

1

0

What do Randomized Studies of Housing Mobility Demonstrate?: Causal Inference in the Face of Interference

Michael E. Sobel
Columbia University

Abstract

During the past 20 years, social scientists using observational studies have generated a large and inconclusive literature on neighborhood effects. Recent workers have argued that estimates of neighborhood effects based on randomized studies of housing mobility, such as the “Moving to Opportunity Demonstration” (MTO), are more credible. These estimates are based on the implicit assumption of no interference between units, that is, a subject’s value on the response depends only on the treatment to which that subject is assigned, not on the treatment assignments of other subjects. For the MTO studies, this assumption is not reasonable. Although little work has been done on the definition and estimation of treatment effects when interference is present, interference is common in studies of neighborhood effects and in many other social settings, for example, schools and networks, and when data from such studies are analyzed under the “no interference assumption”, very misleading inferences can result. Further, the consequences of interference, for example, spillovers, should often be of great substantive interest, though little attention has been paid to this. Using the MTO demonstration as a concrete context, this paper develops a framework for causal inference when interference is present and defines a number of causal estimands of interest. The properties of the usual estimators of treatment effects, which are unbiased and/or consistent in randomized studies without interference, are also characterized. When interference is present, the difference between a treatment group mean and a control group mean (unadjusted or adjusted for covariates) does not estimate an average treatment effect, but rather the difference between two effects defined on two distinct

*I am grateful to Xavier de Souza Briggs, Steven Durlauf, Elizabeth Halloran, Jennifer Hill, Booil Jo, several anonymous reviewers and the members of the McArthur Network on Inequality and Social Interactions, to whom this paper was first presented in 2002, for helpful comments and advice, and to the John D. and Catherine T. MacArthur foundation for financial support.

subpopulations. This result is of great importance, for a researcher who fails to recognize this could easily infer that a treatment is beneficial when it is universally harmful.

Keywords causal inference, interference, neighborhood effects, stable unit treatment value assumption

1 Introduction

For more than a century, social scientists and reformers have argued that living in bad neighborhoods intensifies the obstacles faced by the urban poor. A recent resurgence of interest in this idea, in part due to the rise in the spatial concentration of poverty in the U.S. between 1970 and 1990 (Jargowsky 1997) and the publication of *The Truly Disadvantaged* (Wilson 1987), has given rise to a large literature on so-called “neighborhood effects”.

Various processes that may contribute to the accumulation and reproduction of disadvantage among the urban poor have been discussed (Anderson 1999, Crane 1991, Jencks and Mayer 1990, Sampson, Morenoff and Gannon-Rowley 2002, Wilson 1987). In the American context, where schools are locally financed, Bénabou (1993,1996) and Durlauf (1996) construct formal models of neighborhood choice and human capital formation, showing how interdependence can generate persistent (especially intergenerational) social immobility. Formal models of interactive processes have also been proposed (see Blume and Durlauf 2001 and Durlauf 2001), as have some interesting proposals to measure social interactions (Glaeser and Scheinkman 2001).

A large empirical literature has also emerged, with researchers studying the purported effects of neighborhoods on psychological, sociological, economic and health related outcomes. Coulton and Pandey (1992) consider infant death rates and low birth weights. Other outcomes studied include delinquency and crime (Sampson and Groves 1989), dropping out of school and teenage pregnancy (Crane 1991), educational attainment (Dachter 1982), and hours worked in employment (Weinberg, Reagan and Yankow 2004). For reviews of the literature, see Gephart (1997) and Sampson et al. (2002).

Researchers have mainly used observational studies to study the effects of neighborhoods on individuals and families. Many problems with making causal inferences from such studies, especially the self-selection of families into neighborhoods, have been identified. Thus, there has been great interest in the “Moving to Opportunity” (MTO) demonstration sponsored by the U.S. Department of Housing and Urban Development, a housing mobility experiment in five cities in which eligible ghetto residents are randomly assigned to receive (or not) various forms of assistance to relocate. Using data from these studies, researchers have argued that the benefits of “relocation” on safety and health are substantial. Results on economic self-sufficiency and delinquency are mixed (Del Conte and Kling 2001).

This paper has several goals. First, researchers using the MTO data to study the effects of residential mobility equate the parameters of the models they estimate with various effects defined in the statistical literature (Angrist, Imbens and Rubin 1996) under the stable unit treatment value assumption (SUTVA) (Rubin 1980). Under this assumption, the potential responses of units do not depend on a) the mechanism by which treatments are assigned nor b) the treatment received by any other unit, i.e., there is no inter-

ference between units (Cox 1958). However, the no interference assumption is not plausible for the MTO demonstration. Given the importance and cost of this study, it is also important to know what effects, if any, researchers are estimating.

To clarify these issues, I first extend the framework for causal inference developed primarily by Rubin (1974,1977,1978,1980), defining neighborhood (treatment) effects when interference is present. The estimates in the MTO literature, inappropriately interpreted as estimates of average treatment effects, do not estimate these neighborhood effects. In particular, Theorem 4 shows that the mean difference between outcomes in treatment and control groups does not estimate a causal parameter, but rather the difference between two neighborhood effects, each defined on a different subpopulation. The failure to realize this could lead one to conclude that a relocation policy which actually decreases well-being in each subpopulation is beneficial.

Second, there is not much work on causal inference in the presence of interference. Halloran and Struchiner (1995) consider infectious diseases. Whether or not a unit contracts an infection may depend on its exposure status, which depends on the other units vaccinated. This leads naturally to considering “families” of average effects of vaccination. As many writers have noted, one cannot expect estimates from a small scale study to extrapolate well to the larger scale. Given these types of difficulties, Halloran and Struchiner define “conditional direct causal effects” that are averages of the unit effects of vaccination conditional on the subject being exposed to the infection; they assume (p. 145) “all exposures to infection are discrete and equivalent.” This avoids the interference problem and leads to an interesting and useful parameter that is the difference between the “transmission probabilities” for vaccinated and not vaccinated “susceptibles”. They also discuss estimation and contrast their approach with the more common approach in the epidemiological literature that combines empirical information and mathematical modeling without using the potential outcomes notation that statisticians have used to clarify problems of causal inference. Using general equilibrium theory in economics, Heckman, Locher and Taber (1998) employ this latter approach to study the effect of a national tuition subsidy on college enrollments and earnings. They assume enrollment decisions are based on skill prices (wages of workers with high school and college educations, respectively). They assume prices depend on the fraction of workers in each group and that the college premium decreases as the fraction of workers in the college group rises. In a small scale study, enrollment decisions are assumed to be based on current skill prices, yielding “partial equilibrium” estimates of treatment effects. For the larger scale, “general equilibrium” estimates are computed assuming individuals are rational and adjust their enrollment decisions to the skill prices that would result when the national tuition subsidy is implemented.

This paper takes a different approach. Although “conditional” effects of housing assistance could be

defined, this seems less fruitful for studying neighborhood effects than for studying infectious diseases. Nor is it clear that MTO researchers are interested in the effects of a housing policy that offers universal assistance. Such considerations suggest the need for a more general approach. This paper takes a step in that direction, defining unit effects (including “spillovers”) for any allocation of the population units to treatment groups; these are the building blocks for defining average effects in a specific allocation of the units to treatments. Such allocations may be of intrinsic interest and/or effects may be defined over classes of allocations. Though attention centers on the MTO demonstration, which is used to motivate the definitions and results obtained in this paper, these also apply to a number of other settings where interference is the norm, for example, schools, networks and firms.

I proceed as follows. The idea of a neighborhood effect is introduced in section 2. The MTO demonstration is described and the no interference assumption is discussed. Section 3 defines neighborhood effects, including spillovers, when interference is present; these reduce to the usual effects when interference is absent. Section 4 characterizes the (non-causal) parameters that are estimated in randomized studies of housing mobility when interference is present. Section 5 takes up the case of “partial” interference and concludes.

2 Observational and Experimental Studies of Neighborhood Effects

Jencks and Mayer (1990), Duncan, Connell and Kebanov (1997), Manski (1995), Moffit (2001), Sobel (2006) and others discuss problems estimating neighborhood effects using observational studies. A major problem is that neighborhood is a self-selected treatment. Thus, it is necessary to “control” for determinants of neighborhood choice. As an example, Dachter (1982) studies the effect of average neighborhood income on educational attainment. Individual level data on young men in the Panel Study of Income Dynamics are merged with data on the racial mix and average income in their neighborhood (operationalized by zip-code). Educational attainment is then regressed on individual characteristics, for example, income and educational attainment of the family of orientation, and the neighborhood variables. The coefficient for the neighborhood income variable is positive, suggesting higher income neighborhoods are beneficial. A problem is that important control variables may be missing from the regression. Attempting to deal with this, several researchers have used fixed effects models, taking advantage of sibling variation (Aaronson 1998) and longitudinal data structures (Weinberg, Reagan and Yankow 2004). Others have used instrumental variable methods (Cutler and Glaeser 1997, Evans, Oates and Schwab 1992).

In addition, data sets with many observations per neighborhood have become more common, allowing

neighborhood averages to be estimated from individual data. Thus, neighborhood characteristics not reported in official data can be considered and a more flexible definition of neighborhood can be used. Researchers who use such data often estimate hierarchical models and adjust standard errors for clustering. But it is still possible that important control variables have been overlooked. Further, the use of such models does not solve the interference problem that is the subject of this paper.

Given the difficulties above, and others, many investigators share the view in Jencks and Mayer (1990, p. 119): “From a scientific standpoint, the best way to estimate neighborhood effects would be to conduct controlled experiments in which we assigned families randomly to different neighborhoods, persuaded each family to remain in its assigned neighborhood for a protracted period, and then measured each neighborhood’s effect on the children involved”.

Subsequently, an experimental literature on neighborhood effects developed. The Gatreaux program in Chicago was mandated by a ruling in a class action suit. Over more than 20 years, thousands of eligible low income families in primarily black urban neighborhoods registered for the program. Randomization was used to tender offers to families to move either to (predominantly white) suburbs or within the urban area. Rubinowitz and Rosenbaum (2000) compared the two groups of movers; the former appeared to benefit more from relocation. The comparison is problematic. Not all families given offers actually moved and some who did returned to their old neighborhood.

The Gatreaux program gave impetus to the Moving to Opportunity Demonstration, sponsored by the U.S. Department of Housing and Urban Development. At five sites (Baltimore, Boston, Chicago, Los Angeles and New York) selected in spring 1994, families with children residing in public and assisted housing projects in census tracts with a poverty rate of 40 percent or more were recruited. Approximately 5000 participants (Goering, Feins and Richardson 2002) completed a baseline survey. At each site, randomization was used to assign participants to a control group, an experimental group receiving counseling and housing vouchers that could be used only to move to low poverty neighborhoods (census tracts with poverty rates less than 10 percent), or a section 8 group receiving vouchers that could be used anywhere. For further description of the program, see Goering, Kraft, Feins, McInnis, Holin and Elhassan (1999).

MTO: Research and Methods

At each site, follow up interviews were conducted with participants and the data used to compare the section 8 and experimental groups to controls on various outcomes, for example, neighborhood characteristics, welfare receipt, employment rates, parents’ perceptions of safety and health, crime, social adjustment and isolation, child health and behavioral problems, arrests. While it is difficult to summarize the early results, which vary by site and treatment group, Del Conte and Kling (2001,p. 6) state: “Although

preliminary MTO results are inconclusive on the impact of economic self-sufficiency of moving to lower-poverty neighborhoods, the data consistently suggest that such a transition is associated with significant improvements in safety, child and parental physical and mental health, as well as youth delinquency and behavior problem.” Goering, Feins and Richardson (2002) reach similar conclusions.

Methodological differences may account in part for different findings by site. Rosenbaum, Harris and Denton (2003) and Rosenbaum and Harris (2000, 2001) compare controls with Chicago families in the section 8 and experimental groups who moved. This can be problematic when substantial fractions do not “lease up” (use their vouchers to move); the Chicago lease up rate is 34 % in the experimental group, 66 % in the section 8 group (Goering et al. 1999).

Leventhal and Brooks-Gunn (2003) estimate the intent to treat estimand (ITT), comparing means of New York families in the section 8 and experimental groups to the control group mean on neighborhood characteristics and children’s psychological states and behaviors. Children in both treatment groups have lower expectations of completing college and do not exhibit fewer behavioral problems than controls, but in some comparisons, their mental health is better (less unhappiness and anxiety). Katz, Kling and Leibman (2001) adjust for pre-treatment covariates, using regression analysis to estimate ITT effects for Boston families. On average, members of the treatment groups live in better neighborhoods, perceive their neighborhoods as being safer, and report better health than members of the control group. However, there are no significant differences in employment status or earnings. Experimental group children were also more likely to attend schools with better math and reading scores. Using the same methodology, Ludwig, Duncan and Hirshfield (2001) find that the juvenile arrest rate for violent crimes is lower in the Baltimore treatment group than in the control group; the same is true for Boston (Kling, Ludwig and Katz 2005). Kling et al. (2005) also find that arrest rates are lower for females for property crimes, but the rates for males, which initially drop, increase over time. (For methodologically similar types of analyses, see also Hanratty, McLanahan and Pettit 2003, Katz, Kling and Liebman 2003, Ladd and Ludwig 2003, Ludwig, Duncan and Pinkston 2005).

The ITT measures the effect of assignment, not the effect of voucher induced movement per se. Thus, a number of the investigators above have considered the effect of treatment on the treated (TOT). As a voucher cannot be used to move when not received, the TOT is identical to the local average treatment effect (LATE) defined in Angrist, Imbens and Rubin (1996). As a consequence, under the assumptions in Angrist et al. (1996), the TOT is equal to the instrumental variable (IV) estimand; this is the ITT divided by the lease-up rate. Therefore, the TOT yields substantive results similar to those above, but with effects now greater in magnitude.

MTO researchers have taken great care to adjust for covariates to improve the precision of parameter

estimates and, implicitly recognizing the possibility of interference, have cautioned against extrapolation from the study population of volunteers to the population of those meeting the eligibility requirements for participation in MTO or the broader population of ghetto residents. Nevertheless, they equate the parameters they estimate with the ITT and TOT, which are defined using the no interference assumption.

Social interactions are a primary source of interference in studies with human subjects. At each site, MTO volunteers were recruited from several public housing projects, (8 in Baltimore, 8 in Boston, 6 in Chicago, 11 in Los Angeles, 14 in New York). In each project, a group meeting was held to describe the MTO demonstration; the volunteers are the eligible attendees who wished to participate. Given this method of recruitment, many participants undoubtedly knew other participants, as would be the case, for example, when a participant informed her friends of the meeting and one or more of these participated. Unfortunately, MTO researchers did not collect data on interactions among participants, though Popkin, Harris and Cunningham (2001, p. 70) cite the case of a mother and daughter who participated.

Now consider leasing up. Assuming some interaction among participants, if the no interference assumption holds, no participant is influenced by the behavior of others in her “interaction set” (other participants with whom she interacts). Suppose a family is assigned to the experimental group. Moving to the suburbs alone may mean leaving all one’s friends behind and entering a new and hostile environment, but if all families in the interaction set are assigned to and move to the suburbs, one expects that family is more likely to move. Unfortunately, evidence pertaining to such types of influence was not collected. Given the lack of evidence and that the no interference assumption lacks credibility, it is more reasonable to assume that interference is present than to assume it is not.

Direct social interactions are not the only possible source of interference in leasing up. An experimental group family that relocates when only a few families are assigned to this group may not be able to do so in a tight rental market if many families are assigned to this group.

Interference in location dependent outcomes, for example, neighborhood educational level (Katz et al. 2001), is an immediate consequence of interference in leasing up. Popkin et al (2002, p. 38) also suggest that MTO movers often maintain close contact with project residents. Such social interactions may also generate interference, particularly on outcomes measuring subjective perceptions of neighborhoods.

Because the ITT and TOT are defined assuming no interference and this assumption is not reasonable, it is not clear what parameters MTO researchers are estimating, nor what policies their analyses might support. Building on Rubin’s model for causal inference (Rubin 1974, 1977, 1978, 1980), I now develop an analytical framework that clarifies these issues.

3 Defining Causal Estimands with Interference Between Units

The General Case

Effects are defined when interference is present. To keep matters simple and to minimize assumptions, these are defined as expectations over a finite population, as in Angrist et al.(1996); my definitions reduce to theirs when interference is absent. Estimation is considered in the next section.

At a site, suppose there are $i = 1, \dots, N_v, \dots, N$ families. The first N_v are MTO volunteers ($V_i = 1$). The others ($V_i = 0$) include eligible families who didn't volunteer, other families in the ghetto, families in the suburbs, etc. Because there are two treatment groups and one control group and only volunteers can be treated, there are 3^{N_v} possible allocations.

If participants receiving vouchers do not move, they do not obtain any benefits from relocating and housing mobility programs like MTO would seem to have little point. Thus, researchers are interested in whether or not MTO families use vouchers to move (Shroder 2002). Let r and r' denote two allocations, for example, respectively, the allocation where all units are assigned to the experimental group and the allocation where, assuming N_v is divisible by 3, families $i = 1, \dots, N_v/3$ are assigned to the control group, families $N_v/3 + 1, \dots, 2N_v/3$ to the section 8 group, and the remainder are assigned to the experimental group. Let $\underline{D}(r)$ be the $N \times 1$ vector with i (th) component $D_i(r) = 0$ if unit i does not move using a housing voucher under allocation r , 1 otherwise. The effect of allocation r vs. r' on leasing up for unit i may be defined as: $D_i(r) - D_i(r')$.

As versus the case of no interference, if unit i is allocated to the same group under r and r' , possibly $D_i(r) \neq D_i(r')$, implying the unit effect is not necessarily 0. (Halloran and Struchiner consider the special case of this where unit i is not vaccinated under either allocation r or r' , with $D_i(r) = 1$ if unit i is infected under r , 0 otherwise, and $D_i(r')$ is similarly defined and they (p. 147) call $D_i(r) - D_i(r')$ an “unconditional indirect effect in individual i ”.) Similarly, if unit i is allocated to the same group under allocation r and r^* , this does not imply $D_i(r) - D_i(r') = D_i(r^*) - D_i(r')$.

In general, however, it will be of interest to compare allocation r to a benchmark allocation, hereafter $r' = o$, where no unit receives a voucher, as in this paper (or a standard treatment). Of course, another baseline could be chosen; the subsequent definitions and results are easily modified to handle this.

Because control units cannot use a voucher to move, leasing up effects are defined only for the two treatment groups. Let $j = 0, 1, 2$, index the no voucher, section 8, and experimental groups, respectively, let $\underline{A}(r)$ be a $N \times 1$ vector with i (th) component $A_i(r) = j$, let k_j denote the size of group j , $j = 1, 2$, with $k_0 \equiv N_v - k_1 - k_2$. (Though size is a function of r , there is no need to make this explicit here.) For

allocation r , the average effects in the treatment groups are:

$$E(D(r) - D(o) \mid A(r) = j) = E(D(r) \mid A(r) = j) = k_j^{-1} \sum_{\{i:A_i(r)=j\}} D_i(r), \quad (1)$$

as $D_i(o) = 0$ for all i . (1) is the lease-up rate in group j under allocation r .

In (1), interest centers on a specific allocation, for example, the allocation where all N_v participants receive vouchers and assistance to relocate to low poverty areas. But suppose a researcher is interested in the effect of randomly selecting a participant and giving her a voucher and assistance. Here, the parameter of interest is the average value of (1) over all N_v allocations where one participant is given a voucher and the others are not.

More generally, consider any “mixed” allocation of the participants with group sizes k_0, k_1, k_2 , and let $R(k_0, k_1, k_2)$ denote the set of $N_v!/(k_0!k_1!k_2!)$ allocations with these sizes. Within each treatment group, the effect of randomly selecting k_0 ($k_0/N_v \times 100\%$ of the) families to receive no voucher, k_1 ($k_1/N_v \times 100\%$) to receive a section 8 voucher, etc., is the lease-up rate averaged over the set $R(k_0, k_1, k_2)$:

$$E_{R(k_0, k_1, k_2)} E(D(r) \mid A(r) = j) = \left(\frac{N_v!}{k_0!k_1!k_2!} \right)^{-1} \sum_{r \in R(k_0, k_1, k_2)} [k_j^{-1} \sum_{\{i:A_i(r)=j\}} D_i(r)]. \quad (2)$$

The parameter (2) is a natural parameter to consider when an allocation is (or will be) randomly chosen from the set $R(k_0, k_1, k_2)$, as in the MTO demonstration. Note, however, that because the allocation is randomly selected, the policy relevance of (2) may be diminished when the components of (2) vary greatly over the members of this set. (2) will also be of interest when the scientist or policy maker expects this parameter to be “frequency dependent”, that is, when (2) is a function $f(k_0/N_v, k_1/N_v)$ of the proportions receiving the various treatments.

MTO researchers have also studied outcomes that may depend on relocation status, for example, school characteristics (Ladd and Ludwig 2003). Building on the notation in Angrist et al. (1996), let $\tilde{D}(r)$ be $N \times 1$, where $\tilde{D}_i(r)$ denotes the treatment unit i receives ($\tilde{D}_i(r) = A_i(r) \times D_i(r)$), and let $Y_i(r, \tilde{D}(r))$ denote the value of outcome Y for family i under allocation r . Note that $Y_i(o, \tilde{D}(o))$ is now unknown. The average effect of allocation r on Y among MTO participants is:

$$E(Y(r, \tilde{D}(r)) - Y(o, \tilde{D}(o)) \mid V = 1) = N_v^{-1} \sum_{i=1}^{N_v} (Y_i(r, \tilde{D}(r)) - Y_i(o, \tilde{D}(o))). \quad (3)$$

Equation (3) is a weighted average of three components: a) the average “spillover” effect under allocation r on MTO families who do not receive a voucher under this allocation, b) the average effect among families receiving a section 8 voucher, and c) the average effect among families receiving a voucher and assistance. Unlike effects b) and c), which may be regarded as generalizations of the ITT, treatment status is constant

for the spillover effect a). In the absence of interference, the spillover effect is 0; thus, a) is a “pure effect” of interference.

Typically, the components of (3) will be of interest, as may be the average over the family $R(k_0, k_1, k_2)$:

$$E_{R(k_0, k_1, k_2)} E(Y(r, \tilde{D}(r)) - Y(o, \tilde{D}(o)) | A(r) = j, V = 1) = \left(\frac{N_v!}{k_0!k_1!k_2!}\right)^{-1} \sum_{r \in R(k_0, k_1, k_2)} [k_j^{-1} \sum_{\{i: A_i(r)=j, V_i=1\}} (Y_i(r, \tilde{D}(r)) - Y_i(o, \tilde{D}(o)))]. \quad (4)$$

Thus far, only participants have been considered. Spillovers for non-participants, who also do not receive vouchers, may also be defined, for example:

$$E_{R(k_0, k_1, k_2)} E(Y(r, \tilde{D}(r)) - Y(o, \tilde{D}(o)) | G = 1), \quad (5)$$

where $G = 1$ for non-participants in the origin neighborhoods, 0 otherwise. Implicitly recognizing the possibility of interference, several researchers (for example, Ludwig et al. (2001)) have called attention to the potential effects of MTO on non-participants in both origin and destination neighborhoods. However, data allowing consideration of such issues have not been collected.

The components of (4) are generalizations of the ITT. The TOT, defined using the no interference assumption and the exclusion restriction that the “direct effect” of treatment assignment on Y is 0, is also of interest. The TOT is the average effect, among movers, of moving using a section 8 voucher ($j = 1$) or a voucher plus assistance ($j = 2$).

Under the no interference assumption, the exclusion restriction formalizes the notion that only the actual treatment a unit receives matters. For example, an experimental unit that does not use a voucher to move and whose value on the response is y would have the same value had it been assigned to the control group.

Angrist et al. (1996, p. 447) state the exclusion restriction for the case of one treatment group and a control group, where members of the control group may take up treatment and members of the treatment group may not take up treatment. Here, it is necessary to modify their definition to a) distinguish between the two treatments that may be taken up and b) account for the fact that members of the control group cannot take up either of the treatments.

To that end, let \tilde{D} be a $N \times 1$ vector of potential treatments received, with $\tilde{D}_i \in \{0, 1, 2\}$ for $i = 1, \dots, N$. For a given allocation r , a vector \tilde{d} is said to be “feasible” if, for all i , $A_i(r) = 0$ implies $\tilde{d}_i = 0$, $A_i(r) = 1$ implies $\tilde{d}_i = 0$ or 1, $A_i(r) = 2$ implies $\tilde{d}_i = 0$ or 2. Note that for any allocation, $\tilde{d}_i = 0$ for $i = N_v + 1, \dots, N$. The exclusion restriction may now be stated as:

$$(A1) \text{ For all } i, \text{ and for all } r, r' \text{ and } \tilde{D} \text{ such that } \tilde{D} \text{ is feasible under both } r \text{ and } r', Y_i(r, \tilde{D}) = Y_i(r', \tilde{D}) \equiv$$

$Y_i(\tilde{D})$.

Using (A1), the TOT may be generalized to the case where interference is present:

$$E_{R(k_0, k_1, k_2)} E(Y(\tilde{D}(r)) - Y(\tilde{D}(o)) | A(r) = j, D(r) = 1) = \left(\frac{N_v!}{k_0!k_1!k_2!}\right)^{-1} \sum_{r \in R(k_0, k_1, k_2)} [n_j^{-1}(r) \sum_{\{i: \tilde{D}_i(r)=j\}} (Y_i(\tilde{D}(r)) - Y_i(\tilde{D}(o)))], \quad (6)$$

where $n_j(r) = \sum_{\{i: A_i(r)=j\}} D_i(r)$ is the number of units in group $j = 1, 2$ who move using their vouchers under allocation r .

The Role of the No Interference Assumption

The no interference assumption, with respect to leasing up and an outcome Y , is:

(A2) For all r, r' and i , if $A_i(r) = A_i(r')$, $D_i(r) = D_i(r')$, $Y_i(r, \tilde{D}(r)) = Y_i(r', \tilde{D}(r'))$.

Under (A2), the response of unit i depends on allocation r only through its treatment assignment, precluding spillovers. Causal comparisons now reduce to comparisons of responses under different treatments. This is reflected in the notation by replacing allocations with treatments: $D_i(r) \equiv D_i(A_i(r))$, $\tilde{D}_i(r) \equiv \tilde{D}_i(A_i(r))$, $Y_i(r, \tilde{D}(r)) = Y_i(A_i(r), \tilde{D}_i(A_i(r)))$. In this case,

$$\sum_{r \in R(k_0, k_1, k_2)} \sum_{\{i: A_i(r)=j\}} D_i(r) = \frac{N_v!k_j}{k_0!k_1!k_2!N_v} \sum_{i=1}^{N_v} D_i(j). \quad (7)$$

Thus, for any values of k_0 , k_1 and k_2 , (2) simplifies to:

$$N_v^{-1} \sum_{i=1}^{N_v} D_i(j) = E(D(j) | V = 1) = \Pr(D(j) = 1 | V = 1), \quad (8)$$

the probability of leasing up when all MTO families are assigned to treatment j ; (2) is also the probability a randomly selected MTO family leases up when assigned to group $j = 1, 2$.

Similarly, when (A2) holds, the generalization of the ITT (4) reduces to the usual ITT:

$$E(Y(j, \tilde{D}(j)) - Y(0, \tilde{D}(0)) | V = 1) = N_v^{-1} \sum_{i=1}^{N_v} (Y_i(j, \tilde{D}_i(j)) - Y_i(0, \tilde{D}_i(0))). \quad (9)$$

When the exclusion restriction (A1) holds in addition to (A2), the notation simplifies: $Y_i(j, \tilde{D}_i(j)) \equiv Y_i(\tilde{D}_i(j))$. (9) may then be written as:

$$E(Y(\tilde{D}(j)) - Y(\tilde{D}(0)) | V = 1) = N_v^{-1} \sum_{i=1}^{N_v} (Y_i(\tilde{D}_i(j)) - Y_i(\tilde{D}_i(0))). \quad (10)$$

Because the direct effect of treatment assignment is 0 and $D_i(0) = 0$ for all i , only families who would move if offered a voucher contribute to (10); thus (10) reduces to

$$E(Y(\tilde{D}(j)) - Y(\tilde{D}(0)) | V = 1, D(j) = 1) \Pr(D(j) = 1 | V = 1), \quad (11)$$

suggesting the effect of moving per se can be obtained by adjusting (10) for the lease-up rate (8).

More formally, Angrist et al. (1996) show that under (A1)-(A2), the monotonicity assumption

$$(A3) \ D_i(j) \geq D_i(0) \text{ for } j = 1, 2,$$

and the assumption

$$(A4) \ (8) > 0,$$

the IV estimand, defined as (10)/(8), equals the “Local Average Treatment Effect” (LATE), the average effect for compliers (those who take up treatment when assigned and do not move using a voucher when assigned to the control group). When, in addition, $D_i(0) = 0$ for all i , as here, the compliers constitute the subpopulation of units taking up treatment, so LATE is equal to the TOT:

$$E(Y(\tilde{D}(j)) - Y(\tilde{D}(0)) | V = 1, D(j) = 1). \quad (12)$$

When (A1) and (A2) hold, the generalized TOT (6) reduces to (12).

4 What Parameters Are Estimated in Randomized Studies of Housing Mobility?

I consider the parameters estimated by MTO researchers when interference is present. In keeping with section three, expectations are taken over the distribution of allocations induced by the experimental assignment mechanism. Theorems 5 and 6 complement the results in Angrist et al. (1996), who studied the behavior of the IV estimand a) when the exclusion restriction (A1) does not hold and b) when the monotonicity assumption (A3) does not hold. The results below are also readily modified to handle a model-based approach to inference; the main results would not change, though some simplifications would occur asymptotically. I now explicitly introduce the randomization assumption:

$$(A5) \ \text{Given } k_0, k_1, k_2, \text{ every allocation } r \in R(k_0, k_1, k_2) \text{ has equal probability } (N_v! / (k_0! k_1! k_2!))^{-1}.$$

Researchers have interpreted $\bar{D}_j(r)$, the percentage of MTO families in group $j = 1, 2$ who lease up under assignment r , as an estimate of (8). For the section 8 group, the percentages vary from 47.6 % in Boston to 75.4 % in Los Angeles, and for the experimental group, between 33.8 % in Chicago to 61.2 % in Los Angeles (Goering et al. 2002). But the lease-up rate (8) is well defined only if the no interference assumption (A2) holds. Otherwise it follows, using $\bar{D}_j(r) = E(D(r) | A(r) = j)$:

Theorem 1. Under the randomization assumption (A5), $\bar{D}_j(r)$ is unbiased for the lease-up rate (2).

If the lease-up rate (2) depends on the proportions assigned to each group, the usual interpretation is not only incorrect, as (8) is ill-defined, but misleading. While that is a possibility here, the observed variation in the lease-up rates cannot be accounted for in this way, as the proportions assigned to each group are similar at each site.

Several researchers (Leventhal and Brooks-Gunn 2003, Hanratty et al. 2003) use $\bar{Y}_j(r) - \bar{Y}_0(r)$, the difference between the treatment (for $j = 1, 2$) and control group means on the response under allocation r , to estimate the ITT (9); under assumptions (A2) and (A5), this estimator is unbiased for (9). Others (Katz et al. 2000, 2001, Kling et al. (2005), Ludwig et al. 2001, 2005) use a regression adjusted estimator that is also unbiased for (9) under assumptions (A2) and (A5). However, as the no interference assumption (A2) is untenable, it is important to understand the properties of these estimators when interference is present. This is addressed in the following three theorems.

Theorem 2. Under the randomization assumption (A5), for $j = 1, 2$ $\bar{Y}_j(r) - \bar{Y}_0(r)$ is unbiased for the parameter

$$E_{R(k_0, k_1, k_2)}[E(Y(r, \tilde{D}(r)) | A(r) = j) - E(Y(r, \tilde{D}(r)) | V = 1, A(r) = 0)]. \quad (13)$$

The result follows immediately from $\bar{Y}_j(r) - \bar{Y}_0(r) = E(Y(r, \tilde{D}(r)) | A(r) = j) - E(Y(r, \tilde{D}(r)) | V = 1, A(r) = 0)$.

The regression adjusted estimator is also unbiased for (13) under some additional conditions. Let $\underline{Y}(r, \tilde{D}(r))$ denote the $N_v \times 1$ vector of responses of MTO participants under assignment r , let $X(r)$ be a $N_v \times K$ matrix of predictors, $\underline{1}$ a $N_v \times 1$ vector of 1's, and, for $j = 1, 2$, let $\underline{Z}_j(r)$ be $N_v \times 1$, with $Z_{ij}(r) = 1$ if $A_i(r) = j$, 0 otherwise. Consider the regression:

$$\underline{Y}(r, \tilde{D}(r)) = \underline{1}\alpha(r) + X(r)\underline{\beta}(r) + \underline{Z}_1(r)\gamma_1(r) + \underline{Z}_2(r)\gamma_2(r) + \underline{\epsilon}(r), \quad (14)$$

where $E(\underline{\epsilon}(r) | \underline{X}(r), \underline{Z}_1(r), \underline{Z}_2(r)) = \underline{0}$. In lieu of $\bar{Y}_j(r) - \bar{Y}_0(r)$, some researchers have used the ordinary

least squares coefficients $\hat{\gamma}_j(r)$, $j = 1, 2$, from (14) to estimate the ITT (9). Note that with all N_v observations in the MTO subpopulation observed, $\hat{\gamma}_j(r) = \gamma_j(r)$. Averaging the regression function with respect to the conditional distribution of the predictors given treatment assignment yields:

$$\begin{aligned} & E(Y(r, \tilde{D}(r)) | A(r) = j) - E(Y(r, \tilde{D}(r)) | V = 1, A(r) = 0) = \\ & \gamma_j(r) + \underline{\beta}'(r)[E(\underline{X}(r) | A(r) = j) - E(\underline{X}(r) | V = 1, A(r) = 0)]. \end{aligned} \quad (15)$$

Averaging over the set of assignments $R(k_0, k_1, k_2)$ yields:

Theorem 3. Under assumption (A5) and the regression model (14),

$$E_{R(k_0, k_1, k_2)} \gamma_j(r) = (13) - E_{R(k_0, k_1, k_2)} \underline{\beta}'(r)[E(\underline{X}(r) | A(r) = j) - E(\underline{X}(r) | V = 1, A(r) = 0)]. \quad (16)$$

In (14), $X(r)$, α_r , and $\underline{\beta}(r)$ depend on the allocation r . Typically, the predictors are pretreatment covariates, implying $X(r) = X$. If, in addition, $\underline{\beta}(r) = \underline{\beta}$, $\gamma_j(r)$ is unbiased for (13) because then $E_{R(k_0, k_1, k_2)}[E(\underline{X} | A(r) = j) - E(\underline{X} | V = 1, A(r) = 0)] = 0$.

The parameter (13) is the mean difference between the units in treatment group j and the units in the control group, under assignments of the form $r \in R(k_0, k_1, k_2)$. Because different units are compared under the same allocation, (13) does not admit a causal interpretation unless (A2) holds, in which case the control outcomes are by assumption the same under allocations r and o . In contrast, the parameter of interest (4) is the mean difference (over $r \in R(k_0, k_1, k_2)$) between the units in treatment group j , $j = 1, 2$ under allocation r and allocation o ; this is the average effect of implementing at random an assignment from the family $R(k_0, k_1, k_2)$.

While Theorems 2 and 3 show that the usual estimates are biased for the ITT (4), they do not reveal the nature of the bias. I now address this by characterizing the relationship between (13) and the causal parameter (4).

Theorem 4. The noncausal parameter (13) is the difference between the ITT (4) for treatment group j and the spillover effect (4) for the control group:

$$\begin{aligned} (13) &= E_{R(k_0, k_1, k_2)} E(Y(r, \tilde{D}(r)) - Y(o, \tilde{D}(o)) | A(r) = j, V = 1) - \\ & E_{R(k_0, k_1, k_2)} E(Y(r, \tilde{D}(r)) - Y(o, \tilde{D}(o)) | A(r) = 0, V = 1). \end{aligned}$$

The result follows by observing that both terms $E_{R(k_0, k_1, k_2)} E(Y(o, \tilde{D}(o)) | A(r) = j, V = 1)$ and $E_{R(k_0, k_1, k_2)} E(Y(o, \tilde{D}(o)) | V = 1, A(r) = 0)$ in (4) are equal to $E(Y(o, \tilde{D}(o)) | V = 1)$.

The characterization of bias in theorem 4 is simple and important, revealing that inappropriately assuming no interference can lead to inferences that are extremely misleading. For example, suppose a

researcher finds $\bar{Y}_2(r) - \bar{Y}_0(r) > 0$ at the .01 level of significance, leading him to conclude that giving participants vouchers and assistance to move to a low poverty neighborhood is beneficial, at least with respect to the outcome Y . This is clearly incorrect: $\bar{Y}_2(r) - \bar{Y}_0(r)$ is an unbiased estimate of (13) and the .01 significance level also pertains to this parameter. The correct conclusion is that the average effect in the experimental group is greater than the spillover effect. The effects could be negative in both groups. Similarly, one could conclude that a treatment that is beneficial is harmful.

Because the estimators in theorems 2 and, under additional conditions, theorem 3, are biased for the effect in group j by the spillover effect, substantive assumptions about the latter can be used to bound and/or sign the effect in the treatment groups. For example, Hanratty et al.(2003) find members of the Los Angeles section 8 group worked 10.4 ($p < .05$) hours more per week than controls. While outflow to other neighborhoods may have been accompanied by a reduction in the local demand for services, leading controls to work less than they otherwise would have, it seems more plausible to believe the spillover effect is 0, or if controls took jobs previously held by members of one of the treatment groups, greater than 0; the 10.4 hour difference is then a lower bound estimate. Katz et al. (2003) find members of the Boston experimental group 16% ($p < .05$) less likely than controls to perceive that streets near home are unsafe or very unsafe during the day. As participants in the treatment groups maintained close ties to project residents, they may have communicated positive sentiments about their new neighborhoods that influence controls to perceive their neighborhood as more unsafe than they would have otherwise. If this is the case, 16% would be an upper bound on the reduction.

Many participants receiving vouchers do not move. A number of researchers (Hanratty et al. 2003, Katz et al. 2001, Kling et al. 2005, Ladd and Ludwig 2003, Ludwig et al. 2001,2005) estimate the effect of using a voucher to move in the section 8 group ($j = 1$) and in the experimental group ($j = 2$) using a two stage least squares estimator adjusted for covariates that is asymptotically equivalent to the IV estimator. I now consider the behavior of both estimators when the no interference assumption (A2) does not hold. It is useful to write $\bar{Y}_j(r) - \bar{Y}_0(r)$ as:

$$\begin{aligned} & E(Y(r, \tilde{D}(r)) - Y(o, \tilde{D}(o)) \mid A(r) = j, D(r) = 1) \Pr(D(r) = 1 \mid A(r) = j) + \\ & E(Y(r, \tilde{D}(r)) - Y(o, \tilde{D}(o)) \mid A(r) = j, D(r) = 0) \Pr(D(r) = 0 \mid A(r) = j) - \\ & (E(Y(r, \tilde{D}(r)) \mid V = 1, A(r) = 0) - E(Y(o, \tilde{D}(o)) \mid A(r) = j)). \end{aligned} \quad (17)$$

To study the behavior of the IV estimator over assignments $R(k_0, k_1, k_2)$, assumption (A4) requires modification:

(A4*) For all $r \in R(k_0, k_1, k_2)$, $\bar{D}_j(r) > 0$.

Note that when k_j is “small”, this assumption may not be reasonable.

Theorem 5. Under the exclusion restriction (A1), assumption (A4*) and the randomization assumption (A5), the IV estimator $(\bar{Y}_j(r) - \bar{Y}_0(r))/\bar{D}_j(r)$ has expectation:

$$\begin{aligned} & E_{R(k_0, k_1, k_2)} \{ E(Y(\tilde{D}(r)) - Y(\tilde{D}(o)) \mid A(r) = j, D(r) = 1) + \\ & E_{R(k_0, k_1, k_2)} E(Y(\tilde{D}(r)) - Y(\tilde{D}(o)) \mid A(r) = j, D(r) = 0) \frac{\Pr(D(r) = 0 \mid A(r) = j)}{\Pr(D(r) = 1 \mid A(r) = j)} - \\ & E_{R(k_0, k_1, k_2)} \frac{E(Y(\tilde{D}(r)) \mid V = 1, A(r) = 0) - E(Y(\tilde{D}(o)) \mid A(r) = j)}{\Pr(D(r) = 1 \mid A(r) = j)}. \end{aligned} \quad (18)$$

The first component of (18) is (6), the generalized TOT when interference is present. The second component of (18) is a weighted average of the “pure” effects of interference among the non-movers in group j under assignment r versus the benchmark assignment o . Under the additional assumption (A2), the unit effects for those who do not use a voucher to move are 0, and hence this term vanishes. This term also vanishes when, for every assignment $r \in R(k_0, k_1, k_2)$, $\bar{D}_j(r) = 1$; this latter condition does not hold for either the section 8 or experimental group at any of the MTO sites. Finally, the third term is the inversely weighted average difference (over $R(k_0, k_1, k_2)$) of the control outcomes under assignment r and the baseline outcomes of units assigned to group j .

The two stage least squares estimator is defined as follows. Consider the model:

$$\underline{Y}(r, \tilde{D}(r)) = \underline{1}\tau_0(r) + X(r)\underline{\eta}(r) + \tilde{Z}_1(r)\delta_1(r) + \tilde{Z}_2(r)\delta_2(r) + \varepsilon(r) = \tilde{W}(r)\underline{\lambda}(r) + \varepsilon(r), \quad (19)$$

where, for $j = 1, 2$ $\tilde{Z}_j(r)$ is $N_v \times 1$, with $\tilde{Z}_{ij}(r) = Z_{ij}(r) \times D_i(r)$, and $E(\varepsilon(r) \mid X(r)) = \underline{0}$. Note that $\tilde{Z}_{i1}(r) = 1$ for units that move using a housing voucher when assigned to the section 8 group under assignment r , 0 otherwise; a similar interpretation holds for $\tilde{Z}_{i2}(r)$. To estimate the parameters of (19), for $j = 1, 2$, the variables $Z_{ij}(r)$ are used as instruments for the endogenous variables $\tilde{Z}_{ij}(r)$. Letting $W(r) = (X(r), \underline{1}, \underline{Z}_1(r), \underline{Z}_2(r))$, the estimator is the solution to the set of $K + 3$ equations: $W'(r)\underline{Y}(r, \tilde{D}(r)) = W'(r)\tilde{W}(r)\underline{\lambda}(r)$. Using the last 3 equations of this system leads, for $j = 1, 2$, to the result:

$$\delta_j(r) = \frac{\bar{Y}_j(r) - \bar{Y}_0(r)}{\bar{D}_j(r)} - \frac{(\bar{X}_j(r) - \bar{X}_0(r))'\underline{\eta}(r)}{\bar{D}_j(r)}, \quad (20)$$

where $\bar{X}_j(r) = (E(X_1(r) \mid A(r) = j), \dots, E(X_K(r) \mid A(r) = j))' \equiv E(\underline{X}(r) \mid A(r) = j)$; the vector $\bar{X}_0(r)$ is defined in a similar fashion. That is, $\delta_j(r)$ is the IV estimator $(\bar{Y}_j(r) - \bar{Y}_0(r))/\bar{D}_j(r)$ minus an adjustment for the different values of the covariates in the treatment and control group. This leads immediately to:

Theorem 6. Under the exclusion restriction (A1), assumption (A4*) and the randomization assumption (A5), the two stage least squares estimator $\delta_j(r)$ has expectation:

$$(18) - E_{R(k_0, k_1, k_2)} \frac{(E(\underline{X}_j(r) | A(r) = j) - E(\underline{X}_0(r) | V = 1, A(r) = 0))' \underline{\eta}(r)}{\Pr(D(r) = 1 | A(r) = j)}. \quad (21)$$

If neither $X(r)$ nor $\underline{\eta}_r$ depends on r , the second term in (21) simplifies to

$$E_{R(k_0, k_1, k_2)} \frac{(E(\underline{X}_j | A(r) = j) - E(\underline{X}_0 | V = 1, A(r) = 0))' \underline{\eta}}{\Pr(D(r) = 1 | A(r) = j)}. \quad (22)$$

5 DISCUSSION

The critical role of the no-interference assumption (A2) in causal inference is examined in the context of a randomized study of housing mobility. When interference is present, the lease-up rate for policy r is unbiased for the lease-up rate over the collection of policies $R(k_0, k_1, k_2)$, but there is not a single lease-up rate, as in the case of no interference. When interference is present, $\bar{Y}_j(r) - \bar{Y}_0(r)$, the difference between the treatment and control group means, does not estimate a causal parameter, but rather the difference between two causal parameters: 1) the average effect of treatment j (over $R(k_0, k_1, k_2)$) and 2) the spillover effect on the untreated (over $R(k_0, k_1, k_2)$). When the no interference assumption (A2) is violated and $\bar{Y}_j(r) - \bar{Y}_0(r)$ is viewed as an estimate of the average treatment effect (9), very misleading inferences about the benefits of relocation may be drawn. Similar remarks apply when the alternative estimator $\hat{\gamma}_j(r)$ is used.

In addition, when the exclusion restriction (A1) holds and there is no interference, the unit effects of treatment j are 0 for families who do not move if assigned to this treatment. But (A1) alone does not imply that the effects are 0 for these non-movers. Consequently, neither the IV estimand nor the IV estimand adjusted for covariates equals the generalized TOT (6) when interference is present.

When the no-interference assumption fails, responses from the control group under allocations r and o are not necessarily identical, suggesting the need to also collect data under allocation o . Of course, this is not possible.

In social contexts such as schools and small communities, researchers will often use group randomized studies to circumvent the interference problem. Implicitly, the no interference assumption (A2) is replaced by a partial interference assumption A2* , where the units are grouped into classes and there is no interference between units in different classes:

(A2*) Let $c(i) \in \{1, \dots, C\}$ denote the class to which unit i belongs and decompose $\underline{A}(r) \equiv (\underline{A}'_1(r), \dots, \underline{A}'_C(r))'$, where $\underline{A}_c(r)$ is the column vector of allocations to the members of class c , $c = 1, \dots, C$. Then, for all r, r'

and i, if $\underline{A}_{c(i)}(r) = \underline{A}_{c(i)}(r')$, $D_i(r) = D_i(r')$, $Y_i(r, \underline{D}(r)) = Y_i(r', \underline{D}(r'))$.

Randomization is then used to assign treatments to classes, with all the units in a class receiving the same treatment. In this case, by A2*), the units assigned to treatment group j , $j = 1, 2$ have identical outcomes under the randomized assignment and the allocation $k_j = N_v$ in which all units are assigned to treatment j . Similarly, units in the control group have identical values under the randomized assignment and the baseline allocation o .

Recall that at each of the 5 sites, MTO participants were recruited from housing projects. If participants from different projects do not interact and direct social interactions are the only source of interference, treatments could be randomly assigned to projects. However, when the number of projects is small, as here (ranging from 6 in Chicago to 14 in New York), causal parameters may not be estimated very precisely. This suggests that it would have been better to conduct a group randomized study using more projects. Further, had data on social interactions been collected prior to assignment, the interaction sets of participants could have been used to partition the participants into equivalence classes with no direct social interactions, and treatments could have been randomly assigned to these classes; there will be at least as many classes as projects. Note, however, that when there are other sources of interference, for example, competition for housing units in a tight rental market, it may be necessary to treat each site as a class, reducing the number of classes to 5.

When there are enough groups to estimate parameters with “reasonable” precision, group randomized studies will be useful for making inferences about allocations of the form $k_j = N_v$ or special “mixed” allocations where all the units in a given class receive the same treatment.

But these studies may be considerably less useful for making inferences about allocations (or more generally, policies) in which only a subset of a group receives a particular treatment. In addition, allocations of this nature can be of great theoretical interest. The spillover effect (equation (4) with $j = 0$) compares the outcomes of units who are not treated under either the allocation(s) of interest or the benchmark allocation o . By studying such effects, a researcher might learn that the benefits of treatment accrue to the entire class when only a subset of the class is treated. Such information could then also be used to design policies optimizing the use of scarce resources. Hopefully, students of causal inference and experimental design will devote attention to these important issues in future work.

References

- Aaronson, D. (1998), "Using Sibling Data to Estimate the Impact of Neighborhoods on Children's Educational Outcomes," *Journal of Human Resources*, 33, 915-946.
- Anderson, E. (1999), *The Code of the Street: Decency, Violence and the Moral Life of the Inner City*. New York: W.W. Norton.
- Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996), "Identification of Causal Effects Using Instrumental Variables" (with discussion), *Journal of the American Statistical Association*, 91, 444-472.
- Bénabou, R. (1993), "Workings of a City: Location, Education and Production," *Quarterly Journal of Economics*, 108, 619-652.
- (1996), "Equity and Efficiency in Human Capital Investment: The Local Connection," *Review of Economic Studies*, 62, 237-264.
- Blume, L. E., and Durlauf, S. N. (2001), "The Interactions-Based Approach to Socioeconomic Behavior," in *Social Dynamics*, eds. S. N. Durlauf and H. P. Young, Brookings Institution Press, Washington, D. C., pp. 15-44.
- Coulton, C. J., and Pandey, S. (1992), "Geographic Concentration of Poverty and Risk to Children in Urban Neighborhoods," *American Behavioral Scientist*, 35, 238-257.
- Cox, D. R. (1958), *The Planning of Experiments*, New York: Wiley.
- Crane, J. (1991), "The Epidemic Theory of Ghettos and Neighborhood Effects on Dropping Out and Teenage Childbearing," *American Journal of Sociology*, 96, 1226-1259.
- Cutler, D., and Glaeser, E. (1997), "Are Ghettos Good or Bad?," *Quarterly Journal of Economics*, 112, 827-872.
- Dachter, L. (1982), "Effects of Community and Family Background on Achievement," *Review of Economics and Statistics*, 64, 32-41.

Duncan, G. J., Connell, J. P., and Klebanov, P. K. (1997), "Conceptual and Methodological Issues in Estimating Causal Effects of Neighborhoods and Family Conditions on Individual Development," in *Neighborhood Poverty* (Vol. 1), eds. J. Brooks-Gunn, G. J. Duncan, and J. L. Aber, Russell Sage Foundation, New York, pp. 219-250.

Del Conte, A., and Kling, J. (2001), "A Synthesis of MTO Research on Self-Sufficiency, Safety and Health, and Behavior and Delinquency," *Poverty Research News*, 5, pp. 3-6.

Durlauf, S. (1996), "A Theory of Persistent Income Inequality," *Journal of Economic Growth*, 1, 75-93.

——— (2001), "A Framework for the Study of Individual Behavior and Social Interactions" (with discussion), in *Sociological Methodology 2001*, eds. M. E. Sobel and M. P. Becker, Oxford: Blackwell, pp. 47-128.

Evans, W., Oates, W., and Schwab, R. (1992), "Measuring Peer Group Effects: A Study of Teenage Behavior," *Journal of Political Economy*, 100, 966-999.

Gephart, M. A. (1997), "Neighborhoods and Communities as Contexts for Development," in *Neighborhood Poverty* (Vol. 1), eds. J. Brooks-Gunn, G. J. Duncan, and J. L. Aber, Russell Sage Foundation, New York, pp. 1-43.

Glaeser, E. L., and Scheinkman, J. A. (2001), "Measuring Social Interactions," in *Social Dynamics*, eds. S.N. Durlauf and H.P. Young, Brookings Institution Press, Washington, D.C., pp. 83-131.

Goering, J., Feins, J., and Richardson, T. M. (2002), "A Cross-Site Analysis of Initial Moving to Opportunity Demonstration Results," *Journal of Housing Research*, 13, 1-30.

Goering, J., Kraft, J., Feins, J., McInnis, D., Holin, M.J., Elhassan, H. (1999), "Moving to Opportunity for Fair Housing Demonstration Program: Current Status and Initial Findings," U.S. Department of Housing and Urban Development, Washington, D.C.

Halloran, M. E., and Struchiner, C. J. (1995), "Causal Inference in Infectious Diseases," *Epidemiology*, 6,

142-151.

Hanratty, M. H., McLanahan, S. A., and Pettit, B. (2003), "Los Angeles Site Findings ", in *Choosing A Better Life*, eds. J. Goering and J. Feins, The Urban Institute Press, Washington, D. C., pp. 245-274.

Heckman, J., Lochner, L., and Taber, C. (1998), "General Equilibrium Treatment Effects: A Study of Tuition Policy," *The American Economic Review (Papers and Proceedings)*, 88(2), 381-386.

Imbens, G. W., and Angrist, J. (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467-476.

Jargowsky, P. A. (1997) *Poverty and Place: Ghettos, Barrios, and the American City*, New York: Russell Sage Foundation.

Jencks, C., and Mayer, S. E. (1990), "The Social Consequences of Growing Up in a Poor Neighborhood," in *Inner-City Poverty in the United States*, eds. L. E. Lynn, Jr., and M. G. H. McGeary, National Academy Press, Washington, D. C., pp. 111-186.

Katz, L. F., Kling, J., and Liebman, J. B. (2001), "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment," *Quarterly Journal of Economics*, 116, 607-654.

———(2003), " Boston Site Findings", in *Choosing a Better Life*, eds. J. Goering and J. Feins, The Urban Institute Press, Washington, D. C., pp. 177-212.

Kling, J., Ludwig, J., and L. F. Katz (2005), "Neighborhood Effects on Crime for Female and Male Youth: Evidence from a Randomized Housing Voucher Experiment," *Quarterly Journal of Economics* 120, 87-130.

Ladd, H. F., and Ludwig, J.(2003), "The Effects of MTO on Educational Opportunities in Baltimore," in *Choosing a Better Life*, eds. J. Goering and J. Feins, The Urban Institute Press, Washington, D. C., pp. 117-151.

Leventhal, T., and Brooks-Gunn, J. (2003), " New York Site Findings," in *Choosing a Better Life*, eds. J. Goering and J. Feins, The Urban Institute Press, Washington, D. C., 213-244.

- Ludwig, J. Duncan, G. J., and Pinkston, J. (2005), "Housing Mobility Programs and Economic Self-Sufficiency: Evidence from a Randomized Experiment," *Journal of Public Economics*, 89,131-156.
- Ludwig, J., Hirschfield, P., and G. J. Duncan (2001), "Urban Poverty and Juvenile Crime: Evidence from a Randomized Housing Mobility Experiment," *Quarterly Journal of Economics*, 116, 665-679.
- Manski, C. F. (1995), *Identification Problems in the Social Sciences*, Cambridge, MA: Harvard University Press.
- Moffitt, R. A. (2001), "Policy Interventions, Low-Level Equilibria, and Social Interactions," in *Social Dynamics*, eds. S.N. Durlauf and H.P. Young, Brookings Institution Press, Washington, D.C., pp. 45-82.
- Murray, D. (1998), *Design and Analysis of Group-Randomized Trials*, New York: Oxford University Press.
- Pettit, B., McLanahan, S., and M. Hanratty (2000), "Moving to Opportunity: Benefits and Hidden Costs", unpublished manuscript.
- Popkin, S. J., Harris, L. E., and M. K. Cunningham (2002), "Families in Transition: A Qualitative Analysis of the MTO Experience". Washington, D.C.: The Urban Institute.
- Rosenbaum, E., and Harris, L. E., (2000), "Low Income Families in their New Neighborhoods: The Short Term Effects of Moving from Chicago's Public Housing," *Journal of Family Issues*, 22, 187-210.
- (2001), "Residential Mobility and Opportunities: Early Impacts of the Moving to Opportunity Demonstration Program in Chicago," *Housing Policy Debate*, 12, 321-346.
- Rosenbaum, E., Harris, L. E., and Denton, N. A. (2003), "New Places, New Faces," in *Choosing a Better Life*, eds. J. Goering and J. Feins, The Urban Institute Press, Washington, D. C., pp. 275-310.
- Rubin, D. B. (1974), "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688-701.

—— (1977), “Assignment to Treatment Groups on the Basis of a Covariate,” *Journal of Educational Statistics*, 2, 1-26.

—— (1978), “Bayesian Inference for Causal Effects: The Role of Randomization,” *Annals of Statistics*, 6, 34-58.

—— (1980), Comment on “Randomization Analysis of Experimental Data: The Fisher Randomization Test,” by D. Basu, *Journal of the American Statistical Association*, 75,591-593.

Rubinowitz, L. S., and J. E. Rosenbaum (2000), “Crossing the Class and Color Lines: From Public Housing to White Suburbia,” Chicago: University of Chicago Press.

Sampson, R. J., and W. B. Groves (1989), “Community Structure and Crime: Testing Social Disorganization Theory,” *American Journal of Sociology*, 94, 774-802.

Sampson, R. J., Morenoff, J., and T. Gannon-Rowley (2001), “Assessing Neighborhood Effects: Social Processes and New Directions in Research,” *Annual Review of Sociology*, 28, 443-478.

Shroder, M. (2002), “Locational Constraint, Housing Counseling, and Successful Lease-Up in a Randomized Housing Voucher Experiment,” *Journal of Urban Economics* 51, 315-338.

Sobel, M. E. (2006), “Spatial Concentration and Social Stratification: Does the Clustering of Disadvantage ‘Beget’ Bad Outcomes?,” in *Poverty Traps*, eds. S. Bowles, S. Durlauf, and K. Hoff, Princeton University Press, Princeton, N.J., pages 204-229.

Weinberg, B. A., Reagan, P. B., and J. J. Yankow (2004), “Do Neighborhoods Affect Hours Worked: Evidence from the NLSY79,” *Journal of Labor Economics* 24:891-924.

Wilson, W. J. (1987), *The Truly Disadvantaged: The Inner City, The Underclass, and Public Policy*, Chicago: University of Chicago Press.