



California Center for Population Research
University of California - Los Angeles

Sequential Neighborhood Effects: The Effect of Long-Term Exposure to Concentrated Disadvantage on Children's Reading and Mathematical Skills

Andrew Hicks, Mark Handcock, Nayaran Sastry, Anne R.
Pebbley

PWP-CCPR-2016-030

March 14, 2016

*California Center for Population Research
On-Line Working Paper Series*

**Sequential Neighborhood Effects: The Effect of Long-Term Exposure to
Concentrated Disadvantage on Children's Reading and Mathematical Skills**

Andrew L. Hicks

Department of Health Care Policy

Harvard Medical School

180 Longwood Ave., Boston, MA 02115

E-mail: Hicks@hcp.med.harvard.edu

Phone: 617-432-3459

Fax: 617-432-0713

Mark S. Handcock

California Center for Population Research and Department of Statistics

University of California Los Angeles

Box 951554, 8105C Math Sciences Bldg., Los Angeles, CA 90095-1554

E-mail: handcock@stat.ucla.edu

Phone: 310-817-6778

Fax: 206-457-1953

Narayan Sastry

Population Studies Center and Survey Research Center

Institute for Social Research, University of Michigan

426 Thompson Street, Ann Arbor, MI 48106

E-mail: nsastry@umich.edu

Phone/Fax: 734-936-0462

Anne R. Pebley

California Center for Population Research and Fielding School of Public Health

University of California Los Angeles

Box 957236, 4284 Public Affairs Bldg., Los Angeles, CA 90095

E-mail: pebley@ucla.edu

Phone: 310-825-4505

Fax: 310-825-8762

Acknowledgements

The authors are grateful for support from the California Center for Population Research's Population Research Infrastructure Grant (NICHD: R24-HD041022) and from NICHD grant R25-HD076814.

Abstract

Disadvantaged children's poorer reading and math skills perpetuate socioeconomic and racial/ethnic inequality in the U.S. Although prior research suggest that poor neighborhoods may contribute to low skill levels, it typically relies on cross-sectional data and does not account for neighborhood choice. We examine the cumulative effects of neighborhood characteristics on reading and math skills in Los Angeles County using longitudinal data and a new propensity function approach in which cumulative neighborhood effects are modeled as a continuous treatment variable. This approach offers several advantages compared to the marginal structural modeling approach used in recent research. Our results indicate that the recency of exposure to poor neighborhoods plays a more significant role in children's skills acquisition than the average exposure across childhood. These findings suggest that studies should consider average cumulative exposure, the timing of exposure to neighborhoods of different types, and life-cycle patterns of neighborhood exposure for children's development.

Introduction

Low achievement in reading and math among children from poor, black, and Latino families plays a central role in perpetuating socioeconomic and racial/ethnic inequality across many domains in the U.S. Early and continuing development of reading and math skills during childhood is essential for school success, a college education, well-paying and stable employment, and navigating through life (Dobbie, Fryer and Fryer Jr 2011; Downey 2008; Duncan and Murnane 2011). There is growing consensus in the scientific literature that differences in the family, learning, and social environments between disadvantaged and advantaged children play a central role in causing the academic achievement gap (Duncan and Murnane 2011; Heckman 2011; Roksa and Potter 2011; Shonkoff et al. 2012). For example, evidence indicates that children's and adolescents' brain development is negatively affected by factors such as exposure to violence and stress and positively affected by cognitive stimulation, opportunities for learning, and warm and supportive relationships with parents and others in their surroundings (Eiland and Romeo 2013; Fox, Levitt and Nelson III 2010; Hackman, Farah and Meaney 2010; Shonkoff et al. 2012; Walker et al. 2011).

Although research on the determinants of academic achievement typically focuses on family and school characteristics (Phillips 2011a, 2011b), a sizable literature has also examined the effects of neighborhoods in which children grow up while taking family characteristics into account (Burdick-Will et al. 2011; Dupere et al. 2010; Greenman, Bodovski and Reed 2011; Jargowsky and El Komi 2011; Kling, Liebman and Katz 2007; Sampson 2011; Sampson, Sharkey and Raudenbush 2008; Sanbonmatsu et al. 2006;

Sastry and Pebley 2010; Sharkey and Faber 2014). This research suggests that residential segregation has a direct effect in creating the socioeconomic and racial/ethnic achievement gap by exposing disadvantaged children to neighborhoods with, for example, lower levels of safety and trust; fewer role models; greater violence, drug use, and other social problems; and poorer social institutions like schools and enrichment activities. These studies focus to a large extent on the effects of neighborhood concentrated disadvantage (Sampson, Raudenbush and Earls 1997) (Wilson 1987), a composite indicator that summarizes several highly correlated neighborhood-level indicators of family or population characteristics, such as poverty, low income, welfare receipt, and presence of minority populations. Recent studies—including randomized experiments and observational studies—have produced mixed results, but they generally support the importance of neighborhood effects on children’s achievement, particularly the effects of neighborhood concentrated disadvantage (Sastry 2012; Sharkey and Faber 2014). However, as Burdick-Will et al. (2011) point out, much of the evidence comes from studies based in a single city, Chicago. A potentially important limitation of most studies in the neighborhood effects literature is that they consider the effects of neighborhood characteristics at a single point in time on children’s concurrent or subsequent outcomes. However, children are exposed to their current neighborhood environments for different amounts of time and may have previously lived in neighborhoods with quite different characteristics. The attributes of a neighborhood can also change markedly over time. Cognitive development and skills acquisition are cumulative processes and likely to be influenced by the cumulative effects of environments throughout childhood, rather than at a single point in time. For

this reason, Jackson and Mare (2007), Kunz, Page and Solon (2003), and Crowder and South (2011) examined whether cumulative neighborhood quality measures from children's residential histories are more strongly associated with children's developmental outcomes than single point-in-time measures. In general, prior studies found that differences in the effects of the two types of measures were relatively modest, although Crowder and South (2011) found stronger effects of neighborhood SES on high school graduation when measured for all years of a child's life rather than at a single point in time. None of these studies, however, controlled for the endogeneity of neighborhood characteristics.

Neighborhood characteristics are likely to be endogenous because parents not only choose which type of neighborhood to live in initially, but also whether to stay or move (and where to) throughout their sons' and daughters' childhood. Furthermore, decisions about residential mobility and neighborhood choice at one point in time are likely to be influenced by previous neighborhood choices—and their effects on the child's and family's time-varying characteristics (Harding et al. 2011; Wodtke, Harding and Elwert 2011). For example, parents who moved to a neighborhood because of low housing costs and a shorter commute to work, but then find the local school is not good for their child, may subsequently choose a neighborhood with a longer commute but better schools. Lengthy exposure to a disadvantaged neighborhood (or to a sequence of such neighborhoods) is likely to have a far greater effect on skills acquisition than short durations of exposure (Sampson et al., 2008). In addition, the deleterious effects of exposure to disadvantaged neighborhoods at a particular time may subsequently decline if children move to better neighborhoods. Hence it is potentially important to

examine not just cumulative exposure to disadvantaged neighborhoods, but also how recently this exposure occurred.

To advance research on the effects of neighborhood contextual exposures on children's academic skills, it is also important to account for the endogenous selection processes that lead to different levels of cumulative exposure. Two recent studies have demonstrated the importance of examining the cumulative effects of neighborhood conditions throughout childhood on children's education outcomes and have used new statistical methods for causal inference (Sampson et al. 2008; Wodtke et al. 2011). Both studies employed a propensity score technique for longitudinal data, known as a marginal structural model with inverse probability of treatment weighting.

In this paper, we develop an alternative statistical approach, also involving an extension of propensity score techniques, to examine the cumulative effects of neighborhood characteristics on children's reading and math achievement scores in Los Angeles County, California, using data from the Los Angeles Family and Neighborhood Survey. We make two key contributions. First, our approach addresses several limitations and disadvantages of the marginal structural modeling approach. We compare estimates using this new approach to those from the model proposed by Wodtke, Harding and Elwert (2011), and show that while the results are qualitatively similar, our approach has certain distinct strengths. Our approach allows us to examine two dimensions of neighborhood exposures—an average treatment effect of living in disadvantaged neighborhoods and an effect of the recency of the treatment. Second, we provide new evidence on the cumulative effects of neighborhood characteristics on reading and math achievement for Los Angeles—an urban area substantially different from Chicago,

which Sampson et al. (2008) consider, and from the national U.S. population, which Wodtke et al. (2011) examine.

In the next section, we briefly review the mechanisms through which neighborhood characteristics are thought to affect children's achievement, focusing especially on the role of cumulative exposures. Then we discuss the methodological and data challenges of studying neighborhood effects on children's outcomes, before describing the data, statistical methods, and results of the analysis.

Neighborhood Effects on Children's Achievement

Why would neighborhood conditions affect children's development? Most research on the effects of neighborhood conditions on children's achievement and other outcomes has focused on whether or not neighborhood effects are causal—or due instead to characteristics and choices of families that lead to particular types of neighborhood exposures—and whether these neighborhood effects can be distinguished from family and school effects. Fewer studies have examined the mechanisms through which neighborhood effects may operate (Sampson 2012; Sastry 2012; Sharkey and Faber 2014). Sastry (2012) and Sharkey and Faber (2014) provide recent reviews of hypothesized mechanisms in the literature. We describe them briefly here to explain how cumulative exposure to neighborhood disadvantage—the focus of our analysis—may affect children's development. Testing specific hypotheses about these mechanisms, however, is outside of the scope of this paper.

These explanations can be divided into four interrelated groups: (1) resources and institutions, (2) social processes, (3) norms and expectations about schooling and employment, and (4) violence and stress. Disadvantaged neighborhoods are likely to

have *weaker institutions and poorer resources* for children and families—for example, the quality of child care and schools may be poor and after school homework programs, play groups, libraries, and recreation activities may be scarce or non-existent. Parents in more affluent neighborhoods typically demand more and better public resources and can afford to pay for private resources, if necessary. Children in poorer neighborhoods may also be worse off because the greater needs of all residents may overtax the existing resources (Pebley and Sastry 2004). Many of these resources directly affect children’s learning environment and skills acquisition.

Social processes include collective social interactions of neighborhood residents, such as whether residents share basic levels of trust and expectations for behavior in the neighborhood, and are willing to take collective action; and whether adults collectively monitor children’s behavior and safety and serve as role models for children (Coleman 1988; Sampson, Morenoff and Felton 1999; Wilson 2012). Residents’ social isolation from, or integration with, the outside world is also important for children’s development. Residents with extralocal social ties can provide information and access to opportunities, services, and ideas from those outside the neighborhood (Coleman 1988, Edin 1991, Tiggs, Brown, and Green 1998), but they may be less available in disadvantaged neighborhoods.

Norms and expectations about schooling and employment—akin to what Harding et al. (2011) refer to as “culture”—are likely to affect children’s attitudes toward the importance of academic achievement and the likelihood of eventually getting a job that requires higher-level skills. Neighborhoods in which adults hold only unskilled or low skilled jobs and/or experience high unemployment rates are likely to convey the

message that investment in academic achievement is of less importance than work experience. Educational norms and expectations may also affect the neighborhood language environment—i.e., the frequency with which children are exposed to standard English and to the use of verbal communication in ways that will be useful to them in an academic setting.

Violence and stress have received considerable attention as potential factors affecting children's well-being, through mechanisms such as mental health (Aneshensel and Sucoff 1996; Wheaton and Clarke 2003) and parenting practices (Morrison Gutman, McLoyd and Tokoyawa 2005). For example, Sharkey and colleagues have shown that acute exposure to violence can affect children's attention, impulse control, and academic performance (Sharkey 2010; Sharkey et al. 2012).

These mechanisms are likely to affect children's achievement throughout childhood and adolescence. Our analysis is based on two general hypotheses. First, neighborhood disadvantage is likely to have greater effects on children who have greater average exposure to disadvantage, through the types of mechanisms outlined above. For example, a child living exclusively in neighborhoods with poor resources such as child care, preschools, schools, libraries, and after school programs as well as norms discouraging academic achievement are less likely to be exposed to and interested in reading and math-related skills than those in less disadvantaged neighborhoods.

Second, children more *recently* exposed to disadvantaged neighborhoods are likely to have lower levels of academic achievement than those exposed in the more distant past. In other words, some effects of neighborhood disadvantage may fade over time. For example, a child who lived in a less safe and more stressful neighborhood five

years ago and has subsequently moved to a better neighborhood may have fewer problems than a child living in a disadvantaged neighborhood now. Below we described our methods for testing these two hypotheses.

Methodological and Data Challenges

Investigating the effects of neighborhood characteristics over time is complex and challenging for three reasons. First, the data requirements are substantial because information is needed on residential locations over a considerable period of time. Second, as described above, parents' decisions about where to live reflect each family's choices and constraints—and hence the characteristics of the neighborhoods in which children reside are determined endogenously (Duncan and Raudenbush 1999; Manski 1995). Third, the effects of children's exposure to disadvantaged neighborhoods are likely to be cumulative and lagged (Sampson et al. 2008). We discuss the methodological implications of each of these challenges, as well as the data requirements and statistical options for addressing them.

Data Challenges: Measuring Exposure to Disadvantaged Neighborhoods

A challenge in examining the effects of cumulative neighborhood exposures on children's outcomes is the need for prospective longitudinal data that includes frequent or continuous measures of residential location and moves, family processes, and children's outcomes. In practice, this is a very high bar, and only a handful of studies are able to meet these requirements. Among these studies are the Project on Human Development in Chicago Neighborhoods (PHDCN), the Panel Study of Income Dynamics (PSID), and the Los Angeles Family and Neighborhood Survey (L.A.FANS).

An important first step in analyzing the effects of exposure to disadvantaged neighborhoods on children's outcomes is to characterize children's residential histories and to understand the factors associated with residential moves. Several studies have demonstrated that children's exposure to poor neighborhoods varies considerably by characteristics such as race and income. For example, African American children are much more likely than whites to spend long, uninterrupted periods in poor neighborhoods and more likely to return to poor neighborhoods if their families succeed in moving out (Briggs and Keys 2009; Quillian 2003; Timberlake 2009). Residential moves are highly selective and more likely to involve white and upwardly socioeconomically mobile families (Geronimus, Bound and Ro 2014; Pettit and McLanahan 2003). Socioeconomic, racial/ethnic, and other characteristics which influence children's residential histories may also affect their outcomes -- making it important to consider selection into and out of neighborhoods in studies of the effects of disadvantaged neighborhoods on children's well-being.

Statistical Modeling Challenges

Accounting for the endogenous selection processes that determine children's exposure to a particular set of neighborhood characteristics has challenged researchers for a considerable time. There are no widely accepted solutions to this challenge, although there have been many different statistical modeling approaches employed, each of which has a variety of data requirements—along with specific strengths and limitations. The most common approach for research examining neighborhood effects on children's outcomes is to attempt to control for all possible observed child and family

characteristics that shape both neighborhood selection and children's outcomes (Sastry and Pebley 2010). This approach responds to the critique that endogenous selection mechanism for neighborhood of residence is a problem insofar as it is the result of *unmeasured* characteristics. Many recent studies—including L.A.FANS—have, as a consequence, collected detailed information about child and family processes. These measures cover aspects of children's and parents' characteristics that were not previously measured well or at all, and include such indicators as cognitive assessments of parents themselves and more detailed indicators of family income and wealth. Incorporating such measures into studies of neighborhood effects on children's outcomes provides clearer and more convincing estimates of these effects.

Despite better and more comprehensive measurement of previously omitted child and family variables, unmeasured or unmeasurable characteristics remain an important concern. An ambitious attempt to address this issue directly used an experimental design to randomly assign families to neighborhoods through a lottery for a housing subsidy. The Moving to Opportunity (MTO) study uncovered mixed effects of the program, with small or no significant effects on test scores for children after four to seven years (Sanbonmatsu et al., 2006; Burdick-Will et al., 2011) or a range of other schooling outcomes in the final impact evaluation of the program (Gennetian et al. 2012). There have been numerous critiques of the MTO experiment. Among the most salient from our perspective is that MTO did not experimentally randomize neighborhood quality, but rather reflected a specific intervention focused on receipt of a housing voucher and a particular type of move from a high-poverty to a low-poverty neighborhood (Clampet-Lundquist and Massey 2008; Sampson 2008). Furthermore,

neighborhood changes occurred along with residential moves, making it difficult to disentangle the two effects.

The rest of the neighborhood effects literature has used a variety of statistical techniques to account for neighborhood endogeneity, including family fixed effects models, instrumental variable approaches, and propensity score matching (e.g., Aarronson 1998; Foster and McLanahan 1996; Evans et al. 1992; Plotnick and Hoffman 1992; Solon et al. 2000; Harding 2003). None of these approaches are wholly satisfactory, for reasons described elsewhere (Dietz 2002).

The newest statistical approach to be applied to studying neighborhood effects on children's outcomes is marginal structural models that incorporate inverse probability of treatment weighting. Two recent studies have applied this approach, which draws on statistical techniques from epidemiology (Robbins et al., 2000; Hernán et al., 2000).

Using longitudinal data from the Project on Human Development in Chicago Neighborhoods and from the Panel Study of Income Dynamics, respectively, Sampson et al. (2008) and Wodtke et al. (2011), employed marginal structural models based on trajectories of exposure among children to different neighborhood characteristics over time. These models have a number of strengths—for instance, they directly address the issue of neighborhood selection as it unfolds over time which creates different trajectories of exposure. However, these models also suffer from a number of limitations. First, like most other approaches based on propensity score techniques, they focus exclusively on accounting for observed child and family measures, and hence are susceptible to the potential effects of unmeasured or unmeasurable factors that shape neighborhood exposures and children's outcomes. Second, neighborhood

characteristics are multidimensional but these studies treat exposure in a circumscribed way. Although much of the literature has focused to a significant extent on neighborhood concentrated disadvantage, Sampson et al. (2008) further restrict attention to a dichotomous indicator of exposure to a neighborhood with high concentrated disadvantage. Wodtke et al. (2011) consider five ordinal “treatments,” based on levels of exposure to neighborhood disadvantage in each year. It is not possible to examine non-linear effects with two categories of exposure, and Wodtke et al. (2011) include only a limited investigation of non-linear effects of exposure. Although neighborhood exposures are difficult to conceptualize in terms of discrete categories, propensity score methods are generally confined to binary or ordinal treatment scenarios. However, Imai and van Dyk (2004) have generalized Rosenbaum and Rubin (1983)’s propensity score by developing the *propensity function* to allow for continuous treatment regimes—an approach that we adopt in this study. Third, Wodtke et al. do not examine the sequence or timing of exposure to neighborhood disadvantage. Sampson et al. (2008), on the other hand, examine specific sequences of exposure which is made possible by their examining just two types of neighborhood environments across three time points. The long observation period of 17 years and the use of five treatment categories in Wodtke et al. make it difficult for them to use Sampson et al.’s approach to examining specific sequences of exposures because of their multiplicity. Wodtke et al. instead summarize exposure using a mean of each period’s treatment category across all observation periods. A consequence is that children who spend the first year of the observation period in a disadvantaged neighborhood and those who spend the most recent year in that environment are classified as having equivalent

exposure. In this study, we incorporate an indicator of the recency of exposure to neighborhood disadvantage. Finally, inverse probability of treatment weights are susceptible to imprecise estimation, which can lead to unstable or large weights—an experience reported by Wodtke et al. (2011).

In this paper, in order to benchmark our results, we replicate the marginal structural models with inverse probability of treatment weighting using L.A.FANS data following the approach used by Wodtke et al. (2011). This replication differs only in our use of L.A.FANS data with its shorter observation period of 6 years (rather than 17 years) and in the outcome measure (we consider reading and math test scores while Wodtke et al. considered high school graduation). We also propose and apply a new statistical approach to characterizing exposure to disadvantaged neighborhood environments that does not rely on arbitrary categories of neighborhood conditions, as has been the case in prior research. Rather, we allow exposure to disadvantaged neighborhoods to be continuous, which provides a more realistic characterization of exposures and greater potential scope for being able to discern their causal effects on children's achievement. One implication of our novel approach is that it mitigates certain methodological limitations of the standard marginal structural modeling approach which often result in large and unstable weights. Our approach also allows us to more naturally consider non-linear effects of neighborhood exposures over time which we extend further by using generalized additive modeling (Hastie and Tibshirani 1990). Our approach conceptualizes neighborhood exposure as a function of time and allows us to take into account more than one dimension of exposure to neighborhood conditions. We consider the joint effects of six years of exposure and use dimensional reduction to

summarize these effects as functions of the cumulative exposure and the recency of exposure. The new framework proposed can be used in similar studies where treatments are longitudinal and have complex effects.

METHODS

Data

Data used in this analysis are from the Los Angeles Family and Neighborhood Survey (L.A.FANS), a longitudinal study of households and neighborhoods in Los Angeles County, California, for which two waves of data collection have been completed. In Wave 1, a stratified, multistage, clustered random-sample survey of 3,100 households in 65 neighborhoods (defined as Census tracts) in Los Angeles County was conducted between April 2000 and December 2001 (Sastry et al. 2006). In households with children (70% of the sample), one child was chosen at random from all household members 17 years of age and younger. If the child had siblings, one was chosen at random as a second sampled child. The sampled children's primary caregivers (PCG; usually the mother) were asked to complete questions about their own marital and work history, the family's income and home ownership, the family unit head's education, and the sampled child's health and residence history. The response rate the Wave 1 sample of children was 86% (Sastry and Pebley 2003).

Wave 2 of L.A.FANS was fielded between August 2006 and December 2008, and followed all respondents from Wave 1, as well as adding a sample of new entrants who moved into the 65 sampled neighborhoods between the two waves using similar rules as in Wave 1. In Wave 2 respondents completed a similar set of questionnaire instruments, except that the duration of retrospective reports was expanded in order to

obtain a complete residential history for all respondents between 2001 and 2006. The Wave 2 response rate for children in the panel sample was 63%, while the response rate for children in the new entrant sample was 82% (Peterson et al. 2012).

Our analysis sample comprises 568 children for the math score analysis and 573 children for the reading score analysis. In order to be included in our analysis sample, children had to: participate in Waves 1 and 2 of L.A.FANS (and hence were aged 3–11 years in Wave 1 and 9–17 years in Wave 2); have been eligible for and completed a test in Wave 1; and have co-resided with their PCGs in Los Angeles County between waves in order to accurately assign neighborhood characteristics over the full study period. A total of 259 children for whom we did not have a residential history because they did meet the PCG coresidence requirement were omitted from the analysis sample; we also dropped 185 children who did not complete a test in Wave 2 (but completed other study components). We account for children who were lost to follow-up in Wave 2 through our use of the L.A.FANS panel weights. By design, the L.A.FANS panel weights do not apply to children who moved into or out of Los Angeles County between Waves 1 and 2 because they are not members of the L.A.FANS in-person longitudinal sample (see Peterson et al., 2012).

Child Achievement Outcomes

Child reading and math skills were assessed using two subtests of the Woodcock-Johnson Revised (WJ-R) Tests of Achievement, administered to children during the second wave of L.A.FANS. The Letter-Word Identification subtest assesses symbolic learning and reading identification skills. The Applied Problems test assesses mathematics reasoning. Raw scores were converted to standardized scores based on

the child's age and national norms (McGrew, Werder and Woodcock 1991).

Standardizing scores by age allow outcomes to be compared across children of different ages. The standard scores have a population mean of 100 and standard deviation of 15. The mean standardized scores for the math and reading subtests for children in this sample were 104.7 and 106.9, respectively, which were slightly higher than the national norms of 100 for each test. The sample standard deviations of 19.0 for reading and 20.6 for math were slightly higher than the national standard deviation of 15.

Cumulative Exposure to Neighborhood Concentrated Disadvantage

Our key independent variable of interest is children's cumulative long-term exposure to neighborhood concentrated disadvantage, a structural measure of several associated dimensions of neighborhood poverty, economic disadvantage, racial segregation, and family structure that has been widely used in previous research as a focal neighborhood characteristic (Sampson et al. 1997; Wilson 1987; Sastry 2012). Factor analysis was used to reduce the dimensionality of the multiple tract-level indicators of neighborhood disadvantage to a single variable. The individual indicators include the race and age composition of the population and the prevalence of female-headed households, along with traditional indicators of economic disadvantage comprising of family-level indicators of poverty rates, public assistance receipt, and low income. These specific tract-level indicators were chosen based on previous research to characterize neighborhood disadvantage that also found a disproportionate exposure to disadvantage among particular population segments based on the included demographic variables (Sampson et al. 1997).

We constructed an indicator of exposure to neighborhood concentrated disadvantage for each child using their residential histories and information from the census on characteristics of the tracts in which they resided. We first created a complete residential history for each child between 2001 and 2006 using data from Waves 1 and 2 of L.A.FANS. For each calendar year in this period we identified the child's home census tract. Next, we used census data on tract characteristics to compute an annual score of neighborhood concentrated disadvantage for each tract in Los Angeles County for each year of the study period. Finally, we assigned the values of the concentrated disadvantage score by tract and year to each child based on his or her residential history.

Tract-level data came from the 2000 U.S. Census and from the 2006–2010 American Community Survey (ACS) five-year estimates. We estimated values for each of the six individual tract measures for each year from 2001 to 2007 using linear interpolation between the observed values for 2000 and 2008 (the midpoint of the five-year ACS estimates) for all census tracts in Los Angeles County. The six tract measures comprised of the percent of households headed by females with children; percent of families with income less than \$25,000; percent of population that is non-white/non-Asian/non-Pacific Islander; percent of individuals in poverty; percent of population under age 18 years; and percent of households receiving public assistance. Following the approach of Sampson et al. (1997), we performed a factor analysis using annual data on the six neighborhood variables for all census tracts in Los Angeles County from 2000 to 2008 in order to generate a composite score of neighborhood concentrated disadvantage for each tract in each year. The resulting neighborhood concentrated

disadvantage scores ranged from -1.15 to 4.45 , with higher scores indicating higher levels of neighborhood disadvantage.

Table 1 shows summary neighborhood characteristics by quintile of the concentrated disadvantage score across all tracts and years from 2000 to 2008. Moving from the first to the fifth quintile, there is a monotonic increase in each of the six neighborhood indicators of disadvantage. Across all of the measures, tracts in the first quintile are substantially better off than those in the fifth quintile. For instance, 6% of the population was living in poverty for tracts in the first quintile compared to 34% for tracts in the fifth quintile. Less than 1% of households in the first quintile received welfare benefits, compared to 13% in the fifth quintile.

Covariates

Our analysis incorporates an extensive set of baseline and time-varying covariates to control for potential confounding by background individual and family characteristics. Table 2 includes a list of the full set of model covariates along with weighted and unweighted summary statistics.

Individual-level baseline characteristics include children's age, sex, race/ethnicity, and birth weight. The weighted average age of children in the sample in 2001 was 7.2 years. There were slightly more males (51%) in the sample. The majority of children (52%) were Latinos, reflecting the demographic composition of Los Angeles County. Blacks accounted for 10% of the sample. Whites and other smaller racial groups (Asians, Pacific Islanders, and Native Americans) were combined into a third category, representing 37% of the sample. Children with low birthweight (less than 5.5 pounds) comprised 10% of the sample.

Mother and family time-invariant characteristics included the mother's age and marital status at the child's birth, the mother's standardized test score from the WJ-R Passage Comprehension test of reading skills, the family unit head's educational attainment, and whether or not the family owned their home at Wave 1 of the survey. A majority of mothers (60%) were married at time of childbirth and the average age at childbirth was 28 years. Mothers had a mean score of 85 on the reading assessment, which is one standard deviation below the national average. Two-thirds of households were headed by a family member with at least a high school diploma (67%). Most families (58%) did not own a home at the time of the Wave 1 interview.

We also included time-varying family characteristics at annual intervals between 2001 and 2006, and in Table 2 show summary statistics at baseline. These covariates include the mother's marital status, number of children, employment status, and work hours, as well as the log of the family unit's income and whether the family received any public assistance in that year. A majority (69%) of PCGs remained married for all six years of the study. Less than half (48%) remained employed throughout the study. PCGs worked an average of 23 hours per week throughout the course of the study and had an average of 1.8 children. The median family income for this period was \$40,000 and 9% of families received public assistance at some point between 2001 and 2006.

Statistical Methods

We used two approaches to investigate the effects of exposure to neighborhood disadvantage on children's reading and math test scores. First, we employed the propensity function approach developed by Imai and van Dyk (2004) for non-binary treatment regimes, which allowed us to model the effects of neighborhood disadvantage

as a continuous treatment variable. Second, we used inverse probability of treatment (IPT) weighting to examine the effects of exposure to neighborhood disadvantage.¹

Propensity Function Approach

Let \mathbf{X}_i be the set of covariates for respondent i , measured at baseline. Let $T_i = (T_{1i}, T_{2i}, \dots, T_{6i})$ be their neighborhood disadvantage at years 1 – 6 (which correspond to years 2001 to 2006). To apply the propensity function approach of Imai and van Dyk (2004), the first step was to model the propensity function, via the distribution of T_i , $P(T_i|\mathbf{X})$. This function describes children’s exposure to neighborhood disadvantage—measured as a vector of treatment variables at each year—as a function of child- and family-level covariates from that year as well as covariates measured at the baseline wave of the survey. We model the joint distribution of T_i as multivariate Gaussian with distribution expressed as

$$P(T_i|\mathbf{X}) = P(T_{1i}|\mathbf{X})P(T_{2i}|T_{1i}, \mathbf{X})P(T_{3i}|T_{2i}, \mathbf{X})P(T_{4i}|T_{3i}, \mathbf{X})P(T_{5i}|T_{4i}, \mathbf{X})P(T_{6i}|T_{5i}, \mathbf{X}),$$

(1)

where $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ is the combined matrix of baseline covariates for the n respondents. Each component of the above expression is modeled as a separate Gaussian additive model (Hastie and Tibshirani 1990).

Whereas Imai and van Dyk generated a single propensity function from a cross-sectional dataset, we produced six sequential propensity functions that model a child’s propensity for exposure to neighborhood disadvantage in each year of the observation

¹ The R and Stata programs used to implement these procedures are provided in Appendix II.

period between Waves 1 and 2. Intervals of one year were chosen so that we can compare our methods to the IPT weight model proposed by Wodtke et al. (2011). To compute the expected value of the propensity function for each respondent at each year, we average across a large number, m , of simulated responses from each year's distribution. This technique allows us to use expected values of the propensity function, rather than the models' predicted values. Respondent i 's expected propensity for neighborhood disadvantage at each year is then averaged across the six years to obtain an overall mean expected treatment score, \bar{T}_i :

$$\bar{T}_i = E(T_i) = \frac{1}{6} \sum_{t=1}^6 \frac{1}{m} \sum_{k=1}^m T_{itk}, \quad (2)$$

where T_{itk} is the k th simulated response of respondent i 's neighborhood disadvantage distribution at year $t = 1 - 6$. We chose $m = 10,000$ simulated responses in our computations.

Because the overall mean expected treatment score provides a single summary measure of neighborhood exposure across multiple years (2001 to 2006), we investigated whether there was a time trend in the relationship among the neighborhood disadvantage scores within the six-year study period. We performed a principal components analysis of the six annual scores. The results revealed that factor loadings of the first component, which accounted for 95% of the total variance in the six annual neighborhood disadvantage scores, were roughly equal in each of the six years (see Table 3). This finding suggests that there were not differences in exposure to neighborhood disadvantage across the six years in terms of their contributions to the main component of the treatment score, and that the overall mean expected treatment

score for person i over the six years, \bar{T}_i , could be used as an indicator of exposure to neighborhood disadvantage covering the entire six-year study period.

However, the principal components analysis also revealed a second component, which accounted for the majority of the remaining variance (3% of the total variance). Factor loadings for the second component were positive for the most recent scores and negative for the earliest scores, with a nearly linear decline in the weights across the six years (see Figure A1 in Appendix I). We interpret this component as capturing the effect of the recency of exposure to neighborhood disadvantage. The second component, which we call *recency*, is linear, and we incorporate it into our regression models by generating an expected recency score, R_i :

$$R_i = \frac{1}{21} \sum_{t=1}^6 t(T_{it} - \bar{T}_i), \quad (3)$$

where T_{it} is the expected treatment for person i at year $t = 1 - 6$, which corresponds to 2001 to 2006, and \bar{T}_i is the mean expected treatment. The difference between the year-specific treatment and the mean treatment across all years is computed and summed across the six years. A linear weight from $t = 1 - 6$ is included so that later years are given more weight than earlier years. The resulting expected recency measure ranges from -0.45 to 0.41 , with positive scores indicating higher levels of neighborhood disadvantage in later years of the analysis (when compared with the child's overall levels of disadvantage) and negative scores indicating lower levels of disadvantage in later years (compared with overall individual disadvantage). Based on these results, we model children's exposure to neighborhood disadvantage via their mean expected

treatment score and their recency score. The propensity function is then as a function of $\hat{\theta} = (\bar{T}, R)$, the overall mean expected treatment score and the recency score.

The density and scatter plots of mean expected treatment versus recency are displayed in Figure 1. Recency scores have a mean of -0.04 , indicating that the average child experiences relatively lower levels of neighborhood disadvantage in later years of the study. Although there is a negative correlation between mean expected treatment and recency scores, the average recency score for children with the lowest levels of neighborhood disadvantage is approximately zero, suggesting that children in more advantaged neighborhoods have lower mobility across the disadvantage scale and tend to stay in high-advantage neighborhoods throughout the course of the study. Likewise, children experiencing higher overall expected levels of neighborhood disadvantage have, on average, negative recency scores, indicating higher disadvantage levels earlier in the analysis period rather than later. This pattern suggests that neighborhood conditions are improving among children exposed to neighborhood disadvantage as they age, either through residential mobility or neighborhood improvements.

Let $Y(t)$ be the potential outcome for a child exposed to treatment $t = (t_1, t_2, \dots, t_6)$. We have in mind outcomes such as the children's math and reading test scores. Our main results are based on estimating the expected outcome for each treatment exposure and controlling for baseline covariates that describe child and family characteristics.

Explicitly, we model

$$E(Y(t) | X = x) = \int E(Y(t) | \theta(x) = s, X = x) p(\theta(X) = s) ds, \quad (4)$$

where $\theta(x)$ is the propensity function parameter for covariate values $X = x$. Based on the assumption of strong ignorability of the treatment assignment given the propensity function we have

$$E(Y(t) | X = x) = \int E(Y(T) | T = t, \theta(x) = s, X = x) p(\theta(X) = s) ds. \quad (5)$$

We model $E(Y(T) | T = t, \theta(x) = s, X = x)$ as a smooth function of s , and x . Specifically, we use a generalized additive regression model based on a joint tensor-spline function of t and s . The regression models include controls for baseline covariates that describe child and family characteristics. Missing values of covariates describing family, marital, employment, and income characteristics were imputed using other household data from the same year. Twenty five imputed datasets were created. All of the estimates we present were combined from results using the twenty five imputed datasets. Finally, the estimated standard errors account for clustering of observations by family using the robust variance estimation.

Inverse Probability of Treatment Weights

To apply the IPT weighting approach, we first constructed a measure of treatment—which in this study is children’s exposure to neighborhood disadvantage. Using year-specific quintiles of neighborhood disadvantage coded from 1 (lowest quintile of neighborhood disadvantage) to 5 (highest), we constructed a measure of duration-weighted exposure for each child as the mean across the six study years of the neighborhood disadvantage quintile in which a child resided at each year.

Conventional regression models adjust for neighborhood selection by controlling for time-varying individual and family characteristics that may be associated with neighborhood choice. However, these models may over-control the indirect pathways

because these individual and family characteristics are also associated with child outcomes, which can lead to underestimating the true long-term effects of neighborhood disadvantage. The goal of the IPT weighting approach is to create a pseudo-population in which the treatment variable is no longer confounded by measured covariates.

Wodtke et al. (2011) suggest weighting observations by the inverse of the probability that a child was exposed to their actual neighborhood quintile in each year conditional on past treatments and confounders. The IPT weight for the i th child is given by:

$$w_i = \prod_{j=1}^I \frac{1}{P(T_{(j)}=t_{ij}|\bar{T}_{(j-1)}=\bar{t}_{i(j-1)},\bar{L}_j=\bar{l}_{ij})}, \quad (6)$$

where $T_{(j)}$ represents neighborhood exposure status at the j th wave since start of follow-up, t_{ij} represents child i 's neighborhood exposure at wave j , and \bar{l}_{ij} represents child i 's previous confounders since wave j . By weighting regression models with these overall IPT weights, treatment assignments are balanced by giving more weight to children with underrepresented covariate histories.

Applying this technique to data from L.A.FANS resulted in IPT weights that were hugely inflated (see Table 4). More than a quarter of the weights have values larger than 12, and the largest 5% of weights have values well above 1,000. When applying these weights to regression models, parameter estimates become wildly unstable, with most of the model's power being derived from a small number of observations with very large weights. Wodtke et al. (2011) found similar results when constructing IPT weights using PSID data, and suggest stabilizing the weights by including baseline confounders in the numerator of the weighting equation:

$$sw_j = \prod_{i=1}^I \frac{P(T_{(j)}=t_{ij}|\bar{T}_{(j-1)}=\bar{t}_{(j-1)},\bar{L}_0=\bar{l}_0)}{P(T_{(j)}=t_{ij}|\bar{T}_{(j-1)}=\bar{t}_{(j-1)},\bar{L}_i=\bar{l}_{ij})}. \quad (7)$$

Because stabilized weights now contain baseline confounders in both the numerator and the denominator, the outcome regression model must also condition on these confounders. We implemented the stabilized IPT weights using L.A.FANS data and, following Wodtke et al.'s example for PSID, we further refined the weighting procedure by multiplying the stabilized weights with panel weights provided by L.A.FANS to account for the survey's sampling design and attrition. Summary statistics for the component and final weights are shown in Table 4.

Results

We present two sets of results for the effects of neighborhood disadvantage on children's math and reading scores in Los Angeles using data from Waves 1 and 2 of L.A.FANS. The first set of results is based on our new propensity function modeling approach, while the second set uses IPT weighting. The IPT weighting approach has been used in other recent studies and hence provides a useful comparison for the results of our novel approach.

In Table 5 and Figure 2 we present results from our propensity function models for children's math and reading scores. The table includes the full set of parameter estimates for all covariates in four different models, two for math scores and two for reading scores. The first model for each outcome examines the effects of recency of exposure to neighborhood disadvantage while the second model shows the effects of mean exposure to neighborhood disadvantage. We do not present results for models that include the simultaneous effects of overall mean treatment and recency scores, although such a model could be estimated in principle. However, this model is not

statistically identifiable due to the high correlations between the overall mean treatment and recency scores ($r=0.96$).

For all four models, the effects of neighborhood disadvantage are estimated as an interaction with expected average exposure and recency, and all three of these variables (i.e., expected average exposure, expected recency and either observed recency of exposure or observed mean exposure) are specified as flexible, smoothed tensor-spline functions. The bottom panel of Table 5 provides information about the smoothed functions, namely the number of effective degrees of freedom and reference degrees of freedom and an F-test statistic for these functions. For all four models, the two sets of tensor spline functions have F-test values indicating statistically significant contributions to model fit. The effective and reference degrees of freedom measures provide indicators of model complexity, although the large values for the reference degrees of freedom are not an indicator of overfitting (see Janson, Fithian and Hastie, 2015); note also that the non-integer values are due to the partial penalization.

In Figure 2 we present the substantive effects on math and reading scores of average exposure to neighborhood disadvantage and the recency of exposure to neighborhood disadvantage. In each panel, the solid line represents the conditional expectation of each test score while the dashed lines show the 95% pointwise confidence bounds for the expected score. The figures display the estimated dose-response functions, which are interpreted as the average treatment effect given the level of treatment received. The top-left panel shows that there is no effect of average exposure to neighborhood disadvantage on children's math scores, while the bottom-left panel reveals a

suggestive, but non-significant, negative effect of average exposure to neighborhood disadvantage on children's reading scores.

The right two panels show that there are significant negative effects for recency of exposure to neighborhood disadvantage on children's scores for both math (top right) and reading (bottom right). Although the effects of recency of exposure to neighborhood disadvantage on both outcome variables are estimated using flexible, non-linear specifications, the relationships appear nearly linear. The slopes of the effects of recency of exposure are similar for reading and for math, indicating that a one standard deviation (0.094) increase in recency of exposure to neighborhood disadvantage is associated with a decrease of 2.6 points in math scores and 2.7 in reading scores—in both cases, approximately one-fifth of a standard deviation. Thus, we find no effects of average neighborhood disadvantage on children's math or reading scores but a significant negative effect of recency of exposure to neighborhood disadvantage on both math and reading scores. Note that the absence of a statistically significant effect of average neighborhood disadvantage is due to the absence of an effect (i.e., a nearly flat slope for the estimated relationship), particularly for math, rather than due to an apparent relationship but a lack of statistical precision in estimating this effect. Our results suggest that exposure to a high level of neighborhood disadvantage is a modest predictor of lower math and reading scores—unless this exposure occurs in the recent past. These results are perhaps not surprising given that the expected average treatment adjusts for neighborhood exposure at Wave 1 and they are highly correlated ($r=0.995$). As a result, the causal effect of exposure to neighborhood

disadvantage is difficult to identify statistically—which also explains the large confidence regions.

Table 5 shows regression-adjusted estimates for the effects on children’s math and reading scores of the remaining variables in the models (other than mean exposure to neighborhood disadvantage and recency of exposure). We focus our discussion on results for which there is a consistent pattern across models. The results suggest that older children do worse on both standardized tests, based on the significant and negative coefficients for the child’s age. This result could reflect period effects or stronger negative effects of exposure to disadvantaged during later adolescence. Boys have higher test scores than girls for math but lower scores for reading, although these differences are only statistically significant for models that examine the effects of recency. Children whose race is white or other have significantly higher reading scores compared to Latinos or blacks although the result is only statistically significant in the models examining the effects of recency. Children with older mothers have significantly higher reading and math scores, although again the result is only statistically significant for the models examining the effects of recency. Mothers’ reading test scores have a statistically significant positive association with test scores for both outcomes.

However, the effects of mothers’ reading scores are twice as large for children’s reading scores as for math scores, which is presumably due to the correspondence of the test topic for the former outcome. The family head’s level of educational attainment is also associated with significantly higher reading and math test scores for children.

Next we turn to results from the models that use the IPT weights, which are presented in Table 6. In order to compare these models with the preceding results and the

findings from Wodtke et al. (2011), we focus initially on a pair of model specifications for each outcome that consider separately the effects of observed duration-weighted exposure and observed recency. We find negative effects of observed duration-weighted exposure on both math and reading test scores for children, although the effect on reading scores is statistically significant only at the .10 level. For math scores, these results indicate that a one standard deviation increase in observed duration-weighted exposure (1.262) is associated with a decrease of 4.36 points or about one-third of a standard deviation. For reading scores alone, we find a negative effect of observed recency with a one standard deviation increase in (0.143) associated with a decrease of 3.09 points or about one-fifth of a standard deviation.

The results for other covariates from the models with IPT weights are broadly similar to those based on the propensity function approach. Older children have lower test scores, with the estimated effects again twice as large for math compared to reading. Boys have substantially lower test scores than girls for reading, but have similar scores for math. Several variables only have effects on scores for models that examine the effects of observed recency—including mother's age at the child's birth, mother's reading test score, and family head's educational attainment. Recall that we controlled for these variables because they were included in the calculation of the stabilized weights rather than as a focus for our analysis.

Finally, we conducted a sensitivity analysis of our results for the models with the IPT weights by investigating the joint effects of observed duration-weighted exposure and observed recency and examined non-linear effects by estimating models stratified by neighborhood income (see Table 7). We stratified the models based on the same

stratification design that was used for L.A.FANS sampling plan, which divided tracts in Los Angeles County into three strata classified as very poor, poor, and non-poor (see Sastry et al. 2006). Column 1 in Table 7 show results from the full sample. The results for Models 1 and 2 replicate the findings from Table 6, while the results for Model 3 show the joint effects of observed duration-weighted exposure and observed recency. These results for Model 3 reveal that the effects of observed duration-weighted exposure and observed recency do not change when also controlling for the other variable; rather, each of these variables appears to have an effect on children's reading and math scores that is independent of the other variable.

When estimating the models separately by sampling strata, we find the effects of neighborhood disadvantage exposure to vary qualitatively by poverty stratum (Table 7, Columns 2–4). For both outcomes (math and reading), and across all three model specifications (observed duration-weighted exposure alone, observed recency alone, and both variables jointly), substantively large and statistically significant effects generally emerge only in models for the non-poor stratum alone (Column 4). For the models of math scores, there are no statistically significant effects in models for the very poor and poor strata, and even the magnitudes of the (non-significant) estimated parameters are small. However, for the non-poor stratum, there are statistically significant negative effects of both observed duration-weighted exposure and observed recency, whether estimated separately or jointly. For the models of reading scores, the estimated parameters across the three strata are generally similar although the precision of the estimates varies greatly. There are statistically significant negative

effects of observed duration-weighted exposure, but only in the non-poor stratum, and significant negative effects of observed recency, but only in the poor stratum.

Comparing the results from the two modeling approaches, there is a consistent finding for a negative linear effect of recency of exposure to neighborhood disadvantage on children's reading scores. There is less clear evidence of the effect of recency on math scores—the stratified IPT weighted models suggest a non-linear relationship with a strong negative effect only in the non-poor stratum while the propensity function approach suggests a negative linear effect that operates across all three strata.

However, results from the stratified IPT weighted models are imprecisely estimated in the poor and very poor strata, and the apparent inconsistency in the findings across the two modeling approaches may be due to small sample sizes in the two smaller strata (the very poor and poor strata). We suspect that effect sizes in our underlying propensity models used to generate the IPT weights may also vary by stratum and conclude that these models must also be stratified into three separate models based on tract poverty status. However, due to sample size restrictions, separate propensity models by poverty status were not identifiable under these conditions. Under these circumstances, the propensity function technique approach provides an integrated approach that is attractive because it allows us to model the effects of neighborhood exposure and recency as continuous functions and eliminates the need to arbitrarily split the neighborhood disadvantage score into quintiles to fit an ordered logistic propensity function.

Discussion

This study builds on a considerable body of research that has examined the consequences for children's cognitive development and life chances of growing up in disadvantaged neighborhoods. A challenging problem in this area of study is how to estimate the causal effects of neighborhood exposure in the context of families being able to choose where they live based on the characteristics of these neighborhoods and their children's outcomes. We apply an approach recently proposed by Wodtke et al. (2011) that uses a marginal structural modeling technique through the use of inverse probability of treatment weighting. The reweighting approach aims to rebalance individuals in the sample across treatment categories so that treatment is unconfounded by observed covariates. We identified several significant shortcomings of this approach, but applied it in order to compare it with a novel approach we propose that addresses several key shortcomings. The shortcomings center around three issues. The first is creation of arbitrary treatment categories, rather than being able to examine a continuous treatment regime such as one that reflects more realistic nature of exposure to neighborhood disadvantage. The second shortcoming is the difficulty of examining non-linear effects—a problem that is compounded by the difficulty in estimating the underlying weights using an approach that is consistent with the non-linear effects. The third shortcoming is the difficulty of estimating the effects of a treatment that comprises more than one dimension.

Our novel approach addresses each of these three shortcomings. We propose a model that is based on the propensity function approach developed by Imai and van Dyk (2004) that obviates the need to construct arbitrary categories from a continuous distribution. Next, we adopt a generalized additive modeling approach (Hastie and

Tibshirani 1990) to allow for the estimation of flexible, non-linear effects of treatment.

Finally, the specification of a multidimensional interactive effect between the propensity function and the treatment variables of interest allow us to examine the effects of two (or more) continuous treatment variables, although in our application sample size limitations prevented us from fully implementing this last feature of our approach.

Our main substantive finding regarding the effects of children's exposure to neighborhood disadvantage on reading and math test scores is the importance of the timing—and, in particular, the *recency*—of this exposure. Previous research has moved beyond point-in-time measures of concurrent or past neighborhood exposures to consider the cumulative effects of neighborhood exposures over many years. However, these studies have generally not considered how recently children were exposed to disadvantaged neighborhoods. Our results suggest this is an important oversight, and that future research should consider not just children's cumulative exposure to disadvantaged neighborhoods but also the timing of such exposure.

We find that recency of exposure to neighborhood disadvantage is associated with children achieving significantly lower scores on reading and math tests. Furthermore, the effects are nearly linear for both outcomes and are of similar magnitude, with a one standard deviation increase in the recency of neighborhood disadvantage associated with a 0.2 standard deviation decrease in test scores. Our results suggest that average exposure to neighborhood disadvantage is of less significance for children's test scores than the recency of exposure to neighborhood disadvantage. Comparing these results to findings obtained from using the IPT weighting approach, we find a very similar result for the effects of recency on reading scores but not math scores. For math scores, on

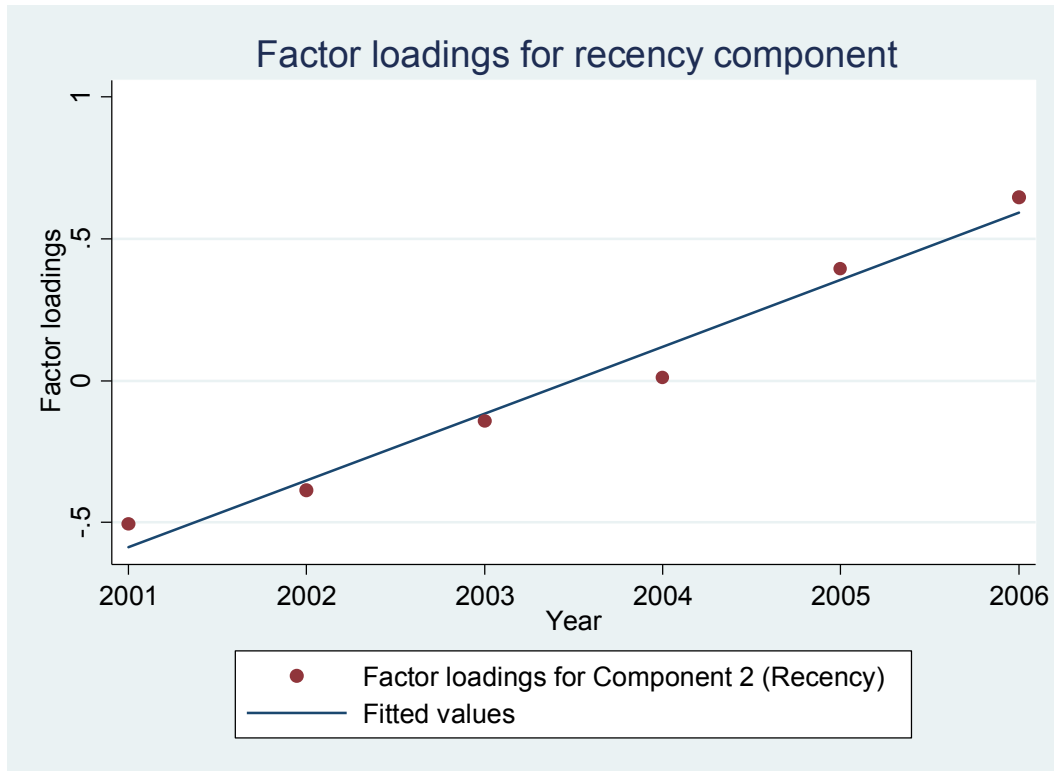
the other hand, there is a strong negative effect of recency only in the non-poor stratum. Results from the IPT weighed models suggest the presence of non-linear effects, but these effects cannot be estimated consistently because we were unable to estimate stratified models to generate the IPT weights.

Another important contribution of this study was to expand the geographic focus of research on neighborhood effects on children's achievement to consider relationships in Los Angeles County, California. This setting is distinct from previous research that has considered almost exclusively data from Chicago or from large national samples. Los Angeles is a location with a distinct race and ethnic composition and spatial structure, but also experiences many demographic and social trends ahead of rest of the nation. Finally, there are a couple of aspects of the exposure of children in L.A.FANS to neighborhood disadvantage that are worth highlighting. First, the average recency score is -0.05 , which suggests that neighborhoods exposures are improving with time. At the same time, child age has a statistically significant association with lower test scores for both math and reading. Together, these two results indicate that there is a family life-cycle process of residing in better neighborhoods that leads to better achievement outcomes among younger children. Second, the distribution of recency is truncated near zero and is at its lowest values when average expected treatment is near its minimum value. This result indicates that few families were moving to more disadvantaged neighborhoods, and that the few moves to disadvantaged neighborhoods that did occur were largely among families already exposed to such neighborhoods. This pattern of moves may reflect life-cycle effects in families'

economic status and the selection processes that reflect the endogeneity of residential choice, as well as the general improvement over time in neighborhood characteristics. There are a few limitations to this study. There are relatively modest sample sizes available from the L.A.FANS data, which made it challenging to estimate certain relationships. In particular, it was difficult to identify the effect of average exposure to neighborhood disadvantage on children's test scores, although this situation was also a consequence of the high correlation with the overall mean expected treatment score from the propensity function. The high correlation between the overall mean treatment and recency scores also meant that we could not model the simultaneous effects of these two variables, and larger sample sizes may have ameliorated this problem. A separate limitation is that the propensity score techniques that we used only consider observed characteristics of children and families. Although there are established approaches to examining the sensitivity of results to the effects of unobserved or unmeasured factors (e.g., Brumback et al. 2004), applying these techniques was beyond the scope of the current study. Finally, we used residential locations as indicators of exposure to concentrated disadvantage at the neighborhood level; a potential improvement for the future would be to use indicators of actual contextual exposures, based on location of schools that children attend, activity patterns (based on where children spend their time), and the locations of where peer group members, friends, and relatives reside. Few studies collect such data, although L.A.FANS is a notable exception.

Appendix I.

Figure A1. Factor Loading for the Second Principal Component from the Principal Components Analysis (Interpreted as a Measure of Recency) by Year for 2001–2006



Appendix II.

R Code Used to Implement Propensity Function Model	1-21
Stata Code Used to Create L.A.FANS Analytic Data Files	22-56
Stata Code to Implement IPTW Approach with L.A.FANS data	57-78

R Code Used to Implement Propensity Function Model in Hicks et al. (2016)

```
#####  
# File: Propensity Function Approach.R  
# Summary: Using LAFans Data implements the  
# propensity function approach found in  
# Sequential Neighborhood Effects: The  
# Effect of Long-Term Exposure to Concentrated  
# Disadvantage on Children's Reading and  
# Mathematical Skills  
#####  
#  
#  
#####  
# Make Data  
#####  
#devtools::install_github("sjewo/readstata13", ref="0.4")  
library(readstata13)  
#p=read.dta13("propensityscoredata_imputed_handcock.dta")  
#saveRDS(p,file="propensityscoredata_imputed_handcock.rds")  
p=read.dta13("propensityscoredata_imputed_handcock25.dta")  
saveRDS(p,file="propensityscoredata_imputed_handcock25.rds")  
n=colnames(p)  
  
#####  
#Fit sequential model with gam term for age, computes expave and recency  
#####  
set.seed(1)  
library(mgcv)  
if(T){  
p=readRDS("propensityscoredata_imputed_handcock25.rds")  
p$lowfaminc2000 <- 1*(p$logfaminc2000 < 9)  
p$llogfaminc2000 <- (p$logfaminc2000 >= 9)*p$logfaminc2000  
p$lowfaminc2001 <- 1*(p$logfaminc2001 < 9)  
p$llogfaminc2001 <- (p$logfaminc2001 >= 9)*p$logfaminc2001  
p$lowfaminc2002 <- 1*(p$logfaminc2002 < 9)  
p$llogfaminc2002 <- (p$logfaminc2002 >= 9)*p$logfaminc2002  
p$lowfaminc2003 <- 1*(p$logfaminc2003 < 9)  
p$llogfaminc2003 <- (p$logfaminc2003 >= 9)*p$logfaminc2003  
p$lowfaminc2004 <- 1*(p$logfaminc2004 < 9)  
p$llogfaminc2004 <- (p$logfaminc2004 >= 9)*p$logfaminc2004  
p$lowfaminc2005 <- 1*(p$logfaminc2005 < 9)  
p$llogfaminc2005 <- (p$logfaminc2005 >= 9)*p$logfaminc2005  
p$lowfaminc2006 <- 1*(p$logfaminc2006 < 9)
```

```

p$llogfaminc2006 <- (p$logfaminc2006 >= 9)*p$logfaminc2006
pim <- p[p[["X_mi_m"]] >0,]
p0 <- p[p[["X_mi_m"]]==0,]
p0 <- p0[match(p[["X_mi_id"]][p[["X_mi_m"]]==1],p0[["X_mi_id"]]),]
nimputes <- 25
npts <- nrow(pim)/nimputes
fit1 <- gam(concdis_2001 ~
  childmale +
  birthweight +
  mommarriedbirth +
  momagebirth +
  s(age01) +
  hhedu +
  momreading_ss +
  ownhome +
  concdis_2000 +
  pcgmstat2000 +
  factor(pcgemploy2000) +
  welfare2000 +
  lowfaminc2000 +
  llogfaminc2000 +
  pcghours2000 +
  numchild2000 +
  pcgmstat2001 +
  factor(pcgemploy2001) +
  welfare2001 +
  lowfaminc2001 +
  llogfaminc2001 +
  pcghours2001 +
  numchild2001 +
  moves2001,
  data=pim)
print(summary(fit1))
p1 <- predict(fit1,type="response",se=T)
sd1 <- p1$se.fit
p1 <- p1$fit
st1 <- p0[match(pim[["X_mi_id"]],p0[["X_mi_id"]]),"concdis_2001"]
pt1 <- dnorm(x=st1,mean=p1,sd=sd1)
sim1<- rnorm(n=p1,mean=p1,sd=sd1)
#
fit2 <- gam(concdis_2002 ~
  childmale +

```

```

    birthweight +
    mommarriedbirth +
    momagebirth +
    s(age02) +
    hhedu +
    momreading_ss +
    ownhome +
    concdis_2000 +
    pcgmstat2000 +
    factor(pcgemploy2000) +
    welfare2000 +
    lowfaminc2000 +
    llogfaminc2000 +
    pcghours2000 +
    numchild2000 +
    concdis_2001 +
    pcgmstat2001 +
    factor(pcgemploy2001) +
    welfare2001 +
    lowfaminc2001 +
    llogfaminc2001 +
    pcghours2001 +
    pcgmstat2002 +
    factor(pcgemploy2002) +
    welfare2002 +
    lowfaminc2002 +
    llogfaminc2002 +
    pcghours2002 +
    numchild2002 +
    moves2002,
    data=pim)
print(summary(fit2))
ps2 <- pim
ps2[["concdis_2001"]] <- st1
p2 <- predict(fit2,type="response",se=T,newdata=ps2)
sd2 <- p2$se.fit
p2 <- p2$fit
st2 <- p0[match(pim[["X_mi_id"]],p0[["X_mi_id"]]),"concdis_2002"]
pt2 <- dnorm(x=st2,mean=p2,sd=sd2)
#
ps2[["concdis_2001"]] <- sim1
p2 <- predict(fit2,type="response",se=T,newdata=ps2)

```

```

sd2 <- p2$se.fit
p2 <- p2$fit
sim2<- rnorm(n=p2,mean=p2,sd=sd2)
#
fit3 <- gam(concdis_2003 ~
  childmale +
  birthweight +
  mommarriedbirth +
  momagebirth +
  s(age03) +
  hhedu +
  momreading_ss +
  ownhome +
  concdis_2000 +
  pcgmstat2000 +
  factor(pcgemploy2000) +
  welfare2000 +
  lowfaminc2000 +
  llogfaminc2000 +
  pcghours2000 +
  numchild2000 +
  concdis_2002 +
  pcgmstat2002 +
  factor(pcgemploy2002) +
  welfare2002 +
  lowfaminc2002 +
  llogfaminc2002 +
  pcghours2002 +
  pcgmstat2003 +
  factor(pcgemploy2003) +
  welfare2003 +
  lowfaminc2003 +
  llogfaminc2003 +
  pcghours2003 +
  numchild2003 +
  moves2003,
  data=pim)
print(summary(fit3))
ps3 <- pim
ps3[["concdis_2002"]] <- st2
p3 <- predict(fit3,type="response",se=T,newdata=ps3)
sd3 <- p3$se.fit

```

```

p3 <- p3$fit
st3 <- p0[match(pim[["X_mi_id"]],p0[["X_mi_id"]]),"concdis_2003"]
pt3 <- dnorm(x=st3,mean=p3,sd=sd3)
#
ps3[["concdis_2002"]] <- sim2
p3 <- predict(fit3,type="response",se=T,newdata=ps3)
sd3 <- p3$se.fit
p3 <- p3$fit
sim3<- rnorm(n=p3,mean=p3,sd=sd3)
#
fit4 <- gam(concdis_2004 ~
  childmale +
  birthweight +
  mommarriedbirth +
  momagebirth +
  s(age04) +
  hhedu +
  momreading_ss +
  ownhome +
  concdis_2000 +
  pcgmstat2000 +
  factor(pcgemploy2000) +
  welfare2000 +
  lowfaminc2000 +
  llogfaminc2000 +
  pcghours2000 +
  numchild2000 +
  concdis_2003 +
  pcgmstat2003 +
  factor(pcgemploy2003) +
  welfare2003 +
  lowfaminc2003 +
  llogfaminc2003 +
  pcghours2003 +
  pcgmstat2004 +
  factor(pcgemploy2004) +
  welfare2004 +
  lowfaminc2004 +
  llogfaminc2004 +
  pcghours2004 +
  numchild2004 +
  moves2004,

```

```

        data=pim)
print(summary(fit4))
ps4 <- pim
ps4[["concdis_2003"]] <- st3
p4 <- predict(fit4,type="response",se=T,newdata=ps4)
sd4 <- p4$se.fit
p4 <- p4$fit
st4 <- p0[match(pim[["X_mi_id"]],p0[["X_mi_id"]]),"concdis_2004"]
pt4 <- dnorm(x=st4,mean=p4,sd=sd4)
#
ps4[["concdis_2003"]] <- sim3
p4 <- predict(fit4,type="response",se=T,newdata=ps4)
sd4 <- p4$se.fit
p4 <- p4$fit
sim4<- rnorm(n=p4,mean=p4,sd=sd4)
#
fit5 <- gam(concdis_2005 ~
  childmale +
  birthweight +
  mommarriedbirth +
  momagebirth +
  s(age05) +
  hhedu +
  momreading_ss +
  ownhome +
  concdis_2000 +
  pcgmstat2000 +
  factor(pcgemploy2000) +
  welfare2000 +
  lowfaminc2000 +
  llogfaminc2000 +
  pcghours2000 +
  numchild2000 +
  concdis_2004 +
  pcgmstat2004 +
  factor(pcgemploy2004) +
  welfare2004 +
  lowfaminc2004 +
  llogfaminc2004 +
  pcghours2004 +
  pcgmstat2005 +
  factor(pcgemploy2005) +

```



```

welfare2005 +
lowfaminc2005 +
llogfaminc2005 +
pcghours2005 +
numchild2005 +
moves2005,
    data=pim)
print(summary(fit5))
ps5 <- pim
ps5[["concdis_2004"]] <- st4
p5 <- predict(fit5,type="response",se=T,newdata=ps5)
sd5 <- p5$se.fit
p5 <- p5$fit
st5 <- p0[match(pim[["X_mi_id"]],p0[["X_mi_id"]]),"concdis_2005"]
pt5 <- dnorm(x=st5,mean=p5,sd=sd5)
#
ps5[["concdis_2004"]] <- sim4
p5 <- predict(fit5,type="response",se=T,newdata=ps5)
sd5 <- p5$se.fit
p5 <- p5$fit
sim5<- rnorm(n=p4,mean=p4,sd=sd4)
#
fit6 <- gam(concdis_2006 ~
    childmale +
    birthweight +
    mommarriedbirth +
    momagebirth +
    s(age06) +
    hhedu +
    momreading_ss +
    ownhome +
    concdis_2000 +
    pcgmstat2000 +
    factor(pcgemploy2000) +
    welfare2000 +
    lowfaminc2000 +
    llogfaminc2000 +
    pcghours2000 +
    numchild2000 +
    concdis_2005 +
    pcgmstat2005 +
    factor(pcgemploy2005) +

```

```

welfare2005 +
lowfaminc2005 +
llogfaminc2005 +
pcghours2005 +
pcgmstat2006 +
factor(pcgemploy2006) +
welfare2005 +
lowfaminc2006 +
llogfaminc2006 +
pcghours2006 +
numchild2006 +
moves2006,
data=pim)
print(summary(fit6))
ps6 <- pim
ps6[["concdis_2005"]] <- st5
p6 <- predict(fit6,type="response",se=T,newdata=ps6)
sd6 <- p6$se.fit
p6 <- p6$fit
st6 <- p0[match(pim[["X_mi_id"]],p0[["X_mi_id"]]),"concdis_2006"]
pt6 <- dnorm(x=st6,mean=p6,sd=sd6)
#
ps6[["concdis_2005"]] <- sim5
p6 <- predict(fit6,type="response",se=T,newdata=ps6)
sd6 <- p6$se.fit
p6 <- p6$fit
sim6<- rnorm(n=p6,mean=p6,sd=sd6)
#
pt <- tapply(pt1*pt2*pt3*pt4*pt5*pt6,rep(1:npts,25),mean)
#expave <- tapply((sim1+sim2+sim3+sim4+sim5+sim6)/6,rep(1:npts,5),mean)
#exprecency <-
tapply((sim1+2*sim2+3*sim3+4*sim4+5*sim5+6*sim6)/21,rep(1:(length(pt1)/5),5),mean)
#pt <- pt1*pt2*pt3*pt4*pt5*pt6
expave <- (sim1+sim2+sim3+sim4+sim5+sim6)/6
exprecency <- (6*sim1+5*sim2+4*sim3+3*sim4+2*sim5+sim6 - 21*expave)/21
expearly <- (6*sim6+5*sim5+4*sim4+3*sim3+2*sim2+sim1 - 21*expave)/21
#pt <- (pt1*pt2*pt3*pt4*pt5*pt6)[1:npts]
#expave <- ((sim1+sim2+sim3+sim4+sim5+sim6)/6)[(1:(length(pt1)/5))]
#exprecency <- ((sim1+2*sim2+3*sim3+4*sim4+5*sim5+6*sim6)/21)[(1:(length(pt1)/5))]
save.image(file="ps.RData")
}else{
load(file="ps.RData")

```

```

}

#library(survey)
#average_treatment <- apply(p0[,66:71],1,mean)
#recency <- apply(sweep(p0[,66:71],1,average_treatment,"-"),1,function(x){sum((1:6)*x)/sum(1:6)})
#cor( average_treatment,recency)
#pim$average_treatment <- average_treatment
#pim$recency <- recency
#pim$expave <- expave
#pim$exprecency <- exprecency
#pim$ps <- pt
#pim$ops <- log(pim$ps)
#pim$im <- rep(1:nimputes,rep(npts,nimputes))
##dp <- svydesign(ids=~1,data=p0,weights=~(pwgt_rscsib/pt))
##
#w <- 1/pt
#w[w < 0.0001] <- 0.0001
#dp <- svydesign(ids=~1,data=pim,weights=~w)
##dp <- svydesign(ids=~1,data=p0,weights=~1)
##
##fit.ss <- svyglm(W2_ap_ss ~ expave + ps +sps+ ps*expave + childmale +
##fit.ss <- svyglm(W2_ap_ss ~ expave + exprecency + average_treatment + recency + childmale +
factor(childrace_latino) +
##fit.w <- svyglm(W2_lwi_ss ~ expave + exprecency + average_treatment + recency + childmale +
factor(childrace_latino) +
## math
#fit.w <- lm(W2_ap_ss ~ expave + exprecency + childmale + factor(childrace_latino) +
#      birthweight+mommarriedbirth+momagebirth+age01+hhedu+momreading_ss+ownhome+
#      pcgmstat2000+pcgemploy2000+welfare2000+logfaminc2000+pcghours2000+numchild2000,
#      data=pim,subset=(im==1))
#summary(fit.w)
## reading
#fit.w <- lm(W2_lwi_ss ~ expave + exprecency + childmale + factor(childrace_latino) +
#      birthweight+mommarriedbirth+momagebirth+age01+hhedu+momreading_ss+ownhome+
#      pcgmstat2000+pcgemploy2000+welfare2000+logfaminc2000+pcghours2000+numchild2000,
#      data=pim,subset=(im==1))
#summary(fit.w)
#fit.w <- svyglm(W2_lwi_ss ~ expave + exprecency + childmale + factor(childrace_latino) +
#      birthweight+mommarriedbirth+momagebirth+age01+hhedu+momreading_ss+ownhome+
#      pcgmstat2000+pcgemploy2000+welfare2000+logfaminc2000+pcghours2000+numchild2000,
#      design=dp)
#summary(fit.w)

```

```

##library(mgcv)
##pdf("ps.pdf")
##fit.ss <- gam(W2_ap_ss ~ s(expave) + ps + childmale + factor(childrace_latino) +
##      birthweight+mommarriedbirth+momagebirth+s(age01)+hhedu+momreading_ss+ownhome+
##      pcgmstat2000+pcgemploy2000+welfare2000+logfaminc2000+pcghours2000+numchild2000,
##      data=p0)
##plot(fit.ss)
##summary(fit.ss)
##fit.ss <- gam(W2_ap_ss ~ s(expave) + s(expave,by=ps) + ps + childmale + factor(childrace_latino) +
##      birthweight+mommarriedbirth+momagebirth+s(age01)+hhedu+momreading_ss+ownhome+
##      pcgmstat2000+pcgemploy2000+welfare2000+logfaminc2000+pcghours2000+numchild2000,
##      data=p0)
##summary(fit.ss)
##plot(fit.ss)
##fit.ss <- gam(W2_ap_ss ~ s(expave) + childmale + factor(childrace_latino) +
##      birthweight+mommarriedbirth+momagebirth+s(age01)+hhedu+momreading_ss+ownhome+
##      pcgmstat2000+pcgemploy2000+welfare2000+logfaminc2000+pcghours2000+numchild2000,
##      data=p0)
##plot(fit.ss)
##summary(fit.ss)
##fit.ss <- gam(W2_ap_ss ~ s(expave) + s(expave,by=ps) + childmale + factor(childrace_latino) +
##      birthweight+mommarriedbirth+momagebirth+s(age01)+hhedu+momreading_ss+ownhome+
##      pcgmstat2000+pcgemploy2000+welfare2000+logfaminc2000+pcghours2000+numchild2000,
##      data=p0)
##summary(fit.ss)
##plot(fit.ss)
####fit.w <- svyglm(W2_aplwi_ss expave + exprecency + childmale +
####      birthweight+mommarriedbirth+momagebirth+age01+hhedu+momreading_ss+ownhome+
####      pcgmstat2000+pcgemploy2000+welfare2000+logfaminc2000+pcghours2000+numchild2000,
####      design=dp)
####summary(fit.w)
####
####fit.ss <- svyglm((as.numeric(childmale)-1) ~ expave + exprecency + ps, family="binomial",
####      design=dp)

#####
# Plot Figures
#####
library(MASS)
library(mgcv)
pdf("psim.pdf")

```

```

#if(T){
set.seed(3141)
load(file="ps.RData")
p1 <- predict(fit1,se=FALSE)
nsim <- 1000
npts <- length(pt1)/nimputes
ns <- array(0,dim=c(nimputes,npts,6,nsim))
for(i in 1:nsim){
  sim1<- rnorm(n=p1,mean=p1,sd=sd1)
  ps2[["concdis_2001"]] <- sim1
  p2 <- predict(fit2,se=FALSE,newdata=ps2)
  sim2<- rnorm(n=p2,mean=p2,sd=sd2)
  ps3[["concdis_2002"]] <- sim2
  p3 <- predict(fit3,se=FALSE,newdata=ps3)
  sim3<- rnorm(n=p3,mean=p3,sd=sd3)
  ps4[["concdis_2003"]] <- sim3
  p4 <- predict(fit4,se=FALSE,newdata=ps4)
  sim4<- rnorm(n=p4,mean=p4,sd=sd4)
  ps5[["concdis_2004"]] <- sim4
  p5 <- predict(fit5,se=FALSE,newdata=ps5)
  sim5<- rnorm(n=p5,mean=p5,sd=sd5)
  ps6[["concdis_2005"]] <- sim5
  p6 <- predict(fit6,se=FALSE,newdata=ps6)
  sim6<- rnorm(n=p6,mean=p6,sd=sd6)
  a <- cbind(sim1,sim2,sim3,sim4,sim5,sim6)
  for(j in 1:nimputes){
    ns[j,,i] <- a[((j-1)*npts+1):(j*npts),]
  }
}
#
cnames <- colnames(a)
nsam=array(0,dim=c(npts,6,nimputes))
expave <- NULL
expearly <- NULL
pim$opca1 <- pim$age01
pim$opca2 <- pim$age01
for(i in 1:6){
  for(j in 1:nimputes){
    nsam[,i,j] <- tapply(ns[j,,i],rep(1:npts,nsim),mean)
  }
}
for(j in 1:nimputes){

```

```

f <- prcomp(nsam[,j])
a=predict(f)
  expa <- apply(nsam[,j],1,mean)
  expr <- as.numeric( sweep(nsam[,j],1,expa,"-") %*% c(6:1) ) / 21)
#   expa <- a[,1]
#   expr <- a[,2]
  expr3 <- a[,3]
  expave <- c(expave,expa)
  expearly <- c(expearly,expr)
od <- pim[((j-1)*npts+1):(j*npts),seq(67,77,by=2)]
colnames(od) <- cnames
pim$opca1[((j-1)*npts+1):(j*npts)]=predict(f,newdata=od)[,1]
pim$opca2[((j-1)*npts+1):(j*npts)]=predict(f,newdata=od)[,2]
}
#save.image(file="psim.RData")
#}else{
#load(file="psim.RData")
#}
#
ns1m=matrix(0,nrow=npts,ncol=6)
for(i in 1:6){
  ns1m[,i] <- tapply(ns[1,,i],rep(1:npts,nsim),mean)
}
expave <- apply(ns1m,1,mean)
expearly <- as.numeric( sweep(ns1m,1,expave,"-") %*% c(6:1) ) / 21)
f <- prcomp(ns1m)
summary(f)
f$rotation
par(mfrow=c(2,2))
plot(f$rotation[,1],main="First PC")
plot(f$rotation[,2],main="Second PC")
plot(f$rotation[,3],main="Third PC")
plot(f$rotation[,4],main="Fourth PC")
cor(ns[1,,1])
#
a=predict(f)
pca1=a[,1]
pca2=a[,2]
pca3=a[,3]
#
#for(i in 1:6){
#   pca1[,i] <- tapply(ns[1,,],i,rep(1:npts,nsim),mean)

```

```

#}
#pca2=pca1 %*% f$rotation
#pca2=apply(matrix(a[,2],nrow=npts,byrow=F),1,mean)
#pca3=apply(matrix(a[,3],nrow=npts,byrow=F),1,mean)
#
cor(a)
cor(cbind(pca1,pca2,pca3))
plot(pca1,pca2)
plot(density(a[,1]))
plot(density(a[,2]))
hist((a[,1]),nclass=100)
hist((a[,2]),nclass=100)
plot(density(pca1))
plot(density(pca2))
hist((pca1),nclass=100)
hist((pca2),nclass=100)
##
#f <- prcomp(ns[1,,])
#summary(f)
#names(f)
#f$rotation
#par(mfrow=c(2,2))
#plot(f$rotation[,1],main="First PC")
#plot(f$rotation[,2],main="Second PC")
#plot(f$rotation[,3],main="Third PC")
#plot(f$rotation[,4],main="Fourth PC")
#cor(ns[1,,])
##
#f <- prcomp(ns2)
#summary(f)
#names(f)
#f$rotation
#par(mfrow=c(2,2))
#plot(f$rotation[,1],main="First PC")
#plot(f$rotation[,2],main="Second PC")
#plot(f$rotation[,3],main="Third PC")
#plot(f$rotation[,4],main="Fourth PC")
#cor(ns2)

#pt <- tapply(pt1*pt2*pt3*pt4*pt5*pt6,rep(1:(length(pt1)/5),5),mean)
#expave <- tapply((sim1+sim2+sim3+sim4+sim5+sim6)/6,rep(1:(length(pt1)/5),5),mean)
#expearly <- tapply((sim1+2*sim2+3*sim3+4*sim4+5*sim5+6*sim6)/21,rep(1:(length(pt1)/5),5),mean)

```

```

#
library(survey)
average_treatment <- apply(pim[,seq(67,77,by=2)],1,mean)
early <- sweep(pim[,seq(67,77,by=2)],1,average_treatment,"-")
early <- apply(early,1,function(x){sum((6:1)*x)/sum(6:1)})
cor( cbind(average_treatment,early,expave,expearly))
pim$average_treatment <- average_treatment
pim$early <- early
pim$expave <- expave
pim$expearly <- expearly
pim$ps <- pt
pim$im <- rep(1:nimputes,rep(npts,nimputes))
pim$pca1 <- pca1
pim$pca2 <- pca2
od <- pim[,seq(67,77,by=2)]
colnames(od) <- cnames
pim$opca1=predict(f,newdata=od)[,1]
pim$opca2=predict(f,newdata=od)[,2]
cor(cbind(pca1,pca2,p0$opca1,p0$opca2))
cor(cbind(pca1,pca2,expave,expearly))
cor(cbind(average_treatment,early,pca2,p0$opca1,p0$opca2))
w <- 1/pt
w[w < 0.0001] <- 0.0001
dp <- svydesign(ids=~1,data=pim,weights=~w)
#
#fit.ss <- svyglm(W2_ap_ss ~ expave + ps +sps+ ps*expave + childmale +
#math
fit.ss <- lm(W2_ap_ss ~ pca1 + opca1 + pca2 +opca2 + childmale + factor(childrace_latino) +
birthweight+mommarriedbirth+momagebirth+age01+hhedu+momreading_ss+ownhome+
pcgmstat2000+pcgemploy2000+welfare2000+lowfaminc2000+llogfaminc2000+pcghours2000+numchild
2000,
data=pim,subset=(im==1))
summary(fit.ss)
#reading
fit.ss <- lm(W2_lwi_ss ~ pca1 + opca1 + pca2 +opca2 + childmale + factor(childrace_latino) +
birthweight+mommarriedbirth+momagebirth+age01+hhedu+momreading_ss+ownhome+
pcgmstat2000+pcgemploy2000+welfare2000+lowfaminc2000+llogfaminc2000+pcghours2000+numchild
2000,
data=pim,subset=(im==1))

```



```

summary(fit.ss)
#
library(mgcv)
par(mfrow=c(2,2))
fit.ss <- gam(W2_lwi_ss ~ s(pca1) + s(pca2) + s(pca1,by=opca1) + s(pca2,by=opca2) + childmale +
factor(childrace_latino) +
      birthweight+mommarriedbirth+momagebirth+s(age01)+hhedu+momreading_ss+ownhome+
      pcgmstat2000+pcgemploy2000+welfare2000+lowfaminc2000+llogfaminc2000+pcghours2000+numchild
      2000,
      data=pim,subset=(im==1))
plot(fit.ss,main="Reading with 2 x PCA")
summary(fit.ss)
par(mfrow=c(2,2))
fit.ss <- gam(W2_lwi_ss ~ s(expave) + s(expave,by=average_treatment) + s(expearly) +
s(expearly,by=early) + childmale + factor(childrace_latino) +
      birthweight+mommarriedbirth+momagebirth+s(age01)+hhedu+momreading_ss+ownhome+
      pcgmstat2000+pcgemploy2000+welfare2000+lowfaminc2000+llogfaminc2000+pcghours2000+numchild
      2000,
      data=pim,subset=(im==1))
summary(fit.ss)
plot(fit.ss, main="Reading with e/o ave e/o R + s(age)")
par(mfrow=c(2,2))
fit.ss <- gam(W2_lwi_ss ~ s(expave) + s(expave,by=average_treatment) + s(expearly) +
s(expearly,by=early) + childmale + factor(childrace_latino) +
      birthweight+mommarriedbirth+momagebirth+age01+hhedu+momreading_ss+ownhome+
      pcgmstat2000+pcgemploy2000+welfare2000+lowfaminc2000+llogfaminc2000+pcghours2000+numchild
      2000,
      data=pim,subset=(im==1))
summary(fit.ss)
plot(fit.ss, main="Reading with e/o ave e/o R")
par(mfrow=c(1,2))
fit.ss <- gam(W2_lwi_ss ~ s(expave) + s(expave,by=average_treatment) + childmale +
factor(childrace_latino) +
      birthweight+mommarriedbirth+momagebirth+(age01)+hhedu+momreading_ss+ownhome+
      pcgmstat2000+pcgemploy2000+welfare2000+lowfaminc2000+llogfaminc2000+pcghours2000+numchild
      2000,
      data=pim,subset=(im==1))
summary(fit.ss)

```

```

plot(fit.ss, main="Reading with e/o ave")
#
rm(ns)
save.image(file="psim.RData")

pdf("pgam.pdf")
set.seed(1)
load(file="H:/CITITraining/Pebley/Sequential Selection/psim.RData")

#
colnames(pim[,seq(67,77,by=2)])
average_treatment <- apply(pim[,seq(67,77,by=2)],1,mean)
recency <- sweep(pim[,seq(67,77,by=2)],1,average_treatment,"-")
recency <- apply(recency,1,function(x){sum((1:6)*x)/sum(1:6)})
cor( cbind(average_treatment,recency,expave,exprecency))
pim$average_treatment <- average_treatment
pim$recency <- recency
pim$expave <- expave
pim$exprecency <- exprecency
pim$im <- rep(1:nimputes,rep(npts,nimputes))
library(mgcv)
#reading
par(mfrow=c(2,2))
b.ss <- NULL
V.ss <- NULL
b.ave <- NULL
V.ave <- NULL
for(imc in 1:nimputes){
par(mfrow=c(2,2))
fit.ss <- gam(W2_lwi_ss ~ te(expave) + te(expave,by=average_treatment) + te(exprecency) +
te(exprecency,by=recency) + childmale + factor(childrace_latino) +
      birthweight+mommarriedbirth+momagebirth+age01+hhedu+momreading_ss+ownhome+
      pcgmstat2000+pcgemploy2000+welfare2000+logfaminc2000+pcghours2000+numchild2000,
      data=pim,subset=(im==imc),method="ML")
summary(fit.ss)
plot(fit.ss, main=paste("Reading with e/o ave e/o R MI=",imc,sep=""))
par(mfrow=c(2,2))
fit.ss <- gam(W2_lwi_ss ~ te(expave) + te(expave,by=average_treatment) + te(exprecency) +
te(exprecency,by=recency) + childmale + factor(childrace_latino) +
      birthweight+mommarriedbirth+momagebirth+age01+hhedu+momreading_ss+ownhome+
      pcgmstat2000+pcgemploy2000+welfare2000+logfaminc2000+pcghours2000+numchild2000,
      data=pim,subset=(im==imc),method="ML")

```

```

summary(fit.ss)
par(mfrow=c(2,2))
#fit.ss <- gam(W2_lwi_ss ~ te(expave,exprecency) + te(expave,by=average_treatment) +
te(exprecency,by=recency) + childmale + factor(childrace_latino) +
fit.ss <- gam(W2_lwi_ss ~ te(expave,exprecency) + te(expave,exprecency,by=average_treatment) +
te(expave,exprecency,by=recency) + childmale + factor(childrace_latino) +
      birthweight+mommarriedbirth+momagebirth+s(age01)+hhedu+momreading_ss+ownhome+
      pcgmstat2000+pcgemploy2000+welfare2000+logfaminc2000+pcghours2000+numchild2000,
      data=pim,subset=(im==imc),method="ML")
print(summary(fit.ss))
plot(fit.ss, main=paste("Reading with e/o ave e/o R + te(age) MI=",imc,sep=""))
b.ss <- cbind(b.ss,coefficients(fit.ss))
if(imc==1){
  V.ss <- summary(fit.ss)$cov.scaled / nimputes
}else{
  V.ss <- V.ss + summary(fit.ss)$cov.scaled / nimputes
}
if(imc==nimputes){
  fit.ss.mi <- fit.ss
  Vmi.ss <- V.ss + (1+1/nimputes)*cov(t(b.ss))
  bmi.ss <- apply(b.ss,1,mean)
  dfmi.ss <- (nimputes-1)*(1+(nimputes*diag(V.ss))/((nimputes+1)*diag(cov(t(b.ss))))))^2
  fit.ss.mi$coefficients <- bmi.ss
  fit.ss.mi$Vp <- Vmi.ss
#save(b.ss,V.ss,bmi.ss,Vmi.ss,fit.ss.mi,file="gamfitsReading.RData")
}
plot(fit.ss, main=paste("Reading with e/o ave e/o R MI=",imc,sep=""))
par(mfrow=c(1,2))
fit.ave <- gam(W2_lwi_ss ~ te(expave) + te(expave,by=average_treatment) + childmale +
factor(childrace_latino) +
      birthweight+mommarriedbirth+momagebirth+s(age01)+hhedu+momreading_ss+ownhome+
      pcgmstat2000+pcgemploy2000+welfare2000+logfaminc2000+pcghours2000+numchild2000,
      data=pim,subset=(im==imc),method="ML")
summary(fit.ave)
plot(fit.ave, main=paste("Reading with e/o ave MI=",imc,sep=""))
b.ave <- cbind(b.ave,coefficients(fit.ave))
if(imc==1){
  V.ave <- summary(fit.ave)$cov.scaled / nimputes
}else{
  V.ave <- V.ave + summary(fit.ave)$cov.scaled / nimputes
}
if(imc==nimputes){

```

```

fit.ave.mi <- fit.ave
Vmi.ave <- V.ave + (1+1/nimputes)*cov(t(b.ave))
bmi.ave <- apply(b.ave,1,mean)
dfmi.ave <- (nimputes-1)*(1+(nimputes*diag(V.ave))/((nimputes+1)*diag(cov(t(b.ave))))))^2
fit.ave.mi$coefficients <- bmi.ave
fit.ave.mi$Vp <- Vmi.ave
save(b.ave,V.ave,bmi.ave,Vmi.ave,fit.ave.mi,
     b.ss,V.ss,bmi.ss,Vmi.ss,fit.ss.mi,file="gamfitsReading.RData")
}
}
par(mfrow=c(2,2))
summary(fit.ss.mi)
plot(fit.ss.mi, main=paste("Reading with e/o ave e/o R combined",sep=""))
par(mfrow=c(2,2))
summary(fit.ave.mi)
plot(fit.ave.mi, main=paste("Reading with e/o ave combined",sep=""))
#math
par(mfrow=c(2,2))
b.ss <- NULL
V.ss <- NULL
b.ave <- NULL
V.ave <- NULL
for(imc in 1:nimputes){
par(mfrow=c(2,2))
fit.ss <- gam(W2_ap_ss ~ te(expave) + te(expave,by=average_treatment) + te(exprecency) +
te(exprecency,by=recency) + childmale + factor(childrace_latino) +
      birthweight+mommarriedbirth+momagebirth+s(age01)+hhedu+momreading_ss+ownhome+
      pcgmstat2000+pcgemploy2000+welfare2000+logfaminc2000+pcghours2000+numchild2000,
      data=pim,subset=(im==imc),method="ML")
summary(fit.ss)
plot(fit.ss, main=paste("Math with e/o ave e/o R + te(age) MI=",imc,sep=""))
par(mfrow=c(2,2))
fit.ss <- gam(W2_ap_ss ~ te(expave,exprecency) + te(expave,exprecency,by=average_treatment) +
te(expave,exprecency,by=recency) + childmale + factor(childrace_latino) +
      birthweight+mommarriedbirth+momagebirth+s(age01)+hhedu+momreading_ss+ownhome+
      pcgmstat2000+pcgemploy2000+welfare2000+logfaminc2000+pcghours2000+numchild2000,
      data=pim,subset=(im==imc),method="ML")
summary(fit.ss)
b.ss <- cbind(b.ss,coefficients(fit.ss))
if(imc==1){
  V.ss <- fit.ss$Vp / nimputes
}else{

```

```

      V.ss <- V.ss + fit.ss$Vp / nimputes
    }
    if(imc==nimputes){
      fit.ss.mi <- fit.ss
      Vmi.ss <- V.ss + (1+1/nimputes)*cov(t(b.ss))
      #Vmi.ss <- V.ss
      bmi.ss <- apply(b.ss,1,mean)
      dfmi.ss <- (nimputes-1)*(1+(nimputes*diag(V.ss))/((nimputes+1)*diag(cov(t(b.ss))))))^2
      fit.ss.mi$coefficients <- bmi.ss
      fit.ss.mi$Vp <- Vmi.ss
      #save(b.ss,V.ss,bmi.ss,Vmi.ss,fit.ss.mi,file="gamfitsReading.RData")
    }
    plot(fit.ss, main=paste("Math with e/o ave e/o R MI=",imc,sep=""))
    par(mfrow=c(1,2))
    fit.ave <- gam(W2_ap_ss ~ te(expave) + te(expave,by=average_treatment) + childmale +
      factor(childrace_latino) +
        birthweight+mommarriedbirth+momagebirth+s(age01)+hhedu+momreading_ss+ownhome+
        pcgmstat2000+pcgemploy2000+welfare2000+logfaminc2000+pcghours2000+numchild2000,
        data=pim,subset=(im==imc),method="ML")
    summary(fit.ave)
    plot(fit.ave, main=paste("Math with e/o ave MI=",imc,sep=""))
    b.ave <- cbind(b.ave,coefficients(fit.ave))
    if(imc==1){
      V.ave <- summary(fit.ave)$cov.scaled
    }else{
      V.ave <- V.ave + summary(fit.ave)$cov.scaled
    }
    if(imc==nimputes){
      fit.ave.mi <- fit.ave
      V.ave <- V.ave / nimputes
      Vmi.ave <- V.ave + (1+1/nimputes)*cov(t(b.ave))
      bmi.ave <- apply(b.ave,1,mean)
      dfmi.ave <- (nimputes-1)*(1+(nimputes*diag(V.ave))/((nimputes+1)*diag(cov(t(b.ave))))))^2
      fit.ave.mi$coefficients <- bmi.ave
      fit.ave.mi$Vp <- Vmi.ave
      save(b.ave,V.ave,bmi.ave,Vmi.ave,fit.ave.mi,
        b.ss,V.ss,bmi.ss,Vmi.ss,fit.ss.mi,file="gamfitsMaths.RData")
    }
  }
  par(mfrow=c(2,2))
  summary(fit.ss.mi)
  plot(fit.ss.mi, main=paste("Math with e/o ave e/o R combined",sep=""))

```

```
par(mfrow=c(2,2))
summary(fit.ave.mi)
plot(fit.ave.mi, main=paste("Math with e/o ave combined",sep=""))
```

Stata Code Used to Create L.A.FANS Analytic Data Files for Hicks et al. (2016)

```
*****
* File:          Create Analytic File.do
* Summary:       Uses LAFANS data to create an analytic file
*               similar to that of Wodtke et al. (2011)
* Author:        Andrew Hicks
* Institution:   California Center for Population Research
*               UCLA
* Last Updated:  9/29/2014
*****

cd C:\Users\ahicks\Documents\Stata_data\lafans

***** make headstart variables *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\parent1.dta", clear
gen waveldate=date(pdate, "MDY")
gen hsattend =3
replace hsattend =1 if (pg40==1)
replace hsattend =2 if (pg68==1)
label var hsattend "current or ever headstart attend"
recode hsattend (1 2=1) (3=0)
tab hsattend

***** make preschool variables *****
gen ccattend =3
replace ccattend =1 if (pg4a_a==21 | pg4b_a==21 | pg4c_a==21 |
pg11a_a==21 | pg11b_a==21 | pg11c_a==21)
replace ccattend =2 if (pg72==1)
label var ccattend "current or ever preschl/childcare attend"

recode ccattend (1 2=1) (3=0)
tab ccattend

***** make other non-parental variables *****
gen nonparentcare =3
replace nonparentcare =1 if ((pg4a_a<=15 & pg4a_a>=1)|(pg4b_a<=15 &
pg4b_a>=1)|(pg4c_a<=15 & pg4c_a>=1)|(pg11a_a<=15 &
pg11a_a>=1)|(pg11b_a<=15 & pg11b_a>=1)|(pg11c_a<=15 & pg11c_a>=1) |
(pg11a_a==23) | (pg11b_a==23) | (pg11c_a==23) | (pg11a_a==24) |
(pg11b_a==24) | (pg11c_a==24))
tab nonparentcare
label var nonparentcare "current other nonparental care"

label define care 1 "Yes" 0 "No"

label values hsattend ccattend nonparentcare care

save C:\Users\ahicks\Documents\Stata_data\lafans\parent1_childcare.dta,
replace

***** flag headstart and non-parental combined *****
gen headst_nonparental =0
replace headst_nonparental =1 if (nonparentcare==1 & hsattend==1)
```

```

tab headst_nonparental

***** flag preschool and non-parental combined *****
gen presch_nonparental =0
replace presch_nonparental =1 if (nonparentcare==1 & ccattend==1)
tab presch_nonparental

*****generate dates between headstart begin and interview
*****
gen int_date=date(pdate, "MDY")
format int_date %d
gen int_date_mo=mofd(int_date)
gen hs_date=mdy(pg43_m,1,pg43_y)
format hs_date %d
gen hs_date_mo=mofd(hs_date)
gen hsmonths=int_date_mo-hs_date_mo

*****generate dates between child care begin and interview
*****
gen cc1_date=mdy(pg55a_m,1,pg55a_y)
gen cc2_date=mdy(pg55b_m,1,pg55b_y)
format cc1_date cc2_date %d
gen cc_date=min(cc1_date,cc2_date)
format cc_date %d
gen cc_date_mo=mofd(cc_date)
gen ccmonths=int_date_mo-cc_date_mo

save C:\Users\ahicks\Documents\Stata_data\lafans\parent1_childcare.dta,
replace

***** sort parent1 *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\parent1_childcare.dta",
clear
sort sampleid kidtype
rename pd3_a childcitizen
rename pd3a childvisa
rename pd3_s child_otherstatus
replace childcitizen=1 if pd2==1
keep hhid pid childid sampleid kidtype page pid childid childcitizen
childvisa child_otherstatus bpi* pg43_a-pg45 pg50-pg51 pg55a_a-pg57a
pg62a pg63a pg55b_a-pg57b pg62b pg63b pg69-pg70_s pg73-pg74_s hsattend
ccattend nonparentcare pc7* pc9* pc11* pc13* pc4 pc5 pf15 pf16* pf11*
pf12-pf14_c pf20_a pf20_b pf4_a hsmonths ccmonths waveldate pk5_lb pk5_oz
pk5_k pk5_a rpcgs age_yr pe49 mom_pcg
save C:\Users\ahicks\Documents\Stata_data\lafans\parent1_sorted.dta,
replace

***** sort assess1 *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\assess1.dta", clear
rename resptype kidtype

```



```

sort sampleid kidtype
keep sampleid kidtype lwi_ss ap_ss age_a yrborn
save C:\Users\ahicks\Documents\Stata_data\lafans\assess1_sorted.dta,
replace

***** merge parent and assess *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\assess1_sorted.dta",
clear
merge 1:1 sampleid kidtype using
"C:\Users\ahicks\Documents\Stata_data\lafans\parent1_sorted.dta"
drop _merge
sort hhid pid kidtype
quietly by hhid pid kidtype: gen dup = cond(_N==1,0,_n)
drop if dup>0
save C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta,
replace

***** merge marital status at birth
use "C:\Users\ahicks\Documents\Stata_data\lafans\parent1_rev.dta", clear
merge n:1 hhid pid using
"C:\Users\ahicks\Documents\Stata_data\lafans\adult1_rev.dta"
drop if _merge==2

lab def marital_status 1 "Currently Married" 2 "Separated" 3 "Widowed" 4
"Divorced" 5 "Never Married"
lab values ael marital_status

lab def Parent_HH 1 "Mother and Father in HH" 2 "Mom Only, No Dad in HH"
3 "Dad Only, No Mom in HH" 4 "Neither Mom Nor Dad in HH"
lab values pel_rev Parent_HH

gen married_date=mdy(ae2_1,1,ae2_3)
gen born_date=mdy(pa7_m,1,pa7_y)

keep hhid pid pel_rev ael married_date born_date kidtype pe9 pe8 pe3
merge 1:1 hhid pid kidtype using
"C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta"
drop _merge
save C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta,
replace

***** sort parents roster1 *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\roster1.dta", clear
sort sampleid pid
keep sampleid pid rb1 rb2*
rename rb1 pcg_edu
rename rb2* pcg_race*
save C:\Users\ahicks\Documents\Stata_data\lafans\roster1_sorted.dta,
replace

***** sort household roster for pov cat *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\rosthhl.dta", clear
sort sampleid

```

```

keep sampleid povcat
merge 1:m sampleid using
"C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta,
replace

***** add pcg roster data to wavel *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta",
clear
drop _merge
merge m:1 sampleid pid using
"C:\Users\ahicks\Documents\Stata_data\lafans\roster1_sorted.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta,
replace

***** sort head of household roster1 *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\roster1.dta", clear
keep if ra7_a==0
sort sampleid pid
keep sampleid pid rb1
rename rb1 hhedu
save C:\Users\ahicks\Documents\Stata_data\lafans\hhedu_sorted.dta,
replace

***** add head of household roster data to wavel *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta",
clear
drop _merge
merge m:1 sampleid using
"C:\Users\ahicks\Documents\Stata_data\lafans\hhedu_sorted.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta,
replace

*****collapse HOME subscales*****
use "C:\Users\ahicks\Documents\Stata_data\lafans\parent1.dta", clear
keep sampleid ph11-ph13 ph27-ph29 ph44-ph47 ph64 ph65
foreach v of var * {
local l`v' : variable label `v'
if `"'l`v'"' == "" {
local l`v' "`v'"
}
}
collapse (lastnm) ph11-ph13 ph27-ph29 ph44-ph47 ph64 ph65, by(sampleid)
foreach v of var * {
label var `v' "'l`v'"
}
save C:\Users\ahicks\Documents\Stata_data\lafans\HomeCollapse_sorted.dta,
replace

***** merge Home Collapse Sorted *****

```

```

use "C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta",
clear
drop _merge
merge n:1 sampleid using
"C:\Users\ahicks\Documents\Stata_data\lafans\HomeCollapse_sorted.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta,
replace

***** sort kid roster1 *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\roster1.dta", clear
sort sampleid pid
rename pid childid
keep sampleid childid rb2*
rename rb2* child_race*
save C:\Users\ahicks\Documents\Stata_data\lafans\roster1child_sorted.dta,
replace

***** add kid roster data to wavel *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta",
clear
drop _merge
merge m:1 sampleid childid using
"C:\Users\ahicks\Documents\Stata_data\lafans\roster1child_sorted.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta,
replace

***** sort pcg1 *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\pcg1.dta", clear
sort sampleid pid
keep sampleid pid rsc_age gc5-gc9 ge2* ge3* caseness probcase gd*
save C:\Users\ahicks\Documents\Stata_data\lafans\pcg1_sorted.dta, replace

***** add pcg1 to wavel *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta",
clear
drop _merge
merge m:1 sampleid pid using
"C:\Users\ahicks\Documents\Stata_data\lafans\pcg1_sorted.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta,
replace

***** sort hof1 *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\hof1.dta", clear
sort sampleid
save C:\Users\ahicks\Documents\Stata_data\lafans\hof1_sorted.dta, replace

***** add hof1 to wavel *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta",
clear

```

```

drop _merge
merge m:1 sampleid using
"C:\Users\ahicks\Documents\Stata_data\lafans\hof1_sorted.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta,
replace

***** sort mom's assess1 *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\assess1.dta", clear
rename pc_raw momreading_raw
rename pc_w momreading_w
rename pc_ss momreading_ss
keep sampleid pid momreading_raw momreading_w momreading_ss
sort sampleid pid
save C:\Users\ahicks\Documents\Stata_data\lafans\assess1_mom_reading.dta,
replace

***** add mom reading to wavel *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta",
clear
drop _merge
merge m:1 sampleid pid using
"C:\Users\ahicks\Documents\Stata_data\lafans\assess1_mom_reading.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta,
replace

***** sort immigration adult1 *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\adult1.dta", clear
keep sampleid pid ac34* ac38-ac42
rename ac34_4 pcgborn
rename ac38 pcgcitizen
rename ac39 pcggreencard
rename ac40 pcgasyllum
rename ac41 pcgvisa
rename ac42 pcgvisa_valid
replace pcgcitizen=1 if pcgborn==1
sort sampleid pid
save C:\Users\ahicks\Documents\Stata_data\lafans\adult1_immigr.dta,
replace

***** add pcg immigration to wavel *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta",
clear
drop _merge
merge m:1 sampleid pid using
"C:\Users\ahicks\Documents\Stata_data\lafans\adult1_immigr.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta,
replace

***** sort income variables *****

```

```

use
"C:\Users\ahicks\Documents\Stata_data\lafans\Restricted\Wave1\VER1\impinc
1r.dta", clear
keep hhid hhrf faminc d_house
sort hhid hhrf
quietly by hhid hhrf: gen dup = cond(_N==1,0,_n)
drop if dup>0
drop dup
sort hhid
quietly by hhid : gen dup = cond(_N==1,0,_n)
drop if dup==1
save C:\Users\ahicks\Documents\Stata_data\lafans\impinc1_sorted.dta,
replace

```

```

***** merge Wave1 income variables *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wave1preschool.dta",
clear
drop _merge
merge m:1 hhid using
"C:\Users\ahicks\Documents\Stata_data\lafans\impinc1_sorted.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wave1preschool.dta,
replace

```

```

***** make Wave 1 employment variables *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\ehc1.dta", clear
/*
gen dasdate1 = date(asdate1,"YMD")
gen daedate1 = date(aedate1,"YMD")
gen dasdate2 = date(asdate2,"YMD")
gen daedate2 = date(aedate2,"YMD")
gen dasdate3 = date(asdate3,"YMD")
gen daedate3 = date(aedate3,"YMD")
gen dasdate4 = date(asdate4,"YMD")
gen daedate4 = date(aedate4,"YMD")
gen dasdate5 = date(asdate5,"YMD")
gen daedate5 = date(aedate5,"YMD")
gen dasdate6 = date(asdate6,"YMD")
gen daedate6 = date(aedate6,"YMD")
gen dasdate7 = date(asdate7,"YMD")
gen daedate7 = date(aedate7,"YMD")
gen dasdate8 = date(asdate8,"YMD")
gen daedate8 = date(aedate8,"YMD")
gen dasdate9 = date(asdate9,"YMD")
gen daedate9 = date(aedate9,"YMD")
gen dasdate10 = date(asdate10,"YMD")
gen daedate10 = date(aedate10,"YMD")
gen dasdate11 = date(asdate11,"YMD")
gen daedate11 = date(aedate11,"YMD")
gen msdate1=month(dasdate1)
gen medate1=month(daedate1)

```

```

gen msdate2=month(dasdate2)
gen medate2=month(daedate2)
gen msdate3=month(dasdate3)
gen medate3=month(daedate3)
gen msdate4=month(dasdate4)
gen medate4=month(daedate4)
gen msdate5=month(dasdate5)
gen medate5=month(daedate5)
gen msdate6=month(dasdate6)
gen medate6=month(daedate6)
gen msdate7=month(dasdate7)
gen medate7=month(daedate7)
gen msdate8=month(dasdate8)
gen medate8=month(daedate8)
gen msdate9=month(dasdate9)
gen medate9=month(daedate9)
gen msdate10=month(dasdate10)
gen medate10=month(daedate10)
gen msdate11=month(dasdate11)
gen medate11=month(daedate11)
gen ysdate1=real(asyear1)
gen yedate1=real(aeyear1)
gen ysdate2=real(asyear2)
gen yedate2=real(aeyear2)
gen ysdate3=real(asyear3)
gen yedate3=real(aeyear3)
gen ysdate4=real(asyear4)
gen yedate4=real(aeyear4)
gen ysdate5=real(asyear5)
gen yedate5=real(aeyear5)
gen ysdate6=real(asyear6)
gen yedate6=real(aeyear6)
gen ysdate7=real(asyear7)
gen yedate7=real(aeyear7)
gen ysdate8=real(asyear8)
gen yedate8=real(aeyear8)
gen ysdate9=real(asyear9)
gen yedate9=real(aeyear9)
gen ysdate10=real(asyear10)
gen yedate10=real(aeyear10)
gen ysdate11=real(asyear11)
gen yedate11=real(aeyear11)
*/
gen pcgempl=1
replace pcgempl=0 if (missing(nemprec))
label var pcgempl "Ever employed before W1"
keep sampleid pid pcgempl stater* cntryr* asmth* aemth* asyear* aeyear*
save C:\Users\ahicks\Documents\Stata_data\lafans\ehc1_employ.dta, replace

***** merge Wavel employment variables *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta",
clear
drop _merge

```

```

merge m:1 sampleid pid using
"C:\Users\ahicks\Documents\Stata_data\lafans\ehc1_employ.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta,
replace

/*
***** sort residential not pcg variables *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta",
clear
replace stater2=pc7_str if (pc4==5 | pc5==5)
replace cntryr2=pc7_cyr if (pc4==5 | pc5==5)
replace asmth2=pc9_m if (pc4==5 | pc5==5)
replace ysdate2=pc9_y if (pc4==5 | pc5==5)
replace aemth2=pc6_m if (pc4==5 | pc5==5)
replace yedate2=pc6_y if (pc4==5 | pc5==5)
replace stater3=pc11_str if (pc4==5 | pc5==5)
replace cntryr3=pc11_cyr if (pc4==5 | pc5==5)
replace asmth3=pc13_m if (pc4==5 | pc5==5)
replace ysdate3=pc13_y if (pc4==5 | pc5==5)
replace msdate3=pc9_m if (pc4==5 | pc5==5)
replace yedate3=pc9_y if (pc4==5 | pc5==5)
*if sib lives with rsc (pc5==1) replace get data from rsc

*/

***** generate wave dummy variables *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta",
clear
gen wave=1
save C:\Users\ahicks\Documents\Stata_data\lafans\WAVE1preschool.dta,
replace

***** Create Preschool Wave1 *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wavelpreschool.dta",
clear
keep if (page>=3 & page<=17)
keep sampleid hhid pid childid kidtype age_a yrborn page lwi_ss ap_ss
bpi* hsattend ccattend nonparentcare pg43_m-pg45 pg50 pg51 pg69 pg70*
pg55a_m-pg57b pg62a-pg63b pg73 pg74* gc5-gc9 ph11-ph13 ph27-ph29 ph44-
ph47 ph64 ph65 hof1-hof21 faminc d_house pcgempl ge2* ge3 pcg_edu
momreading* child* pcg* wave pc7* pc9* pc11* pc13* pc4 pc5 pf15 pf16*
pf11* pf12-pf14_c pf20_a pf20_b pf4_a hsmmonths ccmonths povcat waveldate
pk5_lb pk5_oz pk5_k pk5_a rpcgs age_yr caseness probcase gd* hhedu pe49
pe1_rev ael married_date born_date pe9 pe8 pe3 mom_pcg
drop if missing(hhid)
drop if missing(childid)
drop if missing(pid)
gen notcur_hsmmonths=pg70_b+(pg70_c*12)
replace hsmmonths=notcur_hsmmonths if hsmmonth==.
gen notcur_ccmonths=pg74_b+(pg74_c*12)
replace ccmonths=notcur_ccmonths if ccmonths==.
replace ccmonth=. if ccattend==0

```

```

*sort hhid childid
*quietly by hhid childid: gen dup = cond(_N==1,0,_n)
*tab dup
*list hhid pid childid if dup>0
save C:\Users\ahicks\Documents\Stata_data\lafans\Wave1preschool.dta,
replace

*****
*****
***** add wave 2 variables *****
*****
*****

***** sort parent2 *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\parent2.dta", clear
sort hhid2 kidtype
rename pd3_a childcitizen
rename pd3_s child_otherstatus
replace childcitizen=1 if pd2==1
gen wave2date=date(comp_date, "MDY")
keep hhid2 kidtype chldage pid child_pid w1_hhid w1_pid w1_child_pid
childcitizen child_otherstatus bpi* pf7* pf12-pf14c pf20a pf20b pf17-pf19
pf15 pf16* wave2date child_sex hhdadid
save C:\Users\ahicks\Documents\Stata_data\lafans\parent2_w2sorted.dta,
replace

***** sort child assess2 *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\wjrchild2.dta", clear
rename resp_type kidtype
sort hhid2 kidtype
keep hhid2 kidtype lwi_ss ap_ss resp_age ap_w lwi_w
save C:\Users\ahicks\Documents\Stata_data\lafans\wjrchild2_sorted.dta,
replace

***** merge parent and assess *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\wjrchild2_sorted.dta",
clear
merge 1:1 hhid2 kidtype using
"C:\Users\ahicks\Documents\Stata_data\lafans\parent2_w2sorted.dta"
replace kidtype="RSC" if kidtype=="RSC_P"
replace kidtype="SIB" if kidtype=="SIB_P"
drop if (kidtype=="OTH" | kidtype=="RSC_NE" | kidtype=="SIB_NE")
save C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta,
replace

***** sort adult +18 *****
*****
use "C:\Users\ahicks\Documents\Stata_data\lafans\adult2.dta", clear
gen wave2date=date(comp_date, "MDY")
keep hhid2 pid2 resp_type w1_hhid w1_pid ac34* ac38-ac42 wave2date
resp_sex

```



```

rename resp_sex child_sex
rename resp_type kidtype
rename ac38 childcitizen
rename ac39 childgreencard
rename ac40 childasylum
rename ac41 childvisa
rename ac42 childvisa_valid
replace childcitizen=1 if ac34_4==1
keep if (kidtype=="RSC18+" | kidtype=="SIB18+")
replace kidtype="RSC" if kidtype=="RSC18+"
replace kidtype="SIB" if kidtype=="SIB18+"
sort hhid2 kidtype
save C:\Users\ahicks\Documents\Stata_data\lafans\adult2_18.dta, replace

```

```

***** sort adult 18+ assess2 *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\wjradult2.dta", clear
rename resp_type kidtype
keep if (kidtype=="RSC18+" | kidtype=="SIB18+")
replace kidtype="RSC" if kidtype=="RSC18+"
replace kidtype="SIB" if kidtype=="SIB18+"
sort hhid2 kidtype
keep hhid2 kidtype lwi_ss ap_ss resp_age ap_w lwi_w
gen adulttag=1
save C:\Users\ahicks\Documents\Stata_data\lafans\wjradult2_sorted.dta,
replace

```

```

***** merge adult and assess *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\wjradult2_sorted.dta",
clear
merge 1:1 hhid2 kidtype using
"C:\Users\ahicks\Documents\Stata_data\lafans\adult2_18.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool_18.dta,
replace

```

```

***** append parent2 and adult 18+ *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta",
clear
append using
"C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool_18.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta,
replace

```

```

***** sort kid roster1 *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\roster2.dta", clear
rename w2_resptype kidtype
replace kidtype="RSC" if kidtype=="RSC18+"
replace kidtype="SIB" if kidtype=="SIB18+"
replace kidtype="RSC" if kidtype=="RSC_P"
replace kidtype="SIB" if kidtype=="SIB_P"
sort hhid2 kidtype
keep hhid2 kidtype rb1 hhrf w1_age pid2
rename pid2 child_pid2
rename rb1 hsgrad

```

```

drop if missing(hhid2)
drop if missing(kidtype)
save C:\Users\ahicks\Documents\Stata_data\lafans\roster2_childsorted.dta,
replace

```

```

***** add kid roster data to wavel *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta",
clear
drop _merge
merge m:1 hhid2 kidtype using
"C:\Users\ahicks\Documents\Stata_data\lafans\roster2_childsorted.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta,
replace

```

```

***** sort immigration adult1 *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\adult2.dta", clear
keep if pcg==1
keep hhid2 pid2 ac34_4 ac38-ac42 ae100 ae1
rename ac34_4 pcgborn
rename ac38 pcgcitizen
rename ac39 pcggreencard
rename ac40 pcgasylum
rename ac41 pcgvisa
rename ac42 pcgvisa_valid
replace pcgcitizen=1 if pcgborn==1
gen hasadultrecord=1
sort hhid2
drop if missing(hhid2)
sort hhid2
quietly by hhid2: gen dup = cond(_N==1,0,_n)
drop if dup>0
save C:\Users\ahicks\Documents\Stata_data\lafans\adult2_immigr.dta,
replace

```

```

***** add pcg immigration to wavel *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta",
clear
drop _merge
merge m:1 hhid2 using
"C:\Users\ahicks\Documents\Stata_data\lafans\adult2_immigr.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta,
replace

```

```

***** sort income variables *****
use
"C:\Users\ahicks\Documents\Stata_data\lafans\Restricted\Wave2\VER1\impinc
2r.dta", clear
sort hhid2
keep faminc hhid2 hhrf d_house

```

```
save C:\Users\ahicks\Documents\Stata_data\lafans\inpinc2_sorted.dta,  
replace
```

```
***** merge Wave1 income variables *****  
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta",  
clear  
drop _merge  
merge m:1 hhid2 hhrf using  
"C:\Users\ahicks\Documents\Stata_data\lafans\inpinc2_sorted.dta"  
save C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta,  
replace
```

```
***** make Wave 1 employment variables *****  
use "C:\Users\ahicks\Documents\Stata_data\lafans\ehcresp2.dta", clear  
/*  
gen dasdate1 = date(asdate1,"YMD")  
gen daedate1 = date(aedate1,"YMD")  
gen dasdate2 = date(asdate2,"YMD")  
gen daedate2 = date(aedate2,"YMD")  
gen dasdate3 = date(asdate3,"YMD")  
gen daedate3 = date(aedate3,"YMD")  
gen dasdate4 = date(asdate4,"YMD")  
gen daedate4 = date(aedate4,"YMD")  
gen dasdate5 = date(asdate5,"YMD")  
gen daedate5 = date(aedate5,"YMD")  
gen dasdate6 = date(asdate6,"YMD")  
gen daedate6 = date(aedate6,"YMD")  
gen dasdate7 = date(asdate7,"YMD")  
gen daedate7 = date(aedate7,"YMD")  
gen dasdate8 = date(asdate8,"YMD")  
gen daedate8 = date(aedate8,"YMD")  
gen dasdate9 = date(asdate9,"YMD")  
gen daedate9 = date(aedate9,"YMD")  
gen dasdate10 = date(asdate10,"YMD")  
gen daedate10 = date(aedate10,"YMD")  
gen dasdate11 = date(asdate11,"YMD")  
gen daedate11 = date(aedate11,"YMD")  
gen msdate1=month(dasdate1)  
gen medate1=month(daedate1)  
gen msdate2=month(dasdate2)  
gen medate2=month(daedate2)  
gen msdate3=month(dasdate3)  
gen medate3=month(daedate3)  
gen msdate4=month(dasdate4)  
gen medate4=month(daedate4)  
gen msdate5=month(dasdate5)  
gen medate5=month(daedate5)  
gen msdate6=month(dasdate6)  
gen medate6=month(daedate6)  
gen msdate7=month(dasdate7)  
gen medate7=month(daedate7)  
gen msdate8=month(dasdate8)  
gen medate8=month(daedate8)
```

```

gen msdate9=month(dasdate9)
gen medate9=month(daedate9)
gen msdate10=month(dasdate10)
gen medate10=month(daedate10)
gen msdate11=month(dasdate11)
gen medate11=month(daedate11)
gen ysdate1=real(asyear1)
gen yedate1=real(aeyear1)
gen ysdate2=real(asyear2)
gen yedate2=real(aeyear2)
gen ysdate3=real(asyear3)
gen yedate3=real(aeyear3)
gen ysdate4=real(asyear4)
gen yedate4=real(aeyear4)
gen ysdate5=real(asyear5)
gen yedate5=real(aeyear5)
gen ysdate6=real(asyear6)
gen yedate6=real(aeyear6)
gen ysdate7=real(asyear7)
gen yedate7=real(aeyear7)
gen ysdate8=real(asyear8)
gen yedate8=real(aeyear8)
gen ysdate9=real(asyear9)
gen yedate9=real(aeyear9)
gen ysdate10=real(asyear10)
gen yedate10=real(aeyear10)
gen ysdate11=real(asyear11)
gen yedate11=real(aeyear11)
*/
gen pcgempl =1
replace pcgempl=0 if nemploy==0
gen pcgunempl =1
replace pcgunempl=0 if nunemp==0
*keep if resp_type=="PCG_P"
keep hhid2 pid pcgempl pcgunempl
save C:\Users\ahicks\Documents\Stata_data\lafans\ehcresp2_empl.dta,
replace
/*
***** merge Wave1 employment variables *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta",
clear
drop _merge
merge m:1 hhid2 using
"C:\Users\ahicks\Documents\Stata_data\lafans\ehcresp2_empl.dta"
sort hhid2 kidtype
drop dup
quietly by hhid2 kidtype: gen dup = cond(_N==1,0,_n)
drop if dup>0
save C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta,
replace

***** find number of times lost job *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\unemp2.dta", clear
recode reason_une_n (1=1) (else=.)

```

```

collapse (count) nlostjob=reason_une_n, by(hhid2 pid2)
label var nlostjob "Number of time PCG lost job between W1 and W2"
save C:\Users\ahicks\Documents\Stata_data\lafans\lostjob.dta, replace

***** merge Wave1 employment variables *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta",
clear
drop _merge
merge m:1 hhid2 pid using
"C:\Users\ahicks\Documents\Stata_data\lafans\lostjob.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta,
replace
*/
***** merge Wave1 employment variables *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta",
clear
drop _merge
merge m:1 hhid2 pid using
"C:\Users\ahicks\Documents\Stata_data\lafans\ehcresp2_empl.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta,
replace

***** find number of times lost job *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\unemp2.dta", clear
recode reason_une_n (1=1) (else=.)
collapse (count) nlostjob=reason_une_n, by(hhid2 pid2)
label var nlostjob "Number of time PCG lost job between W1 and W2"
save C:\Users\ahicks\Documents\Stata_data\lafans\lostjob.dta, replace

***** merge Wave1 employment variables *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta",
clear
drop _merge
merge m:1 hhid2 pid2 using
"C:\Users\ahicks\Documents\Stata_data\lafans\lostjob.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta,
replace

**** get pcg2 reported partner/spouse relations *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\pcg2.dta", clear
sort hhid2 pid2
keep hhid2 pid2 ge2* ge3* gd*
quietly by hhid2: gen dup = cond(_N==1,0,_n)
drop if dup>0
save C:\Users\ahicks\Documents\Stata_data\lafans\pcg2_sorted.dta, replace

***** merge pcg2 variables *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta",
clear
drop _merge
merge m:1 hhid2 using
"C:\Users\ahicks\Documents\Stata_data\lafans\pcg2_sorted.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta,
replace

```

```

/*
***** sort residential not pcg variables *****
use "C:\Users\ahicks\Documents\Stata_data\lafans1dta\Wave1.dta", clear
replace stater2=pc7_str if (pc4==5 | pc5==5)
replace cntryr2=pc7_cyr if (pc4==5 | pc5==5)
replace asmth2=pc9_m if (pc4==5 | pc5==5)
replace ysdate2=pc9_y if (pc4==5 | pc5==5)
replace aemth2=pc6_m if (pc4==5 | pc5==5)
replace yedate2=pc6_y if (pc4==5 | pc5==5)
replace stater3=pc11_str if (pc4==5 | pc5==5)
replace cntryr3=pc11_cyr if (pc4==5 | pc5==5)
replace asmth3=pc13_m if (pc4==5 | pc5==5)
replace ysdate3=pc13_y if (pc4==5 | pc5==5)
replace msdate3=pc9_m if (pc4==5 | pc5==5)
replace yedate3=pc9_y if (pc4==5 | pc5==5)
*if sib lives with rsc (pc5==1) replace get data from rsc

***** sort schools2 *****
use
"C:\Users\ahicks\Documents\Stata_data\lafans\Restricted\Wave2\VER25\parv2
5_2.dta", clear
replace kidtype="RSC" if kidtype=="RSC_P"
replace kidtype="SIB" if kidtype=="SIB_P"
sort hhid2 kidtype
keep hhid2 kidtype pf5* pf7*
save C:\Users\ahicks\Documents\Stata_data\lafans\schools2_sorted.dta,
replace

***** merge schools *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta",
clear
drop dup
drop if missing(hhid2)
drop if missing(kidtype)
sort hhid2 kidtype
quietly by hhid2 kidtype: gen dup = cond(_N==1,0,_n)
tab dup
replace resp_age=16 if (hhid2=="14781A" & kidtype=="SIB" & dup==2)
replace ap_ss=100 if (hhid2=="14781A" & kidtype=="SIB" & dup==2)
replace lwi_ss=94 if (hhid2=="14781A" & kidtype=="SIB" & dup==2)
drop if(hhid2=="14781A" & kidtype=="SIB" & dup==1)
replace resp_age=17 if (hhid2=="51377A" & kidtype=="RSC" & dup==2)
replace ap_ss=100 if (hhid2=="51377A" & kidtype=="RSC" & dup==2)
replace lwi_ss=94 if (hhid2=="51377A" & kidtype=="RSC" & dup==2)
drop if(hhid2=="51377A" & kidtype=="RSC" & dup==1)
drop _merge
merge 1:1 hhid2 kidtype using
"C:\Users\ahicks\Documents\Stata_data\lafans\schools2_sorted.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta,
replace

```

```

*/
***** merge schools *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta",
clear
drop _merge
drop dup
replace kidtype="RSC_P" if kidtype=="RSC"
replace kidtype="SIB_P" if kidtype=="SIB"
sort hhid2 kidtype
quietly by hhid2 kidtype: gen dup = cond(_N==1,0,_n)
drop if dup>0
merge 1:1 hhid2 kidtype using
"C:\Users\ahicks\Documents\Stata_data\lafans\AverageSchooling.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta,
replace

***** merge residential moves 33 day gap
*****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta",
clear
drop _merge
sort hhid2 kidtype
merge 1:1 hhid2 kidtype using
"C:\Users\ahicks\Documents\Stata_data\lafans\resid_collapse.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta,
replace

***** merge residential history *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta",
clear
drop _merge
sort hhid2 kidtype
merge 1:1 hhid2 kidtype using
"C:\Users\ahicks\Documents\Stata_data\lafans\Restricted\Wave2\VER2\reside
ntial_history.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta,
replace

***** merge residential moves 8 day gap
*****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta",
clear
drop _merge
sort hhid2 kidtype
merge 1:1 hhid2 kidtype using
"C:\Users\ahicks\Documents\Stata_data\lafans\resid_collapse8.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta,
replace

***** merge household welfare *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta",
clear
drop _merge

```

```

merge m:1 hhid2 pid2 using
"C:\Users\ahicks\Documents\Stata_data\lafans\welfare.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta,
replace

***** merge pcg employment time varying *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta",
clear
drop _merge
merge m:1 hhid2 pid2 using
"C:\Users\ahicks\Documents\Stata_data\lafans\pcgemploy.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta,
replace

/*
***** merge dad employment time varying *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta",
clear
drop _merge
merge m:1 hhid2 hhdadid using
"C:\Users\ahicks\Documents\Stata_data\lafans\pcgemploy_dad.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta,
replace
*/

***** merge marriage status time varying *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta",
clear
drop _merge
merge m:1 hhid2 pid2 using
"C:\Users\ahicks\Documents\Stata_data\lafans\mar_time_var.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta,
replace

***** merge number children time varying *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta",
clear
drop _merge
merge m:1 hhid2 pid2 using
"C:\Users\ahicks\Documents\Stata_data\lafans\numchildren.dta"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta,
replace

***** create wave2 dummy variable and suffix *****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta",
clear
gen wave=2
keep hhid2 kidtype w1_age w1_hhid w1_pid w1_child_pid lwi_ss ap_ss ap_w
lwi_w pf15 pf16* hsgrad child_pid2 pf13-pf15 pf17-pf19 pf20* faminc
d_house pcgempl pcgunempl childcitizen childvisa child_otherstatus
adulttag bpi* childage pf12-pf14c pf15-pf16c average_api residmoves_33gap
nlostjob ge2* ge3 pcgborn pcgcitizen pcggreencard pcg asylum pcgvisa
pcgvisa_valid wave2date child_sex gd* welfare* pcgemploy* salary2*

```



```

salaryunit* emplrtyp2* hrswk2* addr* fromparent* numchild* moves* mar2*
hhdadid adulttag ael100 hasadultrecord ael
rename * W2_=
rename W2_kidtype kidtype
rename W2_w1_hhid hhid
rename W2_w1_pid pid
rename W2_w1_child_pid childid
rename W2_residmoves_33gap residmoves
drop if missing(hhid)
drop if missing(kidtype)
sort hhid kidtype
quietly by hhid kidtype: gen dup = cond(_N==1,0,_n)
tab dup
drop if dup>1
replace kidtype="RSC" if kidtype=="RSC_P"
replace kidtype="SIB" if kidtype=="SIB_P"
save C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta,
replace

```

```

**** merge with Wave1 0-6 ****
use "C:\Users\ahicks\Documents\Stata_data\lafans\Wave2Preschool.dta",
clear
merge 1:1 hhid kidtype using
"C:\Users\ahicks\Documents\Stata_data\lafans\Wave1preschool.dta"
tab W2_w1_age if _merge==2
keep if _merge==3

```

```
label var pf4_a "W1: grade attended last spring"
```

```

label var hsmmonths "Months attended Headstart"
label var ccmonths "Months attended Childcare"
rename W2_hsgrad W2_grade
gen W2_hsgrad=0
replace W2_hsgrad = 1 if W2_grade>=12
label var W2_hsgrad "did child graduate from high school"

```

```

drop W2_pf16a1
rename W2_average_api api_ave
rename childcitizen pd3_a
rename childvisa pd3a
rename W2_childcitizen W2_pd3_a
rename W2_childvisa W2_pd3a
rename W2_wave2date wave2date
rename W2_child_sex child_sex
recode W2_nlostjob (.=0)

```

```

keep W2_hhid2 kidtype page age_a yrborn lwi_ss ap_ss bpi* hsattend
hsmmonths ccattend ccmonths gc5-gc9 ph11-ph13 ph27-ph29 ph44-ph47 ph64
ph65 hof1-hof21 pf15 pf16* ///
pf11* pf12-pf14_c pf20* W2_lwi_ss W2_ap_ss W2_ap_w W2_lwi_w W2_hsgrad
W2_child_pid2 W2_ael100 W2_hasadultrecord W2_ael W2_bpi* W2_pf12-W2_pf14c
W2_pf17-W2_pf20b api_ave residmoves ///

```

```
faminc d_house W2_faminc W2_d_house pcgempl W2_pcgempl W2_nlostjob ge2a-
ge3 W2_ge2a-W2_ge3 pd3_a pd3a W2_pd3a W2_pd3_a child_race* pcgborn
pcgcitizen ///
```

```
pcggreencard pcg asylum pcgvisa pcgvisa_valid W2_pcg* pcg_race*
momreading* pcg_edu W2_pf15 W2_pf16 W2_pf16a W2_pf16b pf4_a W2_grade
W2_pf16c povcat waveldate wave2date child_sex pk5_lb pk5_oz pk5_k pk5_a
rpcgs age_yr caseness probcase gd* W2_gd* hhedu pe49 mom_pcg W2_welfare*
W2_addr* W2_fromparent* W2_moves* pel_rev ael married_date born_date pe9
pe8 pe3 W2_salary* W2_emplrt* W2_hrswk2* W2_numchild* W2_mar2* W2_hhdadid
W2_adulttag
```

```
drop W2_pcgunempl

label define citizen 1 "US citizen" 2 "Permanent Resident (Green Card)" 3
"VISA" 4 "Does not have papers" 5 "Other status"
label values pd3_a W2_pd3_a citizen
```

```
label define fives 1 "yes" 5 "no" 2 "expired"
label values pcgcitizen pcggreencard pcg asylum pcgvisa pcgvisa_valid
W2_pcgcitizen W2_pcggreencard W2_pcg asylum W2_pcgvisa W2_pcgvisa_valid
fives
format waveldate wave2date %d
```

```
gen childrace=0
replace childrace=1 if child_race_1==1
replace childrace=2 if child_race_2==1
replace childrace=3 if child_race_3==1
replace childrace=4 if child_race_4==1
replace childrace=5 if child_race_5==1
replace childrace=6 if child_race_6==1
lab def race 1 "Latino" 2 "White" 3 "Black" 4 "Asian" 5 "Pacific
Islander" 6 "Native American"
lab val childrace race
```

```
gen pcgrace=0
replace pcgrace=1 if pcg_race_1==1
replace pcgrace=2 if pcg_race_2==1
replace pcgrace=3 if pcg_race_3==1
replace pcgrace=4 if pcg_race_4==1
replace pcgrace=5 if pcg_race_5==1
replace pcgrace=6 if pcg_race_6==1
lab val pcgrace race
```

```
recode W2_welfare* (.=0)
*recode W2_pcgemploy* (.=0)
```

```
destring W2_addr1998, replace
destring W2_addr1999, replace
destring W2_addr2000, replace
destring W2_addr2001, replace
destring W2_addr2002, replace
destring W2_addr2003, replace
destring W2_addr2004, replace
```

```

destring W2_addr2005, replace
destring W2_addr2006, replace
destring W2_addr2007, replace
destring W2_addr2008, replace

*merge disadvantage quintile onto 1998 census tracts
sort W2_addr1998
rename W2_addr1998 geoid90
merge n:1 geoid90 using
"C:\Users\ahicks\Documents\Stata_data\lafans\Neighborhood\Disadvantage_Scores1998.dta"
drop if _merge==2
drop _merge
rename concdis_quintile concdis_quintile1998
rename concdis concdis_1998
rename geoid90 W2_addr1998

*merge disadvantage quintile onto 1999 census tracts
sort W2_addr1999
rename W2_addr1999 geoid90
merge n:1 geoid90 using
"C:\Users\ahicks\Documents\Stata_data\lafans\Neighborhood\Disadvantage_Scores1999.dta"
drop if _merge==2
drop _merge
rename concdis_quintile concdis_quintile1999
rename concdis concdis_1999
rename geoid90 W2_addr1999

*merge disadvantage quintile onto 2000 census tracts
sort W2_addr2000
rename W2_addr2000 geoid90
merge n:1 geoid90 using
"C:\Users\ahicks\Documents\Stata_data\lafans\Neighborhood\Disadvantage_Scores2000.dta"
drop if _merge==2
drop _merge
rename concdis_quintile concdis_quintile2000
rename concdis concdis_2000
rename geoid90 W2_addr2000

*merge disadvantage quintile onto 2001 census tracts
sort W2_addr2001
rename W2_addr2001 geoid90
merge n:1 geoid90 using
"C:\Users\ahicks\Documents\Stata_data\lafans\Neighborhood\Disadvantage_Scores2001.dta"
drop if _merge==2
drop _merge
rename concdis_quintile concdis_quintile2001
rename concdis concdis_2001
rename geoid90 W2_addr2001

*merge disadvantage quintile onto 2002 census tracts

```

```

sort W2_addr2002
rename W2_addr2002 geoid90
merge n:1 geoid90 using
"C:\Users\ahicks\Documents\Stata_data\lafans\Neighborhood\Disadvantage_Sc
ores2002.dta"
drop if _merge==2
drop _merge
rename concdis_quintile concdis_quintile2002
rename concdis concdis_2002
rename geoid90 W2_addr2002

*merge disadvantage quintile onto 2003 census tracts
sort W2_addr2003
rename W2_addr2003 geoid90
merge n:1 geoid90 using
"C:\Users\ahicks\Documents\Stata_data\lafans\Neighborhood\Disadvantage_Sc
ores2003.dta"
drop if _merge==2
drop _merge
rename concdis_quintile concdis_quintile2003
rename concdis concdis_2003
rename geoid90 W2_addr2003

*merge disadvantage quintile onto 2004 census tracts
sort W2_addr2004
rename W2_addr2004 geoid90
merge n:1 geoid90 using
"C:\Users\ahicks\Documents\Stata_data\lafans\Neighborhood\Disadvantage_Sc
ores2004.dta"
drop if _merge==2
drop _merge
rename concdis_quintile concdis_quintile2004
rename concdis concdis_2004
rename geoid90 W2_addr2004

*merge disadvantage quintile onto 2005 census tracts
sort W2_addr2005
rename W2_addr2005 geoid90
merge n:1 geoid90 using
"C:\Users\ahicks\Documents\Stata_data\lafans\Neighborhood\Disadvantage_Sc
ores2005.dta"
drop if _merge==2
drop _merge
rename concdis_quintile concdis_quintile2005
rename concdis concdis_2005
rename geoid90 W2_addr2005

*merge disadvantage quintile onto 2006 census tracts
sort W2_addr2006
rename W2_addr2006 geoid90
merge n:1 geoid90 using
"C:\Users\ahicks\Documents\Stata_data\lafans\Neighborhood\Disadvantage_Sc
ores2006.dta"
drop if _merge==2

```

```

drop _merge
rename concdis_quintile concdis_quintile2006
rename concdis concdis_2006
rename geoid90 W2_addr2006

*merge disadvantage quintile onto 2007 census tracts
sort W2_addr2007
rename W2_addr2007 geoid90
merge n:1 geoid90 using
"C:\Users\ahicks\Documents\Stata_data\lafans\Neighborhood\Disadvantage_Scores2007.dta"
drop if _merge==2
drop _merge
rename concdis_quintile concdis_quintile2007
rename concdis concdis_2007
rename geoid90 W2_addr2007

*merge disadvantage quintile onto 2008 census tracts
sort W2_addr2008
rename W2_addr2008 geoid90
merge n:1 geoid90 using
"C:\Users\ahicks\Documents\Stata_data\lafans\Neighborhood\Disadvantage_Scores2008.dta"
drop if _merge==2
drop _merge
rename concdis_quintile concdis_quintile2008
rename concdis concdis_2008
rename geoid90 W2_addr2008

/*
*merge affluence quintile onto 1998 census tracts
sort W2_addr1998
rename W2_addr1998 geoid90
merge n:1 geoid90 using
"C:\Users\ahicks\Documents\Stata_data\lafans\Neighborhood\Affluence_Score1998.dta"
drop if _merge==2
drop _merge
rename concafl_quintile concafl_quintile1998
rename geoid90 W2_addr1998

*merge affluence quintile onto 1999 census tracts
sort W2_addr1999
rename W2_addr1999 geoid90
merge n:1 geoid90 using
"C:\Users\ahicks\Documents\Stata_data\lafans\Neighborhood\Affluence_Score1999.dta"
drop if _merge==2
drop _merge
rename concafl_quintile concafl_quintile1999
rename geoid90 W2_addr1999

*merge affluence quintile onto 2000 census tracts
sort W2_addr2000

```

```
rename W2_addr2000 geoid90
merge n:1 geoid90 using
"C:\Users\ahicks\Documents\Stata_data\lafans\Neighborhood\Affluence_Score
s2000.dta"
drop if _merge==2
drop _merge
rename concafl_quintile concafl_quintile2000
rename geoid90 W2_addr2000
```

```
*merge affluence quintile onto 2001 census tracts
sort W2_addr2001
rename W2_addr2001 geoid90
merge n:1 geoid90 using
"C:\Users\ahicks\Documents\Stata_data\lafans\Neighborhood\Affluence_Score
s2001.dta"
drop if _merge==2
drop _merge
rename concafl_quintile concafl_quintile2001
rename geoid90 W2_addr2001
```

```
*merge affluence quintile onto 2002 census tracts
sort W2_addr2002
rename W2_addr2002 geoid90
merge n:1 geoid90 using
"C:\Users\ahicks\Documents\Stata_data\lafans\Neighborhood\Affluence_Score
s2002.dta"
drop if _merge==2
drop _merge
rename concafl_quintile concafl_quintile2002
rename geoid90 W2_addr2002
```

```
*merge affluence quintile onto 2003 census tracts
sort W2_addr2003
rename W2_addr2003 geoid90
merge n:1 geoid90 using
"C:\Users\ahicks\Documents\Stata_data\lafans\Neighborhood\Affluence_Score
s2003.dta"
drop if _merge==2
drop _merge
rename concafl_quintile concafl_quintile2003
rename geoid90 W2_addr2003
```

```
*merge affluence quintile onto 2004 census tracts
sort W2_addr2004
rename W2_addr2004 geoid90
merge n:1 geoid90 using
"C:\Users\ahicks\Documents\Stata_data\lafans\Neighborhood\Affluence_Score
s2004.dta"
drop if _merge==2
drop _merge
rename concafl_quintile concafl_quintile2004
rename geoid90 W2_addr2004
```

```
*merge affluence quintile onto 2005 census tracts
```

```

sort W2_addr2005
rename W2_addr2005 geoid90
merge n:1 geoid90 using
"C:\Users\ahicks\Documents\Stata_data\lafans\Neighborhood\Affluence_Score
s2005.dta"
drop if _merge==2
drop _merge
rename concafl Quintile concafl_Quintile2005
rename geoid90 W2_addr2005

*merge affluence Quintile onto 2006 census tracts
sort W2_addr2006
rename W2_addr2006 geoid90
merge n:1 geoid90 using
"C:\Users\ahicks\Documents\Stata_data\lafans\Neighborhood\Affluence_Score
s2006.dta"
drop if _merge==2
drop _merge
rename concafl_Quintile concafl_Quintile2006
rename geoid90 W2_addr2006

*merge affluence Quintile onto 2007 census tracts
sort W2_addr2007
rename W2_addr2007 geoid90
merge n:1 geoid90 using
"C:\Users\ahicks\Documents\Stata_data\lafans\Neighborhood\Affluence_Score
s2007.dta"
drop if _merge==2
drop _merge
rename concafl_Quintile concafl_Quintile2007
rename geoid90 W2_addr2007

*merge affluence Quintile onto 2008 census tracts
sort W2_addr2008
rename W2_addr2008 geoid90
merge n:1 geoid90 using
"C:\Users\ahicks\Documents\Stata_data\lafans\Neighborhood\Affluence_Score
s2008.dta"
drop if _merge==2
drop _merge
rename concafl_Quintile concafl_Quintile2008
rename geoid90 W2_addr2008
*/

*create birth married variable
gen birth_married=.
* mother and father in HH & currently married & married before birth
replace birth_married=1 if pel_rev==1 & ael==1 & married_date<=born_date
* mother and father in HH & currently married & married after birth
replace birth_married=0 if pel_rev==1 & ael==1 & married_date>born_date
* mother and father in HH & never married
replace birth_married=0 if pel_rev==1 & ael==5

```

```

* figure out what to do if separated (19) widowed (6) divorced (4)
* (I think we have no information)
* child lives with mom only, no dad in HH, mom never married
replace birth_married=0 if pel_rev==2 & pe8==1
* child lives with mom only, no dad in HH, was married to biodad, married
before born
replace birth_married=1 if pel_rev==2 & pe8==2 & pe9==1 &
married_date<=born_date
* child lives with mom only, no dad in HH, was married to biodad, married
AFTER born
replace birth_married=0 if pel_rev==2 & pe8==2 & pe9==1 &
married_date>born_date
* child lives with mom only, no dad in HH, not married to biodad
replace birth_married=0 if pel_rev==2 & pe8==2 & pe9==2
* child lives with mom only, no dad in HH, child adopted
replace birth_married=0 if pel_rev==2 & pe8==2 & pe9==3

*create mom age at birth
gen age_birth=.
replace age_birth=pe49 if mom_pcg>1
replace age_birth=age_yr-page if mom_pcg==1
recode age_birth (-9 -8=.)

*labels employer type
lab def emtype 1 "Government" 2 "Private company" 3 "Self-employed" 4
"Work without pay" 9 "Unemployed"
lab val W2_emplrtyp* emtype

sort W2_hhid2 kidtype

save
C:\Users\ahicks\Documents\Stata_data\lafans\Longitudinal_Modeling.dta,
replace

*****
* Income interpolation *
*****
use
C:\Users\ahicks\Documents\Stata_data\lafans\Longitudinal_Modeling.dta,
clear
keep W2_hhid2 kidtype faminc W2_faminc wave2date waveldate
gen wavelyear=year(waveldate)
gen wave2year=year(wave2date)

gen faminc_2000=.
gen faminc_2001=.
gen faminc_2002=.
gen faminc_2003=.
gen faminc_2004=.
gen faminc_2005=.
gen faminc_2006=.
gen faminc_2007=.

```



```

gen faminc_2008=.

replace faminc_2000=faminc if wavelyear==2000
replace faminc_2001=faminc if wavelyear==2001
replace faminc_2002=faminc if wavelyear==2002

replace faminc_2006=W2_faminc if wave2year==2006
replace faminc_2007=W2_faminc if wave2year==2007
replace faminc_2008=W2_faminc if wave2year==2008

reshape long faminc_, i(W2_hhid2 kidtype) j(year)
by W2_hhid2 kidtype: ipolate faminc_ year, gen (ipolfaminc_)
reshape wide ipolfaminc_ faminc_, i(W2_hhid2 kidtype) j(year)
keep W2_hhid2 kidtype ipolfaminc_*
merge 1:1 W2_hhid2 kidtype using
"C:\Users\ahicks\Documents\Stata_data\lafans\Longitudinal_Modeling.dta"

*set marriage status to missing if no adult record is found
forval x=0/8 {
replace W2_mar200`x'=. if W2_hasadultrecord==.
}

*set pcg salary to 0 if pcg is unemployed
forval x=0/8 {
replace W2_salary200`x'=0 if W2_pcgemploy200`x'==0
}

*set pcg type to 9 if pcg is unemployed
forval x=0/8 {
replace W2_emplrtyp200`x'=9 if W2_pcgemploy200`x'==0
}

*set hours/wk to 0 if pcg is unemployed
forval x=0/8 {
replace W2_hrswk200`x'=0 if W2_pcgemploy200`x'==0
}

* replace marriage from roster and update roster if marriage status is
missing
drop _merge
merge n:1 W2_hhid2 using
"C:\Users\ahicks\Documents\Stata_data\lafans\update_marriage.dta"
drop if _merge==2
forval x=0/8 {
replace W2_mar200`x'=1 if update_marstat==1 & W2_mar200`x'==.
replace W2_mar200`x'=0 if update_marstat==0 & W2_mar200`x'==.
}

* Add panel RSC/SIB weights
rename W2_child_pid2 pid2
rename W2_hhid2 hhid2
drop _merge

```

```

merge 1:1 hhid2 pid2 using
"C:\Users\ahicks\Documents\Stata_data\lafans\indivwgts2.dta",
keepusing(pwgt_rscsib)

save
C:\Users\ahicks\Documents\Stata_data\lafans\Longitudinal_Modeling.dta,
replace

*****
*****
*****
*****
*           Cleaning
*****
*****
*****
*****
use
"C:\Users\ahicks\Documents\Stata_data\lafans\Longitudinal_Modeling.dta",
clear

*UNDER 18 IN WAVE 2
drop if W2_adulttag==1
drop if W2_hasadultrec==.

quietly: tab childrace
replace childrace=. if childrace==0
quietly: tab childrace

label define femalemale 0 "F" 1 "M"
quietly: tab child_sex
encode child_sex, gen(childmale)
recode childmale (1=0) (2=1)
label val childmale femalemale

rename page childage
quietly: tab childage
keep if childage<18

recode pk5_lb pk5_oz pk5_k (-9 -8=.)
gen birthweight=(pk5_lb*16)+pk5_oz
replace birthweight=pk5_k*35.274 if !missing(pk5_k)
drop pk5*
quietly: tab birthweight

rename pd3_a childcitizen
quietly: tab childcitizen
recode childcitizen(-5 -8=.) (1=1) (else=0)
quietly: tab childcitizen

quietly: tab pcgrace

```

```

recode pcgrace (0=.)
quietly: tab pcgrace

rename rpcgs pcggender
recode pcggender (1=1) (2=0)
quietly: tab pcggender

rename age_yr pcgage
quietly: tab pcgage
recode pcgage (-8=.)
quietly: tab pcgage

quietly: tab pcgcitizen
recode pcgcitizen (-5=.) (5=0)
quietly: tab pcgcitizen

rename pcg_edu pcgedu
quietly: tab pcgedu
recode pcgedu (-8=.)
quietly: tab pcgedu

rename age_birth momagebirth
quietly: tab momagebirth

rename birth_married mommarriedbirth
quietly: tab mommarriedbirth

quietly: tab hhedu
recode hhedu (-8=.)
quietly: tab hhedu

rename ge3 abuse
quietly: tab abuse
recode abuse (1=1) (5=0) (else=.)
quietly: tab abuse

rename probcase pcgdepression

gen social_support=1
replace social_support=0 if gd4_a==1
replace social_support=. if gd4_a<0 | missing(gd4_a)
drop gd4*
quietly: tab social_support

rename d_house ownhome
quietly: tab ownhome

rename W2_mar2000 pcgmstat2000
quietly: tab pcgmstat2000
rename W2_mar2001 pcgmstat2001
quietly: tab pcgmstat2001
rename W2_mar2002 pcgmstat2002
quietly: tab pcgmstat2002
rename W2_mar2003 pcgmstat2003

```

```

quietly: tab pcgmstat2003
rename W2_mar2004 pcgmstat2004
quietly: tab pcgmstat2004
rename W2_mar2005 pcgmstat2005
quietly: tab pcgmstat2005
rename W2_mar2006 pcgmstat2006
quietly: tab pcgmstat2006
rename W2_mar2007 pcgmstat2007
quietly: tab pcgmstat2007
rename W2_mar2008 pcgmstat2008
quietly: tab pcgmstat2008

rename W2_pcgemploy2000 pcgemploy2000
quietly: tab pcgemploy2000
rename W2_pcgemploy2001 pcgemploy2001
quietly: tab pcgemploy2001
rename W2_pcgemploy2002 pcgemploy2002
quietly: tab pcgemploy2002
rename W2_pcgemploy2003 pcgemploy2003
quietly: tab pcgemploy2003
rename W2_pcgemploy2004 pcgemploy2004
quietly: tab pcgemploy2004
rename W2_pcgemploy2005 pcgemploy2005
quietly: tab pcgemploy2005
rename W2_pcgemploy2006 pcgemploy2006
quietly: tab pcgemploy2006
rename W2_pcgemploy2007 pcgemploy2007
quietly: tab pcgemploy2007
rename W2_pcgemploy2008 pcgemploy2008
quietly: tab pcgemploy2008
label define employ 0 "Unempl" 1 "Full" 2 "Part"
label val pcgemploy* employ

rename ipolfaminc_2000 faminc2000
quietly: tab faminc2000
rename ipolfaminc_2001 faminc2001
quietly: tab faminc2001
rename ipolfaminc_2002 faminc2002
quietly: tab faminc2002
rename ipolfaminc_2003 faminc2003
quietly: tab faminc2003
rename ipolfaminc_2004 faminc2004
quietly: tab faminc2004
rename ipolfaminc_2005 faminc2005
quietly: tab faminc2005
rename ipolfaminc_2006 faminc2006
quietly: tab faminc2006
rename ipolfaminc_2007 faminc2007
quietly: tab faminc2007
rename ipolfaminc_2008 faminc2008
quietly: tab faminc2008

rename W2_salary2000 pcgsalary2000
replace pcgsalary2000=. if pcgsalary2000<0

```

```

quietly: tab pcgsalary2000
rename W2_salary2001 pcgsalary2001
replace pcgsalary2001=. if pcgsalary2001<0
quietly: tab pcgsalary2001
rename W2_salary2002 pcgsalary2002
replace pcgsalary2002=. if pcgsalary2002<0
quietly: tab pcgsalary2002
rename W2_salary2003 pcgsalary2003
replace pcgsalary2003=. if pcgsalary2003<0
quietly: tab pcgsalary2003
rename W2_salary2004 pcgsalary2004
replace pcgsalary2004=. if pcgsalary2004<0
quietly: tab pcgsalary2004
rename W2_salary2005 pcgsalary2005
replace pcgsalary2005=. if pcgsalary2005<0
quietly: tab pcgsalary2005
rename W2_salary2006 pcgsalary2006
replace pcgsalary2006=. if pcgsalary2006<0
quietly: tab pcgsalary2006
rename W2_salary2007 pcgsalary2007
replace pcgsalary2007=. if pcgsalary2007<0
quietly: tab pcgsalary2007
rename W2_salary2008 pcgsalary2008
replace pcgsalary2008=. if pcgsalary2008<0
quietly: tab pcgsalary2008

```

```

rename W2_emplrtyp2000 pcgemptytype2000
quietly: tab pcgemptytype2000
rename W2_emplrtyp2001 pcgemptytype2001
quietly: tab pcgemptytype2001
rename W2_emplrtyp2002 pcgemptytype2002
quietly: tab pcgemptytype2002
rename W2_emplrtyp2003 pcgemptytype2003
quietly: tab pcgemptytype2003
rename W2_emplrtyp2004 pcgemptytype2004
quietly: tab pcgemptytype2004
rename W2_emplrtyp2005 pcgemptytype2005
quietly: tab pcgemptytype2005
rename W2_emplrtyp2006 pcgemptytype2006
quietly: tab pcgemptytype2006
rename W2_emplrtyp2007 pcgemptytype2007
quietly: tab pcgemptytype2007
rename W2_emplrtyp2008 pcgemptytype2008
quietly: tab pcgemptytype2008

```

```

rename W2_hrswk2000 pcghours2000
quietly: tab pcghours2000
rename W2_hrswk2001 pcghours2001
quietly: tab pcghours2001
rename W2_hrswk2002 pcghours2002
quietly: tab pcghours2002
rename W2_hrswk2003 pcghours2003
quietly: tab pcghours2003

```

```
rename W2_hrswk2004 pcghours2004
quietly: tab pcghours2004
rename W2_hrswk2005 pcghours2005
quietly: tab pcghours2005
rename W2_hrswk2006 pcghours2006
quietly: tab pcghours2006
rename W2_hrswk2007 pcghours2007
quietly: tab pcghours2007
rename W2_hrswk2008 pcghours2008
quietly: tab pcghours2008
```

```
rename W2_welfare2000 welfare2000
quietly: tab welfare2000
rename W2_welfare2001 welfare2001
quietly: tab welfare2001
rename W2_welfare2002 welfare2002
quietly: tab welfare2002
rename W2_welfare2003 welfare2003
quietly: tab welfare2003
rename W2_welfare2004 welfare2004
quietly: tab welfare2004
rename W2_welfare2005 welfare2005
quietly: tab welfare2005
rename W2_welfare2006 welfare2006
quietly: tab welfare2006
rename W2_welfare2007 welfare2007
quietly: tab welfare2007
rename W2_welfare2008 welfare2008
quietly: tab welfare2008
```

```
rename W2_numchild2000 numchild2000
quietly: tab numchild2000
rename W2_numchild2001 numchild2001
quietly: tab numchild2001
rename W2_numchild2002 numchild2002
quietly: tab numchild2002
rename W2_numchild2003 numchild2003
quietly: tab numchild2003
rename W2_numchild2004 numchild2004
quietly: tab numchild2004
rename W2_numchild2005 numchild2005
quietly: tab numchild2005
rename W2_numchild2006 numchild2006
quietly: tab numchild2006
rename W2_numchild2007 numchild2007
quietly: tab numchild2007
rename W2_numchild2008 numchild2008
quietly: tab numchild2008
```

```
rename W2_moves2000 moves2000
quietly: tab moves2000
rename W2_moves2001 moves2001
quietly: tab moves2001
rename W2_moves2002 moves2002
```

```

quietly: tab moves2002
rename W2_moves2003 moves2003
quietly: tab moves2003
rename W2_moves2004 moves2004
quietly: tab moves2004
rename W2_moves2005 moves2005
quietly: tab moves2005
rename W2_moves2006 moves2006
quietly: tab moves2006
rename W2_moves2007 moves2007
quietly: tab moves2007
rename W2_moves2008 moves2008
quietly: tab moves2008

recode concdis_quintile2000 (5=1) (.=) (else=0), gen(concdis2000)
quietly: tab concdis2000, missing
recode concdis_quintile2001 (5=1) (.=) (else=0), gen(concdis2001)
quietly: tab concdis2001, missing
recode concdis_quintile2002 (5=1) (.=) (else=0), gen(concdis2002)
quietly: tab concdis2002, missing
recode concdis_quintile2003 (5=1) (.=) (else=0), gen(concdis2003)
quietly: tab concdis2003, missing
recode concdis_quintile2004 (5=1) (.=) (else=0), gen(concdis2004)
quietly: tab concdis2004, missing
recode concdis_quintile2005 (5=1) (.=) (else=0), gen(concdis2005)
quietly: tab concdis2005, missing
recode concdis_quintile2006 (5=1) (.=) (else=0), gen(concdis2006)
quietly: tab concdis2006, missing
recode concdis_quintile2007 (5=1) (.=) (else=0), gen(concdis2007)
quietly: tab concdis2007, missing
recode concdis_quintile2008 (5=1) (.=) (else=0), gen(concdis2008)
quietly: tab concdis2008, missing

quietly: tab birthweight
recode birthweight (min/88=1) (else=0)
quietly: tab birthweight

save
"C:\Users\ahicks\Documents\Stata_data\lafans\propensityscoredata.dta",
replace

```

```

*****
*****
*****
*****
*
*           Imputation
*****
*****
*****
*****
* Impute data for propensity estimation

```

```

* First run MModeling_Longitudinal.do
* Then run Longitudinal_cleaing.do

use
"C:\Users\ahicks\Documents\Stata_data\lafans\propensityscoredata.dta",
clear

/* profile plot of disadvantage
generate random=runiform()
generate obs=_n
profileplot concdis_2001 concdis_2002 concdis_2003 concdis_2004
concdis_2005 concdis_2006 if random<.05, by(obs) ///
    xlabel(1 "2001" 2 "2002" 3 "2003" 4 "2004" 5 "2005" 6 "2006")
msymbol(i) legend(off) xtitle("Year") ///
    ytitle("Disadvantage Score") ylab(-.85 -.45 .1 .87,
glcolor(black) glwidth(thick))

* quintiles 1: -.85 2: -.45 3: 0.10 4: .87
*/

gen dayage01= mofd(mdy(6,1,2001))-mofd(born_date)
gen age01= floor(dayage01/12)
gen dayage02= mofd(mdy(6,1,2001))-mofd(born_date)
gen age02= floor(dayage01/12)
gen dayage03= mofd(mdy(6,1,2001))-mofd(born_date)
gen age03= floor(dayage01/12)
gen dayage04= mofd(mdy(6,1,2001))-mofd(born_date)
gen age04= floor(dayage01/12)
gen dayage05= mofd(mdy(6,1,2001))-mofd(born_date)
gen age05= floor(dayage01/12)
gen dayage06= mofd(mdy(6,1,2001))-mofd(born_date)
gen age06= floor(dayage01/12)
drop dayage*

quietly: misstable sum, gen(miss_)

*Birth Imputation
drop if missing(childrace)

drop if missing(concdis_quintile2000) | missing(concdis_quintile2001) |
missing(concdis_quintile2002) | missing(concdis_quintile2003) ///
    | missing(concdis_quintile2004) | missing(concdis_quintile2005) |
missing(concdis_quintile2006)

*codebook  childmale birthweight mommarriedbirth momagebirth childrace

mi set mlong
mi register imputed birthweight momagebirth mommarriedbirth momreading_ss
/// birth imputation
    faminc* pcgsalary* numchild* pcghours*    /// time
varying impute
    pcgemploy* pcgemptype* hhedu ownhome moves*

```



```

mi register regular childmale childrace pcggender welfare* pcgmstat*
mi describe
set seed 1984
mi impute chained (regress) momagebirth (logit) mommarriedbirth
birthweight = childmale i.childrace, add(5)

* 2000 Imputation
drop if missing(concdis_quintile2000)
mi impute chained (pmm) momreading_ss faminc2000 pcgsalary2000
pcghours2000 numchild2000 moves2000(pmm, omit(concdis_quintile2000))
hhedu (logit) ownhome ///
(mlogit, omit(faminc2000 pcgsalary2000 moves2000))
pcgemploy2000 pcgemptype2000= welfare2000 pcgmstat2000
concdis_quintile2000, replace augment

* 2001 Imputation
drop if missing(concdis_quintile2001)
mi impute chained (pmm) faminc2001 pcgsalary2001 pcghours2001
numchild2001 moves2001 ///
(mlogit) pcgemploy2001 (mlogit, omit(i.pcgemploy2001
moves2001)) pcgemptype2001= welfare2001 pcgmstat2001
concdis_quintile2001, replace augment

* 2002 Imputation
drop if missing(concdis_quintile2002)
mi impute chained (pmm) faminc2002 pcgsalary2002 pcghours2002
numchild2002 moves2002 ///
(mlogit) pcgemploy2002 (mlogit, omit(i.pcgemploy2002
moves2002)) pcgemptype2002= welfare2002 pcgmstat2002
concdis_quintile2002, replace augment

* 2003 Imputation
drop if missing(concdis_quintile2003)
mi impute chained (pmm) faminc2003 pcgsalary2003 pcghours2003
numchild2003 moves2003 ///
(mlogit, omit(faminc2003 pcgsalary2003 moves2003))
pcgemploy2003 pcgemptype2003= welfare2003 pcgmstat2003
concdis_quintile2003, replace augment

* 2004 Imputation
drop if missing(concdis_quintile2004)
mi impute chained (pmm) faminc2004 pcgsalary2004 pcghours2004
numchild2004 moves2004 ///
(mlogit, omit(faminc2004 pcgsalary2004 moves2004))
pcgemploy2004 pcgemptype2004= welfare2004 pcgmstat2004
concdis_quintile2004, replace augment

* 2005 Imputation
drop if missing(concdis_quintile2005)

```

```

mi impute chained (pmm) faminc2005 pcgsalary2005 pcghours2005
numchild2005 moves2005 ///
      (mlogit, omit(faminc2005 pcgsalary2005 moves2005))
pcgemploy2005 pcgemptype2005= welfare2005 pcgmstat2005
concdis Quintile2005, replace augment

* 2006 Imputation
drop if missing(concdis Quintile2006)
mi impute chained (pmm) faminc2006 pcgsalary2006 pcghours2006
numchild2006 moves2006 ///
      (mlogit, omit(faminc2006 pcgsalary2006 pcghours2006
moves2006)) pcgemploy2006 pcgemptype2006= welfare2006 pcgmstat2006
concdis Quintile2006, replace augment

recode hhedu (0/11=0) (else=1)

mi passive:gen logfaminc2000=log(faminc2000+.0001)
mi passive:gen logfaminc2001=log(faminc2001+.0001)
mi passive:gen logfaminc2002=log(faminc2002+.0001)
mi passive:gen logfaminc2003=log(faminc2003+.0001)
mi passive:gen logfaminc2004=log(faminc2004+.0001)
mi passive:gen logfaminc2005=log(faminc2005+.0001)
mi passive:gen logfaminc2006=log(faminc2006+.0001)

save
"C:\Users\ahicks\Documents\Stata_data\lafans\propensityscoredata_imputed.
dta", replace

/*
foreach var of varlist birthweight momagebirth mommarriedbirth {
mi xeq 0: summ `var'
mi xeq 1/5: summ `var' if miss_`var'
}

foreach var of varlist faminc2001 pcgsalary2001 pcghours2001 {
mi xeq 0: summ `var'
mi xeq 1/5: summ `var' if miss_`var'
}

foreach var of varlist numchild2001 pcgemploy2001 pcgemptype2001 {
mi xeq 0: tab `var'
mi xeq 1/5: tab `var' if miss_`var'
}

```

Stata Code to Implement IPTW Approach with L.A.FANS data in Hicks et al. (2016)

```
*****
* File:          IPTW Regression.do
* Summary:       Following the methods of Wodtke et al. (2011)
*               use inverse probability of treatment (IPT)
*               weighting to examine the effects of exposure
*               to neighborhood disadvantage on math and
*               reading scores in LAFans data
* Author:        Andrew Hicks
* Institution:    California Center for Population Research
*               UCLA
*               Department of Health Care Policy
*               Harvard Medical School
* Last Updated:  11/23/2015
*****

*****
*               Create IPT Weights
*****
*****
*****
*****
cd C:\Users\ahicks\Documents\Stata_data\lafans
* Generate propensity score weights
* First run MModeling_Longitudinal.do
* Then run Longitudinal_cleaing.do

/* Use 2000 as baseline and compute treatment probabilities for 2001-2006
Use only variables in Wodtke paper
*/
* Generate 2001 P-weights
*****
*****
use "/home/ahicks/CITITraining/Pebley/propensityscoredata_imputed.dta",
clear

/*
* mean center momreading
summ momreading_ss
replace momreading_ss=momreading_ss-r(mean)
*/

/*
mi estimate: codebook concdis2001
/// current treatment
        concdis2000 childmale birthweight mommarriedbirth momagebirth
/// baseline covariates
        hhedu ownhome childrace
/// extra baseline (didn't have time varying)
        pcgmstat2000 pcgemploy2000 welfare2000 faminc2000 pcghours2000
/// baseline time vary covariates
        numchild2000 pcgsalary2000 pcgemptype2000
/// baseline time varying covariates
```

```

    pcgmsstat2001 pcgemploy2001 welfare2001 faminc2001 pcghours2001
/// current time vary covariates
    numchild2001 pcgsalary2001 pcgemptype2001 moves2001
// current time varying covariates
*/

* Denominator
xi: mi estimate, post saving(pscore2001, replace): ologit
concdis_quintile2001
/// current treatment
    childmale birthweight mommarriedbirth momagebirth age01 hhedu
momreading_ss ownhome /// Time-invariant characteristics
    concdis_quintile2000 pcgmsstat2000 pcgemploy2000 welfare2000
logfaminc2000 pcghours2000 numchild2000 /// Time-dependent Baseline
    pcgmsstat2001 pcgemploy2001 welfare2001 logfaminc2001 pcghours2001
numchild2001 moves2001 // Time-dependent K
*    i.pcgmsstat2000*i.pcgmsstat2001 i.pcgemploy2000*i.pcgemploy2001 //
[pweight=pwgt_rscsib] // K * K-1 Interactions
estimates store p2001denom

/*
xi: mi xeq 1: omodel logit concdis_quintile2001
/// current treatment
    childmale birthweight mommarriedbirth momagebirth age01 hhedu
ownhome /// Time-invariant characteristics
    concdis_quintile2000 pcgmsstat2000 pcgemploy2000 welfare2000
logfaminc2000 pcghours2000 numchild2000 /// Time-dependent Baseline
    pcgmsstat2001 pcgemploy2001 welfare2001 logfaminc2000 pcghours2001
numchild2001 moves2001 /// Time-dependent K
    i.pcgmsstat2000*i.pcgmsstat2001 i.pcgemploy2000*i.pcgemploy2001
// K * K-1 Interactions
*/

* DO WE NEED BIRTHYEAR AND AGE???
```

```

mi predictnl pscore2001a_1=predict(outcome(1)) using pscore2001
mi predictnl pscore2001a_2=predict(outcome(2)) using pscore2001
mi predictnl pscore2001a_3=predict(outcome(3)) using pscore2001
mi predictnl pscore2001a_4=predict(outcome(4)) using pscore2001
mi predictnl pscore2001a_5=predict(outcome(5)) using pscore2001

gen treatprob2001_den=pscore2001a_1 if concdis_quintile2001==1
replace treatprob2001_den=pscore2001a_2 if concdis_quintile2001==2
replace treatprob2001_den=pscore2001a_3 if concdis_quintile2001==3
replace treatprob2001_den=pscore2001a_4 if concdis_quintile2001==4
replace treatprob2001_den=pscore2001a_5 if concdis_quintile2001==5

* Numerator

```

```

xi: mi estimate, post saving(pscore2001, replace): ologit
concdis_quintile2001
/// current treatment
    childmale birthweight mommarriedbirth momagebirth age01 hhedu
momreading_ss ownhome /// Time-invariant characteristics
    concdis_quintile2000 pcgmstat2000 i.pcgemploy2000 welfare2000
logfaminc2000 pcghours2000 numchild2000 // [pweight=pwgt_rscsib] ///
Time-dependent Baseline

estimates store p2001num

```

* DO WE NEED BIRTHYEAR AND AGE???

```

mi predictnl pscore2001b_1=predict(outcome(1)) using pscore2001
mi predictnl pscore2001b_2=predict(outcome(2)) using pscore2001
mi predictnl pscore2001b_3=predict(outcome(3)) using pscore2001
mi predictnl pscore2001b_4=predict(outcome(4)) using pscore2001
mi predictnl pscore2001b_5=predict(outcome(5)) using pscore2001

gen treatprob2001_num=pscore2001b_1 if concdis_quintile2001==1
replace treatprob2001_num=pscore2001b_2 if concdis_quintile2001==2
replace treatprob2001_num=pscore2001b_3 if concdis_quintile2001==3
replace treatprob2001_num=pscore2001b_4 if concdis_quintile2001==4
replace treatprob2001_num=pscore2001b_5 if concdis_quintile2001==5

gen treatprob2001=treatprob2001_num/treatprob2001_den

```

```

* Generate 2002 P-weights
*****
*****

```

```

xi: mi estimate, noisily post saving(pscore2002, replace): ologit
concdis_quintile2002
/// current treatment
    childmale birthweight mommarriedbirth momagebirth age02 hhedu
momreading_ss ownhome /// Time-invariant characteristics
    concdis_quintile2000 pcgmstat2000 i.pcgemploy2000 welfare2000
logfaminc2000 pcghours2000 numchild2000 /// Time-dependent Baseline
    concdis_quintile2001 /// K-1 NH Disadvantage
    pcgmstat2001 i.pcgemploy2001 welfare2001 logfaminc2001
pcghours2001 /// Time-dependent K-1
    pcgmstat2002 i.pcgemploy2002 welfare2002 logfaminc2002
pcghours2002 numchild2002 moves2002 // Time-dependent K
* i.pcgmstat2001*i.pcgmstat2002 i.pcgemploy2001*i.pcgemploy2002 //
[pweight=pwgt_rscsib] // K * K-1 Interactions

estimates store p2002denom

```

```

mi predictnl pscore2002a_1=predict(outcome(1)) using pscore2002
mi predictnl pscore2002a_2=predict(outcome(2)) using pscore2002
mi predictnl pscore2002a_3=predict(outcome(3)) using pscore2002
mi predictnl pscore2002a_4=predict(outcome(4)) using pscore2002
mi predictnl pscore2002a_5=predict(outcome(5)) using pscore2002

gen treatprob2002_den=pscore2002a_1 if concdis_quintile2002==1
replace treatprob2002_den=pscore2002a_2 if concdis_quintile2002==2
replace treatprob2002_den=pscore2002a_3 if concdis_quintile2002==3
replace treatprob2002_den=pscore2002a_4 if concdis_quintile2002==4
replace treatprob2002_den=pscore2002a_5 if concdis_quintile2002==5

* Numerator
xi: mi estimate, noisily post saving(pscore2002, replace): ologit
concdis_quintile2002
/// current treatment
    childmale birthweight mommarriedbirth momagebirth age02 hhedu
momreading_ss ownhome /// Time-invariant characteristics
    concdis_quintile2000 pcgmstat2000 i.pcgemploy2000 welfare2000
logfaminc2000 pcghours2000 numchild2000 /// Time-dependent Baseline
    concdis_quintile2001 // [pweight=pwgt_rscsib] // K-1 NH
Disadvantage

estimates store p2002num

```

* DO WE NEED BIRTHYEAR AND AGE???

```

mi predictnl pscore2002b_1=predict(outcome(1)) using pscore2002
mi predictnl pscore2002b_2=predict(outcome(2)) using pscore2002
mi predictnl pscore2002b_3=predict(outcome(3)) using pscore2002
mi predictnl pscore2002b_4=predict(outcome(4)) using pscore2002
mi predictnl pscore2002b_5=predict(outcome(5)) using pscore2002

gen treatprob2002_num=pscore2002b_1 if concdis_quintile2002==1
replace treatprob2002_num=pscore2002b_2 if concdis_quintile2002==2
replace treatprob2002_num=pscore2002b_3 if concdis_quintile2002==3
replace treatprob2002_num=pscore2002b_4 if concdis_quintile2002==4
replace treatprob2002_num=pscore2002b_5 if concdis_quintile2002==5

gen treatprob2002=treatprob2002_num/treatprob2002_den

* Generate 2003 P-weights
*****
*****

xi: mi estimate, noisily post saving(pscore2003, replace): ologit
concdis_quintile2003
/// current treatment

```

```

    childmale birthweight mommarriedbirth momagebirth age03 hhedu
momreading_ss ownhome /// Time-invariant characteristics
    concdis_quintile2000 pcgstat2000 i.pcgemploy2000 welfare2000
logfaminc2000 pcghours2000 numchild2000 /// Time-dependent Baseline
    concdis_quintile2002          /// K-1 NH Disadvantage
    pcgstat2002 i.pcgemploy2002 welfare2002 logfaminc2002
pcghours2002 /// Time-dependent K-1
    pcgstat2003 i.pcgemploy2003 welfare2003 logfaminc2003
pcghours2003 numchild2003 moves2003 // Time-dependent K
*    i.pcgstat2002*i.pcgstat2003 i.pcgemploy2002*i.pcgemploy2003 //
[pweight=pwgt_rscsib] // K * K-1 Interactions

```

```

estimates store p2003denom

```

```

mi predictnl pscore2003a_1=predict(outcome(1)) using pscore2003
mi predictnl pscore2003a_2=predict(outcome(2)) using pscore2003
mi predictnl pscore2003a_3=predict(outcome(3)) using pscore2003
mi predictnl pscore2003a_4=predict(outcome(4)) using pscore2003
mi predictnl pscore2003a_5=predict(outcome(5)) using pscore2003

```

```

gen treatprob2003_den=pscore2003a_1 if concdis_quintile2003==1
replace treatprob2003_den=pscore2003a_2 if concdis_quintile2003==2
replace treatprob2003_den=pscore2003a_3 if concdis_quintile2003==3
replace treatprob2003_den=pscore2003a_4 if concdis_quintile2003==4
replace treatprob2003_den=pscore2003a_5 if concdis_quintile2003==5

```

```

* Numerator

```

```

xi: mi estimate, noisily post saving(pscore2003, replace): ologit
concdis_quintile2003
/// current treatment

```

```

    childmale birthweight mommarriedbirth momagebirth age03 hhedu
momreading_ss ownhome /// Time-invariant characteristics
    concdis_quintile2000 pcgstat2000 i.pcgemploy2000 welfare2000
logfaminc2000 pcghours2000 numchild2000 /// Time-dependent Baseline
    concdis_quintile2002 // [pweight=pwgt_rscsib] /// K-1 NH
Disadvantage

```

```

estimates store p2003num

```

```

* DO WE NEED BIRTHYEAR AND AGE???

```

```

mi predictnl pscore2003b_1=predict(outcome(1)) using pscore2003
mi predictnl pscore2003b_2=predict(outcome(2)) using pscore2003
mi predictnl pscore2003b_3=predict(outcome(3)) using pscore2003
mi predictnl pscore2003b_4=predict(outcome(4)) using pscore2003
mi predictnl pscore2003b_5=predict(outcome(5)) using pscore2003

```

```

gen treatprob2003_num=pscore2003b_1 if concdis_quintile2003==1
replace treatprob2003_num=pscore2003b_2 if concdis_quintile2003==2
replace treatprob2003_num=pscore2003b_3 if concdis_quintile2003==3

```

```

replace treatprob2003_num=pscore2003b_4 if concdis_quintile2003==4
replace treatprob2003_num=pscore2003b_5 if concdis_quintile2003==5

gen treatprob2003=treatprob2003_num/treatprob2003_den

* Generate 2004 P-weights
*****
*****

xi: mi estimate, noisily post saving(pscore2004, replace): ologit
concdis_quintile2004
/// current treatment
    childmale birthweight mommarriedbirth momagebirth age04 hhedu
momreading_ss ownhome /// Time-invariant characteristics
    concdis_quintile2000 pcgmstat2000 i.pcgemploy2000 welfare2000
logfaminc2000 pcghours2000 numchild2000 /// Time-dependent Baseline
    concdis_quintile2003 /// K-1 NH Disadvantage
    pcgmstat2003 i.pcgemploy2003 welfare2003 logfaminc2003
pcghours2003 /// Time-dependent K-1
    pcgmstat2004 i.pcgemploy2004 welfare2004 logfaminc2004
pcghours2004 numchild2004 moves2004 // Time-dependent K
* i.pcgmstat2003*i.pcgmstat2004 i.pcgemploy2003*i.pcgemploy2004 //
[pweight=pwgt_rscsib] // K * K-1 Interactions

    estimates store p2004denom

mi predictnl pscore2004a_1=predict(outcome(1)) using pscore2004
mi predictnl pscore2004a_2=predict(outcome(2)) using pscore2004
mi predictnl pscore2004a_3=predict(outcome(3)) using pscore2004
mi predictnl pscore2004a_4=predict(outcome(4)) using pscore2004
mi predictnl pscore2004a_5=predict(outcome(5)) using pscore2004

gen treatprob2004_den=pscore2004a_1 if concdis_quintile2004==1
replace treatprob2004_den=pscore2004a_2 if concdis_quintile2004==2
replace treatprob2004_den=pscore2004a_3 if concdis_quintile2004==3
replace treatprob2004_den=pscore2004a_4 if concdis_quintile2004==4
replace treatprob2004_den=pscore2004a_5 if concdis_quintile2004==5

* Numerator
xi: mi estimate, noisily post saving(pscore2004, replace): ologit
concdis_quintile2004
/// current treatment
    childmale birthweight mommarriedbirth momagebirth age04 hhedu
momreading_ss ownhome /// Time-invariant characteristics
    concdis_quintile2000 pcgmstat2000 i.pcgemploy2000 welfare2000
logfaminc2000 pcghours2000 numchild2000 /// Time-dependent Baseline
    concdis_quintile2003 // [pweight=pwgt_rscsib] /// K-1 NH
Disadvantage

    estimates store p2004num

```



```

* DO WE NEED BIRTHYEAR AND AGE???

mi predictnl pscore2004b_1=predict(outcome(1)) using pscore2004
mi predictnl pscore2004b_2=predict(outcome(2)) using pscore2004
mi predictnl pscore2004b_3=predict(outcome(3)) using pscore2004
mi predictnl pscore2004b_4=predict(outcome(4)) using pscore2004
mi predictnl pscore2004b_5=predict(outcome(5)) using pscore2004

gen treatprob2004_num=pscore2004b_1 if concdis_quintile2004==1
replace treatprob2004_num=pscore2004b_2 if concdis_quintile2004==2
replace treatprob2004_num=pscore2004b_3 if concdis_quintile2004==3
replace treatprob2004_num=pscore2004b_4 if concdis_quintile2004==4
replace treatprob2004_num=pscore2004b_5 if concdis_quintile2004==5

gen treatprob2004=treatprob2004_num/treatprob2004_den

* Generate 2005 P-weights
*****
*****

xi: mi estimate, noisily post saving(pscore2005, replace): ologit
concdis_quintile2005
/// current treatment
    childmale birthweight mommarriedbirth momagebirth age05 hhedu
momreading_ss ownhome /// Time-invariant characteristics
    concdis_quintile2000 pcgmstat2000 i.pcgemploy2000 welfare2000
logfaminc2000 pcghours2000 numchild2000 /// Time-dependent Baseline
    concdis_quintile2004 /// K-1 NH Disadvantage
    pcgmstat2004 i.pcgemploy2004 welfare2004 logfaminc2004
pcghours2004 /// Time-dependent K-1
    pcgmstat2005 i.pcgemploy2005 welfare2005 logfaminc2005
pcghours2005 numchild2005 moves2005 // Time-dependent K
*    i.pcgstat2004*i.pcgstat2005 i.pcgemploy2004*i.pcgemploy2005 //
[pweight=pwgt_rscsib] // K * K-1 Interactions

estimates store p2005denom

mi predictnl pscore2005a_1=predict(outcome(1)) using pscore2005
mi predictnl pscore2005a_2=predict(outcome(2)) using pscore2005
mi predictnl pscore2005a_3=predict(outcome(3)) using pscore2005
mi predictnl pscore2005a_4=predict(outcome(4)) using pscore2005
mi predictnl pscore2005a_5=predict(outcome(5)) using pscore2005

gen treatprob2005_den=pscore2005a_1 if concdis_quintile2005==1
replace treatprob2005_den=pscore2005a_2 if concdis_quintile2005==2
replace treatprob2005_den=pscore2005a_3 if concdis_quintile2005==3
replace treatprob2005_den=pscore2005a_4 if concdis_quintile2005==4
replace treatprob2005_den=pscore2005a_5 if concdis_quintile2005==5

```

```

* Numerator
xi: mi estimate, noisily post saving(pscore2005, replace): ologit
concdis_quintile2005
/// current treatment
    childmale birthweight mommarriedbirth momagebirth age05 hhedu
momreading_ss ownhome /// Time-invariant characteristics
    concdis_quintile2000 pcgmstat2000 i.pcgemploy2000 welfare2000
logfaminc2000 pcghours2000 numchild2000 /// Time-dependent Baseline
    concdis_quintile2004 // [pweight=pwgt_rscsib]      /// K-1 NH
Disadvantage

```

```
estimates store p2005denom
```

```
* DO WE NEED BIRTHYEAR AND AGE???
```

```

mi predictnl pscore2005b_1=predict(outcome(1)) using pscore2005
mi predictnl pscore2005b_2=predict(outcome(2)) using pscore2005
mi predictnl pscore2005b_3=predict(outcome(3)) using pscore2005
mi predictnl pscore2005b_4=predict(outcome(4)) using pscore2005
mi predictnl pscore2005b_5=predict(outcome(5)) using pscore2005

```

```

gen treatprob2005_num=pscore2005b_1 if concdis_quintile2005==1
replace treatprob2005_num=pscore2005b_2 if concdis_quintile2005==2
replace treatprob2005_num=pscore2005b_3 if concdis_quintile2005==3
replace treatprob2005_num=pscore2005b_4 if concdis_quintile2005==4
replace treatprob2005_num=pscore2005b_5 if concdis_quintile2005==5

```

```
gen treatprob2005=treatprob2005_num/treatprob2005_den
```

```
* Generate 2006 P-weights
```

```

*****
*****

```

```

xi: mi estimate, noisily post saving(pscore2006, replace): ologit
concdis_quintile2006
/// current treatment
    childmale birthweight mommarriedbirth momagebirth age06 hhedu
momreading_ss ownhome /// Time-invariant characteristics
    concdis_quintile2000 pcgmstat2000 i.pcgemploy2000 welfare2000
logfaminc2000 pcghours2000 numchild2000 /// Time-dependent Baseline
    concdis_quintile2005      /// K-1 NH Disadvantage
    pcgmstat2005 i.pcgemploy2005 welfare2005 logfaminc2005
pcghours2005 /// Time-dependent K-1
    pcgmstat2006 i.pcgemploy2006 welfare2005 logfaminc2006
pcghours2006 numchild2006 moves2006 // Time-dependent K
*    i.pcgemploy2005*i.pcgemploy2006 //
[pweight=pwgt_rscsib]      // K * K-1 Interactions

```

```
estimates store p2006denom
```

```
mi predictnl pscore2006a_1=predict(outcome(1)) using pscore2006
mi predictnl pscore2006a_2=predict(outcome(2)) using pscore2006
mi predictnl pscore2006a_3=predict(outcome(3)) using pscore2006
mi predictnl pscore2006a_4=predict(outcome(4)) using pscore2006
mi predictnl pscore2006a_5=predict(outcome(5)) using pscore2006
```

```
gen treatprob2006_den=pscore2006a_1 if concdis_quintile2006==1
replace treatprob2006_den=pscore2006a_2 if concdis_quintile2006==2
replace treatprob2006_den=pscore2006a_3 if concdis_quintile2006==3
replace treatprob2006_den=pscore2006a_4 if concdis_quintile2006==4
replace treatprob2006_den=pscore2006a_5 if concdis_quintile2006==5
```

```
* Numerator
```

```
xi: mi estimate, noisily post saving(pscore2006, replace): ologit
concdis_quintile2006
/// current treatment
    childmale birthweight mommarriedbirth momagebirth age06 hhedu
momreading_ss ownhome /// Time-invariant characteristics
    concdis_quintile2000 pcgmstat2000 i.pcgemploy2000 welfare2000
logfaminc2000 pcghours2000 numchild2000 /// Time-dependent Baseline
    concdis_quintile2005 // [pweight=pwgt_rscsib]    /// K-1 NH
Disadvantage
```

```
estimates store p2006num
```

```
* DO WE NEED BIRTHYEAR AND AGE???
```

```
mi predictnl pscore2006b_1=predict(outcome(1)) using pscore2006
mi predictnl pscore2006b_2=predict(outcome(2)) using pscore2006
mi predictnl pscore2006b_3=predict(outcome(3)) using pscore2006
mi predictnl pscore2006b_4=predict(outcome(4)) using pscore2006
mi predictnl pscore2006b_5=predict(outcome(5)) using pscore2006
```

```
gen treatprob2006_num=pscore2006b_1 if concdis_quintile2006==1
replace treatprob2006_num=pscore2006b_2 if concdis_quintile2006==2
replace treatprob2006_num=pscore2006b_3 if concdis_quintile2006==3
replace treatprob2006_num=pscore2006b_4 if concdis_quintile2006==4
replace treatprob2006_num=pscore2006b_5 if concdis_quintile2006==5
```

```
gen treatprob2006=treatprob2006_num/treatprob2006_den
```

```
gen
```

```
ipt_weight=treatprob2001*treatprob2002*treatprob2003*treatprob2004*treatp
rob2005*treatprob2006
```

```
gen
```

```
unstab_ipt_weight=(1/treatprob2001_den)*(1/treatprob2002_den)*(1/treatpro
```

```
b2003_den)*(1/treatprob2004_den)*(1/treatprob2005_den)*(1/treatprob2006_d
en)
```

```
*****
*****
*****
*****
*
*           Duration Weighted Models
*****
*****
*****
*****
```

```
mi convert wide, clear
```

```
recode pcgemploy2000 (2=1)
recode pcgemploy2001 (2=1)
recode pcgemploy2002 (2=1)
recode pcgemploy2003 (2=1)
recode pcgemploy2004 (2=1)
recode pcgemploy2005 (2=1)
recode pcgemploy2006 (2=1)
```

```
summ unstab_ipt_weight, det
hist unstab_ipt_weight
hist unstab_ipt_weight if unstab_ipt_weight<40, bin(100)
title("Unstabilized weights truncated at 40")
*hangroot unstab_ipt_weight, dist(lognormal)
*lognfit unstab_ipt_weight
```

```
* truncate unstabilized ipt weights at 1st and 99th pct
egen unstab_ipt_5th= pctlile(unstab_ipt_weight), p(5)
egen unstab_ipt_95th= pctlile(unstab_ipt_weight), p(95)
gen unstab_trunc=unstab_ipt_weight
replace unstab_trunc=unstab_ipt_5th if unstab_ipt_weight<unstab_ipt_5th
replace unstab_trunc=unstab_ipt_95th if unstab_ipt_weight>unstab_ipt_95th
sum unstab_trunc, det
*hangroot unstab_trunc, dist(lognormal)
*lognfit unstab_trunc
```

```
summ pwgt_rscsib, det
gen aipt_weight=ipt_weight*pwgt_rscsib
summ aipt_weight, det
```

```
hist ipt_weight, bin(100) xlabel(0(1)30) title("Stabilized Weights
without LAFans weights in PS Model")
```

```
* truncate weights at 1st and 99th pct
summ ipt_weight, det
```

```

gen ipt_notrunc=ipt_weight
egen ipt_1st= pctlile(ipt_weight), p(1)
egen ipt_99th= pctlile(ipt_weight), p(99)
replace ipt_weight=ipt_1st if ipt_weight<ipt_1st
replace ipt_weight=ipt_99th if ipt_weight>ipt_99th
sum ipt_weight, det

* truncate weights at 1st and 99th pct
egen aipt_1st= pctlile(aipt_weight), p(1)
egen aipt_99th= pctlile(aipt_weight), p(99)
replace aipt_weight=aipt_1st if aipt_weight<aipt_1st
replace aipt_weight=aipt_99th if aipt_weight>aipt_99th
sum aipt_weight, det
hist aipt_weight, bin(100)

drop if missing(concdis_quintile2001) | missing(concdis_quintile2002) |
missing(concdis_quintile2003) | ///
      missing(concdis_quintile2004) | missing(concdis_quintile2005) |
missing(concdis_quintile2006)

gen dwe=
(concdis_quintile2001+concdis_quintile2002+concdis_quintile2003+concdis_q
uintile2004+concdis_quintile2005+concdis_quintile2006)/6
gen recency=((concdis_quintile2001-dwe))+((concdis_quintile2002-
dwe)*2)+((concdis_quintile2003-dwe)*3) ///
      +((concdis_quintile2004-dwe)*4)+((concdis_quintile2005-
dwe)*5)+((concdis_quintile2006-dwe)*6)
replace recency=recency/21
summ recency

tab dwe

* generate moves between exposure levels.
gen exposure_moves=0
forvalues i=2002(1)2006 {
local j=`i'-1
replace exposure_moves=exposure_moves+1 if
concdis_quintile`i'!=concdis_quintile`j'
}

** compare this tab to Wodtke list of exposures..

** Tables
gen dwe_table = 5 if dwe<=5
replace dwe_tab = 4 if dwe<4.5
replace dwe_tab = 3 if dwe<3.5
replace dwe_tab = 2 if dwe<2.5
replace dwe_tab = 1 if dwe<1.5
*label def dwe 1 "1.0 to 1.4 (least disadvantaged NHs)" 2 "1.5 to 2.4" 3
"2.5 to 3.4" 4 "3.5 to 4.4" 5 "4.5 to 5.0 (most disadvantaged NHs)"
label val dwe_table dwe

```

```

tab dwe_table childrace if !missing(W2_ap_ss)
tab dwe_table childrace, col

tab exposure_moves childrace if !missing(W2_ap_ss)
tab exposure_moves childrace, col

summ aipt_weight, det

hist aipt_weight, bin(100) normal

** Unadjusted Models
mi estimate, post saving(ap_unadjusted, replace): regress W2_ap_ss dwe
estimates store ap_unadjusted

mi estimate, post saving(lwi_unadjusted, replace): regress W2_lwi_ss dwe
estimates store lwi_unadjusted

/* unstab-IPT-adjusted Models

mi estimate, post saving(ap_ipt, replace): regress W2_ap_ss dwe ///
                                     childmale
birthweight mommarriedbirth momagebirth age01 hhedu momreading_ss ownhome
/// Time-invariant characteristics
                                     pcgmstat2000
i.pcgemploy2000 welfare2000 logfaminc2000 pcghours2000 numchild2000 ///
Baseline time varying

                                     [pweight=unstab_ipt_weight]
estimates store ap_unstab_ipt
*/

/*
mi estimate, post saving(lwi_ipt, replace): regress W2_lwi_ss dwe ///
                                     childmale
birthweight mommarriedbirth momagebirth age01 hhedu momreading_ss ownhome
/// Time-invariant characteristics
                                     pcgmstat2000
i.pcgemploy2000 welfare2000 logfaminc2000 pcghours2000 numchild2000 ///
Baseline time varying

                                     [pweight=unstab_ipt_weight]
estimates store lwi_ipt
*/

** Regression-adjusted Models
mi passive: egen pcgmstat=rowmean(pcgstat2001 pcgmstat2002 pcgmstat2003
pcgmstat2004 pcgmstat2005 pcgmstat2006)
mi passive: gen pcgemploy2001_bin=0

```

```

mi passive: replace pcgemploy2001_bin=1 if (pcgemploy2001==1 |
pcgemploy2001==2)
mi passive: gen pcgemploy2002_bin=0
mi passive: replace pcgemploy2002_bin=1 if (pcgemploy2002==1 |
pcgemploy2002==2)
mi passive: gen pcgemploy2003_bin=0
mi passive: replace pcgemploy2003_bin=1 if (pcgemploy2003==1 |
pcgemploy2003==2)
mi passive: gen pcgemploy2004_bin=0
mi passive: replace pcgemploy2004_bin=1 if (pcgemploy2004==1 |
pcgemploy2004==2)
mi passive: gen pcgemploy2005_bin=0
mi passive: replace pcgemploy2005_bin=1 if (pcgemploy2005==1 |
pcgemploy2005==2)
mi passive: gen pcgemploy2006_bin=0
mi passive: replace pcgemploy2006_bin=1 if (pcgemploy2006==1 |
pcgemploy2006==2)
mi passive: egen pcgemploy=rowmean(pcgemploy2001_bin pcgemploy2002_bin
pcgemploy2003_bin pcgemploy2004_bin pcgemploy2005_bin pcgemploy2006_bin)
mi passive: egen welfare=rowmean(welfare2001 welfare2002 welfare2003
welfare2004 welfare2005 welfare2006)
mi passive: egen logfaminc=rowmean(logfaminc2001 logfaminc2002
logfaminc2003 logfaminc2004 logfaminc2005 logfaminc2006)
mi passive: egen pcghours=rowmean(pcghours2001 pcghours2002 pcghours2003
pcghours2004 pcghours2005 pcghours2006)
mi passive: egen numchild=rowmean(numchild2001 numchild2002 numchild2003
numchild2004 numchild2005 numchild2006)

```

```

mi estimate, post saving(ap_regression, replace): regress W2_ap_ss dwe
/// outcome
    childmale birthweight mommarriedbirth momagebirth age01 hhedu
momreading_ss ownhome /// Time-invariant characteristics
    pcgmstat2000 i.pcgemploy2000 welfare2000 logfaminc2000
pcghours2000 numchild2000 /// Baseline time varying
    pcgmstat pcgemploy welfare logfaminc pcghours numchild
[pweight=pwgt]
estimates store ap_regression

```

```

mi estimate, post saving(lwi_regression, replace): regress W2_lwi_ss dwe
/// outcome
    childmale birthweight mommarriedbirth momagebirth age01 hhedu
momreading_ss ownhome /// Time-invariant characteristics
    pcgmstat2000 i.pcgemploy2000 welfare2000 logfaminc2000
pcghours2000 numchild2000 /// Baseline time varying
    pcgmstat pcgemploy welfare logfaminc pcghours numchild
[pweight=pwgt]
estimates store lwi_regression

```

```

lab var dwe "Duration-Weighted Exposure"

```

```

lab var childmale "Gender (ref: Female)"
lab var birthweight "Birthweight (ref: >5.5 lbs)"
lab var mommarriedbirth "Mother's marital status at birth (ref: Unmarr)"
lab var momagebirth "Mother's age at birth"
lab var age01 "Child's age at baseline"
lab var hhedu "FU head's education (ref: no HS diploma)"
lab var momreading_ss "Mother's Reading Score"
lab var ownhome "Home ownership (ref: do not own)"
lab var pcgmstat2000 "Baseline: PCG marital status (ref: Unmarr.)"
lab var pcgemploy2000 "Baseline: PCG employment status (ref: Unempl.)"
lab var welfare2000 "Baseline: Public assistance receipt (ref: No TANF)"
lab var logfaminc2000 "Baseline: FU income (log $)"
lab var pcghours2000 "Baseline: PCG work hours"
lab var numchild2000 "Baseline: PCG number of children"
lab var pcgmstat "Averaged: PCG marital status"
lab var pcgemploy "Averaged: PCG employment status"
lab var welfare "Averaged: Public assistance receipt"
lab var logfaminc "Averaged: FU income"
lab var pcghours "Averaged: PCG work hours"
lab var numchild "Averaged: PCG number of children"

** AIPT-adjusted Models (IPT x LAFANS Sample Weights)

recode childrace (2=1) (3=2) (else=3), gen(childrace_latino)
label var childrace_lat "Child's race (ref: latino)"
label define race_lat 1 "Latino" 2 "Black" 3 "Other"
label values childrace_lat race_lat

*****
*****      Stratified Models - Math
*****

*****
***      average_treatment and Recency
*****

*****
** All Povcats
*****

mi estimate, post saving(ap_aipt, replace): regress W2_ap_ss dwe recency
///
                                age01 childmale
i.childrace_lat birthweight momagebirth mommarriedbirth momreading_ss
hhedu ownhome /// Time-invariant characteristics

```



```

                                                                    pcgmstat2000
numchild2000 pcgemploy2000 pcghours2000 welfare2000 logfaminc2000  ///
Baseline time varying

                                [pweight=aipt_weight], vce(cluster hhid2)
*****
** Povcat=1
*****
mi estimate, post saving(ap_aipt, replace): regress W2_ap_ss dwe recency
///

                                                                    age01 childmale
i.childrace_lat birthweight momagebirth mommarriedbirth momreading_ss
hhedu ownhome /// Time-invariant characteristics

                                                                    pcgmstat2000
numchild2000 pcgemploy2000 pcghours2000 welfare2000 logfaminc2000  ///
Baseline time varying

                                [pweight=aipt_weight]                if povcat==1 , vce(cluster
hhid2)
*****
** Povcat=2
*****
mi estimate, post saving(ap_aipt, replace): regress W2_ap_ss dwe recency
///

                                                                    age01 childmale
i.childrace_lat birthweight momagebirth mommarriedbirth momreading_ss
hhedu ownhome /// Time-invariant characteristics

                                                                    pcgmstat2000
numchild2000 pcgemploy2000 pcghours2000 welfare2000 logfaminc2000  ///
Baseline time varying

                                [pweight=aipt_weight]                if povcat==2      ,
vce(cluster hhid2)

*****
** Povcat=3
*****
mi estimate, post saving(ap_aipt, replace): regress W2_ap_ss dwe recency
///

                                                                    age01 childmale
i.childrace_lat birthweight momagebirth mommarriedbirth momreading_ss
hhedu ownhome /// Time-invariant characteristics

                                                                    pcgmstat2000
numchild2000 pcgemploy2000 pcghours2000 welfare2000 logfaminc2000  ///
Baseline time varying

                                [pweight=aipt_weight]                if povcat==3      ,
vce(cluster hhid2)

*****
***   average_treatment
*****

```

```

*****
** All Povcats
*****
mi estimate, post saving(ap_aipt, replace): regress W2_ap_ss dwe ///
                                                age01 childmale
i.childrace_lat birthweight momagebirth mommarriedbirth momreading_ss
hhedu ownhome /// Time-invariant characteristics
                                                pcgmstat2000
numchild2000 pcgemploy2000 pcghours2000 welfare2000 logfaminc2000 ///
Baseline time varying

                [pweight=aipt_weight] , vce(cluster hhid2)
*****
** Povcat=1
*****
mi estimate, post saving(ap_aipt, replace): regress W2_ap_ss dwe ///
                                                age01 childmale
i.childrace_lat birthweight momagebirth mommarriedbirth momreading_ss
hhedu ownhome /// Time-invariant characteristics
                                                pcgmstat2000
numchild2000 pcgemploy2000 pcghours2000 welfare2000 logfaminc2000 ///
Baseline time varying

                [pweight=aipt_weight]          if povcat==1 , vce(cluster
hhid2)
*****
** Povcat=2
*****
mi estimate, post saving(ap_aipt, replace): regress W2_ap_ss dwe ///
                                                age01 childmale
i.childrace_lat birthweight momagebirth mommarriedbirth momreading_ss
hhedu ownhome /// Time-invariant characteristics
                                                pcgmstat2000
numchild2000 pcgemploy2000 pcghours2000 welfare2000 logfaminc2000 ///
Baseline time varying

                [pweight=aipt_weight]          if povcat==2          ,
vce(cluster hhid2)

*****
** Povcat=3
*****
mi estimate, post saving(ap_aipt, replace): regress W2_ap_ss dwe ///
                                                age01 childmale
i.childrace_lat birthweight momagebirth mommarriedbirth momreading_ss
hhedu ownhome /// Time-invariant characteristics
                                                pcgmstat2000
numchild2000 pcgemploy2000 pcghours2000 welfare2000 logfaminc2000 ///
Baseline time varying

                [pweight=aipt_weight]          if povcat==3          ,
vce(cluster hhid2)

```

```

*****
***   Recency
*****

*****
** All Povcats
*****
mi estimate, post saving(ap_aipt, replace): regress W2_ap_ss recency ///
                                                age01 childmale
i.childrace_lat birthweight momagebirth mommarriedbirth momreading_ss
hhedu ownhome /// Time-invariant characteristics
                                                pcgmstat2000
numchild2000 pcgemploy2000 pcghours2000 welfare2000 logfaminc2000 ///
Baseline time varying

                [pweight=aipt_weight] , vce(cluster hhid2)
*****
** Povcat=1
*****
mi estimate, post saving(ap_aipt, replace): regress W2_ap_ss recency ///
                                                age01 childmale
i.childrace_lat birthweight momagebirth mommarriedbirth momreading_ss
hhedu ownhome /// Time-invariant characteristics
                                                pcgmstat2000
numchild2000 pcgemploy2000 pcghours2000 welfare2000 logfaminc2000 ///
Baseline time varying

                [pweight=aipt_weight]          if povcat==1      ,
vce(cluster hhid2)
*****
** Povcat=2
*****
mi estimate, post saving(ap_aipt, replace): regress W2_ap_ss recency ///
                                                age01 childmale
i.childrace_lat birthweight momagebirth mommarriedbirth momreading_ss
hhedu ownhome /// Time-invariant characteristics
                                                pcgmstat2000
numchild2000 pcgemploy2000 pcghours2000 welfare2000 logfaminc2000 ///
Baseline time varying

                [pweight=aipt_weight]          if povcat==2      ,
vce(cluster hhid2)

*****
** Povcat=3
*****
mi estimate, post saving(ap_aipt, replace): regress W2_ap_ss recency ///
                                                age01 childmale
i.childrace_lat birthweight momagebirth mommarriedbirth momreading_ss
hhedu ownhome /// Time-invariant characteristics
                                                pcgmstat2000
numchild2000 pcgemploy2000 pcghours2000 welfare2000 logfaminc2000 ///
Baseline time varying

```

```

                [pweight=aipt_weight]          if povcat==3          ,
vce(cluster hhid2)

*****
*****      Stratified Models - Reading
*****

*****
***   average_treatment and Recency
*****

*****
** All Povcats
*****
mi estimate, post saving(ap_aipt, replace): regress W2_lwi_ss dwe recency
///
                                                age01 childmale
i.childrace_lat birthweight momagebirth mommarriedbirth momreading_ss
hhedu ownhome /// Time-invariant characteristics
                                                pcgmstat2000
numchild2000 pcgemploy2000 pcghours2000 welfare2000 logfaminc2000   ///
Baseline time varying

                [pweight=aipt_weight] , vce(cluster hhid2)
*****
** Povcat=1
*****
mi estimate, post saving(ap_aipt, replace): regress W2_lwi_ss dwe recency
///
                                                age01 childmale
i.childrace_lat birthweight momagebirth mommarriedbirth momreading_ss
hhedu ownhome /// Time-invariant characteristics
                                                pcgmstat2000
numchild2000 pcgemploy2000 pcghours2000 welfare2000 logfaminc2000   ///
Baseline time varying

                [pweight=aipt_weight]          if povcat==1 , vce(cluster
hhid2)
*****
** Povcat=2
*****

```

```

mi estimate, post saving(ap_aipt, replace): regress W2_lwi_ss dwe recency
///
                                                    age01 childmale
i.childrace_lat birthweight momagebirth mommarriedbirth momreading_ss
hhedu ownhome /// Time-invariant characteristics
                                                    pcgmstat2000
numchild2000 pcgemploy2000 pcghours2000 welfare2000 logfaminc2000 ///
Baseline time varying

                [pweight=aipt_weight]                if povcat==2 , vce(cluster
hhid2)

*****
** Povcat=3
*****
mi estimate, post saving(ap_aipt, replace): regress W2_lwi_ss dwe recency
///
                                                    age01 childmale
i.childrace_lat birthweight momagebirth mommarriedbirth momreading_ss
hhedu ownhome /// Time-invariant characteristics
                                                    pcgmstat2000
numchild2000 pcgemploy2000 pcghours2000 welfare2000 logfaminc2000 ///
Baseline time varying

                [pweight=aipt_weight]                if povcat==3      ,
vce(cluster hhid2)

*****
***   average_treatment
*****

*****
** All Povcats
*****
mi estimate, post saving(ap_aipt, replace): regress W2_lwi_ss dwe ///
                                                    age01 childmale
i.childrace_lat birthweight momagebirth mommarriedbirth momreading_ss
hhedu ownhome /// Time-invariant characteristics
                                                    pcgmstat2000
numchild2000 pcgemploy2000 pcghours2000 welfare2000 logfaminc2000 ///
Baseline time varying

                [pweight=aipt_weight] , vce(cluster hhid2)
*****
** Povcat=1
*****
mi estimate, post saving(ap_aipt, replace): regress W2_lwi_ss dwe ///
                                                    age01 childmale
i.childrace_lat birthweight momagebirth mommarriedbirth momreading_ss
hhedu ownhome /// Time-invariant characteristics

```

```

                                                                    pcgmstat2000
numchild2000 pcgemploy2000 pcghours2000 welfare2000 logfaminc2000  ///
Baseline time varying

                [pweight=aipt_weight]                if povcat==1                ,
vce(cluster hhid2)
*****
** Povcat=2
*****
mi estimate, post saving(ap_aipt, replace): regress W2_lwi_ss dwe ///
                                                                    age01 childmale
i.childrace_lat birthweight momagebirth mommarriedbirth momreading_ss
hhedu ownhome /// Time-invariant characteristics

                                                                    pcgmstat2000
numchild2000 pcgemploy2000 pcghours2000 welfare2000 logfaminc2000  ///
Baseline time varying

                [pweight=aipt_weight]                if povcat==2                ,
vce(cluster hhid2)

*****
** Povcat=3
*****
mi estimate, post saving(ap_aipt, replace): regress W2_lwi_ss dwe ///
                                                                    age01 childmale
i.childrace_lat birthweight momagebirth mommarriedbirth momreading_ss
hhedu ownhome /// Time-invariant characteristics

                                                                    pcgmstat2000
numchild2000 pcgemploy2000 pcghours2000 welfare2000 logfaminc2000  ///
Baseline time varying

                [pweight=aipt_weight]                if povcat==3                ,
vce(cluster hhid2)

*****
*** Recency
*****

*****
** All Povcats
*****
mi estimate, post saving(ap_aipt, replace): regress W2_lwi_ss recency
///
                                                                    age01 childmale
i.childrace_lat birthweight momagebirth mommarriedbirth momreading_ss
hhedu ownhome /// Time-invariant characteristics

                                                                    pcgmstat2000
numchild2000 pcgemploy2000 pcghours2000 welfare2000 logfaminc2000  ///
Baseline time varying

                [pweight=aipt_weight] , vce(cluster hhid2)
*****
** Povcat=1

```

```

*****
mi estimate, post saving(ap_aipt, replace): regress W2_lwi_ss recency ///
                                                age01 childmale
i.childrace_lat birthweight momagebirth mommarriedbirth momreading_ss
hhedu ownhome /// Time-invariant characteristics
                                                pcgmstat2000
numchild2000 pcgemploy2000 pcghours2000 welfare2000 logfaminc2000 ///
Baseline time varying

                [pweight=aipt_weight]                if povcat==1 , vce(cluster
hhid2)
*****
** Povcat=2
*****
mi estimate, post saving(ap_aipt, replace): regress W2_lwi_ss recency
///
                                                age01 childmale
i.childrace_lat birthweight momagebirth mommarriedbirth momreading_ss
hhedu ownhome /// Time-invariant characteristics
                                                pcgmstat2000
numchild2000 pcgemploy2000 pcghours2000 welfare2000 logfaminc2000 ///
Baseline time varying

                [pweight=aipt_weight]                if povcat==2      ,
vce(cluster hhid2)

*****
** Povcat=3
*****
mi estimate, post saving(ap_aipt, replace): regress W2_lwi_ss recency ///
                                                age01 childmale
i.childrace_lat birthweight momagebirth mommarriedbirth momreading_ss
hhedu ownhome /// Time-invariant characteristics
                                                pcgmstat2000
numchild2000 pcgemploy2000 pcghours2000 welfare2000 logfaminc2000 ///
Baseline time varying

                [pweight=aipt_weight]                if povcat==3      ,
vce(cluster hhid2)

```

REFERENCES

- Aaronson, D. 1998. "Using sibling data to estimate the impact of neighborhoods on children's educational outcomes," *Journal of Human Resources* 33(4):915-946.
- Anderson, S., T. Leventhal, and V. Dupéré. 2014a. "Exposure to neighborhood affluence and poverty in childhood and adolescence and academic achievement and behavior," *Applied Developmental Science* 18:123–138.
- Anderson, S., T. Leventhal, and V. Dupéré. 2014b. "Residential mobility and the family context: a developmental approach," *Journal of Applied Developmental Psychology* 35(2):70-78.
- Aneshensel, C.S. and C.A. Sucoff. 1996. "The neighborhood context of adolescent mental health." *Journal of Health and Social Behavior* 37(4):293-310.
- Been, V., I.G. Ellen, A.E. Schwartz, L. Stiefel, and M. Weinstein. 2011. "Does losing your home mean losing your school?: Effects of foreclosures on the school mobility of children," *Regional Science and Urban Economics* 41(4):407-414.
- Briggs, X.D.S. and B.J. Keys. 2009. "Has exposure to poor neighbourhoods changed in America? Race, risk and housing locations in two decades." *Urban Studies* 46(2):429-458.
- Brumback, B.A., Hernán, M.A., Haneuse, S.J., and Robins, J.M. 2004. Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. *Statistics in Medicine* 23(5):749-767.
- Burdick-Will, J., J. Ludwig, S. Raudenbush, R.J. Sampson, L. Sanbonmatsu, and P. Sharkey. 2011. "Converging evidence for neighborhood effects on children's test scores: An experimental, quasi-experimental, and observational comparison." Pp.

- 255-276 in *Whither opportunity? Rising inequality, schools, and children's life chances*, edited by G.J. Duncan and R. Murnane. New York: Russell Sage.
- Clampet-Lundquist, S., and D.S. Massey. 2008. "Neighborhood effects on economic self-sufficiency: A reconsideration of the moving to opportunity experiment." *American Journal of Sociology* 114(1):107-143.
- Clampet-Lundquist, S., K. Edin, J.R. Kling, and G.J. Duncan. 2011. "Moving teenagers out of high-risk neighborhoods: How girls fare better than boys." *American Journal of Sociology* 116(4):1154-1189.
- Coleman, J.S. 1988. "Social capital in the creation of human capital." *American Journal of Sociology* S95-S120.
- Crowder, K. and S.J. South. 2011. "Spatial and temporal dimensions of neighborhood effects on high school graduation." *Social Science Research* 40(1):87-106.
- Desmond, M., A. Weihua, W. Richelle, and T. Ferriss. 2013. "Evicting children." *Social Forces* 92(1):303-327.
- Desmond, M. 2012. "Eviction and the reproduction of urban poverty." *American Journal of Sociology* 118:88-133.
- Dietz, R.D. 2002. "The estimation of neighborhood effects in the social sciences: An interdisciplinary approach." *Social Science Research*. 31(4):539-575.
- Dobbie, W., R.G. Fryer, and G. Fryer Jr. 2011. "Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone." *American Economic Journal: Applied Economics* 3(3):158-187.
- Downey, D.B. 2008. "Black/White differences in school performance: The oppositional culture explanation." *Annual Review of Sociology* 34(1):107-126.

- Duncan, G.J. and S.W. Raudenbush. 1999. "Assessing the effects of context in studies of child and youth development." *Educational Psychologist* 34:29-41.
- Duncan, G.J. and R. Murnane. 2011. "*Whither Opportunity?: Rising inequality, schools, and children's life chances.*" New York: Russell Sage.
- Dupere, V., T. Leventhal, R. Crosnoe, and E. Dion. 2010. "Understanding the positive role of neighborhood socioeconomic advantage in achievement: the contribution of the home, child care, and school environments." *Developmental Psychology* 46(5):1227.
- Eiland, L. and R. Romeo. 2013. "Stress and the developing adolescent brain." *Neuroscience* 249:162-171.
- Evans, W.N., W.E. Oates, and R.M. Schwab. 1992. Measuring peer group effects: A study of teenage behavior. *Journal of Political Economy* 966-991.
- Foster, E.M. and S. McLanahan. 1996. "An illustration of the use of instrumental variables: Do neighborhood conditions affect a young person's chance of finishing high school?" *Psychological Methods* 1(3):249.
- Fox, S.E., P. Levitt, and C.A. Nelson III. 2010. "How the timing and quality of early experiences influence the development of brain architecture." *Child Development* 81(1):28-40.
- Gennetian, L.A., L. Sanbonmatsu, L.F. Katz, J.R. Kling, M. Sciandra, J. Ludwig, G.J. Duncan, and R.C. Kessler. 2012. "The long-term effects of moving to opportunity on youth outcomes." *Cityscape* 137-167.

- Geronimus, A., J. Bound, and A. Ro. 2014. "Residential mobility across local areas in the United States and the geographic distribution of the healthy population." *Demography* 51(3):777-809.
- Greenman, E., K. Bodovski, and K. Reed. 2011. "Neighborhood characteristics, parental practices and children's math achievement in elementary school." *Social Science Research* 40(5):1434-1444.
- Hackman, D.A., M.J. Farah, and M.J. Meaney. 2010. "Socioeconomic status and the brain: mechanistic insights from human and animal research." *Nature Reviews Neuroscience* 11(9):651-659.
- Hanushek, E.A., J.F. Kain, and S.G. Rivkin. 2004. "Disruption versus Tiebout improvement: the costs and benefits of switching schools," *Journal of Public Economics* 88:1721–1746.
- Harding, D. J. 2003. "Counterfactual models of neighborhood effects: the effect of neighborhood poverty on dropping out and teenage pregnancy¹." *American Journal of Sociology* 109(3)676-719.
- Harding, D.J., L. Gennetian, C. Winship, L. Sanbonmatsu, and J.R. Kling. 2011. "Unpacking Neighborhood Influences on Education Outcomes: Setting the Stage for Future Research." Pp. 277-298 in *Whither Opportunity? Rising inequality, schools, and children's life chances*, edited by G.J. Duncan and R. Murnane. New York: Russell Sage Foundation.
- Hastie, T. J. and Tibshirani, R. J. 1990. *Generalized additive models*. Chapman and HaLL/CRC.

- Heckman, J.J. 2011. "The economics of inequality: The value of early childhood education." *American Educator* 35(1):31-35.
- Hernán, M.Á., B. Brumback and J.M. Robins 2000. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* 11(5):561-570.
- Hutchings, H.A., A. Evans, P. Barnes, et al. 2013. "Do children who move home and school frequently have poorer educational outcomes in their early years at school? An anonymised cohort study," *PLoS ONE* 8(8):e70601.
- Imai, K. and D.A. van Dyk. 2004. "Causal inference with general treatment regimes: Generalizing the propensity score." *Journal of the American Statistical Association* 99(September):854-866.
- Jackson, M.I. and R.D. Mare. 2007. "Cross-sectional and longitudinal measurements of neighborhood experience and their effects on children." *Social Science Research* 36(2):590-610.
- Janson, Lucas, W. Fithian, and T.J. Hastie. 2015. "Effective degrees of freedom: A flawed metaphor." *Biometrika* asv019.
- Jargowsky, P.A. and M. El Komi. 2011. "Before or after the bell? School context and neighborhood effects on student achievement." *Neighborhood and life chances: How place matters in modern America* 50-72.
- Jelleyman, T. and N. Spencer. 2008. "Residential mobility in childhood and health outcomes: A systematic review," *Journal of Epidemiology and Community Health* 62:584-592.

- Kling, J.R., J.B. Liebman, and L.F. Katz. 2007. "Experimental analysis of neighborhood effects." *Econometrica* 75(1):83-119.
- Kunz, J., M.E. Page, and G. Solon. 2003. "Are point-in-time measures of neighborhood characteristics useful proxies for children's long-run neighborhood environment?" *Economics Letters* 79(2):231-237.
- Manski, C. F. 1999. Identification problems in the social sciences. Harvard University Press.
- McGrew, K.S., J.K. Werder, and R.W. Woodcock. 1991. WJ-R technical manual: A reference on theory and current research. Chicago: Riverside.
- Morrison Gutman, L., V.C. McLoyd, and T. Tokoyawa. 2005. "Financial strain, neighborhood stress, parenting behaviors, and adolescent adjustment in urban African American families." *Journal of Research on Adolescence* 15(4):425-449.
- National Research Council and Institute of Medicine. 2010. *Student mobility: Exploring the impact of frequent moves on achievement: summary of a workshop*. Washington, DC: The National Academies Press.
- Pebley, A.R. and N. Sastry. 2004. "Neighborhoods, poverty and children's well-being." Pp. 119-146 in *Social dimensions of inequality*, edited by K. Neckerman. New York: Russell Sage Foundation.
- Peterson, C.E., A.R. Pebley, N. Sastry, K. Yuhas, B. Ghosh-Dastidar, A.C. Haas, J. Gregory, and M.P. Bitler. 2012. "The Los Angeles Family and Neighborhood Survey, Wave 2: Users' guide and codebook." Report WR-240/20-LAFANS, Labor and Population Program, RAND Corporation.

- Pettit, B. and S. McLanahan. 2003. "Residential mobility and children's social capital: Evidence from an experiment." *Social Science Quarterly* 84(3):632-649.
- Phillips, M. 2011a. "Ethnic and social class disparities in academic skills." Diversity in American higher education: Toward a more comprehensive approach:7.
- Phillips, M. 2011b. "Parenting, time use, and disparities in academic outcomes." *Whither Opportunity* 207-228.
- Plotnick, R.D. and S.D. Hoffman, 1999. The effect of neighborhood characteristics on young adult outcomes: Alternative estimates. *Social Science Quarterly* 1-18.
- Pribesh, S. and D.B. Downey. 1999. "Why Are residential and school moves associated with poor school performance?" *Demography* 36(4):521-534.
- Quillian, L. 2003. "How long are exposures to poor neighborhoods? The long-term dynamics of entry and exit from poor neighborhoods." *Population Research and Policy Review* 22(3):221-249.
- Robins, J.M., M.A. Hernan, and B.Brumbach 2000. Marginal structural models and causal inference in epidemiology. *Epidemiology* 11(5):550-560.
- Roksa, J. and D. Potter. 2011. "Parenting and academic achievement intergenerational transmission of educational advantage." *Sociology of Education* 84(4):299-321.
- Rosenbaum, P.R. and D.B. Rubin. 1983. "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70:41-55.
- Sampson, R. J. 2008. Moving to inequality: Neighborhood effects and experiments meet structure. *American Journal of Sociology* 114(11):189.
- Sampson, R., S. Raudenbush, and F. Earls. 1997. "Neighborhoods and violent crime: A multilevel study of collective efficacy." *Science* 277:918-924.

- Sampson, R.J. 2011. "Neighborhood effects, causal mechanisms, and the social structure of the city." Pp. 227-250 in *Analytical sociology and social mechanisms*, edited by P. Demeulenaere. Cambridge and New York: Cambridge University Press.
- Sampson, R.J. 2012. *Great American city: Chicago and the enduring neighborhood effect*. Chicago: University of Chicago Press.
- Sampson, R.J., J.D. Morenoff, and E. Felton. 1999. "Beyond social capital: Spatial dynamics of collective efficacy for children." *American Sociological Review* 64(5):633-660.
- Sampson, R.J., P. Sharkey, and S.W. Raudenbush. 2008. "Durable effects of concentrated disadvantage on verbal ability among African-American children." *Proceedings of the National Academy of Sciences* 105(3):845-852.
- Sampson, R.J., S.W. Raudenbush, and F. Earls. 1997. "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy." *Science* 277: 918–24.
- Sanbonmatsu, L., J.R. Kling, G.J. Duncan, and J. Brooks-Gunn. 2006. "Neighborhoods and academic achievement: Results from the moving to opportunity experiment." *Journal of Human Resources* 41(1):649-691.
- Sastry, N. 2012. "Neighborhood effects on children's achievement: A review of recent research." in *Oxford Handbook on Child Development and Poverty*, edited by V. Maholmes and R.B. King. New York: Oxford University Press.
- Sastry, N., and A.R. Pebley. 2003. "Non-response in the Los Angeles Family and Neighborhood Survey." Report DRU-2400/7-LAFANS, Labor and Population Program, RAND Corporation.

- Sastry, N. and A.R. Pebley. 2010. "Family and neighborhood sources of socioeconomic inequality in children's achievement." *Demography* 47(3):777-800.
- Sharkey, P. 2010. "The acute effect of local homicides on children's cognitive performance." *Proceedings of the National Academy of Sciences* 107(26):11733-11738.
- Sharkey, P. and J.W. Faber. 2014. "Where, when, why, and for whom do residential contexts matter? Moving away from the dichotomous understanding of neighborhood effects." *Annual Review of Sociology* 40:559-579.
- Sharkey, P.T., N. Tirado-Strayer, A.V. Papachristos, and C.C. Raver. 2012. "The effect of local violence on children's attention and impulse control." *American Journal of Public Health* 102(12):2287-2293.
- Shonkoff, J.P., A.S. Garner, B.S. Siegel, M.I. Dobbins, M.F. Earls, L. McGuinn, J. Pascoe, and D.L. Wood. 2012. "The lifelong effects of early childhood adversity and toxic stress." *Pediatrics* 129(1):e232-e246.
- Solon, G., M.E., Page, and G.J. Duncan. 2000. Correlations between neighboring children in their subsequent educational attainment. *Review of Economics and Statistics* 82(3):383-392.
- Timberlake, J.M. 2009. "Effects of household and neighborhood characteristics on children's exposure to neighborhood poverty and affluence." *Social Science Research* 38(2):458-476.
- Walker, S.P., T.D. Wachs, S. Grantham-McGregor, M.M. Black, C.A. Nelson, S.L. Huffman, H. Baker-Henningham, S.M. Chang, J.D. Hamadani, B. Lozoff, J.M.M. Gardner, C.A. Powell, A. Rahman, and L. Richter. 2011. "Inequality in early

childhood: risk and protective factors for early child development.” *The Lancet* 378(9799):1325-1338.

Wheaton, B. and P. Clarke. 2003. “Space meets time: Integrating temporal and contextual influences on mental health in early adulthood.” *American Sociological Review* 68(5):680-706.

Wilson, W.J. 1987. *The Truly Disadvantaged*. Chicago: University of Chicago Press.

Wilson, W.J. 2012. *The truly disadvantaged: The inner city, the underclass, and public policy*. University of Chicago Press.

Wodtke, G.T., D.J. Harding, and F. Elwert. 2011. “Neighborhood effects in temporal perspective.” *American Sociological Review* 76(5):713-736.

Ziol-Guest, K.M., and C.C. McKenna, C. C. 2014. “Early childhood housing instability and school readiness,” *Child Development* 85(1):103-113.

Table 1. Neighborhood Characteristics by Quintile of Concentrated Disadvantage Score for Los Angeles County Census Tracts, 2000 to 2008

Variable	1 st Quintile		2 nd Quintile		3 rd Quintile		4 th Quintile		5 th Quintile	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Percent female-headed households	2.46	(1.70)	5.17	(2.01)	7.76	(3.08)	11.31	(3.86)	16.29	(6.14)
Percent families with income <\$25k	5.68	(4.30)	11.80	(4.80)	18.89	(6.12)	29.46	(7.35)	46.53	(11.36)
Percent population non-white race	18.57	(12.26)	38.10	(17.52)	51.57	(15.73)	59.80	(15.24)	69.04	(12.53)
Percent population in poverty	5.59	(3.56)	7.77	(3.62)	12.65	(4.77)	20.78	(7.26)	34.00	(9.39)
Percent population age <18 years	15.77	(8.66)	21.74	(5.41)	24.71	(5.67)	28.92	(5.90)	33.68	(6.57)
Percent households receiving welfare	0.96	(0.87)	2.02	(1.15)	4.02	(1.79)	7.12	(2.79)	12.94	(5.53)
Concentrated disadvantage score	-1.12	(0.26)	-0.66	(0.11)	-0.19	(0.16)	0.47	(0.23)	1.47	(0.55)

Table 2. Means (and Standard Deviations) for L.A.FANS Children in Analysis Sample

Variable	(1) Unweighted	(2) Weighted
Expected average exposure	0.480 (0.974)	0.272 (0.859)
Expected recency	-0.048 (0.086)	-0.039 (0.094)
Observed duration-weighted exposure	3.684 (1.338)	3.498 (1.262)
Observed recency	-0.039 (0.130)	-0.040 (0.143)
Child's age at baseline (years)	7.255 (2.522)	7.229 (2.514)
Child sex is male (ref.: female)	0.513 (0.500)	0.517 (0.500)
Child's race: Latino	0.643 (0.480)	0.531 (0.499)
Child's race: black	0.068 (0.252)	0.104 (0.306)
Child's race: white or other	0.289 (0.454)	0.365 (0.482)
Child low birthweight (ref.: normal birthweight)	0.063 (0.244)	0.093 (0.291)
Mother's age at child's birth	28.159 (6.500)	28.229 (6.019)
Mother was married at child's birth (ref.: no married)	0.588 (0.493)	0.600 (0.49)
Mother's reading achievement score	84.633 (19.011)	85.021 (18.947)
Family head is HS graduate (ref.: not HS graduate)	0.614 (0.487)	0.673 (0.469)
Family owns home (ref.: does not own)	0.385 (0.487)	0.425 (0.495)
Baseline: mother was married (ref.: not married)	0.700 (0.459)	0.710 (0.454)
Baseline: mother's number of children	1.740 (1.001)	1.854 (1.032)
Baseline: mother was employed (ref.: not employed)	0.667 (0.472)	0.675 (0.469)
Baseline: mother's hours of work	24.154 (18.923)	25.029 (19.076)
Baseline: receiving public assistance (ref.: no receipt)	0.054 (0.225)	0.049 (0.217)
Baseline: family income (\$)	17,102 (29.666)	17,890 (32.656)

Note: Standard deviation in parentheses.

Table 3. Component Weights and Variances from Principal Component Analysis of Neighborhood Disadvantage Scores between 2001 and 2006

Year of disadvantage score	1 st PC weight	2 nd PC weight
2001	0.40	-0.51
2002	0.41	-0.39
2003	0.41	-0.14
2004	0.41	0.01
2005	0.41	0.39
2006	0.40	0.65
Proportion of total variance explained	0.95	0.03

Table 4. IPT Weights, Stabilized IPT Weights, Attrition Weights, Stabilized IPT-Attrition Weights

Variable	Mean	SD	Percentiles			
			1st	25th	75th	99th
IPT weight	1.75e+08	4.52e+09	1.10	1.58	12.14	467295.9
Stabilized IPT weight	1.04	1.16	0.13	0.85	1.04	5.26
Attrition weights	0.88	0.80	0.07	0.38	1.11	4.10
Stabilized IPT weight x attrition weight	0.97	2.21	0.20	0.32	1.02	5.38

Table 5. Regression Model Results for Effects of Neighborhood Disadvantage
Mean Exposure and Recency of Exposure on Reading and Math Scores

Variable	Math		Reading	
	Exposure	Recency	Exposure	Recency
Child's age at baseline	-1.419*** (0.278)	-1.382*** (0.142)	-0.797** (0.307)	-0.798*** (0.151)
Child is male (ref.: female)	1.333 (1.419)	1.628* (0.720)	-1.877 (1.586)	-1.593* (0.794)
Child's race: black (ref.: Latino)	-0.972 (2.236)	-0.722 (1.173)	0.117 (2.482)	0.058 (1.220)
Child's race: white or other (ref.: Latino)	-1.188 (3.313)	-0.743 (1.708)	5.310 (3.663)	5.350** (1.818)
Child low birthweight (ref.: normal birthweight)	0.719 (2.897)	0.457 (1.446)	2.757 (3.230)	2.623 (1.560)
Mother's age at child's birth	0.152 (0.126)	0.148* (0.070)	0.274 (0.140)	0.273*** (0.072)
Mother married at child's birth (ref.: not married)	-0.615 (2.166)	-0.258 (1.409)	0.385 (2.226)	0.467 (1.485)
Mother's reading achievement score	0.099* (0.050)	0.099*** (0.027)	0.185** (0.056)	0.184*** (0.030)
Family head is HS graduate (ref.: not HS graduate)	3.171† (1.858)	3.320*** (0.961)	8.142*** (2.072)	8.106*** (1.016)
Family owns home (ref.: does not own)	0.182 (1.594)	-0.204 (0.815)	-1.602 (1.781)	-1.710 (0.902)
Baseline: mother was married (ref.: not married)	2.467 (1.989)	2.213* (1.079)	-0.064 (2.168)	-0.294 (1.161)
Baseline: mother's number of children	-0.186 (0.835)	0.010 (0.451)	0.578 (0.919)	0.702 (0.460)
Baseline: mother was employed (ref.: not employed)	0.232 (1.532)	0.221 (0.928)	-0.072 (1.642)	0.141 (0.896)
Baseline: mother's hours of work	0.060 (0.055)	0.054 (0.031)	-0.014 (0.059)	-0.034 (0.030)
Baseline: receiving public assistance (ref.: no receipt)	2.744 (4.166)	1.988 (2.458)	0.315 (4.486)	-0.059 (2.225)
Baseline: family income (log \$)	-0.078 (0.263)	-0.080 (0.215)	0.037 (0.296)	-0.016 (0.235)
Constant	97.045*** (7.114)	96.167*** (4.268)	86.146*** (8.008)	86.643*** (4.380)
Smoothed expected average exposure and recency				
Effective degrees of freedom	6.397	6.277	6.058	5.897
Reference degrees of freedom	7.246	7.105	6.870	6.712
F-test	3.578***	9.293***	2.526*	7.060***
Smoothed effects of exposure and of expected average exposure and recency				
Effective degrees of freedom	2.660		3.324	
Reference degrees of freedom	96.000		97.000	
F-test	188.382***		106.533***	

continued

Variable	Math		Reading	
	Exposure	Recency	Exposure	Recency
Smoothed effects of recency and of expected average exposure and recency				
Effective degrees of freedom		3.419		3.545
Reference degrees of freedom		97.000		97.000
F-test		751.359***		1.444***
Number of observations	611	611	616	616
R ²	0.190	0.195	0.172	0.177

Note: Standard errors in parentheses; † p<0.10, * P<0.05, ** p<0.01, *** p<0.001.

Table 6. IPTW-Weighted Regression Model Results for Effects of Duration-Weighted Exposure and Recency of Exposure to Neighborhood Disadvantage on Math and Reading Scores

Variable	Math		Reading	
	Exposure	Recency	Exposure	Recency
Observed duration-weighted exposure	-3.456** (1.019)	.	-2.486+ (1.315)	.
Observed recency	.	-11.463 (7.879)	.	-21.592** (7.344)
Child's age at baseline	-1.647*** (0.372)	-1.635*** (0.393)	-0.732+ (0.384)	-0.732+ (0.381)
Child is male (ref.: female)	-0.212 (1.708)	-0.047 (1.733)	-4.744* (1.924)	-4.458* (1.919)
Child's race: black (ref.: Latino)	3.274 (3.582)	2.163 (4.121)	9.014 (5.332)	8.371 (5.584)
Child's race: white or other (ref.: Latino)	3.205 (2.997)	0.291 (2.943)	2.550 (3.146)	0.050 (3.153)
Child low birthweight (ref.: normal birthweight)	1.196 (3.198)	1.406 (3.736)	-1.099 (3.226)	-1.104 (3.092)
Mother's age at child's birth	0.223 (0.155)	0.367* (0.159)	0.375+ (0.193)	0.497* (0.192)
Mother married at child's birth (ref.: not married)	1.233 (2.773)	1.094 (2.951)	0.936 (2.965)	0.819 (3.115)
Mother's reading achievement score	0.113 (0.060)	0.142* (0.057)	0.135+ (0.070)	0.154* (0.067)
Family head is HS graduate (ref.: not HS graduate)	4.546 (2.205)	6.272** (2.211)	7.018** (2.636)	8.463** (2.460)
Family owns home (ref.: does not own)	-1.810 (2.081)	-0.482 (2.038)	-2.975 (2.222)	-1.467 (2.187)
Baseline: mother was married (ref.: not married)	2.907 (2.961)	2.949 (3.071)	-0.533 (3.124)	-0.434 (3.235)
Baseline: mother's number of children	0.142 (0.912)	0.411 (0.963)	-0.528 (1.122)	-0.339 (1.142)
Baseline: mother was employed (ref.: not employed)	0.924 (2.511)	0.238 (2.516)	0.335 (3.366)	-0.169 (3.309)
Baseline: mother's hours of work	-0.033 (0.074)	-0.011 (0.078)	-0.066 (0.091)	-0.044 (0.092)
Baseline: receiving public assistance (ref.: no receipt)	3.640 (3.404)	1.129 (3.606)	-1.425 (3.596)	-2.485 (4.242)
Baseline: family income (log \$)	0.318 (0.283)	0.4331 (0.258)	-0.113 (0.305)	-0.031 (0.278)
Constant	102.318*** (7.717)	81.795*** (8.736)	97.892*** (13.023)	82.080*** (9.386)
Number of observations	568	568	573	573
F-test	6.19***	4.90***	6.10***	5.55***

Note: Standard errors in parentheses; † p<0.10, * P<0.05, ** p<0.01, *** p<0.001.

Table 7. IPTW-Weighted Regression Model Results for Effects of Duration-Weighted Exposure and Recency of Exposure to Neighborhood Disadvantage on Math and Reading Test Scores for the Full Sample (1) and Stratified by Poverty Status (Cols. 2–4)

Variable	(1) Full sample	(2) Very Poor	(3) Poor	(4) Not Poor
<u>Math scores</u>				
Model 1				
Observed duration-weighted exposure	-3.456*** (1.019)	0.678 (4.229)	-0.717 (1.447)	-4.767** (1.666)
Model 2				
Observed recency	-11.463 (7.879)	2.057 (11.481)	-7.941 (8.927)	-45.976** (16.856)
Model 3				
Observed duration-weighted exposure	-3.466** (1.007)	0.326 (4.799)	-0.542 (1.442)	-4.789** (1.655)
Observed recency	-11.685+ (7.152)	1.410 (12.574)	-7.653 (8.943)	-46.186** (16.013)
Observations	568	193	176	199
<u>Reading scores</u>				
Model 1				
Observed duration-weighted exposure	-2.486+ (1.315)	-6.197 (7.648)	-1.963 (1.715)	-5.718** (1.950)
Model 2				
Observed recency	-21.592** (7.344)	-28.552 (20.846)	22.326* (8.861)	-26.973 (19.257)
Model 3				
Observed duration-weighted exposure	-2.511+ (1.315)	2.355 (7.838)	-1.516 (1.740)	-5.733** (2.015)
Observed recency	-21.792** (7.051)	-33.493 (20.566)	-21.577+ (8.882)	-27.338 (17.683)
Observations	573	194	180	199

Note: All models also control for child's age at baseline, child sex, child's race, child's low birthweight, mother's age at child's birth, mother's marital status at child's birth, mother's reading achievement score, family head's education, family homeownership, and baseline mother's marital status, mother's number of children, mother's employment, mother's hours of work, family receipt of public assistance, and family income. Standard errors in parentheses; + p<0.10, * P<0.05, ** p<0.01, *** p<0.001.

Figure 1. Scatter and Density Plots for Average Expected Treatment vs. Recency of Treatment

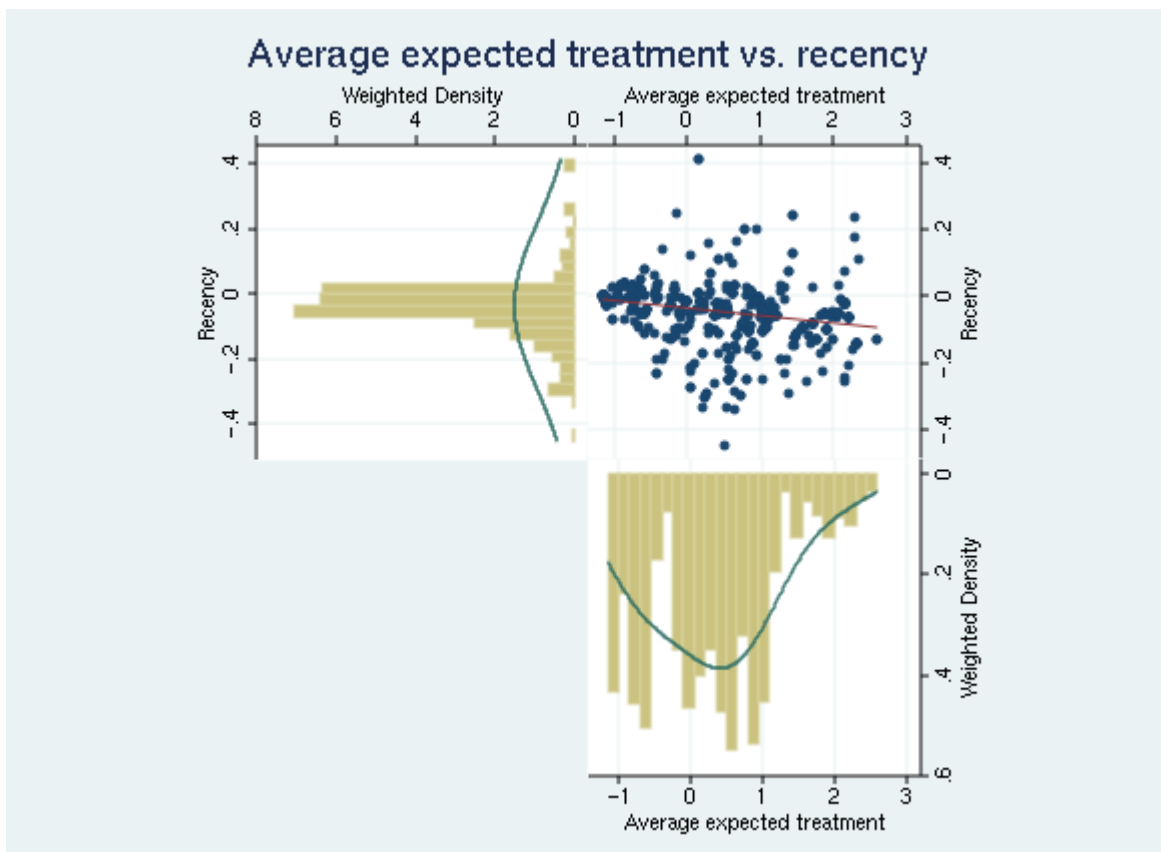
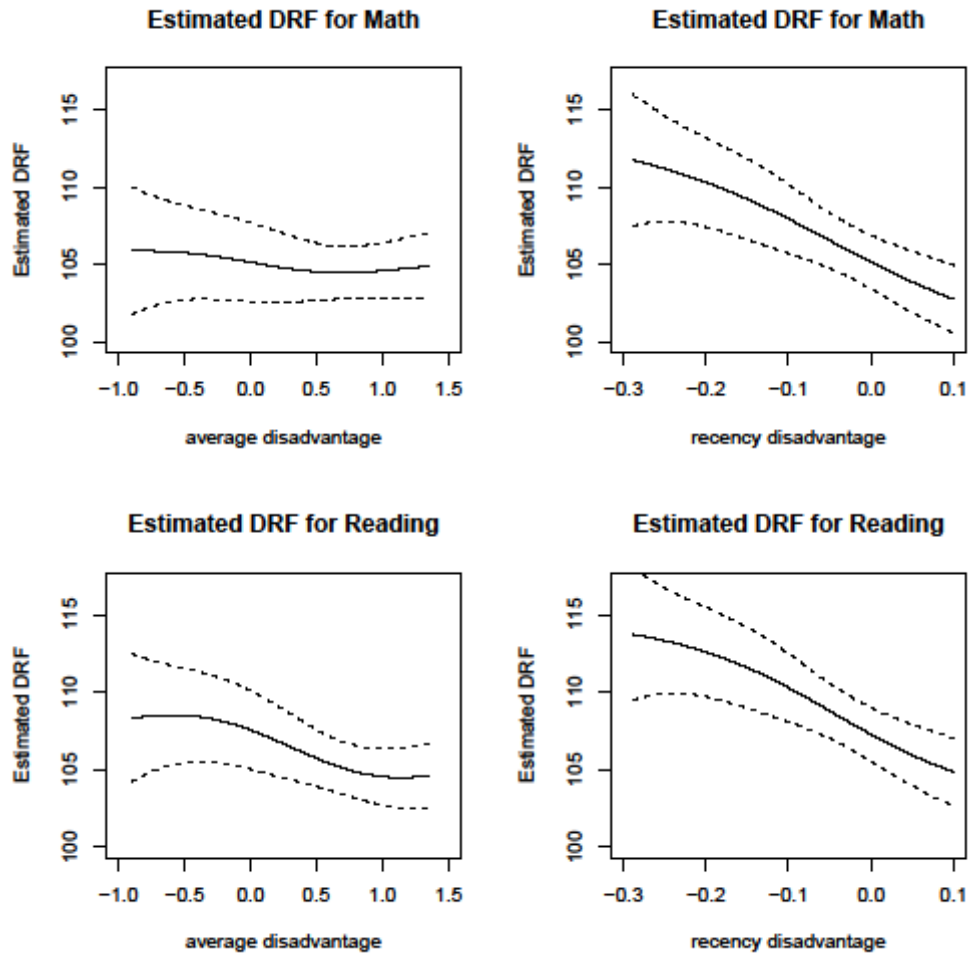


Figure 2. Estimated Effects of Average Expected Neighborhood Disadvantage and Average Expected Recency of Exposure to Neighborhood Disadvantage on Reading and Math Scores



Note: DRF = dose-response function.