Composition or context: using transportability to understand drivers of site differences in a large-scale housing experiment

Kara E. Rudolph[*,1], **Nicole M. Schmidt**[2], **M. Maria Glymour**[3], **Rebecca Crowder**[1], **Jessica Galin**[1], **Jennifer Ahern**[1], and **Theresa L. Osypuk**[2]

[1]School of Public Health, University of California, Berkeley, California

[2]Department of Epidemiology and Community Health, University of Minnesota School of Public Health, Minneapolis, Minnesota

[3]Department of Epidemiology and Biostatistics, University of California, San Francisco, California

Abstract

Background—The Moving to Opportunity (MTO) experiment manipulated neighborhood context by randomly assigning housing vouchers to volunteers living in public housing to use to move to lower poverty neighborhoods in five US cities. This random assignment overcomes confounding limitations that challenge other neighborhood studies. However, differences in MTO's effects across the five cities have been largely ignored. Such differences could be due to population composition (e.g., differences in the racial/ethnic distribution) and/or to context (e.g., differences in the economy).

Methods—Using a nonparametric omnibus test and a multiply robust, semiparametric estimator for transportability, we assessed the extent to which differences in individual-level compositional characteristics that may act as effect modifiers can account for differences in MTO's effects across sites. We examined MTO's effects on marijuana use, behavioral problems, major depressive disorder, and generalized anxiety disorder among black and Latino adolescent males, where housing voucher receipt was harmful for health in some sites but beneficial in others.

Results—Comparing point estimates, differences in composition partially explained site differences in MTO effects on marijuana use and behavioral problems but did not explain site differences for major depressive disorder or generalized anxiety disorder.

Conclusions—Our findings provide quantitative, rigorous evidence for the importance of context or unmeasured individual-level compositional variables in modifying MTO's effects.

Keywords

housing policy; Moving to Opportunity; generalizability; transportability; multiply robust; machine learning

By randomizing receipt of a housing voucher that could be used to move into a lower poverty neighborhood, the Moving to Opportunity (MTO) experiment essentially randomized neighborhood context for families in public housing in five US cities: Baltimore, Boston, Chicago, Los Angeles, and New York (1). As such, it provides strong evidence on the potential for housing policy to affect economic, education, and health outcomes by improving neighborhood and housing environments. Because MTO is the only study with this experimental design, it has received extensive research attention. However, despite the large number of MTO publications, effects have rarely been compared across cities and site differences in effects have largely been ignored. Evaluating whether MTO had similar effects in all five cities, and if not, understanding drivers of those site differences, are important for understanding the generalizability of responses to housing policy.

Site differences in MTO effects could be due to 1) differences in the distribution of individual-level compositional factors across sites that modify intervention effectiveness, such as participants' race/ethnicity or motivations for enrolling in the study (henceforth referred to as "compositional"); and/or 2) differences in site-level contextual factors that modify intervention effectiveness, such as local economic or housing market conditions (henceforth referred to as "contextual"). Focusing on the latter explanation, MTO researchers concluded: "With only five sites, which differ in innumerable potentially relevant ways, it was simply not possible to disentangle the underlying factors that cause impacts to vary across sites" (2, p. B11), although qualitative research, including an examination of intervention implementation differences, has explored city-level underlying factors contributing to differences in MTO effects (3–5). Because five sites is too small a number to use traditional multi-level methods to quantitatively examine contextual drivers of site differences, and because other statistical tools to examine compositional drivers of site differences were not available, nearly all MTO analyses have reported results pooled across cities, controlling for city as a fixed effect in a regression model (6–8, e.g.). This strategy implicitly assumes that the MTO intervention effect in one city is the same as in another city, because the city fixed effect changes the intercept of the regression model but not the treatment effect coefficient.

However, the status quo of assuming a constant treatment effect across sites (and considering few if any compositional effect modifiers) may not be appropriate for the two reasons given above, and may even result in wasted resources if policies or programs are implemented in populations unlikely to benefit. Recent statistical advances in the subfield of transportability (which is related to generalizability or external validity) offer the opportunity to improve upon this status quo by evaluating the contribution of differences in population composition to site differences (9–15). Instead of predicting that an intervention will have the same effect in a new setting, we can flexibly incorporate numerous compositional effect modifiers into the prediction, resulting in potentially more accurate predictions that are "personalized" for place. While this will not allow us to fully disentangle the underlying compositional and contextual factors that cause impacts to vary across sites, it will allow us to assess the extent to which these site differences are due to specific sets of individual compositional characteristics.

Our objective is to use a recently developed flexible and robust transport estimator (10) to better understand site differences in MTO's effects. We focus on MTO's mental health and risk behavior effects among adolescent boys, because previous research unexpectedly found harmful effects on these outcomes among this subgroup (though effects for diagnostic mental health disorders were nonsignificant) (2, 7, 8, 16–18). We first test for site differences for each outcome. Focusing on outcomes with qualitative site differences—marijuana use, behavioral problems, major depressive disorder, and generalized anxiety disorder—we identify sites that differ in terms of whether receipt of a housing voucher was beneficial or harmful. Then, we employ the transport estimator to assess the extent to which those site differences could be explained by differences in the distribution of compositional factors. Differences that cannot be explained suggest that macro-level, contextual factors (or unmeasured compositional factors) may be critical to determining whether the intervention was harmful.

METHODS

Data

We used data on male youth who were enrolled in MTO at the baseline and interim visits. The baseline visit occurred 1994–1998, and we use outcomes measured at the follow-up visit, which occurred 4–7 years later when the youth were 12–19 years. MTO has been described previously (2, 6, 16). Briefly, it was a randomized control trial conducted by the U.S. Department of Housing and Urban Development enrolling 4,600 families living in public housing with children under 18. Weights accounted for changing random assignment ratios, sampling of children within households, and loss to follow up. (16). We excluded the Baltimore site from all analyses, because voucher receipt was not associated with a subsequent move to a low-poverty neighborhood (<25% of persons in poverty; results available upon request). We restricted to black and Hispanic/Latino youth as there were too few participants of other racial/ethnic groups to control for race/ethnicity without extrapolation concerns. Finally, we limited our analysis to those individuals with at least one nonmissing outcome. These exclusions resulted in a sample size of N=1,094–1,095 (depending on imputed dataset). This study was determined to be nonhuman subjects research by the University of California, Berkeley.

Variables

MTO randomly assigned families into one of three groups: 1) receipt of a Section 8 housing voucher to be used to move to a low-poverty neighborhood and assistance finding housing, 2) receipt of a Section 8 housing voucher without assistance finding housing, and 3) no intervention. Effect estimates comparing the two intervention groups to the control were similar, as has been shown previously (2) and as supported by a partial F test (p-values: 0.61, 0.21, 0.49, and 0.81 for marijuana use, behavioral problems, major depressive disorder, and generalized anxiety disorder, respectively). Therefore, we combined the two intervention groups and defined our instrument as randomization to receive a housing voucher versus not, as has been done previously (8). To allow for differences in intervention "take-up" across sites, we incorporated moving to a 2000 Census tract with less than 25% of persons in poverty as a causal intermediate. This poverty level was chosen as it represented a natural

breakpoint in the distributions of residential poverty at follow-up for each site. The distribution of Census tract poverty levels by site is shown in eFigure 1.

We tested several binary adolescent self-reported mental health and risk behavior outcomes. major depressive disorder and generalized anxiety disorder correspond to *DSM-IV* diagnoses, which have been shown to align with clinical diagnoses (2, 19). Factor scores from an abbreviated Behavioral Problems Index (20) were estimated using item response theory following previous work (8) and then dichotomized at the 90th percentile (21). Marijuana use was defined as any lifetime use.

Because use of marijuana and diagnoses of generalized anxiety disorder and major depressive disorder are lifetime measures, we cannot definitively establish temporality for these outcomes. However, we believe temporality is likely because the ages at baseline—5–16 years old—are younger than the typical ages at onset for major depressive disorder and generalized anxiety disorder (median ages at onset are 31 and 32, respectively (22)) and also before most try marijuana (median age of 16) (23). Moreover, although these outcomes were not measured at baseline, we would expect them to be balanced between treatment groups by virtue of the random assignment.

Baseline covariates included sociodemographic characteristics of the adolescent and family members, behavior and learning characteristics of the adolescent, neighborhood characteristics at baseline, and reasons for participation in MTO. A full list of covariates is provided in the eAppendix.

Statistical Analysis

We estimate the intent-to-treat average treatment effect, which is the average effect of being randomized to receive a housing voucher versus control on each health outcome considered. These measures are risk differences; for example, for major depressive disorder, the intent-to-treat average treatment effect, would be the difference in risk of major depressive disorder at follow-up comparing those randomized to the housing voucher group versus the control group. We restricted our analysis to those observations with nonmissing outcomes. This resulted in a sample size of 1,018–1,019 for generalized anxiety disorder, 1,077–1,078 for marijuana, and 1,094–1,095 for major depressive disorder and behavioral problems. We used multiple imputation by chained equations to impute covariate values (race/ethnicity and ever repeating a grade, missing for <1% and 6%, respectively), making 30 imputed datasets (24).

For each outcome, the analysis proceeded as follows. We estimated site-specific intent-to-treat average treatment effects, and identified sites with qualitatively different estimates (i.e., one site had an increase in risk while another had a decrease) that also demonstrated quantitative differences using a partial F test and an alpha level of 0.15. We aimed to transport the treatment effect estimate from the group of site(s) with the more extreme estimate, S=1, to the site(s) with the less extreme estimate, S=0. (Note that we could transport in either direction; we chose to transport from the site with the more extreme estimate to facilitate the inclusion of treatment in the outcome model when using data-adaptive methods in model fitting.) The transport estimator is premised on the assumption of a shared outcome model across sites. We tested this using a nonparametric omnibus test of

equality of two functions in distribution (25), which uses the same flexible machine learning approach (26, 27) as is used in the transport estimator. Specifically, we test the null hypothesis, H_0, that E(Y | S=0, W, A, Z) = E(Y | S=1, W, A, Z), where Y is the outcome, W is the vector of covariates, A is housing voucher randomization, and Z is intervention take-up of moving to a lower poverty neighborhood. (Note that we would not need to incorporate Z to estimate a site-specific intent-to-treat average treatment effect. Z is incorporated into estimation of the transported treatment effect to account for compositional differences in take-up (9, 10).) If we did not have evidence to reject H_0, then we proceeded with estimating the transported treatment effect. Scenarios that result in the inequality $E(Y|S = 0, W, A, Z)$ $E(Y|S = 1, W, A, Z)$ would include contextual differences or differences in unmeasured individual-level confounders. Using Pearl and Bareinboim's transport notation, this would be depicted by an S node pointing into Y node (9), and could include scenarios such as differences in intervention implementation across sites or differences in other sources of bias in modeling Y across sites, such as measurement error.

We used a semiparametric transport estimator (10) to predict the intent-to-treat average treatment effect for S=0, accounting for differences in population composition and intervention uptake across sites. Practically, this means flexibly modeling several relationships: 1) how S=1 and S=0 differ in the distribution of covariates, 2) the S-specific model of moving to a low-poverty neighborhood (Z) as a function of covariates and randomization, and 3) the model of the outcome as a function of covariates and the take-up in the S=1 group. Using the outcome model for S=1, we predict outcome values for those in the S=0 group, using their covariate and take-up values and take-up model. For example, in the case of marijuana use, we predict the intent-to-treat average treatment effect for Los Angeles (S=0) using covariate, randomization, take-up, and outcome data from Boston (S=1) but no outcome data from Los Angeles. This is similar to re-weighting the intent-to-treat average treatment effect for S=1 using the composition and intervention take-up of S=0. We call the predicted treatment effect for S=0 the "transported intent-to-treat average treatment effect". The transport estimator uses targeted maximum likelihood estimation, which is a multiply robust approach (meaning that it is consistent even if some of the relationships described above are misspecified) compatible with flexible, machine learning approaches for modeling the relationships. The ensemble machine learning algorithm we used allowed for a very flexible model, including complex interactions between covariates, treatment, and take-up and incorporated cross-validation to avoid overfitting. Targeted maximum likelihood estimation differs from other estimation frameworks in that it targets the specific effect of interest and optimizes its estimation in terms of bias and variance (28).

Lastly, we compare the predicted intent-to-treat average treatment effect for S=0 to the observed estimate for S=0. The amount by which the predicted effect reduces the difference between the site-specific treatment effects represents the degree to which the site differences may be due to population composition. The portion unexplained represents the extent to which the site difference may be due to context or unmeasured individual characteristics. We used R version 3.3.1 for all analyses.

RESULTS

Table 1 describes the analytic sample by site, which included black and Hispanic/Latino male adolescents at follow-up. Boston, Los Angeles, and New York City sites had roughly equal proportions of black and Hispanic/Latino male teens, while Chicago had very few Hispanic/Latino participants (1.36%). Ages of the participants at follow-up were similar across sites with a mean of 15 years.

There is substantial variation across the sites at baseline in terms of family sociodemographics, neighborhood perceptions, and school-related experiences. These individual-level compositional differences could be interacting with the MTO intervention to produce different effects across sites.

Rates of moving to a low-poverty neighborhood following randomization were highest for Chicago (32%), followed by Los Angeles (30%), Boston (24%), and New York City (18%). At follow-up, teens in Chicago reported slightly higher rates of marijuana use (30%) and lower rates of behavioral problems (7%). Those in Boston reported the highest rates of behavioral problems (18%). The proportion of teens with generalized anxiety disorder was lowest in New York City (3%). The rates of major depressive disorder were similar across the sites.

Each outcome showed evidence of qualitative and quantitative site differences. The site-specific estimates and 95% CIs are shown in Figure 1. Receipt of a housing voucher increased marijuana use among boys in Boston but slightly decreased use in New York City (p-value for site difference = 0.03). Housing voucher receipt decreased behavioral problems among boys in Los Angeles but increased problems in New York City (p-value for site difference = 0.12). Housing voucher receipt increased risk of major depressive disorder among boys in New York City but decreased risk in Chicago (p-value for site difference = 0.10). Finally, voucher receipt increased risk of generalized anxiety disorder among boys in Boston and New York City but decreased risk in Los Angeles (p-value for site difference = 0.04).

We tested whether we had evidence of a common outcome model across sites, an assumption required to identify the intent-to-treat average treatment effect for S=0 (10). For marijuana use, behavioral problems, and major depressive disorder we found no evidence against the shared outcome model assumption (p-values: 0.72, 0.44, and 0.74, respectively). However, we rejected the shared outcome model assumption for generalized anxiety disorder (p-value = 0.04). This indicates that the intent-to-treat average treatment effect for generalized anxiety disorder could not be transported based on measured individual-level covariates, but we proceeded with transporting the treatment effect estimates for the remaining outcomes.

Figure 1 shows the estimates for the effect of being randomized to the housing voucher group on risk of marijuana use (panel A), behavioral problems (panel B), major depressive disorder (panel C), and generalized anxiety disorder (panel D) among black and Hispanic/Latino male youth. Comparing the transported estimates (dashed lines) to observed estimates (solid lines) for site S=0 allows us to assess the extent to which site differences

could be attributed to differences in the distribution of compositional characteristics between the sites. If the site effects are transportable using measured individual-level characteristics, then the transported effect for S=0 will coincide with the observed effect for S=0. This would suggest that differences in the site effects are entirely due to differences in compositional factors between the sites. If, on the other hand, the transported effect is no closer to the observed estimate for S=0 than the observed estimate for S=1, then it suggests that the MTO's effect on that outcome is not transportable based on these characteristics. If the transported estimate is between the two observed estimates, it suggests a partially transportable effect—that the site differences in effects are partially but not entirely attributable to measured compositional differences.

Figure 1 panel A shows that the effect of housing voucher receipt on risk of using marijuana is partially transportable between Boston and Los Angeles using measured individual-level characteristics. The transported effect estimate for Los Angeles is 0.04 (95% CI: −0.11, 0.19). So, if Boston, counter to fact, had the population distribution of covariates shown in Los Angeles, we would predict that being randomized to the housing voucher group would increase risk of marijuana use at follow-up by 0.04. The transported estimate is closer to the observed Los Angeles estimate (−0.05, 95% CI: −0.19, 0.09) than is the observed Boston estimate (0.15, 95% CI: 0.04, 0.25); differences in measured compositional factors explained 52% of the difference between sites. The t-statistic of the difference between the observed and transported Los Angeles estimates is −0.92 (95% CI: −2.89, 1.04).

There is also evidence for partial transportability in the case of behavioral problems (Figure 1, panel B). The transported effect estimate for Los Angeles (0.00, 95% CI: −0.11, 0.19) is closer to the observed Los Angeles estimate (−0.04, 95% CI: −0.14, 0.04) than is the observed New York City estimate (0.05, 95% CI: −0.00, 0.11), explaining 57% of the site difference. The t-statistic of the difference between the observed and transported Los Angeles estimates is −1.02 (95% CI: −2.98, 0.95).

In contrast, the wide CI for the transported major depressive disorder estimate in Figure 1 panel C precludes any clear conclusion about transportability. The the transported point estimate for Chicago (0.02, 95% CI: −0.11, 0.15) is essentially no closer to the observed estimate (−0.02, 95% CI: −0.07, 0.03) than was the observed New York City estimate (0.02, 95% CI: 0.01, 0.04)— the transported estimate only explained 9% of the site difference. The t-statistic of the difference between the observed and transported Chicago estimates is −0.62 (95% CI: −2.59, 1.34).

As previously noted, we found evidence against the common outcome model assumption for generalized anxiety disorder, which precluded estimating the transported effect (panel D). This is quantitative evidence suggesting that differences in measured aspects of population composition across sites did not contribute to the differences in site effects for generalized anxiety disorder, and potentially also for major depressive disorder; instead, differences in context or in unmeasured individual-level compositional factors may play an important role.

DISCUSSION

We found evidence of site differences in MTO's effects on marijuana use, behavioral problems, major depressive disorder, and generalized anxiety disorder among black and Hispanic/Latino adolescent boys. These site differences were qualitative—receiving a housing voucher appeared to be harmful for the health of adolescent males in some sites but beneficial in other sites. Since population composition also varied by site and such compositional factors could act as effect modifiers, we assessed the extent to which the site differences could be explained by differences in population composition. Differences in composition partially explained site differences in marijuana use and behavioral problems but did not appear to explain site differences for major depressive disorder and generalized anxiety disorder. Thus, the effects of housing vouchers on mental health and risk behaviors in adolescent males do not appear to be fully transportable across sites even after flexibly accounting for numerous baseline characteristics, thereby providing quantitative evidence for the importance of context (or unmeasured individual-level compositional factors) in modifying MTO's effects.

To our knowledge, this is the first time that quantitative evidence has been brought to bear in the "composition vs. context" debate in settings where there are too few sites to use multi-level methods to estimate contextual effects (see (29) and (30) for an introduction to this debate). When there was evidence of differences in effects between sites, we used a transport estimator (10) to predict the effect of the MTO intervention in one of the sites based on differences in individual-level characteristics between sites and the outcome model from the other site. The transport estimator is flexible, data-adaptive, and multiply robust; it can simultaneously account for numerous compositional factors that may modify the treatment effect without concerns about cherry-picking or misspecifying possibly complex relationships in parametric models. In addition, it results in accurate inference when incorporating machine learning algorithms, which is a challenge of other estimation strategies.

The transported point estimates did not equal the observed point estimates in the target site for any of the outcomes, suggesting unmeasured factors at the contextual- or individual-level partially contributed to site differences. However, with only five cities (four of which were used for this analysis), we do not have the sample size necessary at the site level to empirically identify which contextual-level variables contribute to these site differences. Such identification would require a larger number of sites as part of a multi-level experimental design (31).

A limitation of our analysis is small sample size in terms of the number of adolescents. Given the high-dimensional vector of baseline characteristics coupled with the small number of cases for the mental health outcomes, a larger sample size of adolescents would likely improve the precision of our transport estimates. In addition, small sample size coupled with a rare outcome could also negatively affect estimator performance in terms of bias and coverage (32).

Another limitation further compounded by the small sample size limitation is the presence of practical violations of the positivity assumption (32). In this application, practical positivity violations mean that there are certain combinations of covariate values that nearly determine site membership and/or intervention take-up. This is a problem, because our estimator must then rely on extrapolation over areas where there are little or no data. Positivity violations can adversely affect estimator performance in terms of bias, variance, and confidence interval coverage (32–34). Although the transport estimator we use here is not very sensitive to practical violations of the positivity assumption, it nonetheless demonstrated slight increases in bias and variance and loss of confidence interval coverage under such violations (10). Coupled with small sample size, practical positivity violations could pose an additional problem if the site that we are transporting *to* has participants with combinations of baseline characteristics not observed in the site that we are transporting *from*. If those baseline characteristics are also important for modifying the effect of moving to a low poverty neighborhood on the outcome, then the transported estimate may be biased because of its inability to account for those modifiers.

Lastly, MTO's effects on marijuana use and behavioral problems were partially transportable, and it might be of interest to know which aspects of population composition were most important in explaining these site differences. Unfortunately, no appealing method exists for identifying the most relevant modifiers. It is an area of future work to develop a variable importance algorithm for the machine learning approach used in our analysis, similar to other variable importance metrics (35).

In summary, we found that a large number of baseline characteristics—including individual and family sociodemographics, experiences at school and in the neighborhood, and motivations to move—partially explained site differences in MTO effects for risk behavior outcomes but could not explain site differences in mental health outcomes among adolescent boys. This suggests that context or unmeasured compositional variables are important in modifying MTO's effects. For example, qualitative evidence suggests that city-level differences in aspects of the economy, housing market, racial mixing, and school choice policies may have modified MTO's effectiveness (3–5). Thus, even if great care is taken to implement an intervention the same way across cities (2) and all relevant individual-level variables are accounted for, an intervention may still not be transportable due to differences in context. Given that social experiments like MTO are embedded within complex social settings, the potential for nontransportable effects due to differences in context may not be surprising, and is aligned with sociological research demonstrating the importance of considering multiple levels of contextual influence (36–39).

Our findings have two implications that are marked departures from the current understanding of MTO results. First, our finding suggests that the current practice of analyzing MTO data by pooling across sites and including a dummy variable for site in a regression equation may be inappropriate. Ideally, one would first check for site differences in treatment effects and, if site differences are found, whether they could be explained by measured baseline covariates. If not, as in our examples, then a site-specific analysis should be conducted (e.g., using interaction terms between site and treatment) or an analysis that only pools sites with similar effect estimates. The second, more general implication is that

interventions should be considered in the context of where they are implemented and developed or modified with potentially relevant macro-level factors in mind. Previous MTO research demonstrated the importance of the neighborhood context on health outcomes; this study suggests widening the contextual lens to encompass city-level factors.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Shroder MD, Orr LL. Moving to opportunity: Why, how, and what next? Cityscape. 2012:31–56.

2. Orr, L., Feins, J., Jacob, R., et al. Moving to opportunity: Interim impacts evaluation. 2003.

3. de Souza Briggs, X., Popkin, SJ., Goering, J. Moving to opportunity: The story of an American experiment to fight ghetto poverty. Oxford University Press; 2010.

4. de Souza Briggs X, Ferryman KS, Popkin SJ, et al. Why did the moving to opportunity experiment not get young people into better schools? Housing Policy Debate. 2008; 19(1):53–91.

5. Briggs, XdS, Comey, J., Weismann, G. Struggling to stay out of high-poverty neighborhoods: housing choice and locations in moving to opportunity's first decade. Housing Policy Debate. 2010; 20(3):383–427.

6. Kling JR, Liebman JB, Katz LF. Experimental analysis of neighborhood effects. Econometrica. 2007; 75(1):83–119.

7. Kling JR, Ludwig J, Katz LF. Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment. The Quarterly Journal of Economics. 2005:87–130.

8. Osypuk TL, Schmidt NM, Bates LM, et al. Gender and crime victimization modify neighborhood effects on adolescent mental health. Pediatrics. 2012; 130(3):472–481. [PubMed: 22908105]

9. Pearl, J., Bareinboim, E. Transportability of causal and statistical relations: A formal approach. Twenty-Fifth AAAI Conference on Artificial Intelligence; 2011;

10. Rudolph KE, Laan MJ. Robust estimation of encouragement design intervention effects transported across sites. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2016; Epub ahead of print. doi: 10.1111/rssb.12213

11. Stuart EA, Cole SR, Bradshaw CP, et al. The use of propensity scores to assess the generalizability of results from randomized trials. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2011; 174(2):369–386.

12. Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations the actg 320 trial. American journal of epidemiology. 2010; 172(1):107–115. [PubMed: 20547574]

13. Frangakis C. The calibration of treatment effects from clinical trials to target populations. Clinical trials (London, England). 2009; 6(2):136.

14. Ogburn EL, Rotnitzky A, Robins JM. Doubly robust estimation of the local average treatment effect curve. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2015; 77(2):373–396. [PubMed: 25663814]

15. Westreich D, Edwards JK. Invited commentary: every good randomization deserves observation. American journal of epidemiology. 2015; 182(10):857–860. [PubMed: 26487342]

16. Sanbonmatsu L, Ludwig J, Katz LF, et al. Moving to opportunity for fair housing demonstration program–final impacts evaluation. 2011

17. Leventhal T, Dupéré V. Moving to opportunity: Does long-term exposure to 'low-poverty' neighborhoods make a difference for adolescents? Social Science & Medicine. 2011; 73(5):737–743. [PubMed: 21821323]

18. Osypuk TL, Tchetgen EJT, Acevedo-Garcia D, et al. Differential mental health effects of neighborhood relocation among youth in vulnerable families: results from a randomized trial. Archives of general psychiatry. 2012; 69(12):1284–1294. [PubMed: 23045214]

19. Kessler RC, Avenevoli S, Green J, et al. National comorbidity survey replication adolescent supplement (ncs-a): Iii. concordance of dsm-iv/cidi diagnoses with clinical reassessments. Journal of the American Academy of Child & Adolescent Psychiatry. 2009; 48(4):386–399. [PubMed: 19252450]

20. Zill, N. Behavior problems index based on parent report (publication no. 9103). Washington, DC: Child Trends; 1990.

21. McDermott S, Coker AL, Mani S, et al. A population-based analysis of behavior problems in children with cerebral palsy. Journal of pediatric psychology. 1996; 21(3):447–463. [PubMed: 8935244]

22. Kessler RC, Berglund P, Demler O, et al. Lifetime prevalence and age-of-onset distributions of dsm-iv disorders in the national comorbidity survey replication. Archives of general psychiatry. 2005; 62(6):593–602. [PubMed: 15939837]

23. Swendsen J, Anthony JC, Conway KP, et al. Improving targets for the prevention of drug use disorders: sociodemographic predictors of transitions across drug use stages in the national comorbidity survey replication. Preventive medicine. 2008; 47(6):629–634. [PubMed: 18926848]

24. Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in r. Journal of statistical software. 2011; 45(3)

25. Luedtke AR, Carone M, van der Laan MJ. An omnibus nonparametric test of equality in distribution for unknown functions. 2015 arXiv preprint arXiv:151004195.

26. van der Laan MJ, Polley EC, Hubbard AE. Super learner. Statistical applications in genetics and molecular biology. 2007; 6(1)

27. Sapp S, van der Laan MJ, Canny J. Subsemble: an ensemble method for combining subset-specific algorithm fits. Journal of applied statistics. 2014; 41(6):1247–1259. [PubMed: 24778462]

28. van der Laan MJ, Rubin D. Targeted maximum likelihood learning. The International Journal of Biostatistics. 2006; 2(1)

29. Macintyre S, Ellaway A, Cummins S. Place effects on health: how can we conceptualise, operationalise and measure them? Social science & medicine. 2002; 55(1):125–139. [PubMed: 12137182]

30. Oakes JM. The (mis) estimation of neighborhood effects: causal inference for a practicable social epidemiology. Social science & medicine. 2004; 58(10):1929–1952. [PubMed: 15020009]

31. Pals SL, Murray DM, Alfano CM, et al. Individually randomized group treatment trials: a critical appraisal of frequently used design and analytic approaches. American Journal of Public Health. 2008; 98(8):1418–1424. [PubMed: 18556603]

32. Petersen ML, Porter KE, Gruber S, et al. Diagnosing and responding to violations in the positivity assumption. Statistical methods in medical research. 2010; 21(1):31–54. [PubMed: 21030422]

33. Rudolph KE, Díaz I, Rosenblum M, et al. Estimating population treatment effects from a survey subsample. American journal of epidemiology. 2014; 180(7):737–748. [PubMed: 25190679]

34. Robins J, Sued M, Lei-Gomez Q, et al. Comment: Performance of double-robust estimators when "inverse probability" weights are highly variable. Statistical Science. 2007:544–559.

35. Breiman L. Random forests. Machine learning. 2001; 45(1):5–32.

36. Graif C. Delinquency and gender moderation in the moving to opportunity intervention: The role of extended neighborhoods. Criminology. 2015; 53(3):366–398.

37. Lee BA, Reardon SF, Firebaugh G, et al. Beyond the census tract: Patterns and determinants of racial segregation at multiple geographic scales. American Sociological Review. 2008; 73(5):766–791. [PubMed: 25324575]

38. Massey DS, Fischer MJ, Dickens WT, et al. The geography of inequality in the united states, 1950–2000 [with comments]. Brookings-Wharton papers on urban affairs. 2003:1–40.

39. Ellen IG, Lens MC, O'Regan K. American murder mystery revisited: do housing voucher households cause crime? Housing Policy Debate. 2012; 22(4):551–572.
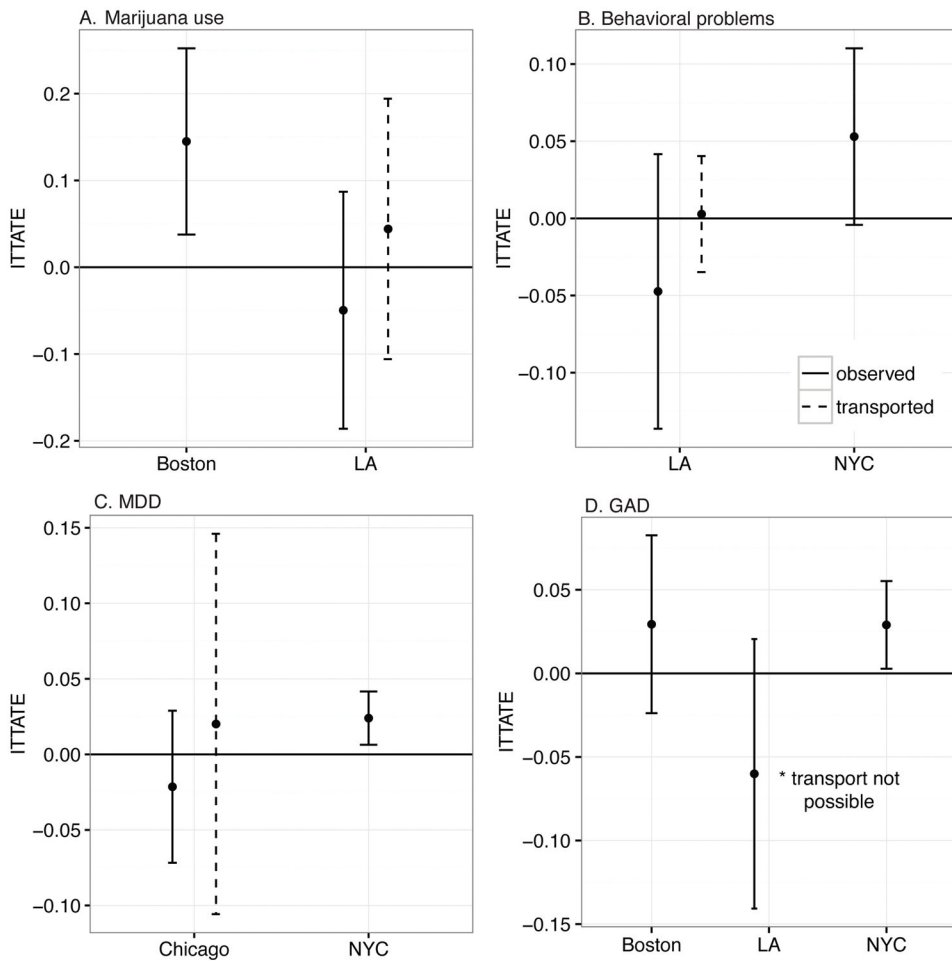
Figure 1.
Estimated intent-to-treat average treatment effects (ITTATE) and 95% CIs by site. Panel A shows the effect of receiving a housing voucher on any marijuana use among males. The ITTATE was transported from Boston to Los Angeles (LA), so the transported estimate for LA should be compared to the observed estimate. Panel B shows the effect of receiving a housing voucher on behavioral problems among males. The ITTATE was transported from New York City (NYC) to LA. Panel C shows the effect of receiving a housing voucher on risk for major depressive disorder among males. The ITTATE was transported from NYC to Chicago. Panel D shows the effect of receiving a housing voucher on risk for generalized anxiety disorder among males. Transport was not possible because the assumption of a common outcome model was not met.

Table 1

Survey-weighted characteristics of black and Hispanic/Latino male adolescents by Moving to Opportunity site combined across 30 imputed datasets (outcomes were not imputed). Numbers are percentages unless otherwise specified.[a]

Characteristic	Boston N=228	Chicago N=289–290[b]	LA N=259	NYC N=318
Adolescent characteristics at baseline				
Black	38	99	42	49
Hispanic/Latino	62	1	58	51
School help for behavioral or emotional problems	12	11	9	14
Suspended or expelled from school	10	18	17	9
In a gifted program in school	14	16	15	22
Received help for learning problems	22	23	22	26
School problems with work or behavior	33	36	30	41
Problems that prevented going to school or active play	8	7	7	9
Ever repeated a grade	35	24	17	34
Characteristics of parent/head of household at baseline				
High school graduate	39	39	31	37
Never married	54	70	48	45
Teen parent when had first child	20	38	27	14
Working	33	29	25	22
Welfare receipt	60	82	76	78
Family has car	22	15	48	7
Family member with disability	17	14	10	28
Household size				
2	7	6	3	8
3	28	15	16	27
4	25	20	22	26
5	40	59	59	39
Member of household was the victim of a crime, past 6 months	27	42	51	50
Lived in baseline neighborhood 5 years	50	72	72	73
Stopped to chat with neighbor 1x/week	50	53	53	51
Would tell neighbor if neighbor's child was in trouble	72	56	53	53
No family in neighborhood	74	50	57	71
No friends in neighborhood	39	33	24	48
Neighborhood streets unsafe at night	35	43	60	55
Very dissatisfied with neighborhood	26	49	39	58
Take-up and outcomes at follow-up				
Moved to a low poverty neighborhood	24	32	30	18
Any marijuana use (% missing)	21 (1)	30 (1)	25 (1)	21 (3)
Behavioral problems	18	7	9	7

Characteristic	Boston N=228	Chicago N=289–290[b]	LA N=259	NYC N=318
GAD[b] (% missing)	5 (15)	4 (3)	6 (4)	3 (7)
MDD[c]	1	2	2	2

[a] LA indicates Los Angeles, NYC: New York City, GAD: generalized anxiety disorder, MDD: major depressive disorder.

[b] The number of participants per site differed by one in the Chicago site across the imputed datasets, because we excluded non black or Hispanic/Latino participants and race/ethnicity was an imputed variable.