Social Science & Medicine 73 (2011) 737–743

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Short report

Moving to Opportunity: Does long-term exposure to 'low-poverty' neighborhoods make a difference for adolescents?

Tama Leventhal [a,*], Véronique Dupéré [b]

[a] Tufts University, USA
[b] Université de Montréal, Canada

ARTICLE INFO

Article history:
Available online 23 July 2011

Keywords:
USA
Gender
Neighborhoods
Poverty
Social experiment
Adolescence
Mental health
Propensity score analysis

ABSTRACT

This study re-analyzed data on adolescent health outcomes ($N = 1780$; M age $= 15.15$, $SD = 2.30$) from a 5-year evaluation of the Moving to Opportunity (MTO) Program. The MTO program is a randomized experiment conducted in five cities in the United States (Baltimore, Boston, Chicago, Los Angeles and New York) in which low-income families living in public housing in 'high-poverty' neighborhoods were offered vouchers and assistance to move to 'low-poverty' neighborhoods. The objective was to reexamine program effects as a function of exposure to 'low-poverty' neighborhoods to determine whether beneficial effects reported for girls' mental health and behavior generalized to other outcomes for girls and to boys. As found in previous evaluations, girls in the MTO program group, whose families remained in 'low-poverty' neighborhoods for comparatively long periods (approximately 5 years), had better mental health and engaged in fewer risky behaviors than a matched control sample of girls, whose families stayed in 'high-poverty' neighborhoods. Further, additional benefits for girls were seen in the education domain. Adverse program effects on boys' behavior problems, reported in other MTO research using different methods, were not evident in our analysis. Findings suggest that programs relocating low-income families to 'low-poverty' neighborhoods should provide supports to families and to receiving communities to promote residential stability and social integration.

Mounting non-experimental research documents associations between growing up in a poor neighborhood and a range of adverse adolescent outcomes such as mental health problems, crime and delinquency, risky sexual behavior, and high school dropout (see Leventhal, Dupéré, & Brooks-Gunn, 2009; and Sampson, Morenoff, & Gannon-Rowley, 2002 for reviews). These findings focused attention on the handful of experimental studies attempting to replicate such links because of problems of selection bias inherent in non-experimental research, where families self-select into neighborhoods. In other words, unobserved factors related to neighborhood choice may drive any observed associations between neighborhood poverty and adolescent outcomes. Experimental studies have been conducted in the context of housing programs for low-income families. Because programs cannot serve all eligible or interested families, selection of neighborhoods is rarely based on choice, but instead on housing or subsidy availability. The importance of these experimental studies clearly extends beyond science to their policy relevance because they indicate which policy strategies may best ameliorate urban poverty and its effects.

The Moving to Opportunity for Fair Housing Demonstration (MTO) is the most recent and well-known of the experimental studies and the only one with a true experimental design (for details, see Orr et al., 2003). Families in five US cities (Baltimore, Boston, Chicago, Los Angeles and New York) were recruited from public housing projects located in census tracts with poverty rates in excess of 40%, a cut-off typically considered 'high-poverty' in the US. In the US, family (or individual) poverty is determined by absolute thresholds set by the US government, adjusted annually for the consumer price index, and based on annual income and household size. We define neighborhood 'high-poverty' rates as those where 40% or more of the residents are living below the US poverty threshold. Those who volunteered for the program were assigned to one of three conditions: (a) the *experimental group* who received both housing vouchers (or subsidies) to be applied toward rent in the private housing market and housing counseling to move to 'low-poverty' neighborhoods (defined as those with less than 10% of the residents living below the US poverty threshold); (b) the *voucher comparison group* who received vouchers under the regular, geographically unrestricted program (this group is not

* Corresponding author.
E-mail address: tama.leventhal@tufts.edu (T. Leventhal).

0277-9536/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.socscimed.2011.06.042

a focus of the current study); or (c) the *in-place control group* who did not receive vouchers or counseling but continued to receive project-based support. Although higher than expected, the take-up rate among experimental families was only 47%. That is, not quite half of the families in the experimental group used their vouchers to move to new neighborhoods (or complied with the program). Because experimental complier families were only required to stay in low-poverty neighborhoods for at least one year, 66% of these families moved out of their initial placement after the first year, typically to neighborhoods with higher poverty levels (Orr et al., 2003).

The MTO findings on adolescent outcomes from a 5-year interim evaluation have been widely circulated and quite controversial (Kling, Liebman, & Katz, 2007; Kling, Ludwig, & Katz, 2005; Orr et al., 2003). Adolescent girls in the experimental group reported less psychological distress, anxiety, and substance use and were less likely to be arrested (for both violent and property crimes) than girls in the control group. In contrast, adolescent boys in the experimental group reported more behavior problems and substance use and were more likely to be arrested for property crimes than boys in the control group. Qualitative work attempting to explain these mixed results has focused on the gender differences in program effects, highlighting changes in gender-specific aspects of neighborhood safety (Popkin, Leventhal, & Weismann, 2010) and gender differences in adaptation, particularly around peer networks (Clampet-Lundquist, Duncan, Edin, & Kling, 2006). Quantitative investigations have focused less on gender differences and more on the lack of overall beneficial findings including the modest compliance or take-up rate among the experimental group (i.e., low proportion of eligible families who moved to low-poverty neighborhoods through the MTO program), selection processes among the experimental compliers, cross-site differences, and low prevalence of adverse outcomes among the control group (Kling et al., 2007).

The current study follows in these footsteps by drawing attention to another challenge of interpreting program effects—differential length of exposure to low-poverty neighborhoods within the experimental group. In MTO, most experimental families who complied with the program had only modest (i.e., short-term) exposure to low-poverty neighborhoods. A family of models referred to as propensity score analysis is gaining popularity in the behavioral sciences for addressing this problem (see Guo & Fraser, 2010). The goal of these methods is to rebalance the control group to make it comparable to the subgroup of individuals in the experimental group who self-select into treatment, including varying levels of treatment or, in the case of MTO, varying lengths of exposure to low-poverty neighborhoods. In this study, we used propensity score analysis to examine the association between moving *and* staying in low-poverty neighborhoods (or long-term exposure to low-poverty neighborhoods), compared with staying in high-poverty neighborhoods, and adolescent outcomes. We sought to determine whether beneficial effects reported for experimental girls' mental health and behavior in the 5-year interim evaluation (Kling et al., 2007; Orr et al., 2003) would generalize to other outcomes for girls and to boys in this subgroup of families with long-term exposure to low-poverty neighborhoods.

Method

Design and description of the Moving to Opportunity for Fair Housing Demonstration

Baseline interviews were conducted with all MTO families from 1994 to 1999—prior to random assignment and relocation—with one adult family member identified as the head of household ($N = 4608$). During this interview, demographic information was obtained with limited data gathered on each household member including children. Approximately five to seven years after mover families had relocated, an interim evaluation was conducted in 2002 ($N = 4248$; 90% response rate). Extensive in-home, structured interviews were conducted with heads of households and up to two randomly selected children per household between the ages of 5- and 19-years (Orr et al., 2003). Informed consent was obtained from all participants 18 years of age and older. In addition, participants were made aware of the fact that a Federal Certificate of Confidentiality was obtained for the study, which made it illegal to disclose their responses to anyone (the exception being if participants posed a potential threat to themselves or others). All procedures and forms were approved by human subjects review boards at the appropriate institutions.

Sample

This study focuses only on experimental group and in-place control adolescents between 12 and 19 years of age at the time of the interim evaluation ($N = 2531$). Of these adolescents, 400 did not participate in the evaluation and 351 had missing data and were excluded, leaving a final sample of 1780 ($n = 755$ control and $n = 1025$ experimental). In the experimental group, 478 young people had families who complied with the program and moved to low-poverty neighborhoods (this group is also commonly referred to as "experimental adherent families," but is referred to hereafter as "compliers" to be consistent with past MTO research). About a quarter of this complier group (119/478 = 24.9%) remained in low-poverty neighborhoods for at least 50% of the observation window, for an average duration of 61 months ($SD = 14.4$, hereafter referred to as the "long-term [LT] exposure" group).

Table 1 provides descriptive information on baseline characteristics of control and experimental (full sample, compliers and LT-exposure subgroup) groups. Experimental participants were significantly more likely to have mothers who were Black, who completed secondary school, and who were very dissatisfied with their neighborhoods than mothers in the control group. These differences were magnified in the LT-exposure subgroup, and new differences also emerged, with LT-exposure mothers being less likely to be Latinas and more likely to be enrolled in secondary school (*vs.* control mothers). Significant differences between groups also were seen for site; although LT-exposure families were distributed relatively evenly across the five sites, a significant over-representation of LT-exposure families was from Baltimore and a significant under-representation was from Los Angeles. Boys and girls were equally distributed across the LT-exposure subgroup in each site (results available from authors upon request). In terms of post-randomization contextual characteristics, descriptive statistics reveal that the experimental group spent more time in less disadvantaged neighborhoods and secondary schools than the control group, with differences especially marked in the LT-exposure subgroup (see Table 2).

Analytic plans

Propensity score methods were used to address the research questions. A propensity score is "a conditional probability of a study participant receiving treatment given observed covariates" (Guo & Fraser, 2010). In this study, we estimated the propensity for long-term exposure to low-poverty neighborhoods, that is, for experimental families to stay in a low-poverty neighborhood after being offered a voucher and housing assistance from the MTO program. Following Hill, Brooks-Gunn, and

Table 1
Baseline profiles of control and experimental groups and outcomes descriptive statistics.

Background variables/outcomes (% except otherwise noted)	In-place controls ($n = 755$)	MTO experimental group		
		Full sample ($n = 1025$)	Compliers ($n = 478$)	Long-term exposure ($n = 119$)
	M/%	M/%	M/%	M/%
Mother/household				
Age (years)	40.3	40.3	39.4*	39.7
Hispanic ethnicity	30.1	27.8	27.4	16.8*
Race				
Black	62.5	68.0*	66.9	70.6†
White	7.8	6.8	7.7	8.4
Other	29.7	25.2*	25.3†	21.0†
Completed high school	34.4	39.0*	40.6*	46.2*
Works for pay	24.2	26.0	25.3	27.7
Enrolled in school	12.2	12.3	15.5†	21.0*
Household income ($)	9894	9961	9711	9510
Has a car	20.4	19.4	24.9†	26.9
Household size[a]	1.8	1.8	1.8	1.8
No teen (13–17) in hh	50.1	49.4	53.3	58.0
Very dissatisfied with neigh.	41.1	46.3*	50.2*	49.6†
Moved > 3 times in past 5 years	9.5	7.4	9.0	6.7
Wants to move to get…				
Away from gangs/drugs	79.5	77.6	79.7	83.2
Access to better schools	49.7	49.4	51.7	50.4
Child				
Age (years)	15.1	15.2	15.1	14.8
Male	49.5	48.4	47.7	52.9
Child in a special class for…				
Gifted children	16.0	14.5	14.6	21.8
Learning/behavior problems	17.5	20.3	21.1	19.3
Health problem requiring special medicine/equipment	6.9	9.1†	9.6†	9.2
Site				
Baltimore	17.1	15.8	18.0	26.9*
Boston	21.6	16.6*	16.5*	23.5
Chicago	17.1	28.1*	19.5	16.0
LA	25.4	17.3*	24.7	14.3*
NYC	18.8	22.2†	21.3	19.3
Outcomes				
Internalizing problem	11.4	10.1	10.3	8.4
Distress	0.27	0.25	0.26	0.22†
Marijuana use	24.0	22.7	20.9	19.3
Alcohol use	31.0	31.3	31.4	31.1
Tobacco use	21.5	22.2	21.1	21.0
Sexual intercourse	46.1	48.4	45.4	42.0
Behavior problems	0.36	0.38	0.38	0.38
Delinquency	0.10	0.09	0.09	0.1
Reading	511.0	512.4	511.3	516.2*
Math	513.5	514.6	513.2	516.9*
Hs graduate/in school full-time	81.9	82.5	84.9	89.1†

Note. hh = household; neigh = neighborhood hs = high school.
*Statistically significantly ($p < 0.05$) or †marginally ($p < 0.10$) different from mean (t test) or proportion (χ^2 test) of same variable in in-place control group.
[a] Household size is coded: 0 = 2 or less; 1 = 3; 2 = 4; 3 = 5 or more.

Waldfogel (2003), we conducted a logistic regression to predict LT-exposure within the experimental group from a large set of covariates (listed in Table 1). These covariates were selected from a larger pool, based on theoretical expectations that they were likely to influence selection into the program, as well as on preliminary empirical comparisons. The regression coefficients obtained were then used, along with the values of the background characteristics, to calculate propensity scores for the control group. Then, various matching techniques were used to create balanced groups of control and LT-exposure participants with similar propensity scores. For one-on-one matching techniques, matches were done within site and sex. For matching techniques using more than one match per LT-exposure participant, matching was performed sequentially for the whole sample and for girls and boys separately, but to have sufficient potential matches, matching was not restricted by site.

In this study, we used three variants of propensity score methods to assess the association between long-term exposure to low-poverty neighborhoods via the MTO program (*vs.* staying in high-poverty neighborhoods) and adolescent outcomes. The first, *nearest neighbor matching*, matches every LT-exposure participant with the control participant with the closest propensity score. This control participant is then removed from the pool of potential matches (i.e., not replaced). To avoid matching participants with dissimilar propensity scores, it is possible to define a caliper, this is, a maximum acceptable distance between the propensity scores of two matched participants. If no match falls within the specified caliper, then the LT-exposure participant with no close match is discarded. Once a matched-control group is created, analysis comparing LT-exposure and control groups, such as multivariate regression, may be performed to obtain "treatment effects" (note in this section we refer to LT-exposure estimates as treatment or

Table 2
Neighborhood and school characteristics for control and experimental groups.[a]

Neighborhood and school variables	*N* Valid	In-place controls	MTO experimental group		
			Full sample	Compliers	High dosage
		M (SD)	M (SD)	M (SD)	M (SD)
Months in tract < 10% poor[b]	1780	1.54 (7.65)	10.73*** (21.27)	20.76*** (26.30)	61.29*** (14.38)
Neighborhood problems[c]	1723	3.11 (2.08)	2.30*** (2.14)	1.43*** (1.72)	1.19*** (1.63)
Neighborhood satisfaction[d]	1711	2.83 (1.38)	2.42*** (1.35)	2.02*** (1.19)	1.91*** (1.23)
Safety[e]	1719	0.73 (0.45)	0.86*** (0.35)	0.92*** (0.27)	0.96*** (0.20)
School[f]					
School percentile rank	1069	17.24 (15.50)	18.56*** (14.87)	23.64*** (15.02)	29.83*** (17.32)
% Eligible for free lunch	1102	68.18 (18.57)	62.69*** (20.23)	58.13*** (20.78)	46.29*** (23.82)
% Black students	1501	53.76 (33.17)	57.12* (33.84)	50.29† (32.51)	46.00* (30.46)
% Hispanic students	1501	33.24 (31.03)	27.05*** (28.60)	28.18** (26.94)	20.43*** (24.40)

***$p < 0.001$. **$p < 0.01$. *$p < 0.05$. †$p < 0.10$.
[a] Because of missing data, descriptive statistics are based on varying *n* values.
[b] Number of months since randomization spent in non-poor neighborhoods (census tract with poverty rate under 10%).
[c] Sample adults' reports of neighborhood problems based on a 6-item (dichotomous) measure (trash, graffiti, abandoned buildings, public drinking, groups hanging out, police not responding when called).
[d] Adult-reported 1-item measure of neighborhood satisfaction ranging from 1 = very satisfied to 5 = very dissatisfied.
[e] Adult–reported 1-item dichotomous measure of neighborhood safety during the day.
[f] Weighted average based on the number of years spent at each school attended since random assignment calculated for four school characteristics: percentile rank on state assessments, the percentage of students eligible to participate in the Free Lunch Program, of Black students, and of Hispanic students.

program effects, but acknowledge that they are associations because we are not adhering strictly to the experimental design). For this study, the nearest neighbor matching routine was implemented with the *psmatch2* program in Stata 10 (Leuven & Sianesi, 2003) with a caliper representing one-quarter of a standard deviation of the propensity score, a commonly used threshold.

One-on-one matching strategies such as nearest neighbor matching have been criticized because they discard data and potentially valuable information. To address this problem and increase matching efficiency and statistical power, a method called *optimal or full matching*, which uses information from all available participants has been developed (see Guo & Fraser, 2010). Optimal/full matching creates match sets in which each LT-exposure participant is matched with at least one control participant and potentially many more when many control participants have close propensity scores; conversely, every control participant matches to one or more LT-exposure participants. The matching procedure minimizes the total distance in propensity scores between LT-exposure participants and controls; the distance is minimized across the whole sample, and not based on one-on-one distances, as was the case for nearest neighbor matching. In this study, optimal/full matching was implemented with the *optmatch* procedure in the free software R (Hansen, 2007). Program effects and their significance levels were estimated with the Hodges–Lehman aligned rank test obtained with the *hodgesl* procedure (Guo, 2008), a procedure designed for use with optimal/full matching.

A third approach, *matching estimators,* is a related method for estimating treatment effects, with some departures from the propensity matching methods already described (see Guo & Fraser, 2010). Importantly, matching estimators do not require the use of propensity scores. Instead, they impute, for each LT-exposure participant, an outcome value that would have been observed had the participants not received treatment (the counterfactual). This potential non-observed or counterfactual outcome is imputed from the matched control case with the shortest distance from a given LT-exposure participant on a series of covariates (the same as those used in other propensity score methods). Once counterfactual outcomes are imputed for the entire LT-exposure group, treatment effects are estimated. One important advantage of matching estimators is that they require fewer decisions on the part of the investigator because treatment effects are calculated directly. In this study, program effects based on bias-corrected matching estimators were obtained with the *nnmatch* procedure in Stata 10 (Abadie, Drukker, Herr, & Imbens, 2004). Bias-corrected matching estimators use least-square regression adjustment to correct for any imbalance remaining between LT-exposure and control participants on the covariates used for matching.

Outcome measures

Descriptive statistics for all outcome measures are presented in Table 1.

Internalizing problems

Lifetime internalizing problems were measured with two dichotomous self-reported clinical measures assessing the presence of depression and generalized anxiety disorder (NCSR-AS, see http://www.hcp.med.harvard.edu/ncs/ for details). The depression scale included 14 items, and a young person was considered to be depressed if five criteria were met based on the answers to these items; the anxiety scale included ten items and four such criteria (see Kling et al., 2007 for a full description). We combined the two measures into a single outcome because of their low base rate and the relatively small sample size of the LT-exposure group (0 = absence of internalizing problem, 1 = presence of internalizing problem).

Psychological distress

This outcome was measured with a six-item self-report scale, tapping feelings of depression and nervousness during the past 30 days (Kessler, Andrews, Colpe, Hiripi, Mroczek, Normand et al., 2002). The final scale indicates the proportion of items on which the young person answered at least "some of the time".

Substance use

Three self-reported dichotomous items assessed whether the respondent had ever used marijuana, smoked a cigarette, or had a drink of alcohol (0 = no, 1 = yes, Moore et al., 1999).

Sexual behavior

A single self-report item evaluated whether the respondent ever had sexual intercourse (0 = no, 1 = yes, Moore et al., 1999).

Behavior problems

An 11-item scale measured behavior problems in the past six months (Moore et al., 1999). The scale represents the proportion of

times the respondent reported sometimes or often engaging in a set of behaviors, such as cheating or lying or being restless or overactive (for a full list, see Orr et al., 2003).

Delinquency

The delinquency scale included nine self-reported items, such as carrying a hand gun or attacking someone with idea of hurting them (Moore et al., 1999; for a full list, see Orr et al., 2003). The final score represents the proportion of delinquent behaviors in which the respondent reported ever engaging.

Achievement

Broad reading and broad mathematics scores were assessed by the Woodcock-Johnson Psycho-Educational Battery—Revised (WJ-R, Woodcock & Johnson, 1989), a widely-used standardized test. To conduct analyses across different age groups and grade levels, raw scores were converted to W scores that centered the raw score at a value of 500.

Graduation/school status

This measure was derived by combining two dichotomous items, the first indicating whether the respondent was currently enrolled in secondary school full-time, and the second indicating whether the respondent had received either a high-school diploma or a GED (0 = not in school or graduated; 1 = in secondary school or graduated).

Results

To facilitate comparison with previous MTO studies and provide benchmark estimates of program effects, we first replicated the interim findings with the current sample, and then examined LT-exposure program effects.

Intention-to-treat and treatment-on-treated analyses

We conducted intention-to-treat (ITT) program effects, which compare the entire experimental group with the entire control group based on random assignment status using Ordinary Least Squares Regression, and treatment-on-treated (TOT) analyses, which also are based on the entire experimental and control groups but provide unbiased estimates of treatment effects among experimental compliers using 2-Stage Least Squares Regression (see Table 3 for details). In line with previous MTO studies (e.g., Orr et al., 2003), ITT and TOT treatment estimates showed favorable associations for experimental girls' mental health and substance use and unfavorable associations for boys' substance use and behavior problems compared with their respective peers in the control group who stayed in high-poverty neighborhoods.

Long-term exposure analyses

The main goal of this study was to explore whether beneficial program effects would generalized to a broad array of outcomes and to all adolescents from experimental families who remained in low-poverty neighborhoods for more than half of the observation window—i.e., had long-term exposure to low-poverty neighborhoods.

Effectiveness of matching procedures

Before examining results from propensity score analyses, it is important to first assess if the matching routines were successful in creating comparable groups in terms of baseline characteristics; matching estimators do not use propensity scores *per se*, thus imbalance checks are presented only for nearest neighbor and optimal/full matching. To do so, we first computed balance *t* (for continuous variables) and χ^2 (for dichotomous variables) statistics comparing the full control group to three different treatment groups: the full experimental group, the compliers, and the LT-exposure subgroup (results available from authors upon request). Differences between control and experimental groups were greater for the complier and LT-exposure subgroups than for the full experimental group, indicating that experimental families who complied with the program and stayed in low-poverty neighborhoods for extended periods of time differed from their counterparts in the experimental group who did not use vouchers offered through the program to move. Then, we computed the same balance statistics based on a resampled control group selected with the nearest neighbor matching procedure. No significant differences were found between this matched-control group and the LT-exposure subgroup, and differences between the groups were, on average, much smaller than those when no matching was applied.

For optimal/full matching, imbalance checks were conducted with the imbalance procedure in Stata 10, which calculates standardized differences in covariates' means before and after matching. Again, average standardized mean differences between the matched control subgroup and the LT-exposure subgroup were smaller, by about one-half, after than before optimal/full matching.

Long-term exposure estimates

Table 4 presents associations among long-term exposure to low-poverty neighborhoods in the MTO program and adolescent

Table 3
Intention-to-treat (ITT) and treatment-on-treated (TOT) results for the total sample and for boys and girls.

Outcome variables	Total sample				Girls				Boys			
	ITT		TOT		ITT		TOT		ITT		TOT	
	T.E.	SE	T.E.	SE	T.E.	SE	T.E.	SE	T.E.	SE	T.E.	SE
Internalizing problem%	−0.01	0.02	−0.03	0.04	−0.06*	0.02	−0.13*	0.05	0.03	0.02	0.06	0.05
Distress	−0.01	0.02	−0.01	0.03	−0.04†	0.02	−0.08†	0.04	0.02	0.02	0.05	0.05
Marijuana use%	−0.03	0.02	−0.06	0.05	−0.07*	0.03	−0.15*	0.06	0.01	0.04	0.01	0.08
Alcohol use%	0.00	0.03	0.00	0.05	−0.05	0.03	−0.10	0.07	0.05	0.03	0.11	0.08
Sexual intercourse%	−0.01	0.02	−0.02	0.05	−0.06*	0.03	−0.13*	0.06	0.05	0.04	0.10	0.08
Tobacco use%	0.00	0.02	0.00	0.05	0.02	0.03	0.03	0.06	−0.02	0.03	−0.04	0.08
Behavior problems	0.04*	0.02	0.08*	0.03	−0.00	0.02	−0.01	0.04	0.07**	0.02	0.15**	0.05
Delinquency	0.00	0.01	0.00	0.02	−0.01	0.01	−0.01	0.02	0.00	0.01	0.00	0.03
Reading	0.73	1.32	1.58	2.85	2.30	1.55	4.75	3.21	−1.70	1.96	−3.77	4.39
Math	0.56	1.15	1.21	2.47	2.21	1.52	4.57	3.16	−1.21	1.49	−2.68	3.32
Hs graduate/in school%	0.01	0.02	0.02	0.04	0.04	0.03	0.08	0.06	−0.02	0.03	−0.05	0.06

Note. ITT treatment effect estimates and associated standard errors (in parentheses) are based on OLS regressions adjusting for all variables used for propensity scoring (listed in Table 1), with robust standard errors. TOT analysis used two-stage instrumental variable regression, where random assignment status is used as an instrument along with covariates listed in Table 1; the second stage regression estimating treatment effects adjusted for variables listed in Table 1. To facilitate comparison with previous MTO studies, analyses were weighted. $^{**}p < 0.01$. $^{*}p < 0.05$. $^{\dagger}p < 0.10$.

Table 4
Associations among long-term exposure to low-poverty neighborhoods in MTO program and adolescent outcomes for total sample, girls, and boys by propensity score method.

Outcome variables	Nearest neighbor matching within calipers		Optimal/full matching			Matching estimator	
	T.E.	SE	T.E.	HL	SE	T.E.	SE
Total							
Internalizing problem%	−0.02	0.04	−0.04	1053	1963	−0.05	0.03
Distress	−0.02	0.03	−0.06	−3054	2508	−0.03	0.03
Marijuana use%	−0.01	0.05	−0.02	−1217	2314	−0.03	0.05
Alcohol use%	−0.01	0.06	0.01	86	2497	0.02	0.05
Sexual intercourse%	0.00	0.06	0.01	−1002	2559	−0.05	0.06
Tobacco use%	0.00	0.05	0.01	273	2359	0.01	0.05
Behavior problems	0.00	0.03	0.00	1041	2666	0.01	0.03
Delinquency	0.02	0.02	0.00	2632	2609	0.00	0.02
Reading	2.18	2.74	5.50†	3707	2571	6.44**	1.85
Math	4.11	2.68	2.59†	4153	2662	3.06†	1.80
Hs graduate/in school%	0.04	0.04	0.01†	2886	2233	0.05	0.04
Girls							
Internalizing problem %	−0.10	0.06	−0.14†	−1073	708	−0.12***	0.03
Distress	−0.10†	0.05	−0.09*	−1839	881	−0.12***	0.04
Marijuana use%	−0.11†	0.06	−0.16*	−1769	824	−0.18***	0.04
Alcohol use%	−0.07	0.09	−0.08	−516	911	0.01	0.06
Sexual intercourse%	−0.15†	0.08	−0.12†	−1255	862	−0.19*	0.07
Tobacco use%	−0.13	0.08	−0.16	−1027	846	−0.12*	0.06
Behavior problems	−0.02	0.05	−0.03	−252	956	−0.05	0.04
Delinquency	−0.02	0.03	−0.01	71	880	−0.01	0.01
Reading	2.25	3.87	3.61	1047	909	5.04*	2.47
Math	6.22	3.70	1.58	983	916	2.87	2.47
Hs graduate/in school%	0.13†	0.07	0.07*	1283	754	0.10*	0.05
Boys							
Internalizing problem%	0.04	0.06	0.02*	1265	734	0.03	0.05
Distress	0.05	0.04	−0.01	1039	894	0.03	0.04
Marijuana use%	0.10	0.08	0.00	218	882	0.09	0.07
Alcohol use%	0.04	0.08	0.04	480	889	0.08	0.08
Sexual intercourse%	0.14†	0.08	0.05	578	897	0.07	0.07
Tobacco use%	0.11	0.08	0.08	854	876	0.11	0.07
Behavior problems	0.04	0.06	0.02	1181	952	0.04	0.04
Delinquency	0.05	0.04	0.01†	1378	949	0.02	0.03
Reading	−1.77	3.69	6.06	958	914	5.59*	2.49
Math	−0.04	3.84	1.08	732	966	3.07	2.22
Hs graduate/in school%	−0.04	0.06	0.03	387	840	0.02	0.06

Note. Within matched samples (nearest neighbor, matching estimator), long-term exposure estimates were obtained via OLS regression on treatment condition, site, ethnicity, race, household income, maternal age, maternal work status, maternal education, child sex, and child age, with robust standard errors. With the full matching procedure, statistical test of significance is based on Hodges-Lehmann aligned rank test, which does not permit covariate adjustment. Because of the resampling involved in matching procedures, weights were not applied, but were used as a matching variable along with covariates listed in Table 1. T.E. = Treatment effect. HL = Hodges–Lehman Test statistic.$^{***}p < 0.001$. $^{**}p < 0.01$. $^{*}p < 0.05$. $^{\dagger}p < 0.10$.

outcomes obtained with nearest neighbor matching, optimal/full matching and matching estimators, respectively, for the full sample and by gender. We note marginally significant results in Table 4, which indicate trends, but they are not discussed here. No significant associations were obtained with the nearest neighbor matching procedure for the full sample or for the gender subgroups.

In contrast with the nearest neighbor results, the two other, more powerful matching procedures yielded significant results. Optimal/full matching resulted in no significant associations for the full sample, but girls in the LT-exposure group had significantly less psychological distress and lower marijuana use than a sample of matched control girls. The magnitude of these associations is comparable to the TOT analyses (Table 3). A new significant, positive association emerged, with girls in the LT-exposure group more likely to attend secondary school/graduate than matched control girls. However, a significant, adverse association was evident for boys; boys in the LT-exposure group were more likely to report internalizing problems than matched control boys.

The strongest results were obtained with matching estimators. Again, significant favorable associations were found for LT-exposure girls' internalizing problems, psychological distress, marijuana use and sexual intercourse as compared with matched control girls. The magnitude of these associations tended to be somewhat higher than those reported in the TOT analysis (Table 3). With this procedure, new significant, favorable program associations were found not only for LT-exposure girls' secondary school attendance/graduation, but also for their tobacco use and reading achievement (both *vs.* matched control girls). Among boys, none of the adverse program effects or associations found with other procedures was observed; rather boys in the LT-exposure group had significantly higher reading achievement than matched control boys. This significant association between LT-exposure and reading also was seen for the full sample.

Discussion

The goal of this study was to reanalyze the MTO interim data to investigate the association between moving *and* staying in low-poverty neighborhoods, compared with remaining in high-poverty neighborhoods, and adolescent outcomes, to determine whether beneficial effects reported for experimental girls' mental health and behavior generalized to their other outcomes and to boys. The results of propensity score analyses, which are a relatively unbiased approach for evaluating long-term exposure to low-poverty neighborhoods in the MTO program, indicate that experimental girls whose families remained in low-poverty neighborhoods for

comparatively long periods, approximately five years, had better mental health (internalizing problems and psychological distress) and engaged in fewer risky behaviors (sexual intercourse and marijuana use) then a matched sample of control girls. These LT-exposure estimates were similar to the program effects found for the full female complier subgroup via TOT analysis. Additional benefits for experimental girls in the LT-exposure group were found in the education domain for secondary school attendance/graduation and reading achievement, although the favorable association with reading was evident with only one of the three techniques used.

For boys, associations between LT-exposure and adolescent outcomes were less consistent. The adverse program effect on experimental boys' behavior problems reported in the TOT analysis was not detectable in the LT-exposure analyses. However, significant associations with LT-exposure, both unfavorable and favorable, were found, but none were replicated with more than one method.

These findings, in part, confirm earlier evaluations of MTO program, but reveal that facilitating adaptation of experimental families (and perhaps their neighborhoods) toward residential stability may confer additional benefits for girls in terms of the range of outcomes affected. Counter to these other studies, we did not find that experimental boys were harmed by the program. It is likely that residential stability among experimental families in the LT-exposure group, coupled with extended time in low-poverty neighborhoods accounts for this pattern of results. Our findings suggest that programs relocating low-income families to low-poverty neighborhoods should provide support to families to help them to remain in these neighborhoods. These supports should not only entail working with relocating families, but also with the receiving communities to promote social integration such as receptivity to diverse residents, social cohesion, and social participation in local institutions (Jackson et al., 2009).

Although results were generally consistent across different matching methods, it is important to note that newer matching methods maximizing data were more robust. Propensity score methods represent state-of-the art strategies for adjusting for selection bias and are superior to simple regression adjustment; however, their effectiveness depends on the quality of the measured pre-treatment characteristics used for matching, the availability of suitable matches from the control group, and the fulfillment of initial assumptions (Guo & Fraser, 2010). All of these conditions appear to be met in this study, but selection bias among LT-exposure experimental families cannot be definitely ruled out as an explanation of the study findings. Our matching was based largely on demographic characteristics assessed at the baseline interviews and did not include comprehensive measure of parents' or adolescents' mental health and risky behavior or of families' preferences for different types of neighborhoods, all factors that might remain unobserved in the current study despite efforts to address them via propensity score matching. Nor does our study address selection among families who signed up for MTO in the first place compared with those who did not.

Finally, it is critical to acknowledge that social experiments are very difficult to evaluate (Moffitt, 2003). Because they are not laboratory experiments, or even clinical trials, they tend to be fraught with factors that cannot be controlled by researchers. In addition, the issues they address are complex and not always amenable to a single form of treatment or one-shot treatments (Clark, 2008). Therefore, alternative approaches to estimating their impact, such as the current study, are warranted, but need to be evaluated with caution. Likewise, consideration of alternative policy interventions which address the root causes of urban poverty and their consequences also are merited (Clark, 2008; Jackson et al., 2009).

Acknowledgments

The authors would like to acknowledge the William T. Grant Foundation and the Canadian Institutes of Health Research for their support. Moving to Opportunity was funded by the United States Department of Housing and Urban Development. Additional support was provided by the Russell Sage Foundation, the National Institute of Child Health and Development (R01-HD40444), the National Science Foundation (0091854), the MacArthur Foundation, and the Spencer Foundation. We are especially thankful to our colleagues on this project including Jeanne Brooks-Gunn, Greg Duncan, Judie Feins, John Goering, Larry Katz, Jeff Kling, Jeff Liebman, Jens Ludwig, Sue Popkin, Todd Richardson, Lisa Sanbonmatsu, and Mark Shroder. In addition, the authors wish to thank Sara Anderson for assistance with this manuscript.

References

Abadie, A., Drukker, D. M., Herr, J. L., & Imbens, G. W. (2004). Implementing matching estimators for average treatment effects in Stata. *The Stata Journal, 4*(3), 290–311.

Clampet-Lundquist, S., Duncan, J., Edin, K., & Kling, J. (2006). *Moving at-risk teenagers out of high-risk neighborhoods: Why girls fare better than boys*. Working Papers. Princeton, NJ: Princeton University, Department of Economics, Industrial Relations Section.

Clark, W. A. V. (2008). Reexamining the Moving to Opportunity study and its contribution to changing the distribution of poverty and ethnic concentration. *Demography, 45*(3), 515–535.

Guo, S. (2008). *HODGESL: Stata module to perform Hodges–Lehmann aligned rank test*. Author.

Guo, S., & Fraser, M. W. (2010). *Propensity score analysis: Statistical methods and applications*. Thousand Oaks, CA: Sage.

Hansen, B. B. (2007). Optmatch: Flexible, optimal matching for observational studies. *R News, 7*, 19–24.

Hill, J. L., Brooks-Gunn, J., & Waldfogel, J. (2003). Sustained effects of high participation in an early intervention for low-birth-weight premature infants. *Developmental Psychology, 39*, 730–744.

Jackson, L., Langille, L., Lyons, R., Hughes, J., Martin, D., & Winstanley, V. (2009). Does moving from a high-poverty to a lower-poverty neighborhood improve mental health: A realist review of 'Moving to Opportunity'. *Health and Place, 15*, 961–970.

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L., et al. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine, 32*, 959–976.

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

Kling, J. R., Ludwig, J., & Katz, L. F. (2005). Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment. *Quarterly Journal of Economics, February*, 87–130.

Leuven, E., & Sianesi, B. (2003). *PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing*. Author.

Leventhal, T., Dupéré, V., & Brooks-Gunn, J. (2009). Neighborhood influences on adolescent development. In R. M. Lerner, & L. Steinberg (Eds.), *Handbook of adolescent psychology* (pp. 411–443). Hoboken, NJ: Wiley.

Moffitt, R. (2003). *The role of randomized field trials in social science research: A perspective from evaluations of reforms of social welfare programs*. Discussion Paper no. 1264-03. Madison: University of Wisconsin. Institute for Research on Poverty.

Moore, K. A., McGroder, S., Hair, E. C., Gunnoe, M., Richter, K., Mariner, C., et al. (1999). *NLSY97 codebook supplement main file round 1, appendix 9, family process and adolescent outcome measures*. Washington, DC: U.S. Department of Labor.

Orr, L., Feins, J. D., Jacob, R., Beecroft, E., Sanbonmatsu, L., Katz, L. F., et al. (2003). *Moving to Opportunity for Fair Housing Demonstration: Interim impact evaluation*. Washington, DC: U.S. Department of Housing and Urban Development.

Popkin, S. J., Leventhal, T., & Weismann, G. (2010). Girls in the 'hood: Reframing safety and its impact on health and behavior. *Urban Affairs Review, 45*(6), 715–744.

Sampson, R. J., Morenoff, J. D., & Gannon-Rowley, T. (2002). Assessing "neighborhood effects": social processes and new directions in research. *Annual Review of Sociology, 28*, 442–478.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock–Johnson Psycho-Educational Battery - Revised*. Allen, TX: DLM Teaching Resources.