit) Specification						
Age	✓	✓	✓	✓	✓	✓
AFQT	✓	✓	✓	✓	✓	✓
Mean #Teen Moms	✓	✓	✓	✓	✓	✓
Mean Age		✓		✓		✓
Mean AFQT		✓		✓		✓
Constant	✓	✓	✓	✓	✓	✓

Notes:

1. Models are estimated using the teffects command in Stata 14: "teffects ipwra (…) (…), aeq" with option "ate" or "atet" to estimate the Average Treatment Effect (ATE) or the Average Treatment Effect on the Treated (ATET).
2. Age is measured at baseline (1979). AFQT is the Armed Forces Qualification Test score (based on Armed Services Vocational Aptitude Battery administered in 1980; recoded in 2006) with missing observations imputed as a value of 50. A separate dummy for AFQT non-missing is included.
3. Discordant siblings are siblings from families in which at least one sister had a teenage birth and one did not. Teenage is defined as younger than exact age 19.
4. Model could not be estimated due to insufficient overlap.

Appendix Tables & Figures

Table A1: Conventional PSM Estimates of Teen Birth on Years of Education, Various Samples, Add Health.[1, 2]

	Coefficient (SEs)		
	All Families (N=8,345)	Families w/ ≥ 2 Siblings (N=1,361)	Families w/ Discordant Siblings[3] (N=289)
ATE (Teen Birth)	**-1.254** (0.061)	**-1.100** (0.146)	**-0.233** (0.205)
95% CI	[-1.37, -1.14]	[-1.39, -0.81]	[-0.64, 0.17]
Outcome Model "Untreated"			
Age	**0.002** (0.012)	**0.052** (0.032)	**-0.064** (0.075)
PVT	**0.053** (0.002)	**0.058** (0.005)	**0.050** (0.010)
Constant	**14.558** (0.366)	**13.039** (0.953)	**15.029** (2.132)
Outcome Model "Treated"			
Age	**0.037** (0.032)	**0.117** (0.082)	**0.101** (0.076)
PVT	**0.031** (0.004)	**0.025** (0.011)	**0.022** (0.009)
Constant	**12.323** (0.988)	**10.074** (2.433)	**10.402** (2.289)
Treatment Model (Logit)			
Age	**-0.062** (0.019)	**-0.156** (0.044)	**-0.130** (0.067)
PVT	**-0.029** (0.002)	**-0.036** (0.006)	**-0.005** (0.009)
Constant	**-0.313** (0.546)	**2.221** (1.329)	**3.702** (1.904)

Notes:

1. Models are estimated using the teffects command in Stata 14: "teffects ipwra (…) (…), aeq".
2. Age is measured age at Wave IV. PVT score is the Wave I Add Health Picture Vocabulary Test with missing observations imputed as a value of 50. A separate dummy for PVT non-missing is included.
3. Discordant siblings are siblings from families in which at least one woman had a teenage birth and one did not. Teenage is defined as younger than exact age 19.

Table A2: Conventional PSM Estimates of Teen Birth on Years of Education, Various Samples, NLSY79.[1, 2]

	Coefficient (SEs)		
	All Families (N=5,305)	Families w/ ≥ 2 Siblings (N=1,558)	Families w/ Discordant Siblings[3] (N=383)
ATE (Teen Birth)	**-1.406** (0.078)	**-1.327** (0.131)	**-0.623** (0.169)
95% CI	[-1.56, -1.25]	[-1.57, -1.06]	[-0.95, -0.29]
Outcome Model "Untreated"			
Age	**0.057** (0.014)	**0.120** (0.025)	**0.085** (0.056)
AFQT	**0.047** (0.001)	**0.044** (0.002)	**0.037** (0.007)
Constant	**11.416** (0.317)	**10.423** (0.571)	**12.465** (1.376)
Outcome Model "Treated"			
Age	**0.060** (0.028)	**0.118** (0.059)	**0.085** (0.065)
AFQT	**0.031** (0.003)	**0.023** (0.005)	**0.023** (0.005)
Constant	**9.897** (0.567)	**9.027** (1.104)	**9.717** (1.233)
Treatment Model (Logit)			
Age	**0.003** (0.016)	**-0.121** (0.035)	**0.003** (0.054)
AFQT	**-0.033** (0.002)	**-0.034** (0.003)	**-0.001** (0.005)
Constant	**-1.238** (0.346)	**0.800** (0.740)	**1.044** (1.223)

Notes:

1. Models are estimated using the teffects command in Stata 14: "teffects ipwra (…) (…), aeq".
2. Age is measured at baseline (1979). AFQT is the Armed Forces Qualification Test score (based on Armed Services Vocational Aptitude Battery administered in 1980; recoded in 2006) with missing observations imputed as a value of 50. A separate dummy for AFQT non-missing is included.
3. Discordant siblings are siblings from families in which at least one sister had a teenage birth and one did not. Teenage is defined as younger than exact age 19.

Table A3: Cluster PSM Estimates of Teen Birth on Years of Education, Various Samples, Add Health.[1, 2]

	Coefficient (SEs)					
	All Families (N=8,345)		Families w/ 2 Sibs. (N=1,361)		Fam. w/ Discordant Sibs. (N=289)[3]	
	(1)	(2)	(3)	(4)	(5)	(6)
ATE (Teen Birth)	**-1.176**	**-1.145**	**-0.915**	**-0.877**	**-0.192**	**-0.186**
	(0.082)	(0.090)	(0.169)	(0.176)	(0.205)	(0.203)
95% CI	[-1.34,-1.02]	[-1.32,-0.97]	[-1.25,-0.58]	[-1.22,-0.53]	[-0.59,0.20]	[-0.58,0.21]
Outcome Model "Untreated"						
Age	**0.007**	**0.005**	**0.051**	**0.047**	**-0.073**	**-0.070**
	(0.012)	(0.012)	(0.032)	(0.032)	(0.074)	(0.075)
PVT	**0.056**	**0.056**	**0.062**	**0.061**	**0.049**	**0.046**
	(0.002)	(0.002)	(0.004)	(0.004)	(0.010)	(0.010)
Constant	**14.329**	**14.446**	**12.794**	**13.078**	**15.135**	**15.317**
	(0.386)	(0.367)	(0.993)	(0.954)	(2.080)	(2.150)
Outcome Model "Treated"						
Age	**0.061**	**0.082**	**0.138**	**0.163**	**0.105**	**0.118**
	(0.037)	(0.037)	(0.070)	(0.065)	(0.074)	(0.072)
PVT	**0.027**	**0.028**	**0.020**	**0.020**	**0.020**	**0.020**
	(0.004)	(0.005)	(0.009)	(0.009)	(0.009)	(0.009)
Constant	**11.571**	**11.033**	**9.284**	**8.823**	**10.225**	**10.025**
	(1.163)	(1.118)	(2.202)	(1.982)	(2.280)	(2.170)
Treatment Model (Logit)						
Age	**-0.144**	**-0.327**	**-0.141**	**-0.304**	**-0.132**	**-0.282**
	(0.072)	(0.127)	(0.071)	(0.118)	(0.068)	(0.107)
PVT	**-0.009**	**-0.031**	**-0.008**	**-0.027**	**-0.005**	**-0.024**
	(0.010)	(0.023)	(0.009)	(0.020)	(0.009)	(0.018)
Mean #Teen Moms	**15.461**	**15.618**	**11.872**	**12.153**	**4.257**	**4.433**
	(0.656)	(0.323)	(0.730)	(0.763)	(2.102)	(2.209)
Mean Age		**0.323**		**0.300**		**0.281**
		(0.160)		(0.154)		(0.143)
Mean PVT		**0.027**		**0.025**		**0.024**
		(0.026)		(0.024)		(0.021)
Constant	**-4.217**	**-8.319**	**-2.433**	**-6.467**	**1.447**	**-2.198**
	(2.114)	(2.767)	(2.053)	(2.779)	(2.197)	(2.974)

Notes:

1. Models are estimated using the teffects command in Stata 14: "teffects ipwra (…) (…), aeq".
2. Age is measured age at Wave IV. PVT score is the Wave I Add Health Picture Vocabulary Test with missing observations imputed as a value of 50. A separate dummy for PVT non-missing is included.
3. Discordant siblings are siblings from families in which at least one woman had a teenage birth and one did not. Teenage is defined as younger than exact age 19.

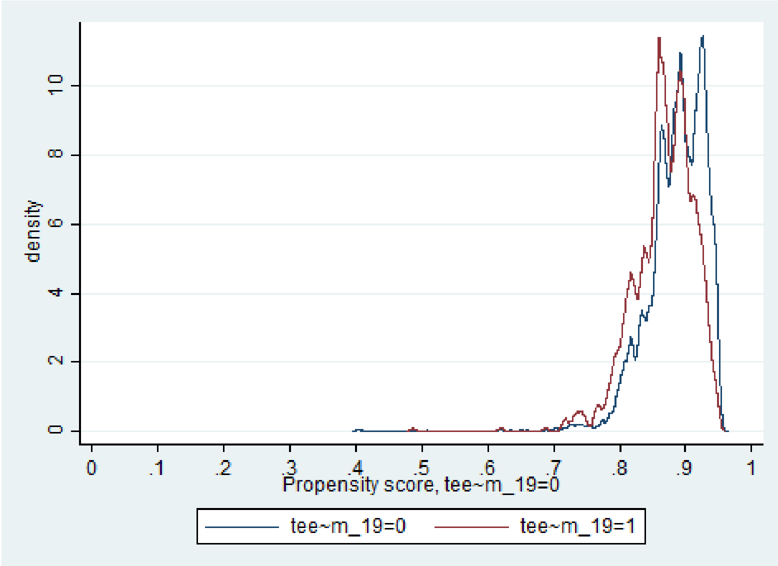
Table A4: Cluster PSM Estimates of Teen Birth on Years of Education, Various Samples, NLSY79.[1, 2]

	Coefficient (SEs)					
	All Families (N=5,305)		Families w/ 2 Sibs. (N=1,558)		Fam. w/ Discordant Sibs. (N=383)[3]	
	(1)	(2)	(3)	(4)	(5)	(6)[4]
ATE (Teen Birth)	**-1.276**	**-1.272**	**-1.246**	**-1.229**	**-0.641**	
	(0.086)	(0.085)	(0.156)	(0.156)	(0.172)	
95% CI	[-1.44,-1.11]	[-1.44,-1.10]	[-1.55,-0.94]	[-1.53,-0.92]	[-0.98,-0.30]	
Outcome Model "Untreated"						
Age	**0.067**	**0.068**	**0.126**	**0.129**	**0.088**	
	(0.014)	(0.014)	(0.026)	(0.026)	(0.057)	
AFQT	**0.047**	**0.047**	**0.044**	**0.044**	**0.037**	
	(0.001)	(0.001)	(0.002)	(0.002)	(0.007)	
Constant	**11.279**	**11.235**	**10.461**	**10.300**	**12.372**	
	(0.322)	(0.319)	(0.621)	(0.589)	(1.397)	
Outcome Model "Treated"						
Age	**0.015**	**0.016**	**0.027**	**0.031**	**0.067**	
	(0.049)	(0.049)	(0.104)	(0.102)	(0.01)	
AFQT	**0.036**	**0.037**	**0.028**	**0.029**	**0.023**	
	(0.003)	(0.003)	(0.005)	(0.005)	(0.005)	
Constant	**10.720**	**10.711**	**10.682**	**10.701**	**10.036**	
	(0.922)	(0.911)	(1.820)	(1.781)	(1.335)	
Treatment Model (Logit)						
Age	**-0.024**	**-0.029**	**0.005**	**0.031**	**0.011**	
	(0.054)	(0.090)	(0.058)	(0.085)	(0.056)	
AFQT	**-0.006**	**-0.002**	-**0.006**	**-0.005**	**-0.002**	
	(0.005)	(0.011)	(0.005)	(0.010)	(0.005)	
Mean #Teen Moms	**12.572**	**12.619**	**10.164**	**10.264**	**4.192**	
	(0.436)	(0.451)	(0.509)	(0.539)	(1.186)	
Mean Age		**0.008**		**-0.052**		
		(0.111)		(0.116)		
Mean AFQT		**-0.004**		**-0.001**		
		(0.012)		(0.012)		
Constant	**-5.107**	**-6.181**	-4.325	**-5.045**	**-1.188**	
	(1.084)	(1.373)	(1.171)	(1.613)	(1.365)	

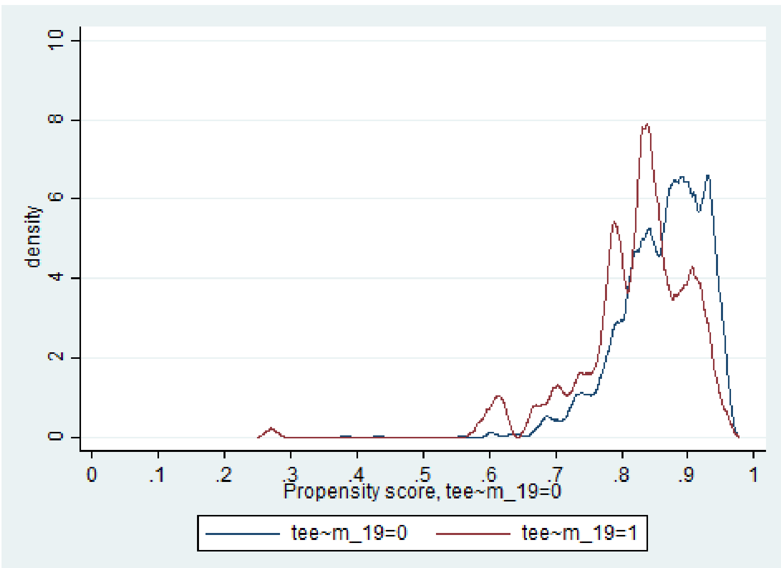
Notes:

1. Models are estimated using the teffects command in Stata 14: "teffects ipwra (…) (…), aeq".
2. Age is measured at baseline (1979). AFQT is the Armed Forces Qualification Test score (based on Armed Services Vocational Aptitude Battery administered in 1980; recoded in 2006) with missing observations imputed as a value of 50. A separate dummy for AFQT non-missing is included.
3. Discordant siblings are siblings from families in which at least one sister had a teenage birth and one did not. Teenage is defined as younger than exact age 19.
4. Model could not be estimated due to insufficient overlap.

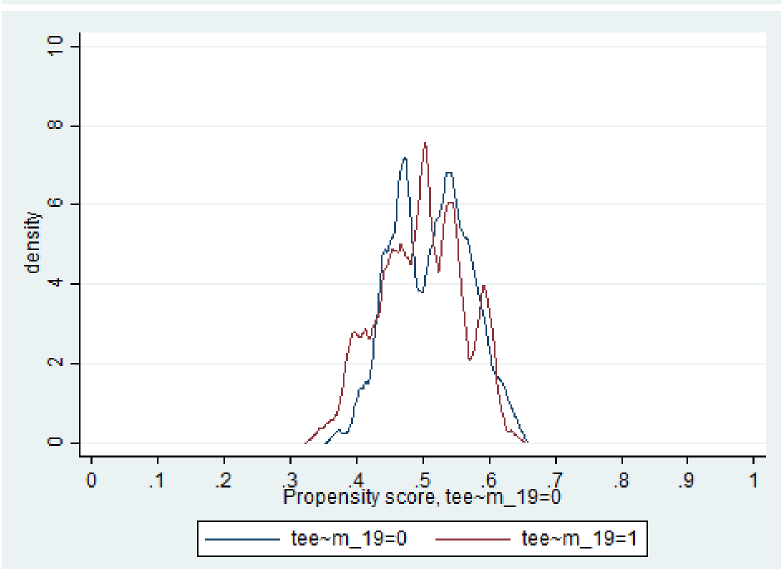
Figure A1: Propensity Score Overlap Plots, Conventional PSM, Add Health (corresp. to models in Table A1)



Panel I. Model 1 (N=8,345)

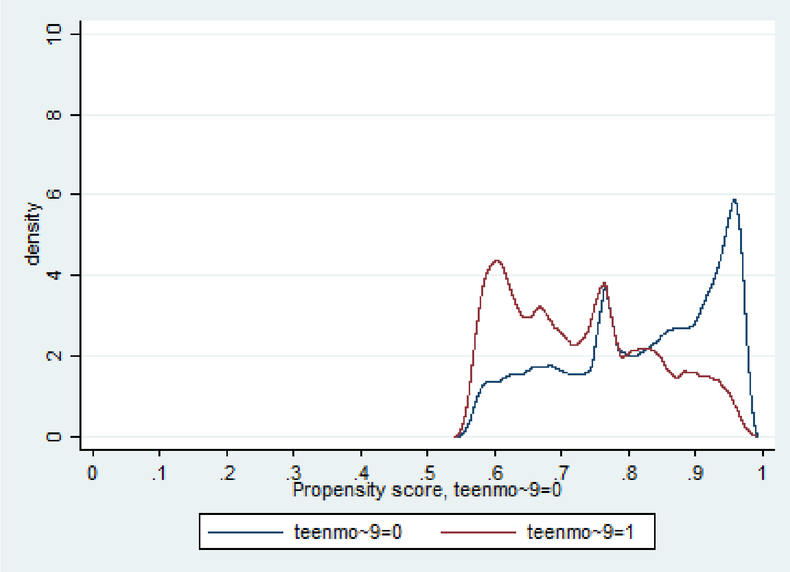


Panel II. Model 2 (N=1,361)

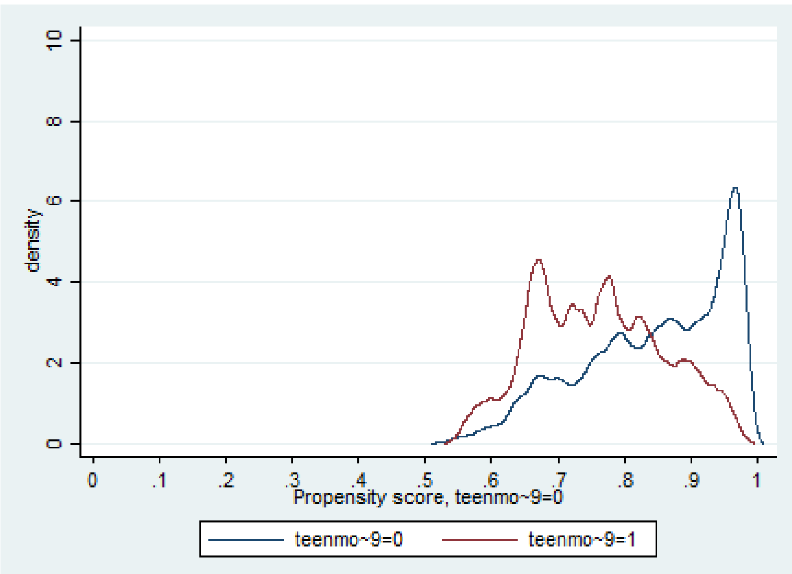


Panel III. Model 3 (N=289)

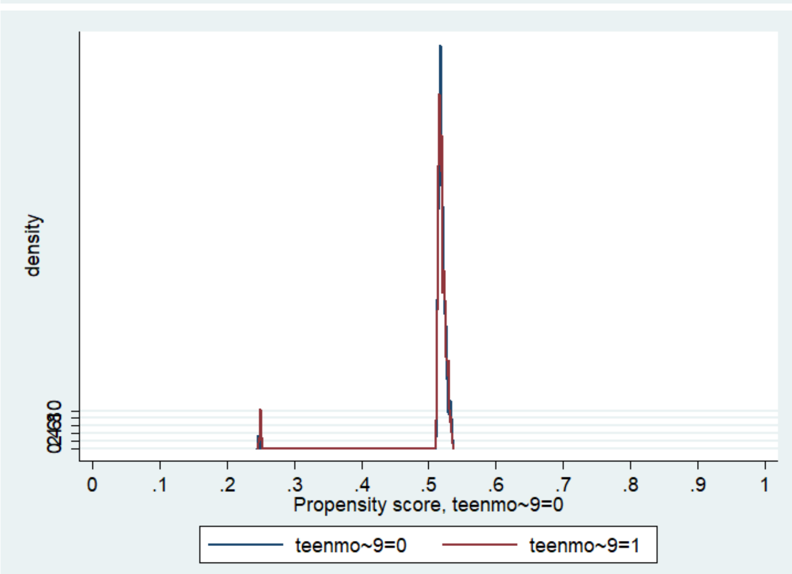
Figure A2: Propensity Score Overlap Plots, Conventional PSM, NLSY79 (corresp. to models in Table A2)



Panel I. Model 1 (N=5,305)

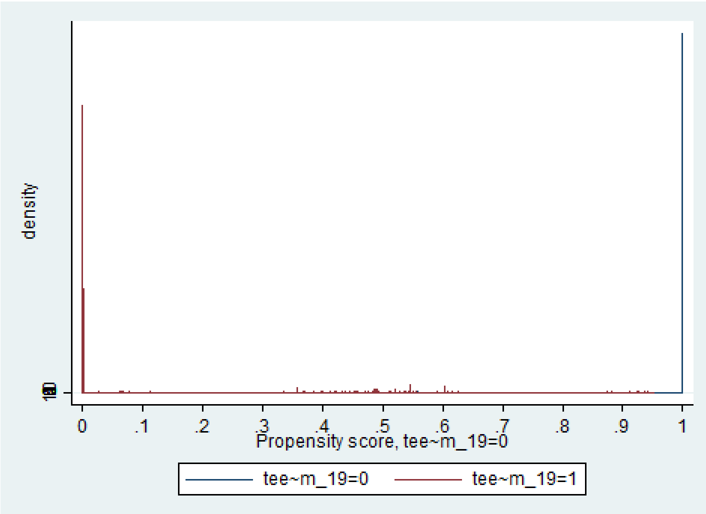


Panel II. Model 2 (N=1,558)

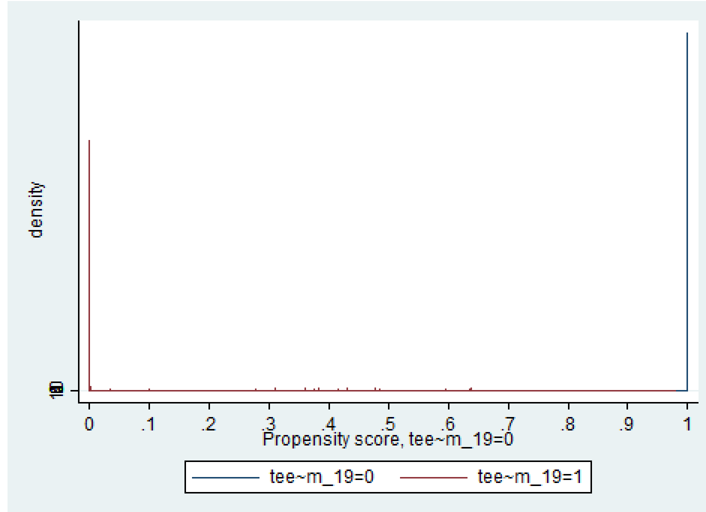


Panel III. Model 3 (N=383)

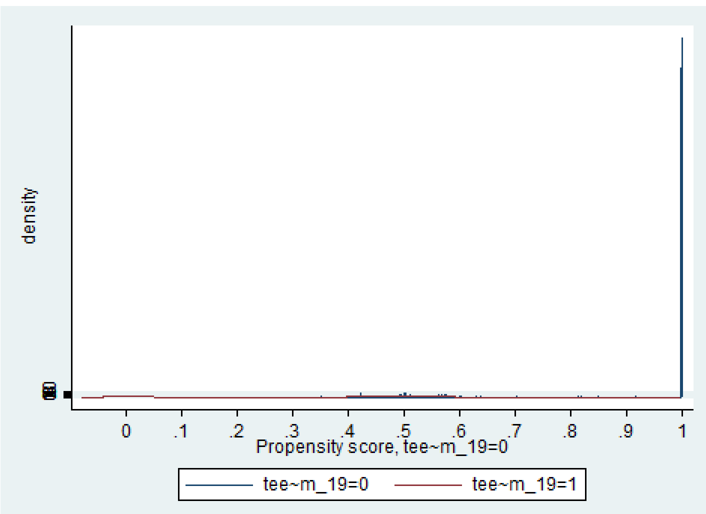
Figure A3: Propensity Score Overlap Plots, Cluster PSM, Add Health (corresp. to models in Table A3)



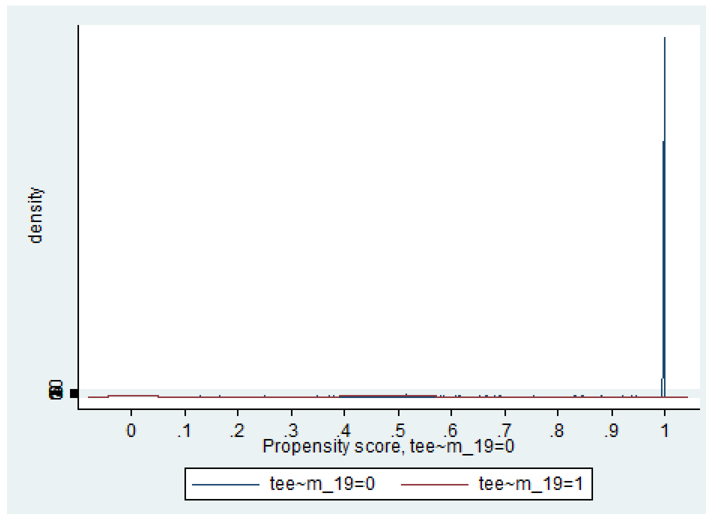
Panel I. Model 1 (N=8,345)



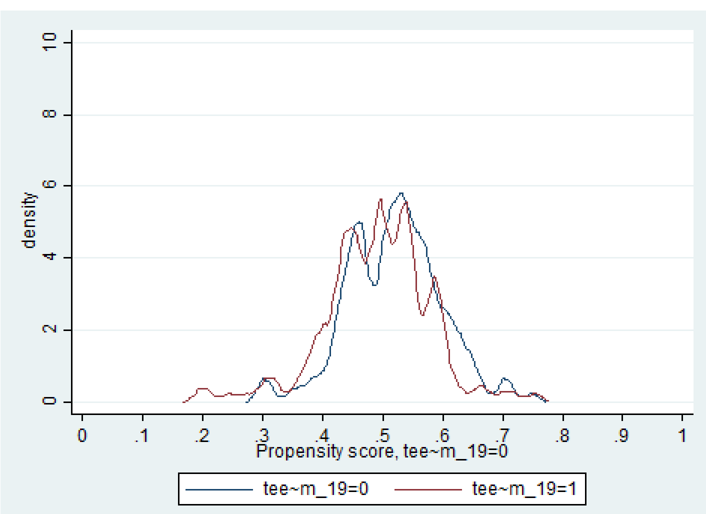
Panel II. Model 2 (N=8,345)



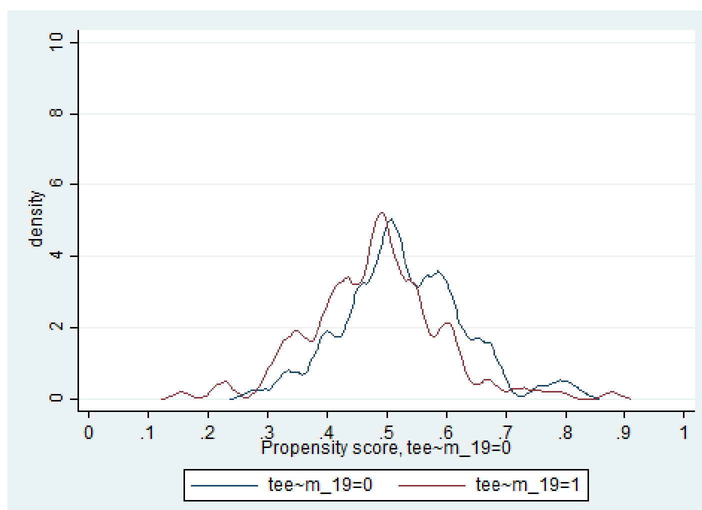
Panel III. Model 3 (N=1,361)



Panel IV. Model 4 (N=1,361)

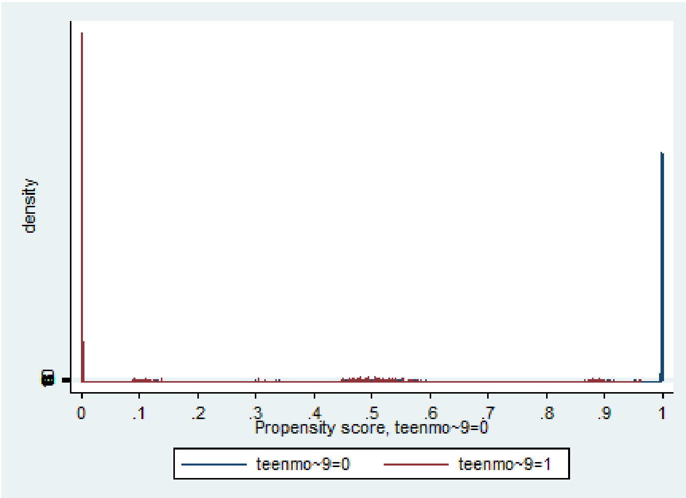


Panel V. Model 5 (N=289)

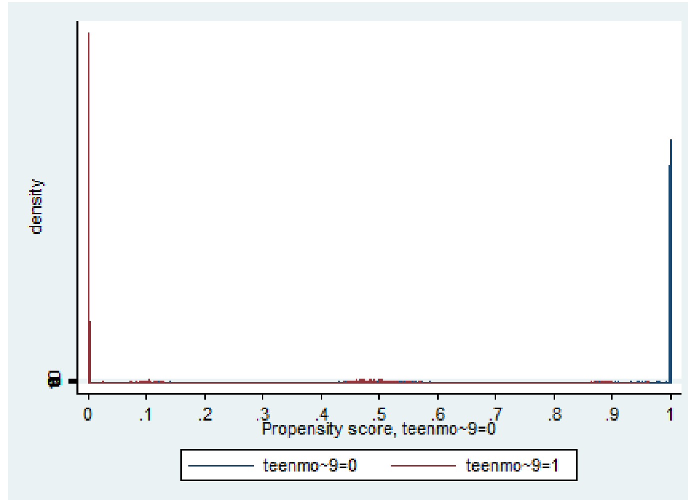


Panel VI. Model 6 (N=289)

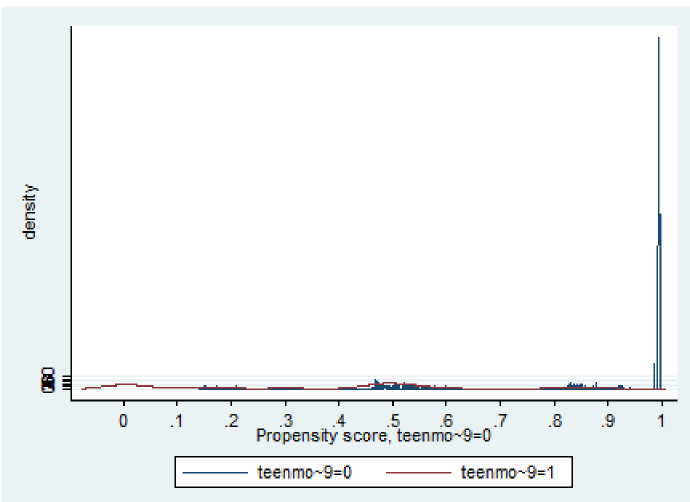
Figure A4: Propensity Score Overlap Plots, Cluster PSM, NLSY79 (corresp. to models in Table A4)



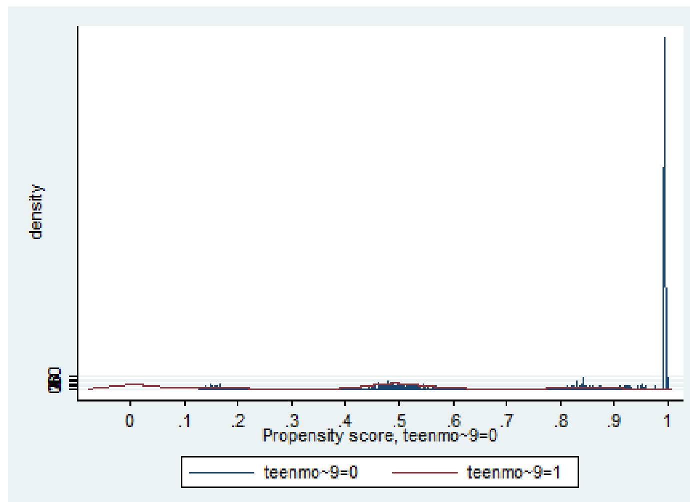
Panel I. Model 1 (N=5.305)



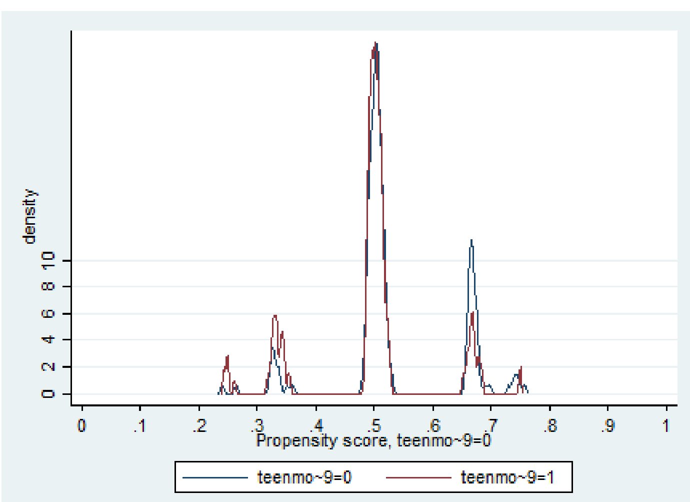
Panel II. Model 2 (N=5,305)



Panel III. Model 3 (N=1,558)



Panel IV. Model 4 (N=1,558)



Panel V. Model 5 (N=383)