

WORKING PAPER · NO. 2021-22

Evaluating Contradictory Experimental and Non-Experimental Estimates of Neighborhood Effects on Economic Outcomes for Adults

David J. Harding, Lisa Sanbonmatsu, Greg J. Duncan, Lisa A. Gennetian, Lawrence F. Katz,

Ronald C. Kessler, Jeffrey R. Kling, Matthew Sciandra, and Jens Ludwig

FEBRUARY 2021

EVALUATING CONTRADICTIONARY EXPERIMENTAL AND NON-EXPERIMENTAL
ESTIMATES OF NEIGHBORHOOD EFFECTS ON ECONOMIC OUTCOMES FOR ADULTS

David J. Harding
Lisa Sanbonmatsu
Greg J. Duncan
Lisa A. Gennetian
Lawrence F. Katz
Ronald C. Kessler
Jeffrey R. Kling
Matthew Sciandra
Jens Ludwig

February 2021

Support for this research was provided by a contract from the U.S. Department of Housing and Urban Development (HUD; C-CHI-00808) and grants from the National Science Foundation (SES-0527615), National Institute of Mental Health (R01-MH077026), National Institute of Aging (P30-AG012810, R01-AG031259, and P01-AG005842-22S1), Smith Richardson Foundation (SRF-20161249), University of Chicago's Center for Health Administration Studies, Russell Sage Foundation, and Robert Wood Johnson Foundation. The collection of the PSID data used in this study was partly supported by the National Institutes of Health under grant number R01 HD069609 and the National Science Foundation under award number 1157698. Outstanding assistance with the data preparation and analysis was provided by Joe Amick, Ryan Gillette, Ray Yun Gou, Ijun Lai, Jordan Marvakov, Nicholas Potter, Nathan Weil, Fanghua Yang, Sabrina Yusuf, and Michael Zabek. The survey data collection effort was led by Nancy Gebler of the University of Michigan's Survey Research Center under subcontract to our research team. The views expressed in this work are those of the authors and should not be interpreted as those of the Congressional Budget Office or HUD. A restricted-access version of the MTO data used in this paper will be provided to the U.S. Department of Housing and Urban Development (HUD). At the time of writing, information about access to the MTO data could be found here: <http://www.nber.org/mtopublic/>. Researchers may apply for access to PSID restricted data on census tracts through the University of Michigan (see <http://simba.isr.umich.edu/restricted/RestrictedUse.aspx> for more information). In the past 3 years, Dr. Kessler was a consultant for Datastat, Inc., RallyPoint Networks, Inc., Sage Pharmaceuticals, and Takeda.

© 2021 by David J. Harding, Lisa Sanbonmatsu, Greg J. Duncan, Lisa A. Gennetian, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling, Matthew Sciandra, and Jens Ludwig. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Evaluating Contradictory Experimental and Non-Experimental Estimates of Neighborhood Effects on Economic Outcomes for Adults

David J. Harding, Lisa Sanbonmatsu, Greg J. Duncan, Lisa A. Gennetian, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling, Matthew Sciandra, and Jens Ludwig

February 2021

JEL No. H0,I3,J0

ABSTRACT

Although non-experimental studies find robust neighborhood effects on adults, such findings have been challenged by results from the Moving to Opportunity (MTO) residential mobility experiment. Using a within-study comparison design, this paper compares experimental and non-experimental estimates from MTO and a parallel analysis of the Panel Study of Income Dynamics (PSID). Striking similarities were found between non-experimental estimates based on MTO and PSID. No clear evidence was found that different estimates are related to duration of adult exposure to disadvantaged neighborhoods, non-linear effects of neighborhood conditions, magnitude of the change in neighborhood context, frequency of moves, treatment effect heterogeneity, or measurement, although uncertainty bands around our estimates were sometimes large. One other possibility is that MTO-induced moves might have been unusually disruptive, but results are inconsistent for that hypothesis. Taken together, the findings suggest that selection bias might account for evidence of neighborhood effects on adult economic outcomes in non-experimental studies.

David J. Harding
Department of Sociology
462 Social Science Building
University of California, Berkeley
Berkeley, CA 94720
dharding@berkeley.edu

Lisa Sanbonmatsu
Center for Education Policy Research at
Harvard University
50 Church Street
4th Floor
Cambridge, MA 02138
lsanbonm@nber.org

Greg J. Duncan
University of California, Irvine
School of Education
2056 Education Building, Mail Code 5500
Irvine, CA 92697
gduncan@uci.edu

Lisa A. Gennetian
Sanford School of Public Policy
Duke University
212 Rubenstein Hall 302 Towerview Road
Durham, NC 27708
United States
lisa.gennetian@duke.edu

Lawrence F. Katz
Department of Economics
Harvard University
Cambridge, MA 02138
and NBER
lkatz@harvard.edu

Ronald C. Kessler
Harvard Medical School
Department of Health Care Policy
180 Longwood Avenue
Boston, MA 02115
kessler@hcp.med.harvard.edu

Jeffrey R. Kling
Congressional Budget Office
3403 Ordway St NW
Washington, DC 20016
and NBER
jeffrey@kling.us

Matthew Sciandra
630 Irving Street, NW
Washington, DC 20010
mattsciandra@gmail.com

Jens Ludwig
Harris School of Public Policy
University of Chicago
1307 East 60th Street
Chicago, IL 60637
and NBER
jludwig@uchicago.edu

At the time of writing, information about access to the MTO data could be found here: <http://www.nber.org/mtopublic/>

Researchers may apply for access to PSID restricted data on census tracts through the University of Michigan:
<http://simba.isr.umich.edu/restricted/RestrictedUse.aspx>

INTRODUCTION

A prolific literature has generated many insights about the nature and magnitude of the effects of living in disadvantaged neighborhoods. That literature has examined outcomes in many domains—including the labor market, health, crime, fertility, and education—not only in the U.S. but throughout the world.¹ Although an emerging consensus suggests that neighborhoods affect children’s outcomes in multiple domains and also affect adults’ mental and physical health, a persistent unresolved question concerns the causal status of the association between neighborhood characteristics and the economic outcomes of adults. A large number of observational studies find robust associations between exposure to disadvantaged neighborhoods during adulthood and adults’ economic outcomes that are interpreted as causal effects after conditioning on individual and family characteristics. But these results have been called into question by contradictory results from the Moving to Opportunity (MTO) residential mobility experiment, which was able to evaluate the effects of random variation in neighborhood contexts created by random assignment of housing vouchers to public housing families in five U.S. cities in the late 1990s who were followed over the next ten to fifteen years (Orr et al. 2003; Sanbonmatsu et al. 2011).

Analyses of MTO data find persistent statistically significant differences in outcomes by neighborhood poverty in adult mental and physical health (Kessler et al. 2014; Ludwig et al. 2011, 2012, 2013; Sanbonmatsu et al. 2011) and find that children who moved through MTO when they are young (prior to adolescence) experienced long-term economic benefits including

¹ While this literature is vast, important reviews or original contributions have been made in numerous disciplines such as sociology (for example, Sampson 2012; Sharkey and Faber 2014), psychology (Leventhal and Brooks-Gunn 2000), economics (Chetty and Hendren 2018a,b), public health (Kawachi and Berkman 2003) and housing (Galster 2017), as well as key reports by the National Academy of Sciences such as Lynn and McGeary (1990) and Shonkoff and Phillips (2000).

higher adult earnings and household incomes (Chetty, Hendren and Katz 2016). But MTO did not find detectable effects on the economic self-sufficiency of those who moved as adults (Ludwig et al. 2012; Orr et al. 2003; Sanbonmatsu et al. 2011). Why do neighborhood effects on economic outcomes for adults differ so markedly from the results in non-experimental studies? One possibility is that the difference is due to selection bias or other differences in study designs. This is an especially challenging problem in non-experimental studies because adults' employment and earnings are directly related to the process of selection into disadvantaged neighborhoods. And as these unobserved variables also are determinants of future labor market outcomes, the causal effects of neighborhood disadvantage experienced in adulthood on adult employment and earnings may be difficult to detect because of these unobserved common causes in non-experimental studies.

In addition to selection bias, several other possible explanations exist for the seemingly contradictory results between the results from the MTO experiment and the results of non-experimental studies of neighborhood effects on the economic outcomes of adults. These include reliance on different outcome measures, neighborhood effect heterogeneity, differential duration of exposure to disadvantaged neighborhoods, non-linear neighborhood effects, the magnitude of changes in neighborhood context, measures of neighborhood context, and residential moves. However, we are unaware of any empirical studies that attempt to adjudicate between these explanations using both experimental and observational data.

The puzzle of contradictory results is important because neighborhood exposure effects are central to theories about the role of neighborhood context in economic and racial inequalities (Massey and Denton 1993; Wilson 1987). The puzzle is also important for public policy in light of the growth in U.S. residential segregation by income that has been occurring since 1970

(Jargowsky 1997, 2015; Kneebone, Nadeau, and Berube 2011; Reardon and Bischoff 2014; Reardon et al. 2018; Watson 2009). Uncertainty about the nature of neighborhood effects on low-income adults is an obstacle to the design of public policies targeting poverty and inequality. Income poverty is in many ways the central focus of social policy discussions in the U.S., and whether adult economic outcomes are affected by neighborhood environments is central to consideration of whether place-based policies—be they residential mobility or direct efforts to improve neighborhoods themselves—have the potential to reduce poverty.

Drawing inspiration from a long tradition of “within study comparison” research designs (e.g. Cook, Shadish, and Wong 2008; LaLonde 1986; Michalopoulos, Bloom, and Hill 2004), this paper sheds new light on the contradictions between MTO and non-experimental studies of neighborhood effects on the economic outcomes of adults through a non-experimental reanalysis of the MTO data compared with a parallel non-experimental analysis of data from the Panel Study of Income Dynamics (PSID). We choose the PSID because it is one of the most popular sources of data for observational studies of neighborhood effects. Our use of multiple samples, including a separate non-experimental analysis of the MTO sample, provides us with analytic leverage to test various hypotheses and propose more general explanations about the conditions under which non-experimental and experimental results in this domain are most likely to differ.

As detailed below, our results are inconsistent with a number of commonly offered hypotheses about neighborhood effect differences arising from MTO and PSID data. We find no clear evidence that differences in neighborhood effects on adult economic outcomes are due to the duration of exposure to different neighborhood contexts, non-linear effects, or to the magnitude of the change in neighborhood context produced by the MTO intervention. We also find little support for the idea that discrepant findings between MTO and observational studies

are due to something about the MTO sample itself, since we can apply non-experimental methods to the MTO dataset (ignoring random assignment and instead focusing on within-randomized-group variation). This is a particularly useful comparison given the difficulty of drawing comparably disadvantaged families to MTO from within the PSID sample. We find the MTO non-experimental analysis typically yields estimates that are fairly similar to those from the PSID. Nor do we find strong evidence that the measurement of neighborhood context or labor market outcomes in MTO and observational studies can reconcile prior divergent estimates.

Our results narrow the set of possible explanations for study differences in the effects of adult neighborhood conditions on economic outcomes down to just two. One hypothesis, derived from qualitative research on housing mobility programs, is that the program timing and location constraints placed on MTO-based moves disrupt the lives of MTO families more than families observed to move in observational studies like the PSID. This hypothesis receives mixed support in further analyses reported below (i.e., it is consistent with some of the analyses we carry out, but not with others). The second hypothesis is essentially a residual explanation and cannot be tested directly with the data: that selection bias accounts for the differences in adult economic outcomes between MTO and observational studies.

We begin with a brief review of the nonexperimental literature on the effects of exposure to neighborhood disadvantage in adulthood on adult labor market outcomes. We then describe the key features of the MTO study design and its impacts on adult outcomes, highlighting similarities and differences between MTO findings and results from observational studies of neighborhood effects. Next we discuss the main hypotheses regarding the sources of these similarities and differences and how we test these hypotheses. After describing the MTO and

PSID data and the subsamples we employ, we present our tests of the hypotheses. In the conclusion, we discuss remaining possible explanations for the discrepancy between MTO and PSID estimates and the implications of our findings for future studies.

NON-EXPERIMENTAL STUDIES OF NEIGHBORHOOD EFFECTS ON ECONOMIC OUTCOMES AMONG ADULTS

The empirical study of “neighborhood effects” dates back to the 17th century (MacIntyre and Ellaway 2003). Hundreds of studies in sociology, epidemiology and other fields have regressed measures of individual outcomes against the attributes of the neighborhood in which the individual resides, adjusting for individual- or family-level background characteristics.² We focus our review here on evidence regarding the effects of adult neighborhood conditions on adult economic outcomes, because it is for adults that prior findings appear to be most divergent between MTO and observational studies.³ It is also the case that adult economic outcomes are a (if not *the*) dominant focus of social policy. For recent reviews of the effects of *childhood* exposure to disadvantaged neighborhoods on outcomes during childhood or adulthood, see Galster and Hedman (2013), Galster and Sharkey (2017), and Sharkey and Faber (2014).

Empirical investigations of neighborhood conditions and labor market outcomes during adulthood are typically motivated by one of three theories. The first is spatial mismatch, in which

² For example see Delgadillo, Coster and Erickson (2006); Diez-Roux (2001); Ellaway, Anderson, and Macintyre (1997); Ellen and Turner (2003); Goldsmith, Holzer, and Manderscheid (1998); Holzer (1991); Kawachi and Berkman (2003); Leventhal and Brooks-Gunn (2000); O’Regan and Quigley (1996); Ross (2000); Ross, Reynolds, and Geis (2000); Ross and Mirowsky (2001); Sampson, Morenoff, and Gannon-Rowley (2002); Silver, Mulvey, and Swanson (2002); Vermilyea and Wilcox (2002); Waitzman and Smith (1998); and Yen and Kaplan (1999a,b).

³ Given our focus on the urban poor in the United States, we do not review literature on neighborhood effects in other countries, nor do we cover literature on the effects of ethnic enclaves or neighborhood social networks on the labor market outcomes of recent immigrants. Galster and Hedman (2013) and Galster and Sharkey (2017) include recent reviews of these issues.

disadvantaged neighborhoods (particularly those in the central city) are thought to be located far from available job opportunities (Kain 1968; Wilson 1987, 1996). The second is the inability to use social networks for job search (Granovetter 1973). Residents of high poverty neighborhoods are socially isolated from mainstream institutions and so are less able to help one another secure employment (Wilson 1987, 1996). An alternative network-based explanation is that social distrust within high-poverty neighborhoods may make residents less willing to help each other find jobs (Smith 2007). The third theory is about social norms around work (Wilson 1987). In neighborhoods with high unemployment rates, formal work may become less normative and other strategies of income generation more socially acceptable. This last explanation is the most controversial (Anderson 1999; Harding 2010). Most empirical research on neighborhood effects on adult outcomes has either attempted to estimate an overall effect of disadvantaged neighborhoods on adult employment, or has focused specifically on testing the spatial mismatch or, more rarely, job-network theories.

Such observational studies have typically attempted to account for the selection of people or families into neighborhoods by including controls for individual and family characteristics or, if the data support it, by using individual or family fixed effects.⁴ These studies often find that more advantaged neighborhoods or lower levels of residential segregation are associated with greater likelihood of employment, higher earnings, and a decreased probability of welfare receipt (Casciano and Massey 2008; Collins and Margo 2000; Cutler and Glaeser 1997; Dawkins, Shen, and Sanchez 2005; Elliott 1999; Fauth, Leventhal and Brooks-Gunn 2004; Hoynes 2000; Parks 2004; Reingold, Van Ryzin, and Ronda 2001; Shang 2014; Vartanian 1997; Weinberg, Reagan,

⁴ That is, allowing for separate intercept terms for each individual or family in the study sample in order to control for time-invariant confounding factors.

and Yankow 2004). Proximity to jobs is also associated with such outcomes (Allard and Danziger 2003; Howell-Moroney 2005; Thompson 1997; Weinberg, Reagan, and Yankow 2004). Greater neighborhood job referral opportunities are associated with greater employment, work hours, and earnings (Bayer, Ross, and Topa 2008). However, Bania, Coulton, and Leete (2003) find that neither neighborhood poverty nor job proximity is associated with employment and earnings among welfare recipients in one Ohio county, and Gurmu, Ihlanfeldt, and Smith (2008) find that geographic location is not associated with employment among Atlanta welfare recipients. Lens and Gabbe (2017) using the MTO data (discussed further below) find no relationship between neighborhood employment access and economic outcomes. Although most studies find links between adult exposure to disadvantaged neighborhoods and adult labor market outcomes, their research designs typically rely on strong assumptions about selection on observables (regression with controls) or selection on time-invariant characteristics (fixed effects), which make it difficult to dismiss selection bias as an alternative explanation.

Another set of studies has used quasi-experimental methods to address selection. A particularly influential study in this literature given its research design was Rosenbaum and Popkin's (1991) quasi-experimental analysis of a survey of female household heads in Chicago's Gautreaux Program (see also Rosenbaum [1995] and Rubinowitz and Rosenbaum [2000]). The Gautreaux intervention stemmed from litigation against the Chicago Housing Authority (CHA), which resulted in an agreement by the CHA to relocate African-American residents of public housing projects to new apartments located in either other parts of the city of Chicago or in low-poverty, predominantly white areas of the Chicago suburbs. Rosenbaum and Popkin found substantially higher employment rates (14 percentage points) for those who moved to the suburbs than for those who moved within the city.

The longer-term and more comprehensive follow-up analysis of Gautreaux Program impacts on low-income, black female household heads by Mendenhall, DeLuca, and Duncan (2006) finds little systematic impact on employment of moves to suburbs vs. the city, which was the initial contrast examined by Rosenbaum and Popkin. Their study also finds evidence that city versus suburban placement in Gautreaux was systematically related to the baseline attributes of the households, suggesting that their “natural experiment” study did not quite mimic the idealized design of a randomized experiment (see also Deluca et al. 2010; Votruba and Kling 2009). The long-term follow-up did find suggestive evidence of persistent positive employment effects of being initially placed in less-segregated (lower percent black) neighborhoods with greater neighborhood resources. Although other natural experiment strategies have been used to study the effects of childhood exposure to neighborhood disadvantage, such as the demolition of public housing (e.g. Jacob 2004; Galster and Santiago 2017), the work by Chyn (2018) is the only study we are aware of that applied that research design to examine the effects of adult exposure to neighborhood disadvantage on labor market outcomes. Chyn’s findings for adults are quite consistent with the experimental findings from MTO of large improvements in neighborhood quality but small and insignificant impacts on the employment and earnings of the displaced adults.

EXPERIMENTAL ESTIMATES FROM MOVING TO OPPORTUNITY

From 1994 to 1998 MTO enrolled 4,604 low-income public housing families living in high-poverty neighborhoods within five U.S. cities: Baltimore, Boston, Chicago, Los Angeles, and New York. Families were randomized into three groups: i) the *low poverty voucher* group, which received housing vouchers that subsidize private-market rents and could only be used in

census tracts with 1990 poverty rates below 10 percent, and additional housing-mobility counseling; ii) the *traditional voucher group*, which received regular Section 8 housing vouchers without any MTO relocation constraint; and iii) a *control* group, which received no assistance through MTO. Some 48% of households assigned to the low-poverty voucher group and 63% of those assigned to the traditional voucher group moved using housing vouchers through MTO (the MTO “compliance rate”). We pay particular attention to the contrast between the low-poverty voucher and control groups, where the MTO-induced difference in neighborhood conditions (and hence power to detect neighborhood effects on adult outcomes) is greatest.

The MTO study sample is unusual within this literature partly because the baseline neighborhoods from which families were drawn were particularly disadvantaged. MTO enrolled families living in some of the most distressed public housing projects in the nation, including the Robert Taylor Homes in Chicago and Lafayette Courts in Baltimore. The average poverty rate in the census tracts in which MTO families lived at baseline was 53 percent. In the Chicago and Baltimore sites the baseline neighborhoods were almost entirely African American, while in the other three cities the neighborhoods were more evenly split between African-Americans and Hispanics. There were almost no non-Hispanic white families in these baseline neighborhoods, and hence virtually no white families in the MTO study sample. Given the high level of disadvantage of the baseline neighborhoods, it is not surprising that data from baseline surveys show that these families were quite economically disadvantaged when they applied for MTO.

In addition, MTO was a *voluntary* mobility program. The best available estimates from HUD housing records suggest that about one-quarter of eligible households that were offered the chance to sign up for the MTO lottery enrolled in it (Goering et al. 1999). These volunteers may differ systematically from other residents of the baseline neighborhoods, but that would not

compromise the internal validity of MTO given the study's experimental design. However, it could have implications for what other populations the experimental results generalize to (external validity), and therefore their comparability with results from other studies—an issue we return to below.

Follow-up data on MTO families show that the demonstration was successful in moving families into lower-poverty areas, particularly those families assigned to the low-poverty voucher group (see Table 2 and the discussion below). Over time the contrast between the neighborhood conditions of the low-poverty voucher and control groups attenuated, partly because the control group itself was moving into somewhat lower-poverty areas either on their own or as a result of other policy changes in the background (such as the demolitions of public housing projects carried out within the five MTO cities over this period under HUD's HOPE VI program). Nevertheless, the effect of winning the lottery on duration-weighted tract poverty rates over the entire 10 to 15 year long-term MTO follow up was quite substantial—equal to 18 percentage points among those who moved with an MTO voucher.⁵ By comparison, over the course of the entire MTO study period control group families lived in census tracts that on average were 40 percent poor. In the national distribution of poverty rates, control group families lived in tracts at the 89th percentile, on average. The effect of winning the lottery was a reduction of 21 percentiles, on average, among those who moved with an MTO voucher. Experimental group members who used their vouchers spent an average of over five more years in lower-poverty

⁵ For our study we use the geo-coded locations of MTO addresses as constructed by HUD, which have the advantage of being consistently available for all of the residential addresses we have available for MTO households, but have the disadvantage of perhaps including some additional measurement error relative to other sources of geo-coded addresses. This would work in the direction of our understating MTO's effects on the neighborhood environments of participating households.

neighborhoods (census tracts with poverty rates below 25 percent) than they otherwise would have.

What does this imply about the magnitude of the MTO treatment “dosage”? On the one hand, by the standards of different social policy interventions that have been tried in the United States, MTO generates very large changes in social conditions for participating families. One year after randomization, the effect of receiving the housing voucher offer and subsequently using the voucher to move on tract poverty rates is about 2.8 standard deviations within the national census tract poverty-rate distribution. Over the entire MTO follow-up period the average family that moved with an experimental voucher lived in census tracts with poverty rates of about 20 percent, nearly 50 percent lower than that of the average control group family. As we will see below, only a small fraction of PSID families improved their neighborhood conditions by anything close to this amount.

MTO generated sizable changes in other measures of neighborhood economic disadvantage, and also led families to live in neighborhoods that were substantially safer, with seemingly higher levels of collective efficacy (Sampson, Raudenbush and Earls 1997), and with more social-network ties to more affluent neighborhoods, such as those with a college education (e.g. Sanbonmatsu et al. 2011). Despite these large impacts on neighborhood poverty, it is still possible that MTO created shorter-duration differences in neighborhood conditions compared with the variation in neighborhood conditions across individuals that is captured in observational datasets like the PSID.

Another possible way in which the MTO demonstration “dose” may be different from that of the PSID and other non-experimental datasets is that MTO families wound up living in lower-poverty neighborhoods that were still predominantly minority. The average control group

family spent the MTO follow-up period in a census tract in which 88 percent of residents were members of a racial or ethnic minority group. While moving with a low-poverty voucher had a statistically significant effect in reducing tract share minority, equal to 12 percentage points, these families still lived in predominantly-minority neighborhoods (see also Clampet-Lundquist and Massey 2008; Sampson 2008). Put differently, the correlation between tract poverty and tract minority share for the MTO sample is quite different from what we see in national data like the PSID (see Appendix Table A1).

For MTO adults perhaps the most important puzzle is why observational studies often find evidence of gains in economic outcomes—employment, earnings, social-program participation, participation in job training or educational activities, etc.—but MTO does not. The absence of detectable MTO effects on the earnings or employment of those who moved during adulthood has been consistently documented in both the interim and long-term follow-ups of MTO using both survey and administrative data measures of economic outcomes (Orr et al. 2003; Sanbonmatsu et al. 2011; Ludwig et al. 2012).⁶ The same finding continues to hold in the longer-term follow-up of MTO participants using IRS data (Chetty, Hendren, and Katz 2016). There is evidence of MTO effects on important health outcomes such as depression, obesity and diabetes (Sanbonmatsu et al. 2011; Ludwig et al. 2011). But the discrepancy in findings for earnings remains a mystery, and a particularly important one given the special emphasis given to earnings in both social science and policy discussions.

HYPOTHESES

⁶ However Pinto (2019) and Aliprantis and Richter (2020) seem to find beneficial MTO effects on adult labor market outcomes for the subset of the sample that experienced particularly large neighborhood changes.

These puzzles motivate a number of hypotheses about the discrepancies between MTO findings and those of previous observational studies regarding neighborhood effects for adults on economic outcomes. We provide a brief description of each hypothesis and indicate how they will be tested.

Outcome Measurement: MTO impacts on economic outcomes may differ substantially from effects found in observational studies because of the ways in which economic outcomes are measured. We will assess this with an examination of whether MTO experimental estimates differ from both MTO nonexperimental estimates and PSID estimates. Similar MTO experimental and nonexperimental estimates based on identical outcome measures would argue against this hypothesis.

Treatment Magnitude: Differences may arise if changes in neighborhood poverty induced by the MTO intervention are smaller than the variation in neighborhood poverty used to estimate neighborhood effects in the PSID. If selecting and weighting the data to align neighborhood poverty comparisons in MTO and the PSID does not reduce discrepancies in estimates, then this hypothesis is unlikely to be correct.

Treatment Duration: Differences in the duration of exposure to high-poverty neighborhoods in MTO and the PSID may explain differences in effects because, as we will see below, the MTO treatment groups tended to move back to somewhat higher poverty neighborhoods over time and the MTO control group experienced some improvements in neighborhood environments over the follow-up period. Weighting the data to align adult exposure to high poverty neighborhoods allows us to test this hypothesis. If such weighted analyses do not reduce the MTO-PSID discrepancy, this hypothesis is unlikely to be correct.

Non-linear effects: Differences in neighborhood effects may arise because most nonexperimental studies of neighborhood effects assume a linear relationship between neighborhood poverty levels and outcomes. If the relationship is non-linear and different studies make comparisons across different ranges of neighborhood poverty, then different estimates will result. Again, selecting and weighting the data to align the range of neighborhood poverty across the different samples used to estimate neighborhood effects allows us to test this hypothesis. This hypothesis is unlikely to be correct if such weighting does not reduce the discrepancy between MTO and PSID estimates.

Treatment effect heterogeneity: This hypothesis suggests that families with different characteristics have different capacities to take full advantage of moves to more advantaged neighborhoods. Because the MTO families are particularly disadvantaged relative to PSID families in high poverty neighborhoods (see Table 1 and our discussion below), we might expect smaller neighborhood effect estimates in MTO. By weighting the data to align observed family characteristics across the samples, we can examine whether effect heterogeneity related to those characteristics can account for the MTO-PSID discrepancies. If discrepancies persist even after family characteristics are aligned, that would suggest this hypothesis is incorrect. We caution, however, that we cannot examine effect heterogeneity by unobserved characteristics, including unobserved characteristics that may have impacted whether families volunteered for the MTO experiment in the first place.

Neighborhood measures: MTO families moved to lower poverty neighborhoods, which, given the high correlation of poverty with other components of composite neighborhood measures like the widely-used “neighborhood disadvantage index” of Sampson, Sharkey and Raudenbush (2008), we typically interpret as a summary proxy for neighborhood SES. But if

these elements do not co-vary the same way within the MTO sample as national samples, we might expect to see few effects from moving to lower poverty neighborhoods that are still disadvantaged on other dimensions. To address this hypothesis, we weight the data to align not only neighborhood poverty but also the neighborhood disadvantage index. Since, as noted above, we already know the MTO intervention was less successful at promoting one specific dimension of that index, racial integration, which could be particularly important for outcomes (Massey and Denton, 1993, Sampson 2008), we pay special attention to that in our analysis. If aligning samples on these other measures of neighborhood SES does not reduce discrepancies between MTO and PSID, then this hypothesis is unlikely to be correct.

Residential Mobility: To improve their neighborhood environments, MTO families also had to move, and residential mobility can itself be harmful if it removes individuals from social networks that support employment or provide job leads. Because the MTO research design cannot separate neighborhood effects from residential mobility effects, null effects on adult economic outcomes in MTO could be due to offsetting positive neighborhood effects and negative residential mobility effects.⁷ To test this hypothesis, we weight the data to align the number of residential moves between census tracts and the distances moved across the MTO and PSID samples. If doing so does not reduce MTO-PSID discrepancies, then this hypothesis is unlikely to be correct.

Selection bias: Because neighborhood effects in MTO are estimated by leveraging random variation in neighborhood poverty generated by the random assignment of housing

⁷ Previous MTO work (Kling, Liebman, and Katz 2007) did try to compare the effects of a discrete 1/0 residential-move variable with a measure of actual neighborhood poverty rates. The analysis suggested neighborhood poverty had much larger effects, although the statistical power of this analysis was somewhat limited.

vouchers, it is unlikely that any effects estimated in MTO are due to unobserved differences between families living in different types of neighborhoods.⁸ In contrast, observational studies rely on observing and conditioning on all pre-existing differences between families living in high and low poverty neighborhoods, a potentially untenable assumption. Because this is a hypothesis about unobserved variables, we cannot directly test this hypothesis. That is, it is a residual explanation that is rendered more plausible by the failure of the other hypotheses above.

DATA AND METHODS

The goal of our analysis is to examine discrepancies in findings between MTO and previous studies. Such an effort to understand what estimation approaches or sample restrictions let us reproduce an experimental impact estimate using non-experimental comparisons has become known as a “within-study comparison.”⁹ We use two data sources: the survey of *MTO participants* conducted as part of the final follow-up of MTO households, which measured outcomes 10-15 years after baseline (Sanbonmatsu et al. 2011); and the *Panel Study of Income Dynamics* (PSID 2014), one of the most important and widely-used social science panel datasets in the literature on neighborhood effects. Our choice of the PSID reflects prior knowledge about requirements for effective within-study comparisons, including the availability of pre-treatment

⁸ There is of course always the chance that in any given sample, random assignment fails to achieve balance between the treatment and control groups in the distributions of all baseline variables that affect the outcomes of interest. Previous MTO research has shown that at least among those baseline variables that are captured in the available data, there does indeed seem to be balance (for example Orr et al. 2003; Sanbonmatsu et al. 2011).

⁹ Within-study comparisons began with LaLonde (1986) and Fraker and Maynard (1987) in the context of job training, and have now been extended to a wide range of other applications such as welfare (Michaelopoulos, Bloom, and Hill 2004), early childhood interventions (Dong and Lipsey 2018), education (Agodini and Dynarski 2004; Angrist et al. 2015; Bifulco 2012; Fortson et al 2015; Hallberg, Wong, and Cook 2016; Jacob et al. 2016; Wilde and Hollister 2007), immigration (McKenzie, Stillman, and Gibson 2010), health (Schneeweiss et al. 2004), and voting and political behavior (Arceneaux, Gerber, and Green 2010). See Wong, Steiner, and Anglin (2018) for a recent overview of within study comparisons and Burdick-Will et al. (2011) for a within-study comparison study that addresses child outcomes using MTO participants in Chicago and PHDCN subjects.

measures of the outcomes, the capacity to measure treatments and outcomes in the same way, and correspondence of geographic scope (Cook, Shadish, and Wong 2008). Although the correspondence between the five MTO cities and the nationwide scope of the PSID is far from perfect, the PSID provides the closest possible geographic correspondence among candidate studies that otherwise meet our need for neighborhood measurements, long-term follow-up, and measurement similarity.

MTO families completed a baseline survey at the time they enrolled in the MTO program between 1994 and 1998. Between 2008 and 2010, the Institute for Social Research (ISR) at the University of Michigan collected in-person survey data on 3,273 adult participants in MTO. ISR achieved an effective response rate of 90% for the adults, with similar response rates across treatment groups. In this paper, we focus exclusively on female adults because females comprised about 98% of the MTO adult sample.

The PSID is a longitudinal household survey which began in 1968 with a nationally representative sample of about 18,000 people—and by 2009 included over 24,000 people. In addition to having been used in numerous previous studies of neighborhood effects, the PSID also has the advantage for our purposes of including outcome measures and time periods that are comparable to MTO. We use 1997 as our “baseline” year of PSID data, examine neighborhood exposures over the 1999-2009 period, and look at adult outcomes for 2009. Restricting the PSID data to females who were household heads or spouses in both 2009 and 1997 yields a sample of 4,299 observations with neighborhood and outcome information.¹⁰ We link public-use PSID data

¹⁰ Unlike with MTO, where HUD set eligibility criteria for the program that included the requirement that heads of household had children, the PSID was intended to be a nationally representative sample, so it includes people without children. (We can see this in the descriptive statistics for the unweighted PSID sample, where the mean number of children is less than 1). The trimming of the PSID sample to make it look more like MTO, which reduces the sample size we use from PSID from 4299 to 850, reduces the share of adults without children from 59% (unweighted) to 2% (trimmed).

on baseline characteristics and subsequent outcomes to the PSID restricted use data that includes residential census tract. The PSID survey conditional response rates in 2009 were about 95% for the main sample and 90% for the Latino/Immigrant refresher sample, although of course those families still in the PSID by the start of our study period and also eligible for the 2009 survey frame are a selected sample.¹¹

For our analysis of neighborhood characteristics, we link MTO address data from self-reports and passive tracking sources to census tract characteristics from the 2000 census. For both the MTO and PSID samples, we use biannual address data for the 1998/99 through 2008/09 time period. For PSID, follow-up surveys occurred biannually beginning in 1997, and we therefore use address data from odd years from 1999 to 2009. The MTO long-term survey included an address history where adults confirmed or reported addresses from their randomization date through their address as of the survey. To construct neighborhood measures parallel to the PSID, MTO address information was limited to biannual years between either 1998 and 2008 or 1999 and 2009.¹²

A key advantage of the randomized experimental design of MTO is that the low-poverty voucher and control groups have background characteristics that are very similar on average, with respect to both observed variables (which we can check) and unobserved variables (given the statistical implications of randomization). These two very similar groups then go on to

¹¹ The methods section below includes additional information about weighting, but our nonexperimental analyses of the PSID and MTO samples are generally unweighted except for our models that use propensity score weights to produce more comparable groups or treatment dosages and some of our PSID sensitivity analyses that apply attrition weights. The experimental analysis of the MTO sample applies the MTO probability weights to account for changes in random assignment ratios across randomization cohorts, survey sample selection, and two-phase interviewing.

¹² The time period used depends on the year that the adult completed her long-term survey interview. For adults interviewed in 2008 (fielding began in June), we use data for even years in 1998-2008 range (excluding 1998 for the small number of adults in the Los Angeles site who were randomized in 1998). For adults interviewed in 2009 (and the first few months of 2010, when fielding ended in April), we use data for odd years in 1999-2009 range.

experience very different neighborhood environments over the course of the post-program period (see Table 1 below). In contrast, in the PSID we can construct sub-samples of people that experience “MTO-like” differences in neighborhood poverty rates, but these high- and low-neighborhood-poverty groups differ substantially from one another with respect to their observed baseline characteristics (and hence potentially with respect to unobserved characteristics as well).

A key challenge in using the PSID for present purposes is that when we define PSID sub-samples that experience contrasts in neighborhood poverty rates that are similar to what we see in the MTO experiment, we end up with a set of PSID respondents whose background characteristics are not nearly as disadvantaged as those for the MTO study sample. This pattern motivates the concerns embedded in our “treatment heterogeneity” hypothesis—the MTO sample is very disadvantaged even compared to samples of people living in high poverty neighborhoods drawn from a nationally representative sample. If, instead, we try to identify a sub-sample of PSID respondents whose own level of household disadvantage is similar to that of the MTO study sample, the PSID group’s neighborhood poverty rates are not nearly as high as what we see among the MTO control group. In some sense this should not be surprising, since MTO limited eligibility to people living in public housing projects with poverty rates above 40 percent, and thereby selects only those households with disadvantaged observable characteristics who were initially living in very high-poverty areas. Whatever the underlying cause, this pattern helps to motivate our treatment magnitude and nonlinearity hypotheses. To overcome these challenges in creating comparable samples in MTO and PSID, we use weighting as described further below.

In our main models, we use the average neighborhood poverty over the follow-up period to construct our nonexperimental “treatment” and “control” groups. We define the groups such

that the difference in poverty between the two groups is similar to the effect of the MTO intervention on neighborhood poverty. In addition to using average poverty, we perform analyses using other measures of neighborhood disadvantage including the duration of time spent in low poverty neighborhoods and the “concentrated disadvantage” index commonly used in neighborhood effects studies (e.g. Sampson, Sharkey and Raudenbush 2008).¹³

A final issue is our choice of outcome measures. The neighborhood effects literature draws on a large number of different data sources that include different specific measures of adult economic self-sufficiency. One common measure is current employment status, which captures the idea that job opportunities themselves may vary across neighborhoods. In principle one could look at current earnings instead, which creates the possibility of also incorporating information about the quality of jobs. However, relative to a binary employment indicator, earnings are harder for subjects to report and more prone to outlier values, so earnings could introduce not just new signal but also additional noise.

Our approach here is essentially to focus on both measures and use the composite “outcome indices” that have been used in many MTO analyses (Kling, Liebman, and Katz 2007; Ludwig et al. 2012, 2013). The use of a single index rather than all of its components reduces the risk of “false positives” from carrying out a large number of different hypothesis tests, improves measurement validity and reliability, and can also reduce the risk of “false negatives” by improving statistical power to detect effects (Kling, Liebman, and Katz 2007). This strategy views the index as a summary of information on self-sufficiency from multiple constructs rather than as a measure of a single latent construct. Because we are limited to variables measured in

¹³ This measure includes tract percent black, welfare receipt, unemployment rate, female-headed households, and share residents under 18, in addition to tract poverty rate.

both MTO and PSID, our economic self-sufficiency index includes individual earnings (in 2009 dollars) and an indicator for current employment. The index is an average of these two measures, both standardized using the PSID female sample into Z-score form (subtracting off the PSID mean and dividing by the PSID standard deviation). It has a Cronbach's alpha reliability of 0.70 for the MTO data and 0.77 for the PSID data.¹⁴ In the MTO data, the two-item index used here has a correlation of 0.91 with the index used in prior MTO studies. The appendix shows results separately for the individual measures of earnings and employment (Tables A8 and A9). Earnings for the PSID sample were top-coded at the 95th percentile, and a small number of outlier values were also top-coded in the MTO data. The results are qualitatively similar for each of the components of our labor market index when analyzed separately.

Statistical Models

We begin our analysis with a careful examination of how participation in MTO affected future exposure to disadvantaged neighborhoods. These estimates provide important background for understanding any neighborhood effects estimated with MTO, as they elucidate the changes in neighborhood context created by the MTO experiment, a key question in the debate over discrepancies between experimental and non-experimental studies of neighborhood effects on adults. We follow prior analyses of the MTO data by estimating MTO's experimental impact in a multi-variate regression framework (Sanbonmatsu et al. 2011). We begin with MTO's impact on neighborhood conditions over the follow-up period. Let D be an indicator variable for use of a voucher to move through the MTO program, or treatment compliance. Let Z be an indicator for

¹⁴ The lower reliability in MTO is due to standardizing based on the PSID. Were the MTO measures standardized on the MTO data, the reliability of the index in the MTO data would be 0.76.

being eligible for an MTO program subsidy, or randomized treatment group assignment. Let subsidy use be a function of a set of observed characteristics from the MTO baseline survey either known prior to randomization or retrospectively reported as existing prior to randomization (X) and other factors ε_1 , as in (1).

$$(1) \quad D = Z\pi_1 + X\beta_1 + \varepsilon_1$$

The “Intention-To-Treat effect” (ITT) is captured by the estimate of the coefficient π_2 in a regression of some neighborhood measure (Y) on an indicator for assignment (Z) to a treatment group as in (2).

$$(2) \quad Y = Z\pi_2 + X\beta_2 + \varepsilon_2$$

We condition on baseline characteristics (X) to improve the precision of our estimates by accounting for residual variation in pre-existing characteristics.

This ITT parameter is an average of the causal effects of MTO randomization on future exposure to neighborhood disadvantage for those who do and do not take-up the “treatment” (that is, to relocate with a MTO voucher). Therefore, ITT estimates will understate the effects of actually moving to a lower-poverty neighborhood after baseline given that not all families “complied” with their MTO treatment assignment. Specifically 49 percent of families randomly assigned to the experimental group relocated to a low-poverty neighborhood through MTO, while 62 percent of those assigned to the regular Section 8 group relocated with a voucher through MTO.¹⁵ Note however that ITT estimates are *not* biased by self-selection in MTO treatment take-up within the treatment groups, because the control group is compared with all

¹⁵ These are the MTO compliance rates among female MTO adult respondents in the long-term follow-up survey.

families assigned to a treatment group whether or not the latter decide to accept the invitation to participate in the voucher program (Ludwig et al. 2008).¹⁶

We then use the randomized treatment assignment as an instrumental variable to calculate the effect of moving with an MTO voucher on future neighborhood context, or what is commonly called “effect of the Treatment on the Treated” (TOT). Under assumptions that (a) treatment group assignment is random, (b) the Control group is prohibited from receiving MTO vouchers, and (c) the effect on outcomes of treatment assignment works entirely through making a subsidized move through MTO, TOT is equal to π_2/π_1 , or ITT divided by the proportion receiving the treatment.¹⁷ The TOT provides an estimate of the effect of an MTO induced move among those who actually moved. We show ITT and TOT estimates on various neighborhood characteristics measured between the MTO baseline and the final follow-up.

Our analysis of economic self-sufficiency first estimates the MTO experimental TOT using the self-sufficiency scale at the end of long-term follow-up as the outcome. We then compare these MTO experimental estimates to neighborhood effects estimates from the PSID.¹⁸ If we let P represent some binary measure of neighborhood conditions, such as low poverty (1)

¹⁶ The greatest threat to internal validity is potential bias from sample attrition. Extensive efforts were made to achieve an effective response rate of over 89% for the long-term MTO follow-up study, which was quite similar for each of the three randomly-assigned MTO groups (Sanbonmatsu et al. 2011). In addition, the results do not seem to be very sensitive to survey non-response as judged by trying alternative techniques such as weighting and multiple imputation and through comparisons to estimates of MTO employment and earnings impacts using administrative unemployment insurance data covering the full MTO sample.

¹⁷ These assumptions imply that the experience of housing counseling and search induced by assignment to a treatment group did not affect later outcomes if that household did not make a program move. We believe that this assumption is probably not strictly true, but we believe that effects of housing counseling are likely to be orders of magnitude smaller than the effects of moving. When equations (1) and (2) are estimated using ordinary least squares (OLS), this is numerically identical to a two-stage least squares regression of Y on D with Z used as an instrumental variable for D.

¹⁸ Note that because moving with a voucher in MTO is highly correlated with future exposure to neighborhood poverty once we weight the data to align neighborhood poverty gaps, the MTO experimental TOT estimate is equivalent to an instrumental variables analysis that uses the MTO randomization to instrument for neighborhood poverty.

and high poverty (0), we use the PSID to estimate the standard regression as in equation (3). We construct PSID high-poverty and low-poverty groups so the size of the difference in tract poverty rate is similar to the MTO TOT effect on average tract poverty rates over the 11 years leading up to and including the MTO follow-up year.¹⁹

$$(3) \quad Y = P\pi_3 + X\beta_3 + \varepsilon_3$$

The key scientific question of interest for our analysis is whether the MTO TOT estimates are distinguishable from estimates obtained by the PSID; that is, whether $\pi_2/\pi_1 = \pi_3$.

We should note that our estimates for both parameters can be somewhat imprecise, given limits on the sample sizes available in both of our datasets—particularly when we weight samples to align them on baseline characteristics and neighborhood contrasts.²⁰ As a result, the confidence intervals around our test of the contrast between π_2/π_1 and π_3 can be sizable. A different way to think about comparing the MTO and PSID estimates would be to ask whether we would draw different inferences about the existence of neighborhood effects if we had relied on the MTO study rather than the PSID. In other words, are there situations where we would reject the null hypothesis that $\pi_3=0$ but would not reject the null hypothesis that $\pi_2/\pi_1=0$, or vice versa? More generally we tend to focus on differences in point estimates even though (particularly when we use weights, discussed below) few of them would pass conventional statistical significance tests at $p<.05$.

¹⁹ One might worry that we are simply matching the “treatment dose” on neighborhood differences for one neighborhood attribute at a time, so that if it is actually the combination of neighborhood attributes that a family experiences, we might miss that in our analysis. But below we also present results that focus on matching the MTO and PSID samples on the “dosage” of the index of multiple neighborhood disadvantage measures suggested by Sampson, Sharkey, and Raudenbush (2008).

²⁰ PSID analyses are unweighted except for the models in which the PSID and MTO data have been propensity-score reweighted. That is, we do not use the PSID sampling weights in our regression analyses.

To explore the question of whether different samples and treatment heterogeneity can help explain the difference in the MTO versus PSID findings, we also replicate the observational estimates in the PSID using the MTO data by treating the MTO data as non-experimental data. That is, we essentially discard the experimentally-induced variation in neighborhood conditions from the MTO experimental design and instead use only within-randomized-group variation in neighborhood conditions, which will be purely observational (non-experimental). For example, we use the MTO control group to estimate an equation like (4):

$$(4) \quad Y = P\pi_4 + X\beta_4 + \varepsilon_4 \text{ for all observations with } Z=0$$

In practice we pool data from both the traditional voucher and control groups in estimating equation (4) to improve sample size and hence statistical power, adding an indicator for the traditional voucher group to the model as a control variable. Even with data from two MTO groups included in the regression, with the inclusion of the traditional voucher indicator all of the variation that is used to estimate the relationship between neighborhood poverty (P) and the outcome (Y) in equation (4) will be within-randomly-assigned-MTO group (that is, “non-experimental”). Generating non-experimental estimates using the MTO data also helps us determine whether something else about the MTO study, such as the way the data were collected or how the outcome measures were constructed, could also explain differences in results with datasets like the PSID.

We explore the degree to which treatment heterogeneity can explain differences in results between MTO and PSID by weighting the data so that the sample characteristics with respect to basic baseline attributes (Xs) are similar across samples. We weight the various groups to look similar to the PSID respondents who lived in high poverty neighborhoods (on average greater than 25% poor). Prior to weighting the different subgroups to look like the PSID high poverty

group, we first trim the samples to exclude observations with little overlap with the MTO sample on basic demographics such as age, race, education, and being above or below the poverty line.²¹ We exclude observations with propensity scores greater than 0.9 from a model predicting whether the respondent is in the PSID vs. MTO low-poverty voucher sample using basic demographics. By using inverse probability weighting together with regression adjustment we are following in the spirit of the double-robust analysis recommended by Imbens (2015).

Finally, we examine how different types of neighborhood exposures, nonlinearities in neighborhood effects, or residential moves might explain differences between the MTO experimental estimates and those from PSID. We re-estimate equations (3) and (4) in different ways. First, we can construct our non-experimental “treatment” and “control” groups to both maintain the same contrast or “gap” in average poverty and the same levels. This allows us to examine if, say, a 17-percentage point reduction in average neighborhood poverty has the same effect whether the contrast is between 15% and 32% average poverty or between 20% and 37% poverty. See Appendix A for more details on the methods used to construct these contrasts. Second, we can use weighting to simultaneously bring the samples closer together on demographics and make the treatment magnitude (or “dosage”) on different neighborhood characteristics more similar to the MTO experimental analysis. We can match the treatment magnitude with respect to poverty, the concentrated disadvantage index, and duration of exposure to high poverty neighborhoods. To examine hypotheses related to the duration of the exposure and mobility, we can also bring the nonexperimental analyses into alignment with the experimental treatment magnitude on years in low poverty neighborhoods and number of moves across census tracts.

²¹ We focus on basic demographics for which we can achieve some balance between the samples.

We use propensity models of the trimmed data to weight each lower-poverty group (<25% poverty) to look like the PSID higher-poverty group (> 25% poverty) on basic demographics.²² Weights are defined as $p/(1-p)$ from a propensity model using a boosted Classification and Regression Tree implemented using the Toolkit for Weighting and Analysis of Nonequivalent Groups (Griffin et al. 2014). Using the weighted lower-poverty group, we then weight each higher-poverty group to look like the corresponding lower-poverty group on basic demographics and at the same time to have average poverty levels that are higher than the lower-poverty group's average poverty rate by the magnitude of the TOT (using the weighted and trimmed data). The result of the weighting is to construct high and low neighborhood poverty groups that are similar on basic demographics and approximate the MTO treatment magnitude as well as the range of neighborhood poverty or disadvantage in which the gap occurs. See Appendix A for details on the propensity score trimming and weighting.

RESULTS

Comparison of MTO and PSID Samples

We begin by presenting descriptive analyses of the MTO and PSID samples we analyze. Recall that the PSID female sample we use is a subset of the larger PSID designed to be most comparable to MTO adults. These results serve to document the similarities and differences between the samples and to demonstrate the initial plausibility of some of the hypotheses we will analyze more rigorously below. Table 1 shows MTO and PSID sample characteristics at baseline and subsequent exposure to neighborhood poverty over the ten-year follow-up period. We show

²² We choose a 25% threshold for distinguishing the lower and higher neighborhood poverty groups because this threshold allows us to better align the MTO and PSID samples through weighting.

descriptive statistics for the PSID calculated without sample weights (which is how we carry out our regression analysis), and for completeness also show these means calculated using the PSID's longitudinal sampling weights and cross-section sampling weights. Note that although the MTO baseline characteristics are stratified by the three arms of the MTO experiment (low-poverty voucher, traditional voucher, and control group), the average characteristics of these three groups are almost identical, which is a direct result of the MTO randomization.

One important result from Table 1 is that the MTO sample is considerably more disadvantaged than the PSID sample.²³ MTO respondents are more likely to be single mothers, to have not graduated from high school, to be not working, to have incomes below the poverty line, to be on welfare, and to have more children. MTO respondents are also living at baseline in poorer neighborhoods with more African-American residents than PSID respondents measured in the same year. These baseline differences motivate the treatment effect heterogeneity hypothesis, which argues that MTO respondents are so much more disadvantaged at baseline that they may be ill-prepared to take advantage of the features of low-poverty neighborhoods that could lead to improvements in employment or earnings.²⁴

Note also that there are important baseline differences within the PSID sample between respondents who will experience low poverty neighborhoods in the future (here, defined as mean yearly poverty rate less than 25% poor) and those who will experience high poverty

²³ This is the case even when we narrow the PSID sample to African American and Hispanic women with incomes below 200% of the poverty line.

²⁴ Table 1 also shows that MTO respondents live in different regions than PSID respondents, especially when we limit the PSID sample to those living in high poverty neighborhoods or to African American or Hispanic women. PSID respondents are more likely to live in the South than MTO respondents. This is a result of city-stratified sampling design of MTO and subsampling by poverty status in the PSID. Unfortunately, our weighting procedures are unable to correct for such large regional differences (see Appendix Table A4). All models control for region. Because the discrepancies between the PSID and experimental MTO estimates are similar to the discrepancies between the MTO experimental and MTO nonexperimental estimates, we are not concerned that the particularities of MTO study cities are driving the PSID-MTO discrepancies.

neighborhoods in the future (greater than 25% poor). The latter group is more likely to be single mothers, has more children, is less likely to have graduated from high school, has lower incomes, and is more likely to be on welfare and less likely to be working. They are also living in neighborhoods with a higher proportion African American population at baseline. It is these observed differences that lend plausibility to the selection bias hypothesis. If observed differences between PSID mothers who experience low and high poverty neighborhoods are stark, then there are potentially unobserved differences as well. The question is whether such unobserved differences are both sufficiently independent of the observed covariates (control variables) and sufficiently important for future employment and earnings to impart serious biases to causal effects estimated from “selection on observables” methods like regression or matching.

The bottom section of Table 1 documents exposure to neighborhood poverty during the follow-up period. Focusing first on the differences across the randomized MTO groups in the left side of the table, we see that there are differences in neighborhood poverty, with the control group experiencing the poorest neighborhoods and the low-poverty voucher group experiencing the least poor neighborhoods. Note that these groups are based on the randomization, not on the actual neighborhoods experienced later, so they demonstrate that the randomization did indeed produce variation in neighborhood poverty. However, the variation looks relatively small, especially in comparison with the naturally occurring variation within the PSID sample. Note also that, in comparison to the PSID, all the MTO randomization groups have large numbers of mothers who experience very high poverty neighborhoods (the 75th percentile for all groups is above 40% poor).

Numbers like these lend some plausibility to the treatment magnitude hypothesis, as they suggest that the change in neighborhood poverty induced by MTO may not be large enough to

affect outcomes appreciably. Furthermore, they are also suggestive of the non-linear effects hypothesis because they indicate that changes in neighborhood poverty in the MTO sample occurred at higher levels of neighborhood poverty than the naturally occurring variation in neighborhood poverty in the PSID, where there are large numbers of PSID mothers in very low poverty neighborhoods (less than 10% poor), thus providing the opportunity to compare mothers who experience very low and very high poverty neighborhoods.

Table 2 provides more details on the changes in neighborhood conditions created by the MTO intervention by examining multiple neighborhood measures and how exposure to neighborhood poverty evolved over time. The mean duration-weighted neighborhood poverty rate in the control group was 40 percent.²⁵ In the experimental group, the mean neighborhood poverty rate was 9.2 percentage points lower (the ITT effect of the MTO experiment on mean neighborhood poverty), but when we use the randomization as an instrument (and isolate impacts on families initially moving to low-poverty neighborhoods), we see that there was an 18.8 percentage point drop, on average, among those who moved with a voucher (the TOT effect on mean neighborhood poverty).

We also examined the degree to which MTO changed other neighborhood characteristics because the “neighborhood measure” hypothesis argues that other measures of disadvantage that might be important for our outcomes were less affected by the MTO intervention. This appears to be the case, especially for percent African-American and percent minority. Percent African-American was reduced from 54.3 percent in the control group by seven percentage points while percent minority was reduced from 88.1 percent by 12.7 percentage points, meaning the average

²⁵ Table 2 uses neighborhood measures that are duration weighted using all MTO addresses whereas for analyses comparing MTO and PSID below we limit the MTO data to biannual measures to match PSID.

MTO mother in the experimental group lived in a neighborhood that was almost majority African-American and decidedly majority minority.

The concentrated disadvantage index shows similar results, falling from 1.855 standard deviations above the national mean by only about half a standard deviation, meaning low-poverty voucher group mothers still lived in neighborhoods far above the national mean on the concentrated disadvantage index after their initial program-related moves. This is consistent with past findings that MTO improved neighborhood poverty far more than neighborhood racial composition, in part due to the challenges of defying the powerful social forces that enforce neighborhood racial segregation (Sampson 2008, 2012). Reductions in neighborhood unemployment and public assistance receipt are similar to those for poverty, while reductions in percent youth and female headed households are similar to those for racial composition.²⁶ Below we will align the MTO and PSID on the concentrated disadvantage index to examine the neighborhood measure hypothesis directly.

The remainder of Table 2 focuses on how the neighborhood poverty rate evolved over time in MTO, showing mean neighborhood poverty rates at 1, 5, and 10 years following randomization and the number of years in tracts with poverty rates below various thresholds. Over time, the MTO induced reduction in neighborhood poverty fell from 33 percentage points one year after randomization to 10.8 percentage points ten years after randomization. This reduction reflects in part lower neighborhood poverty rates in the control group and in part higher poverty rates in the experimental group over time. Concordantly, the mean number of

²⁶ Appendix Table A1 shows correlations between the tract poverty rate and the other neighborhood characteristics. Percent poor is highly correlated with welfare receipt, unemployment, and female headed households, moderately correlated with percent youth, and exhibits a low correlation with percent black in the MTO sample. This explains how MTO reductions in neighborhood poverty are accompanied by reductions in welfare receipt, unemployment, and female-headed households but smaller reductions in percent black.

years in the lowest poverty neighborhoods (less than 10% poor) was increased by MTO by 2.7 years, from half a year (0.5) in the control group. The increase in years in neighborhoods with poverty below 25% was larger at 5.2 years, up from 2.6 in the control group. These results suggest the plausibility of the treatment duration hypothesis, which argues that the reduction in neighborhood poverty created by MTO could have been too short-lived for many participants to substantially improve adult economic outcomes (although little is known in practice about how much time is required in such a neighborhood to change outcomes). They also suggest a potential role for the non-linear effects hypothesis because they show the reductions in neighborhood poverty that did occur still left many MTO participants whose neighborhood environments did not improve in neighborhoods where poverty rates were still fairly high, at 25 percent. (Appendix Table A10 provides parallel results using the percentile rank for each variable.)

Effects on Economic Self-Sufficiency

We now turn to the hypotheses that might explain discrepant results between the MTO experimental estimates and PSID observational estimates of effects of neighborhood poverty on adult economic outcomes (as measured by the economic index described above). Results are presented in Table 3 (parallel results for separate earnings and employment outcome variables are presented in Appendix Tables A8 and A9).²⁷ There is one row for each estimate of the effect of living on average in low poverty neighborhoods rather than high poverty neighborhoods during the follow-up period. That is, the independent variable is a binary indicator for low (1) vs. high (0) mean neighborhood poverty during follow-up. Each row shows the mean neighborhood

²⁷ Parallel analyses of two other adult outcomes, mental and physical health, are provided in Appendix B. These are outcomes for which the existing research finds similar results in experimental and observational studies.

poverty rate in the two groups, the difference between the neighborhood poverty rates, the mean outcome in the high neighborhood poverty group, and the estimated effect along with its standard error, p-value, and the sample size.²⁸ Positive effects indicate improvements in economic outcomes from living in less poor neighborhoods, which corresponds to the analysis framework in MTO of assessing the effects of an intervention to improve the neighborhood environments of participants.

Row A1 shows the TOT estimate that compares the low-poverty voucher group with the MTO control group (while also conditioning on observed covariates).²⁹ Consistent with prior results on employment and earnings from MTO, this effect is very close to zero (slightly negative in fact) and not statistically significant. Row B1 shows estimates from the PSID data comparing low and high neighborhood poverty groups via a regression model controlling for covariates (see Appendix Tables A2 for covariates available in MTO and PSID; covariates used in the main text are dataset-specific covariates, but results are similar when covariates common to both datasets are used as shown in Appendix Table A7).

The PSID estimate is positive and statistically significant, with an estimated improvement in the economic index of about a tenth of a standard deviation. (Note that the mean outcome in the high neighborhood poverty group is very similar in the two samples, -0.229 vs. -0.262, so the difference in effect estimates is due to differences in outcomes between the PSID and MTO low-

²⁸ Table 2 and Table 3 (Row A1) show different mean poverty rate TOTs because Table 2 uses the duration-weighted MTO measure making use of the full MTO address data over the entire follow-up period. Table 3's measure of neighborhood poverty exposure is constructed for the MTO sample to be parallel to the PSID so it uses only biannual addresses during the present study's follow-up time.

²⁹ In Row A1 of Table 3, the low-poverty group mean is the experimental complier mean, the poverty rate differential is the TOT for the effect on the mean neighborhood poverty rate during follow-up, and the high-poverty group mean is the implied experimental control complier mean (the experimental complier mean minus the TOT).

poverty groups). Researchers looking at these estimates would draw different conclusions from the MTO experimental estimates versus the PSID observational estimate about whether neighborhood poverty in adulthood matters for labor market outcomes.³⁰ The differences between these two estimates (A1 and B1) for the full sample and for key subgroups (discussed below) are the puzzles this paper attempts to address.

Row B2 presents the “within-study comparison” estimate from MTO. It uses the MTO traditional voucher group and the control group and analyzes the data as if it were observational data (ignoring the random assignment and simply comparing MTO respondents who lived in low and high poverty neighborhoods during follow-up). These “nonexperimental” MTO estimates are valuable because they rely on the same assumption as the PSID regarding selection on observables while preserving congruence with the MTO experimental estimates on other aspects of study design, including variable measurement, geography, and baseline characteristics. This approach produces an almost identical treatment effect estimate as the PSID, corresponding to a 0.096 standard deviation improvement in the economic index. We interpret the close similarity between the PSID and MTO nonexperimental estimates coupled with their divergence from the MTO experimental estimates to mean that the difference between the PSID and MTO experimental estimates cannot be due to differences in outcome measures.³¹

³⁰ Given the size of the standard errors around the two sets of estimates, the difference between the estimated effects in MTO versus PSID of about 0.11 SD is not statistically significant.

³¹ We made an analogous comparison between experimental and nonexperimental estimates within the MTO study using data from 2002 (4 to 7 years after random assignment) instead of the data collected from 2008 to 2010 (10 to 15 years after random assignment) focused on this paper, and found a similar but weaker pattern of experimental and nonexperimental results. The labor market in 2002 exhibited lower unemployment than from 2008 to 2010, enhancing the external validity of results from that earlier data. We focus on the later data because it has equivalent internal validity and was collected more than twice as long after random assignment—providing more comprehensive information about the effects of living in different neighborhoods over time.

Note that the neighborhood poverty rate differences on which the effects are estimated are smaller in the MTO experimental sample than the PSID sample (16.9 percentage points vs. 23.5 percentage points). The difference in the MTO nonexperimental sample is similar at 23.6 percentage points. Rows C1 and C2 present PSID and MTO nonexperimental estimates from samples selected to achieve a gap similar to the MTO experimental sample, a step in the direction of creating a sort of “common support” on neighborhood conditions. (As is evident from the group mean neighborhood poverty rates, this was accomplished in the PSID sample by selecting low neighborhood poverty cases with higher neighborhood poverty rates and accomplished in the MTO nonexperimental sample by selecting both low neighborhood poverty cases with slightly higher neighborhood poverty and high neighborhood poverty cases with slightly lower neighborhood poverty.³²) Doing so does not entirely remove the discrepancy between the MTO experimental estimates and the two nonexperimental estimates, but it does make the discrepancy between the MTO experimental and MTO nonexperimental smaller than the MTO experimental-PSID discrepancy (and results in neighborhood effect estimates that are no longer significantly different from zero). We interpret these findings to mean that the treatment magnitude hypothesis cannot fully explain the discrepancy between the magnitude of the MTO and PSID estimates (particularly since the discrepancy re-emerges when we weight the data below). Because the MTO samples are also estimating neighborhood effects in almost the exact same range of neighborhood poverty, these results also suggest that the non-linear effects hypothesis is incorrect.

³² Appendix Tables A5 and A6 track the means and 25th and 75th percentiles of the tract poverty rate, concentrated disadvantage index, years in low-poverty areas, and number of moves across census tracts as we limit and weight the samples in Table 3.

The second panel of Table 3 presents estimates from trimming and weighting the samples. We weight the MTO treatment groups from the experiment and the MTO and PSID high and low poverty groups to look like a common reference group: the PSID high poverty group. We choose the PSID high-poverty group as our common reference group, rather than choosing the MTO sample as the reference group, because as noted above the MTO sample is unusually disadvantaged and so is less nationally representative. As also noted above, we do some trimming to avoid problems that arise with people with propensity scores that are very close to 0 or 1, which can yield extreme weights. This process allows us to align the baseline characteristics in the PSID and MTO samples, thereby addressing the effect heterogeneity hypothesis. Furthermore, it will allow us to align the samples on other neighborhood characteristics and the number of residential moves in order to examine additional hypotheses. Appendix Tables A3 and A4 show the balance on baseline covariates both before and after trimming and weighting. The weighting procedure has resulted in good balance on key covariates across samples. The weighting procedure has also increased the standard errors on all effect estimates substantially (as weights always do).

Row D1 of Table 3 shows reweighted MTO experimental estimates and rows E1 and E2 show weighted PSID and MTO nonexperimental estimates, respectively. The neighborhood poverty gaps and levels continue to be closely aligned between the three samples, but the effect estimates are now even more different.³³ The weighting increased somewhat the PSID and MTO

³³ Our analysis purposefully aims to analyze a traditional nonexperimental method in which the comparison is made between people above and below thresholds of the neighborhood poverty rate. By its nature, that method has less variance in the neighborhood poverty rate of the low-poverty group for the nonexperimental method than for the experimental method; the subset of the experimental low-poverty group compliers with neighborhood poverty rate above the threshold have no counterpart in the nonexperimental low-poverty group. The reason this might matter is that the effect of the neighborhood poverty rate might be nonlinear. We address that as one of our candidate hypotheses below, but do not find evidence supporting it. The reason as seen in Appendix Table A6 is that the bulk

nonexperimental effect estimates, bringing them back closer to the levels in Rows B1 and B2. More importantly, the MTO experimental estimate is now negative (but not statistically significant due to the impact of weighting on the standard errors). This is a product of the weighting on baseline covariates, which resulted in upweighting of more advantaged MTO participants to make the MTO experimental sample more similar to the PSID sample. Note that the high poverty group's neighborhood poverty mean is the same pre- and post-weighting (compare rows A1 and D1), but the mean outcome in the high neighborhood poverty group is now higher (better) and the implied mean in the low poverty group is unchanged (add the mean outcome in the high poverty group to the effect estimate).³⁴

Below we return to the question of why the MTO experimental estimate is negative when the sample is weighted to align more closely with the more advantaged PSID sample. Understanding the source of this negative effect estimate may prove important to understanding why the MTO experimental estimate differs from the MTO nonexperimental and PSID estimates. Before we do so, we need to first examine whether the other hypotheses can account for the discrepancy.

The neighborhood measures hypothesis suggests that the reason MTO did not find effects on adult employment or earnings was that MTO improved neighborhood poverty but had only a small effect on other dimensions of neighborhood SES. To examine whether this can explain the divergent outcomes, we added the concentrated disadvantage index to the weighting procedure to

of the distribution is actually pretty similar: The 25th and 75th percentiles for neighborhood poverty in panel D for the MTO TOT are .128 and .250 versus .152 and .219 in panel E for the MTO non-experimental comparison (just a little narrower at the top and bottom but unlikely enough to generate important nonlinear effects).

³⁴ This interpretation is also consistent with the comparison of the unweighted and weighted MTO experimental group means in Appendix Tables A4 and A5, respectively, which show the weighting changed the baseline means on number of children, welfare receipt, education, income, and especially baseline employment. Appendix Table A5 also shows that the weighting has not changed the covariate balance between the experimental and control groups within the MTO experimental estimate sample.

align the disadvantage index in the PSID and MTO nonexperimental samples with that in the MTO experimental sample (Rows F1 and F2 in Appendix Table A5 show that this procedure was successful at aligning the gap on the disadvantage index). Rows F1 and F2 in Table 3 show that this did not appreciably change the estimates in either sample. To the extent that racial isolation may be particularly important for outcomes and varies in a way that is different within MTO compared to national samples, matching on concentrated disadvantage overall in MTO need not match to racial isolation specifically in the PSID. That said, the effect on racial isolation in the MTO experimental versus MTO non-experimental analyses are quite similar (with a percentage point change in tract share black of -0.080 versus -0.069), yet the impact on adult labor market outcomes are quite different and the MTO non-experimental analysis remains similar to what we see in the PSID. So we conclude that the neighborhood measures hypothesis seems unlikely to explain divergent effect estimates.

The treatment duration hypothesis argues that the MTO experiment produced a change in the duration of exposure to low poverty neighborhoods that was too small to lead to improved outcomes for the MTO adults. To examine this hypothesis, we added the number of years in a low poverty neighborhood to the weighting procedure to align that measure in the PSID and MTO nonexperimental samples with that in the MTO experimental sample. However, this procedure fully aligned the gap only in the MTO nonexperimental sample (See rows G1 and G2 in Appendix Table A5). Doing so produces only a small change in the effect estimates (rows G1 and G2 in Table 3), suggesting that the treatment duration hypothesis is unlikely to explain the difference between the MTO experimental and non-experimental results. For the PSID, we cannot fully test this hypothesis because we cannot fully align the years in a low poverty neighborhood between the PSID and MTO experimental results.

Finally, the residential mobility hypothesis argues that the increased residential mobility experienced by the MTO participants who moved to better neighborhoods had the potential to counteract any gains from living in a lower-poverty neighborhood. To examine this hypothesis, we added number of residential moves between census tracts to our weighting procedure to align the number of residential moves in the PSID and MTO nonexperimental samples with that in the MTO experimental sample. The weighting procedure was able to fully align the gap only for the PSID sample (See rows H1 in Appendix Table A5). Because we still see a positive point estimate in the PSID sample (Row H1 in Table 3) even when the gaps in the number of moves between the high neighborhood poverty and low neighborhood poverty groups are the same in the MTO experimental and PSID samples, we conclude that the residential mobility hypothesis cannot explain the divergent estimates between the MTO experiment and the PSID.³⁵

DISCUSSION

While there are limits to the statistical power of our estimates, the analysis above suggests that none of the commonly offered hypotheses that seek to explain the discrepancies in the effects of improvements in neighborhood poverty on adult economic outcomes between existing experimental and non-experimental studies—including outcome measures, treatment magnitude, nonlinear effects, treatment effect heterogeneity, neighborhood measures, treatment duration, and residential mobility—can account for differences in estimates between experimental and non-experimental estimates. When we hold these factors constant, different

³⁵ Throughout our discussion of the results from the weighting procedure we have focused on aligning the gaps and levels of the means on the neighborhood poverty rate, the concentrated disadvantage index, the number of years in low-poverty neighborhoods, and the number of residential moves. Appendix Table A6 shows the 25th and 75th percentiles of the distributions of these four variables when various weighting procedures are applied. This table shows that whenever we are able to balance the means, we are also able to balance those percentiles as well.

point estimates of effects remain.³⁶ We next discuss some remaining possible explanations for the discrepancy in estimates between the MTO experimental estimates and the PSID and MTO non-experimental results.

We begin by returning to an anomalous result in the above analysis that offers a possible clue as to the source of the discrepancy between MTO experimental estimates and the observational estimates from the PSID and MTO nonexperimental analysis. Recall that when we weighted the MTO experimental sample to be similar to the more advantaged PSID sample, the point estimate from the MTO experiment changed from essentially zero (Table 3, row A1) to large and negative (Table 3, row D1). Further inspection of the data shows that this weighting increased the weight given to those employed at baseline in the analysis (although we continue to trim large weights to avoid a small number of observations unduly influencing the results).

In an exploratory analysis of treatment effect heterogeneity by baseline employment and education in the MTO and PSID (Table 4), we see suggestive—but far from definitive—evidence that among individuals working at baseline, the TOT effects on the economic index are negative (but not statistically significant). These negative effects are largest among those working at baseline who were not high school graduates, individuals who might have been most precariously employed at baseline due to their low skill levels and the types of industries and occupations available to them (on precarious employment in the low-skill labor market, see for example Kalleberg [2009]). Notably, these negative effects occur for both the low poverty and traditional voucher groups, both of whom received Section 8 vouchers. In contrast, those with high school diplomas who were not working at baseline appear to benefit economically from the

³⁶ We also examined the results of specifications including fixed effects for quintiles of the propensity score. In most cases, the results were similar to those reported here. In a few cases, the standard errors were much larger and the estimates became uninformative.

MTO intervention, although the effects are smaller in magnitude, are not statistically significant at conventional levels, and could be due to a “floor effects” problem. These results suggest that the standard MTO effect estimates could be a combination of negative effects among those working at baseline and positive effects among those not working at baseline, resulting in an average effect close to zero. In other words, it may be that the negative effects among those employed at baseline are driving any differences in results between MTO and observational studies (although these are just 25% of the MTO study sample).

It is impossible to know whether the negative economic effects of the MTO intervention we see among those employed at baseline are unique to the particular sample that participated in MTO or even whether they are unique to the time and place in which MTO was implemented, but it is worth considering how it might have come about and why we do not see similar effects in the PSID (Table 4). We see two possible explanations. The first is simply selection bias. It may be that the inability of observational studies to condition on unobserved covariates that play a role in both sorting individuals and families into low or high poverty neighborhoods and effect adult economic outcomes generates a positive bias in the PSID and MTO nonexperimental estimates among those employed at baseline. It seems likely that such selection bias is more severe for individuals employed at baseline, as such individuals may have, *ex ante*, more neighborhood options. The PSID estimates are consistent with this explanation, as positive estimates of moving to more advantaged neighborhoods are concentrated among those employed at baseline, who may have been on a better economic trajectory irrespective of residential moves to more advantaged neighborhoods. Because the MTO experimental estimates are based on a randomized experiment with a large sample, selection bias is highly unlikely to bias the MTO

estimates, even in a subgroup that is only a part of the entire sample. However, subgroup analyses are considerably noisier due to smaller sample sizes.

Another possible explanation is that the MTO voucher moves were particularly disruptive for MTO respondents who were employed at baseline. In other words, although the above analysis tries to align the number of residential moves, it may also be that there is a qualitative difference in the type of moves created by the MTO intervention and residential moves experienced by PSID respondents or the MTO control group. Although residential moves by poor families are often involuntary and disruptive (Desmond 2016), moving with an MTO voucher puts particular constraints on the timeframe for moving without losing the voucher, and the low-poverty voucher restricts the neighborhoods in which the voucher can be used.

Qualitative evidence from MTO and other studies of housing voucher programs suggest that “leasing up” with a voucher can be difficult because many landlords do not want to accept Section 8 vouchers. This seems to be particularly true in low-poverty neighborhoods, where landlords have more potential tenants and where housing unit sizes that can accommodate families are in shorter supply (Edin, DeLuca, and Owens 2012; Rosenblatt and DeLuca 2012; Small et al. 2013). One consequence of these constraints is that leasing up with a Section 8 voucher can be very time-consuming. Somewhat counterintuitively, those who are working may face a particularly daunting task, as they will not have the time to engage with these challenges (e.g. Pashup et al. 2005) and may already have stable childcare arrangements that allow them to work. As a result, it is possible that MTO participants working at baseline were forced to accept housing units in locations that were not otherwise ideal for maintaining their employment in the long term. This suggests one possible explanation regarding qualitative differences between moves experienced by members of the low-poverty voucher and traditional voucher groups

compared to PSID respondents: for individuals employed at baseline, their more constrained moves may be disruptive to their employment. They may be unable to lease-up in neighborhoods from which they can continue to commute to their jobs or from which they can access the social supports they need to maintain employment, such as childcare.

However, we also uncovered some evidence that was inconsistent with the disruptive voucher moves explanation. For example, one group that might provide an alternative test of the disruptive voucher moves explanation is MTO control group respondents who lived in housing developments that were scheduled for demolition, since they likely experienced voucher moves similar to those of the low-poverty voucher group. They were likely awarded Section 8 vouchers with a limited timeframe in which to move (although they were unlikely to have been required to move to low poverty neighborhoods as the low-poverty voucher group was). Yet our exploratory analysis of effects of living in a low poverty neighborhood in this subsample exhibit the largest positive coefficients among those who were employed at baseline, the opposite of what the disruptive moves explanation would predict. (We caution, however, that this group is particularly small and concentrated in one MTO city, Chicago.) Moreover, when we examined effect heterogeneity in the MTO experimental sample by the presence of children under age five in the household, who are too young to attend school and therefore might present the most challenges in terms of child care, our exploratory analysis finds no differences in point estimates. And while the disruptive voucher moves explanation would suggest that effects should be largest in the period after MTO randomization, this does not seem to be the case. Finally, for the MTO sample we are examining long-term employment outcomes measured 10-15 years after baseline, which raises the question of whether any disruptive effects of constrained moves might actually persist that long. Experimental effects on the economic index among people working at baseline were

smaller 4 to 7 years after random assignment than 10 to 15 years after random assignment. Oreopoulos (2003) studied neighborhood effects on adult earnings of people whose families were assigned to different public housing units when they were youths—a research design that was like that of MTO in many ways except that it involved a comparison between living in higher- and lower-poverty public housing among people who all moved into public housing. The fact that neither the MTO experiment nor Oreopoulos’s quasi-experiment produced evidence of economic effects when a key difference between them was that Oreopoulos’s quasi-experiment involved changes in neighborhood quality with more moving suggests that the association between neighborhood quality and moving is of limited importance. Further research on housing mobility programs might investigate the possibility of potential disruptive effects more thoroughly than we are able to with these data.

CONCLUSION

The apparent discrepancy in findings between the MTO randomized residential-mobility experiment and previous non-experimental studies of the effects of exposure to poor neighborhoods during adulthood on adult labor market outcomes represents a puzzle for social science with important implications for social policy. This paper has improved our understanding of how to reconcile these estimates by replicating the MTO experimental estimate through application of non-experimental methods to both the PSID data and the MTO data.

While our statistical analysis frequently suffers from limited statistical power, our primary contribution is to at least cast doubt on many of the hypothesized explanations for the discrepancy between MTO and non-experimental studies. Because our PSID and MTO nonexperimental point estimates are so similar to each other but different from MTO

experimental estimates, we view hypotheses about differences in outcome measurement and sample composition as unlikely. We do not find clear evidence that aligning the MTO experimental sample with the PSID and MTO non-experimental samples on observed baseline characteristics, neighborhood poverty gaps, years in high poverty neighborhoods, neighborhood poverty ranges, the concentrated disadvantage index, and number of moves across census tracts removes the discrepancies in neighborhood effects across samples. We therefore conclude that hypotheses related to effect heterogeneity, treatment magnitude, treatment duration, neighborhood effect nonlinearities, residential mobility, and neighborhood measures are unlikely to account for differences between experimental and non-experimental estimates of neighborhood effects on adult economic outcomes, recognizing again the caveat about the limits to our statistical power.

We also find suggestive evidence that null effects in the MTO experimental analyses could be due to a combination of positive effects among those not employed at baseline and negative effects among those employed at baseline. This would narrow the source of the discrepancy between MTO and PSID in particular to those employed at baseline. One possible explanation for the discrepancy between MTO experimental estimates and PSID observational estimates is that MTO voucher moves were particularly disruptive to women employed at baseline. Another is that selection bias is particularly important in this subgroup. To the extent that selection bias with respect to unobserved determinants of economic outcomes is a key explanation for why MTO and non-experimental studies diverge in their findings, future research must address selection more rigorously in order to both test theory and evaluate policy. Previous within-study comparisons in other domains of social policy suggest it may be particularly important to focus data collection on those variables associated with “treatment selection”

(Cook, Steiner, and Pohl 2009; Pohl et al. 2009). Our analyses suggest there may be particular value to focusing such data collection on the detailed factors that help determine economic outcomes and that are often hard to fully proxy with what is available in typical datasets.

Given that true randomized field experiments in the social sciences are practically difficult, expensive, and rare, our findings also suggest that there would be great value to supporting data collection built around “natural experiments” in which policy changes generate exogenous differences in neighborhood environments across similar families. Examples include housing program lotteries to allocate scarce program slots (Jacob and Ludwig 2012; Jacob, Kapustin, and Ludwig 2015), public housing demolitions (Chyn 2018, Jacob 2004), and quasi-random relocation of people displaced by war or natural disasters (Kirk 2009). In principle this could also include examination of natural experiments that leave people in place but change the neighborhood around them.

To the extent that residential moves with housing choice vouchers are disruptive to employment among low-income women with children, effective housing mobility policy would require voucher program implementation that avoids such unintended consequences. Moreover, social science research would need to further understand the sources of these effects. With regard to the latter, the role of social and institutional supports for maintaining employment in an increasingly precarious and unstable low-wage labor market is an important topic of future research. So too is the geography of opportunity in the rental housing market, particularly as it intersects with child care, transportation and neighborhood quality.

One remaining question is why there are discrepancies between MTO and non-experimental studies in adult economic outcomes but not in other adult outcomes or in child outcomes. Our suggestive findings that the discrepancies for adult economic outcomes might be

concentrated among adults employed at baseline may shed some light on this question. We have argued that either selection bias among this subsample of adults in non-experimental studies or the particularly disruptive effects of housing voucher moves on employment in this subsample in MTO may be responsible for the discrepancies. Both of these explanations should lead us to think more carefully about baseline characteristics related to employment, which is likely to be closely related to both selection into neighborhoods and to future adult economic outcomes. Because baseline adult employment is likely to be more closely related to future adult economic outcomes than to either child outcomes or adult health outcomes, both the selection bias and the voucher move disruption explanations should be more important for analyses of adult economic outcomes.

We conclude by reminding the reader of the limitations of this study. First, we cannot account for possible differences between MTO and PSID introduced by differential selection into those studies based on unobserved characteristics, including the decision to volunteer for the MTO randomization or the decision to participate in the PSID survey. Similarly, we cannot account for effect heterogeneity by unobserved characteristics that are independent of observed characteristics. However, the general similarity in non-experimental estimates that we see in MTO and the PSID suggest this may not be a major source of bias in practice. Second and more importantly, the sample sizes available in both the MTO and PSID datasets limit our ability to precisely determine when and under what conditions experimental estimates diverge from those derived using non-experimental approaches. In addition, with only five demonstration cities in MTO, we do not have the statistical power to understand how neighborhood effects vary across different parts of the country. Third, this research focuses on people moving among neighborhoods in the upper half of the distribution of tract poverty rates and does not examine

effects that might occur in the lower half. Fourth, given the nature of the MTO data it is also not possible to fully explore the role played by the neighborhoods that surrounds one's own neighborhood. Previous research suggests that there may be important differences across race and ethnic groups in the U.S. in the characteristics of neighborhoods around where one lives, even conditional on one's immediate neighborhood characteristics, and raises the possibility that those contiguous neighborhoods could independently exert influence over people's life outcomes (Crowder and South 2011; Sharkey 2014). Finally, MTO by its nature is not well-suited to understanding the effects of having neighborhoods differ for very long amounts of time, including across generations (Sharkey 2013).

APPENDIX A: FURTHER INFORMATION ON MODELS

Unweighted Nonexperimental Analyses

For our unweighted non-experimental regression models, we generate low and high poverty groups within each sample such that the contrast in average neighborhood poverty between the high poverty and low poverty groups is the same as the amount by which the MTO low poverty voucher is estimated to have reduced the neighborhood poverty of voucher users (the TOT effect of a voucher move on neighborhood poverty). As a first cut of each sample, we divide by whether respondents' average neighborhood poverty is above or below 25 percent. For the MTO sample, the average neighborhood poverty (for the combined controls and traditional voucher group is 17.7% after limiting to those with less than 25 percent poverty). In combination, that average for the control group and traditional voucher group members with neighborhood poverty above 25% is 41.3% which would produce a contrast (23.6%) that is larger than the MTO TOT of 16.9%. To produce a more comparable contrast, we start with respondents having around 25% percent neighborhood poverty and incrementally (by 0.5 percentage points) expand the pool to encompass higher and higher levels of poverty (e.g., 25 to <25.5%, 25.5 to <26%, 26 to < 26.5%, etc.) until we achieve a gap that matches the MTO TOT of about 17 percentage points. We follow a similar method for the PSID group except that rather than include everyone with neighborhood poverty below 25 percent, we construct a low poverty group that has an average poverty of 15 percent so that it is more similar to the MTO nonexperimental low poverty group (mean of 20 percent) but not so high that we could not construct a high poverty group that achieves the desired gap. To construct the low poverty sample, we incrementally expand the inclusion criteria for respondents in the low poverty group starting at the 25 percent cutoff and include respondents with lower neighborhood poverty

(24.5% to 25%, 24.5% to 24%, 24% to 23.5%, etc.) until we reach a mean of 15 percent. For the PSID high poverty group, we start with those respondents with 25 percent average neighborhood poverty and continuing adding respondents (by 0.5% poverty increments) until we achieve the desired gap (e.g., the MTO TOT).

For each sample (MTO or PSID), we stack the high and low poverty groups and regress the outcome of interest on an indicator for the low poverty group controlling for a set of baseline covariates. The coefficient on the low poverty indicator provides an estimate of the effect of exposure to low poverty neighborhoods. We then compare this estimate to the MTO TOT. In this example we have focused on the average poverty neighborhood measures. A parallel approach was used to construct groups using other characteristics such as concentrated disadvantage.

Weighted Nonexperimental Analyses

We use propensity score models in three steps: 1) to trim the sample to drop PSID records that are the most unlike the MTO sample, 2) to weight the low poverty groups to look more similar on basic demographics to a common reference (the trimmed PSID high poverty subsample), and 3) to simultaneously weight each trimmed high poverty subgroup to look like its low poverty counterpart on basic demographics and to produce the desired contrast on one or more specific neighborhood characteristics. In the first step, we stack the MTO low-poverty voucher group and PSID high neighborhood poverty group and use a boosted CART regression to predict being in the PSID high poverty group versus the MTO low-poverty voucher group. The algorithm takes as input all covariates available and constructs the propensity scores based on the covariates and functional form that provide the most accurate prediction of the propensity score. We use that model to generate propensity scores for the larger PSID and MTO samples

and to drop those that are least likely to be in MTO (e.g., propensity scores of greater than 0.9, drawing on the rule of thumb developed by Crump et al. 2009). This trimming reduces the PSID female sample from 4299 to 1102 respondents and also drops a small number of MTO participants. (Appendix Figure A1 shows the distributions of the propensity scores pre-trimming.) All the weighted analyses use this same trimmed data, regardless of neighborhood characteristics. Trimming the data helps reduce problems with extreme weights created by propensity scores close to zero or one. In the second step, we construct a series of low poverty groups based on specific cut points and desired gaps in the neighborhood measure and we weight this low poverty group to look like the PSID high poverty group. The initial low poverty group consists of everyone with poverty below the 25% threshold level but excludes anyone whose neighborhood poverty is below the threshold by more than the size of the desired gap. For example, if the threshold is 25% poverty and the gap in poverty is 17%, we limit the low poverty group to those with neighborhood poverty rates of 8% to <25%. We then set the target poverty (or another neighborhood characteristic) by adding the value of the target gap to the actual value. For example, a person with a poverty rate of 10% would have a target poverty rate of $10\% + 17\% = 27\%$. High poverty groups are constructed using everyone above the poverty threshold and then weighting them (using a boosted CART regression) so that they look similar to the low poverty group on both basic demographics and the target poverty rate. This produces a gap between the weighted high and low poverty groups that is similar to the target gap and also produces samples that look similar on basic demographics across weighted analyses. (Appendix Figure A2 shows the distributions of the weights post-trimming and after weighting the high poverty group to look like the low poverty group).

Weighted Experimental TOT

We estimate a TOT for a sample that is more similar demographically to the PSID by weighting the low-poverty voucher and control groups to look like the PSID high poverty group on baseline characteristics.

APPENDIX B: EFFECTS ON MENTAL AND PHYSICAL HEALTH

In this appendix we present parallel effect estimates for the mental and physical health indices. The physical health index includes three components: fair or poor health, asthma, and obesity. Fair or poor health is a self-rated health measure capturing “fair” or “poor” health vs. “good”, “very good”, or “excellent” health. We measure asthma as whether the respondent had ever been told by a medical professional that she had asthma.³⁷ We define obesity as a body mass index (BMI) value of 30 or higher. BMI is from measured height and weight for MTO and from self-reports for PSID. For mental health, only one comparable measure, a short-term psychological distress index known as the K6, was available for both studies (Furukawa et al. 2003; Kessler et al. 2003). The K6 consists of six survey items (feeling sad, nervous, restless/fidgety, hopeless, like everything is an effort, and worthless) that are scored and added together to create a scale that ranges from 0 to 24 (highest distress). To make the results easier to interpret, all z-scores are defined such that higher scores indicates more beneficial outcomes.

With regard to mental health (Appendix Table B1), the MTO experimental estimate is positive and statistically significant (row A1). The PSID estimate is positive but small and not statistically significant (row B1), while the MTO nonexperimental estimate (row B2) is positive and statistically significant, but smaller than the MTO experimental estimate. The differences between these three estimates are not statistically significant. Matching on the neighborhood poverty gap does little to change the PSID or MTO nonexperimental estimates (rows C1 and C2). Weighting the MTO experimental sample to be like the PSID high poverty group increases its effect estimate slightly (row D1). This suggests that if there was any effect heterogeneity with

³⁷ The original MTO index included a past year recall measure of asthma instead of the lifetime recall version used here for consistency with the available PSID measure.

regard to mental health, the extremely disadvantaged nature of the MTO experimental sample resulted in a slightly dampened estimate for mental health. Weighting the PSID and MTO nonexperimental samples to be more similar on observed baseline characteristics to the PSID high poverty group and to align the neighborhood poverty gap with the MTO experimental gap changes the PSID estimate considerably, making it negative but not statistically significant (row E1). Consistent with the change that weighting brought about in the MTO experimental results, the weighting also increases the MTO nonexperimental estimate slightly (row E2). Further re-weighting to align concentrated disadvantage (rows F1 and F2), number of years in low poverty neighborhoods (G1 and G2), and residential mobility (H1 and H2) results in only slight changes to the estimates in all three samples.

With regard to physical health (Appendix Table B2), the MTO experimental estimate is small and positive but not statistically significant (row A1). This is also the case for the PSID estimate (row B2), while the MTO nonexperimental estimate is close to zero (row B3). Matching to reduce the neighborhood poverty gap to that of the MTO experimental sample reduces the PSID estimate only slightly (row C1) and increases the MTO nonexperimental estimate (row C2). The weighting in rows D1, E1, and E2 only changes the PSID estimate, making it somewhat larger (row E1). Weighting further to align the concentrated disadvantage (rows F1 and F2), number of years in low poverty neighborhoods (G1 and G2), and residential mobility (H1 and H2) changes the PSID estimate little, but the MTO nonexperimental estimate is more volatile. Nevertheless, once all the weights are applied, the estimates of effects on physical health are very similar in the three samples.

References

- Agodini, Roberto, and Mark Dynarski. 2004. "Are Experiments the Only Option? A Look at Dropout Prevention Programs." *Review of Economics and Statistics* 86(1):180–94. <https://doi.org/10.1162/003465304323023741>.
- Aliprantis, Dionissi and Francisca G. C. Richter. 2020. "Evidence of Neighborhood Effects from Moving to Opportunity: LATEs of Neighborhood Quality," *Review of Economics and Statistics* 102(4):633-647. https://doi.org/10.1162/rest_a_00933.
- Allard, Scott W., and Sheldon Danziger. 2003. "Proximity and Opportunity: How Residence and Race Affect the Employment of Welfare Recipients." *Housing Policy Debate* 13(4): 675–700. <https://doi.org/10.1080/10511482.2002.9521461>.
- Anderson, Elijah. 1999. *Code of the Street: Decency, Violence, and the Moral Life of the Inner City*. New York, NY: WW. Norton.
- Angrist, Josh, David Autor, Sally Hudson, and Amanda Pallais. 2015. "Evaluating Econometric Evaluations of Post-Secondary Aid." *American Economic Review* 105(5):502-07. <https://doi.org/10.1257/aer.p20151025>.
- Arceneaux, Kevin, Alan S. Gerber, and Donald P. Green. 2010. "A Cautionary Note on the Use of Matching to Estimate Causal Effects: An Empirical Example Comparing Matching Estimates to an Experimental Benchmark." *Sociological Methods & Research* 39(2): 256–82. <https://doi.org/10.1177/0049124110378098>.
- Bania, Neil, Claudia Coulton, and Laura Leete. 2003. "Public Housing Assistance, Public Transportation, and the Welfare-to-Work Transition." *Cityscape: A Journal of Policy Development and Research* 6(2): 7–44. www.huduser.gov/portal/periodicals/cityscpe/vol6num2.
- Bayer, Patrick, Stephen L. Ross, and Giorgio Topa. 2008. "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes." *Journal of Political Economy* 116(6):1150-96. <https://doi.org/10.1086/595975>.
- Bifulco, Robert. 2012. "Can Nonexperimental Estimates Replicate Estimates Based on Random Assignment in Evaluations of School Choice? A Within-Study Comparison." *Journal of Policy Analysis and Management* 31(3):729–51. <https://doi.org/10.1002/pam.20637>.
- Burdick-Will, Julia, Jens Ludwig, Stephen W. Raudenbush, Robert J. Sampson, Lisa Sanbonmatsu, and Patrick Sharkey. 2011. "Converging Evidence for Neighborhood Effects on Children's Test Scores: An Experimental, Quasi-experimental, and Observational Comparison." Pp. 255-76 in *Whither Opportunity: Rising Inequality, Schools, and Children's Life Chances*, edited by Greg J. Duncan and Richard J. Murnane. New York, NY: Russell Sage. www.russellsage.org/publications/whither-opportunity.

- Casciano, Rebecca and Douglas S. Massey. 2008. "Neighborhoods, Employment, and Welfare Use: Assessing the Influence of Neighborhood Socioeconomic Composition." *Social Science Research* 37(2):544-58. <https://doi.org/10.1016/j.ssresearch.2007.08.008>.
- Chetty, Raj, and Nathaniel Hendren. 2018a. "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects." *Quarterly Journal of Economics* 133(3):1107-62. <https://doi.org/10.1093/qje/qjy007>.
- Chetty, Raj, and Nathaniel Hendren. 2018b. "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates." *Quarterly Journal of Economics* 133(3):1163-1228. <https://doi.org/10.1093/qje/qjy006>
- Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *American Economic Review* 106(4): 855-902. <https://doi.org/10.1257/aer.20150572>.
- Chyn, Eric. 2018. "Moved to Opportunity: The Long-Run Effect of Public Housing Demolition on Labor Market Outcomes of Children." *American Economic Review* 108(10):3028-56. <https://doi.org/10.1257/aer.20161352>
- Clampet-Lundquist, Susan, and Douglas S. Massey. 2008. "Neighborhood Effects on Economic Self-Sufficiency: A Reconsideration of the Moving to Opportunity Experiment." *American Journal of Sociology* 114(1):107-43. <https://doi.org/10.1086/588740>
- Collins, William J., and Robert A. Margo. 2000. "Residential Segregation and Socioeconomic Outcomes: When did ghettos go bad?" *Economics Letters* 69(2):239-43. [https://doi.org/10.1016/S0165-1765\(00\)00300-1](https://doi.org/10.1016/S0165-1765(00)00300-1)
- Cook, Thomas D., William R. Shadish, and Vivian C. Wong. 2008. "Three Conditions Under Which Experiments and Observational Studies Produce Comparable Causal Estimates: New Findings From Within-Study Comparisons." *Journal of Policy Analysis and Management* 27(4):724-50. <https://doi.org/10.1002/pam.20375>.
- Cook, Thomas D., Peter M. Steiner, and Steffi Pohl. 2009. "Assessing How Bias Reduction is Affected by Covariate Choice, Unreliability, and Mode of Data Analysis: Results from Two Types of Within-Study Comparisons." *Multivariate Behavioral Research* 44(6):828-47. <https://doi.org/10.1080/00273170903333673>.
- Crowder, Kyle, and Scott J. South. 2011. "Spatial and Temporal Dimensions of Neighborhood Effects on High School Graduation." *Social Science Research* 40(1):87-106. <https://doi.org/10.1016/j.ssresearch.2010.04.013>.
- Crump, Richard K., V. Joseph Hotz, Guido W. Imbens, and Oscar A. Mitnik. 2009. "Dealing with Limited Overlap in Estimation of Average Treatment Effects." *Biometrika* 96(1):187-199, <https://doi.org/10.1093/biomet/asn055>.

- Cutler, David M., and Edward L. Glaeser. 1997. "Are Ghettos Good or Bad?" *Quarterly Journal of Economics* 112(3):827-72. <https://doi.org/10.1162/003355397555361>.
- Dawkins, Casey J., Qing Shen, and Thomas W. Sanchez. 2005. "Race, Space, and Unemployment Duration." *Journal of Urban Economics* 58(1):91-113. <https://doi.org/10.1016/j.jue.2005.02.001>.
- Delgadillo, Lucy, Dan Coster, and Luke Erickson. 2006. "Geography of Consumer Bankruptcies and Neighborhood Characteristics." *Consumer Interests Annual* 52:197-209. www.consumerinterests.org/cia2006.
- DeLuca, Stefanie, Greg J. Duncan, Micere Keels, and Ruby M. Mendenhall. 2010. "Gautreaux Mothers and Their Children: An Update." *Housing Policy Debate* 20(1):7-25. <https://doi.org/10.1080/10511481003599829>.
- Desmond, Matthew. 2016. *Evicted: Poverty and Profit in the American City*. New York, NY: Crown Publishers.
- Diez-Roux, Ana V. 2001. "Investigating Neighborhood and Area Effects on Health." *American Journal of Public Health* 91(11):1783-89. <https://doi.org/10.2105/ajph.91.11.1783>.
- Dong, Nianbo, and Mark W. Lipsey. 2018. "Can Propensity Score Analysis Approximate Randomized Experiments Using Pretest and Demographic Information in Pre-K Intervention Research?" *Evaluation Review*, 42(1): 34–70. <https://doi.org/10.1177/0193841X17749824>.
- Edin, Kathryn, Stefanie DeLuca, and Ann Owens. 2012. "Constrained Compliance: Solving the Puzzle of MTO's Lease-Up Rates and Why Mobility Matters." *Cityscape* 14(2):163-78. www.huduser.gov/portal/periodicals/cityscpe/vol14num2/ch7.html.
- Ellaway, A., A. Anderson, and S. Macintyre. 1997. "Does Area of Residence Affect Body Size and Shape?" *International Journal of Obesity* 21(4):304-08. <https://doi.org/10.1038/sj.ijo.0800405>.
- Ellen, I. G., & Turner, M. A. 2003. "Do Neighborhoods Matter and Why?" Pp. 313-338 in *Choosing a Better Life? Evaluating the Moving to Opportunity Social Experiment*, edited by John Goering and Judith D. Feins. Washington, D.C.: Urban Institute Press. <http://webarchive.urban.org/publications/210783.html>.
- Elliott, James R. 1999. "Social Isolation and Labor Market Insulation: Network and Neighborhood Effects on Less-Educated Urban Workers." *Sociology Quarterly* 40(2):199-216. <https://doi.org/10.1111/j.1533-8525.1999.tb00545.x>.
- Fauth, Rebecca C., Tama Leventhal, and Jeanne Brooks-Gunn. 2004. "Short-Term Effects of Moving From Public Housing in Poor to Middle-Class Neighborhoods on Low-Income, Minority Adults' Outcomes." *Social Science & Medicine* 59(11):2271-84. <https://doi.org/10.1016/j.socscimed.2004.03.020>.

- Fortson, Kenneth, Philip Gleason, Emma Kopa, and Natalya Verbitsky-Savitz. 2015. "Horseshoes, Hand Grenades, and Treatment Effects? Reassessing Whether Nonexperimental Estimators Are Biased." *Economics of Education Review* 44:100-113. <https://doi.org/10.1016/j.econedurev.2014.11.001>.
- Fraker, Thomas, and Rebecca Maynard. 1987. "The Adequacy of Comparison Group Designs For Evaluations of Employment-Related Programs." *Journal of Human Resources* 22(2): 194-227. www.jstor.org/stable/145902.
- Furukawa, T. A., R.C. Kessler, T. Slade, and G. Andrews. 2003. "The Performance of the K6 and K10 Screening Scales For Psychological Distress in the Australian National Survey of Mental Health and Well-Being." *Psychological Medicine* 33(2):357-62. <https://doi.org/10.1017/S0033291702006700>.
- Galster, George. 2017. "People Versus Place, People and Place, or More? New Directions For Housing Policy." *Housing Policy Debate* 27(2):261-5. <https://doi.org/10.1080/10511482.2016.1174432>.
- Galster, George, and Lina Hedman. 2013. "Measuring Neighbourhood Effects Non-Experimentally: How Much Do Alternative Methods Matter?" *Housing Studies* 28(3):473-98. <https://doi.org/10.1080/02673037.2013.759544>.
- Galster, George, and Anna Maria Santiago. 2017. "Do Neighborhood Effects on Low-Income Minority Children Depend on Their Age? Evidence from a Public Housing Natural Experiment." *Housing Policy Debate* 27(4):584-610. <https://doi.org/10.1080/10511482.2016.1254098>.
- Galster, George, and Patrick Sharkey. 2017. "Spatial Foundations of Inequality: An Empirical Overview and Conceptual Model." *RSF: The Russell Sage Journal of the Social Sciences* 3(2):1-33. <https://doi.org/10.7758/rsf.2017.3.2.01>.
- Goering, John, Joan Kraft, Judith D. Feins, Debra McInnis, Mary J. Holin, and Huda Elhassan. 1999. *Moving to Opportunity For Fair Housing Demonstration Program: Current Status and Initial Findings*. Washington, D.C.: U.S. Department of Housing and Urban Development. www.huduser.gov/portal/publications/fairhsg/mto.html.
- Goldsmith, Harold F., Charles E. Holzer, and Ronald W. Manderscheid. 1998. "Neighborhood Characteristics and Mental Illness." *Evaluation and Program Planning* 21(2):211-25. [https://doi.org/10.1016/S0149-7189\(98\)00012-3](https://doi.org/10.1016/S0149-7189(98)00012-3).
- Granovetter, Mark S. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78(6): 1360-80. <https://doi.org/10.1086/225469>.
- Griffin, Beth Ann, Greg Ridgeway, Andrew R. Morral, Lane F. Burgette, Craig Martin, Daniel Almirall, Rajeev Ramchand, Lisa H. Jaycox, and Daniel F. McCaffrey. 2014. *Toolkit for Weighting and Analysis of Nonequivalent Groups (TWANG) Website*. Santa Monica, CA: RAND Corporation. www.rand.org/statistics/twang.

- Gurmu, Shiferaw, Keith R. Ihlanfeldt, and William J. Smith. 2008. "Does Residential Location Matter to the Employment of TANF Recipients? Evidence From a Dynamic Discrete Choice Model With Unobserved Effects." *Journal of Urban Economics* 63(1):325-51. <https://doi.org/10.1016/j.jue.2007.02.002>.
- Hallberg, Kelly, Vivian C. Wong, and Thomas D. Cook. 2016. *Evaluating methods for selecting school-level comparisons in quasi-experimental designs: Results from a within-study comparison*. EdPolicy Works Working Paper Series No. 47. https://curry.virginia.edu/uploads/resourceLibrary/47_School_Comparisons_in_Observational_Designs.pdf.
- Harding, David J. 2010. *Living the Drama: Community, Conflict, and Culture Among Inner-City Boys*. Chicago, IL: University of Chicago Press. www.press.uchicago.edu/ucp/books/book/chicago/L/bo8380997.html.
- Holzer, Harry J. 1991. "The Spatial Mismatch Hypothesis: What Has the Evidence Shown?" *Urban Studies* 28(1):105-22. <https://doi.org/10.1080/00420989120080071>.
- Howell-Moroney, Michael. 2005. "The Geography of Opportunity and Unemployment: An Integrated Model of Residential Segregation and Spatial Mismatch." *Journal of Urban Affairs* 27(4):353-77. <https://doi.org/10.1111/j.0735-2166.2005.00234.x-i1>.
- Hoynes, Hilary W. 2000. "Local Labor Markets and Welfare Spells: Do Demand Conditions Matter?" *Review of Economics and Statistics* 82(3):351-68. <https://doi.org/10.1162/003465300558812>.
- Imbens, Guido W. 2015 "Matching approaches in practice: Three examples." *Journal of Human Resources* 50(2): 373-419. <https://doi.org/10.3368/jhr.50.2.373>.
- Jacob, Brian A. 2004. "Public Housing, Housing Vouchers, and Student Achievement: Evidence from Public Housing Demolitions in Chicago." *American Economic Review* 94(1):233-58. <https://doi.org/10.1257/000282804322970788>.
- Jacob, Brian A., Jens Ludwig. 2012. "The Effects of Housing Assistance on Labor Supply: Evidence From a Voucher Lottery." *American Economic Review* 102(1):272-304. <https://doi.org/10.1257/aer.102.1.272>.
- Jacob, Brian A., Max Kapustin, and Jens Ludwig. 2015. "The Impact of Housing Assistance on Child Outcomes: Evidence From a Randomized Housing Lottery." *Quarterly Journal of Economics* 130(1):465-506. <https://doi.org/10.1093/qje/qju030>.
- Jacob, Robin, Marie-Andree Somers, Pei Zhu, and Harold Bloom. 2016. "The Validity of the Comparative Interrupted Time Series Design for Evaluating the Effect of School-Level Interventions." *Evaluation Review* 40(3):167-98. <https://doi.org/10.1177/0193841X16663414>.

- Jargowsky, Paul A. 1997. *Poverty and Place: Ghettos, Barrios and the American City*. New York, NY: Russell Sage Foundation. www.russellsage.org/publications/poverty-and-place-0.
- Jargowsky, Paul A. 2015. *Architecture of Segregation: Civil Unrest, the Concentration of Poverty, and Public Policy*. New York, NY: The Century Foundation. <http://apps.tcf.org/architecture-of-segregation>.
- Kain, John F. 1968. "Housing Segregation, Negro Employment, and Metropolitan Decentralization." *Quarterly Journal of Economics* 82(1):175-97. <https://doi.org/10.2307/1885893>.
- Kalleberg, Arne L. 2009. "Precarious Work, Insecure Workers: Employment Relations in Transition." *American Sociological Review* 74(1):1-22. <https://doi.org/10.1177/000312240907400101>.
- Kawachi, Ichiro, and Lisa F. Berkman, editors. 2003. *Neighborhoods and Health*. New York, NY: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195138382.001.0001>.
- Kessler, Ronald C., Greg J. Duncan, Lisa A. Gennetian, Lawrence F. Katz, Jeffrey R. Kling, Nancy A. Sampson, Lisa Sanbonmatsu, Alan M. Zaslavsky, and Jens Ludwig. 2014. "Associations of Housing Mobility Interventions for Children in High Poverty Neighborhoods With Subsequent Mental Disorders During Adolescence." *Journal of the American Medical Association* 311(9):937-48. <https://doi.org/10.1001/jama.2014.607>.
- Kessler, Ronald C., Patricia Berglund, Olga Demler, Robert Jin, Doreen Koretz, Kathleen R. Merikangas, K.R., A. John Rush, Ellen E. Walters, and Philip S. Wang. 2003. "The Epidemiology of Major Depressive Disorder: Results From the National Comorbidity Survey Replication (NCS-R)." *Journal of the American Medical Association* 289(23): 3095-105. <https://doi.org/10.1001/jama.289.23.3095>.
- Kirk, David S. 2009. "A Natural Experiment on Residential Change and Recidivism: Lessons from Hurricane Katrina." *American Sociological Review* 74(3):484-505. <https://doi.org/10.1177/000312240907400308>.
- Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75(1):83-119. <https://doi.org/10.1111/j.1468-0262.2007.00733.x>.
- Kneebone, Elizabeth, Carey Nadeau, and Alan Berube. 2011. *The Re-Emergence of Concentrated Poverty: Metropolitan Trends in the 2000s*. Washington, DC: Brookings. www.brookings.edu/research/the-re-emergence-of-concentrated-poverty-metropolitan-trends-in-the-2000s.
- LaLonde, Robert J. 1986. "Evaluating the Economic Evaluations of Training Programs with Experimental Data." *American Economic Review* 76(4):604-20. www.jstor.org/stable/1806062.

- Lens, Michael C., and C. J. Gabbe. 2017. "Employment proximity and outcomes for Moving to Opportunity families." *Journal of Urban Affairs* 39(4):547-562. <https://doi.org/10.1080/07352166.2016.1255528>.
- Leventhal, Tama, and Jeanne Brooks-Gunn. 2000. "The Neighborhoods They Live In: The Effects of Neighborhood Residence on Child and Adolescent Outcomes." *Psychological Bulletin* 126(2):309-37. <https://doi.org/10.1037/0033-2909.126.2.309>.
- Ludwig, Jens, Jeffrey B. Liebman, Jeffrey R. Kling, Greg J. Duncan, Lawrence F. Katz, Ronald C. Kessler, and Lisa Sanbonmatsu. 2008. "What Can We Learn About Neighborhood Effects From the Moving to Opportunity Experiment?" *American Journal of Sociology* 114(1):144-88. <https://doi.org/10.1086/588741>.
- Ludwig, Jens, Lisa Sanbonmatsu, Lisa Gennetian, Emma Adam, Greg J. Duncan, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling, Stacey T. Lindau, Robert C. Whitaker, and Thomas W. McDade. 2011. "Neighborhoods, Obesity and Diabetes: A Randomized Social Experiment." *New England Journal of Medicine* 365(16):1509-19. <https://doi.org/10.1056/NEJMsal103216>.
- Ludwig, Jens, Greg J. Duncan, Lisa A. Gennetian, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling, and Lisa Sanbonmatsu. 2012. "Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults." *Science*, 337(6101):1505-10. <https://doi.org/10.1126/science.1224648>.
- Ludwig, Jens, Greg J. Duncan, Lisa A. Gennetian, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling, and Lisa Sanbonmatsu. 2013. "Long-Term Neighborhood Effects on Low-Income Families: Evidence From Moving to Opportunity." *American Economic Review* 103(3):226-31. <https://doi.org/10.1257/aer.103.3.226>.
- Lynn, Lawrence E., and Michael G. H. McGeary, editors. 1990. *Inner-City Poverty in the United States*. Washington, DC: National Academies Press. <https://doi.org/10.17226/1539>.
- MacIntyre, Sally, and Anne Ellaway. 2003. "Neighborhoods and Health: An Overview." Pp. 20-42 in *Neighborhoods and Health*, edited by Ichiro Kawachi and Lisa F. Berkman. New York, NY: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195138382.001.0001>.
- Massey, Douglas S. and Nancy A. Denton. 1993. *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, MA: Harvard University Press. www.hup.harvard.edu/catalog.php?isbn=9780674018211.
- McKenzie, David, Steven Stillman, and John Gibson. 2010. "How Important is Selection? Experimental vs. Non-Experimental Measures of the Income Gains From Migration." *Journal of the European Economic Association* 8(4):913-45. <https://doi.org/10.1111/j.1542-4774.2010.tb00544.x>.

- Mendenhall, Ruby, Stefanie DeLuca, and Greg Duncan. 2006. "Neighborhood Resources, Racial Segregation, and Economic Mobility: Results From the Gautreaux Program." *Social Science Research* 35(4):892-923. <https://doi.org/10.1016/j.ssresearch.2005.06.007>.
- Michalopoulos, Charles, Howard S. Bloom, and Carolyn J. Hill. 2004. "Can Propensity-Score Methods Match the Findings from a Random Assignment Evaluation of Mandatory Welfare-to-Work Programs?" *Review of Economics and Statistics* 86(1):156-79. <https://doi.org/10.1162/003465304323023732>.
- Orr, Larry, Judith D. Feins, Robin Jacob, Erik Beecroft, Lisa Sanbonmatsu, Lawrence F. Katz, Jeffrey B. Liebman, and Jeffrey R. Kling. 2003. *Moving to Opportunity Interim Impacts Evaluation*. Washington, DC: U.S. Department of Housing and Urban Development. www.huduser.gov/portal/publications/fairhsg/mtoFinal.html.
- O'Regan, Katherine M., and John M. Quigley. 1996. "Teenage Employment and the Spatial Isolation of Minority and Poverty Households." *Journal of Human Resources* 31(3):692-702. <https://doi.org/10.2307/146272>.
- Oreopoulos, Philip. 2003. "The Long-Run Consequences of Living in a Poor Neighborhood." *Quarterly Journal of Economics* 118(4):1533-75. <https://doi.org/10.1162/003355303322552865>.
- Panel Study of Income Dynamics. 2014. "Data." Retrieved July 1, 2014 (<https://psidonline.isr.umich.edu>).
- Parks, Virginia. 2004. "Access to Work: The Effects of Spatial and Social Accessibility on Unemployment for Native-Born Black and Immigrant Women in Los Angeles." *Economic Geography* 80(2):141-72. <https://doi.org/10.1111/j.1944-8287.2004.tb00305.x>.
- Pashup, Jennifer, Kathryn Edin, Greg J. Duncan, and Karen Burke. 2005. "Participation in a Residential Mobility Program From the Client's Perspective: Findings from Gautreaux Two." *Housing Policy Debate* 16(3-4): 361-92. <https://doi.org/10.1080/10511482.2005.9521550>.
- Pinto, Rodrigo. 2019. "Noncompliance as a Rational Choice: A Framework that Exploits Compromises in Social Experiments to Identify Causal Effects," unpublished manuscript, UCLA. <https://www.rodrigopinto.net>.
- Pohl, Steffi, Peter M. Steiner, Jens Eisermann, Renate Soellner, and Thomas D. Cook. 2009. "Unbiased Causal Inference From an Observational Study: Results of a Within-Study Comparison." *Educational Evaluation and Policy Analysis* 31(4):463-79. <https://doi.org/10.3102/0162373709343964>.
- Reardon, Sean F., and Kendra Bischoff. 2014. "Residential Segregation by Income, 1970-2009." Pp. 208-233 in *Diversity and Disparities: America Enters a New Century*, edited by John Logan. New York, NY: Russell Sage Foundation. www.russellsage.org/publications/diversity-and-disparities.

- Reardon, Sean F., Kendra Bischoff, Ann Owens, and Joseph B. Townsend. 2018 “Has Income Segregation Really Increased? Bias and Bias-Correction in Sample-Based Segregation Estimates.” *Demography* 55(6):2129-60. <https://doi.org/10.1007/s13524-018-0721-4>.
- Reingold, David A. Reingold, Gregg G. Van Ryzin, and Michelle Ronda. 2001. “Does Urban Public Housing Diminish the Social Capital and Labor Force Activity of Its Tenants?” *Journal of Policy Analysis and Management* 20(3):485–504. <https://doi.org/10.1002/pam.1004>.
- Rosenbaum, James E. 1995. “Changing the Geography of Opportunity by Expanding Residential Choice: Lessons from the Gautreaux Program.” *Housing Policy Debate* 6(1):231-69. <https://doi.org/10.1080/10511482.1995.9521186>.
- Rosenbaum, James E. and Susan J. Popkin. 1991. “Employment and Earnings of Low-Income Blacks Who Move to Middle-Class Suburbs.” Pp. 342-56 in *The Urban Underclass*, edited by Christopher Jencks and Paul E. Peterson. Washington, DC: Brookings. www.brookings.edu/book/the-urban-underclass.
- Rosenblatt, Peter and Stefanie DeLuca. 2012. “We Don't Live Outside, We Live in Here: Neighborhoods and Residential Mobility Decisions Among Low-Income Families.” *City and Community* 11(3):254-84. <https://doi.org/10.1111/j.1540-6040.2012.01413.x>.
- Ross, Catherine E. 2000. “Neighborhood Disadvantage and Adult Depression.” *Journal of Health and Social Behavior* 41(2):177-87. <https://doi.org/10.2307/2676304>.
- Ross, Catherine E. and John Mirowsky. 2001. “Neighborhood Disadvantage, Disorder, and Health.” *Journal of Health and Social Behavior* 42(3):258-76. <https://doi.org/10.2307/3090214>.
- Ross, Catherine E., John R. Reynolds, and Karlyn J. Geis. 2000. “The Contingent Meaning of Neighborhood Stability for Residents’ Psychological Well Being.” *American Sociological Review* 65(4):581-97. <https://doi.org/10.2307/2657384>.
- Rubinowitz, Leonard S., and James E. Rosenbaum. 2000. *Crossing the Class and Color Lines: From Public Housing to White Suburbia*. Chicago, IL: University of Chicago Press. www.press.uchicago.edu/ucp/books/book/chicago/C/bo3639719.html.
- Sampson, Robert J., Stephen W. Raudenbush, and Felton Earls. 1997. “Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy.” *Science* 277(5328):918-24. <https://doi.org/10.1126/science.277.5328.918>.
- Sampson, Robert J. 2012. *Great American City: Chicago and the Enduring Neighborhood Effect*. Chicago, IL: University of Chicago Press. www.press.uchicago.edu/ucp/books/book/chicago/G/bo5514383.html.
- Sampson, Robert J. 2008. “Moving to Inequality: Neighborhood Effects and Experiments Meet Social Structure.” *American Journal of Sociology* 114(1):189-231. <https://doi.org/10.1086/589843>.

- Sampson, Robert J., Jeffrey D. Morenoff, and Thomas Gannon-Rowley. 2002. "Assessing 'Neighborhood Effects': Social Processes and New Directions in Research." *Annual Review of Sociology* 28(1):443-478. <https://doi.org/10.1146/annurev.soc.28.110601.141114>.
- Sampson, Robert J., Patrick Sharkey, and Stephen W. Raudenbush. 2008. "Durable Effects of Concentrated Disadvantage on Verbal Ability Among African-American Children." *Proceedings of the National Academy of Sciences* 105(3):845-52. <https://doi.org/10.1073/pnas.0710189104>.
- Sanbonmatsu, Lisa, Jens Ludwig, Lawrence F. Katz, Lisa A. Gennetian, Greg J. Duncan, Ronald C. Kessler, Emma Adam, Thomas W. McDade, and Stacy T. Lindau. 2011. *Moving to Opportunity for Fair Housing Demonstration Program: Final Impacts Evaluation*. Washington, DC: U.S. Department of Housing and Urban Development. www.huduser.gov/portal/publications/pubasst/MTOFHD.html.
- Schneeweiss, Sebastian, Malcolm Maclure, Bruce Carleton, Robert J Glynn, and Jerry Avorn. 2004. "Clinical and Economic Consequences of a Reimbursement Restriction of Nebulized Respiratory Therapy in Adults: Direct Comparison of Randomised and Observational Evaluations." *British Medical Journal* 328(7439):560. <https://doi.org/10.1136/bmj.38020.698194.F6>.
- Shang, Qingyan. 2014. "Endogenous Neighborhood Effects on Welfare Participation." *Empirical Economics* 47(2):639-67. <https://doi.org/10.1007/s00181-013-0754-x>.
- Sharkey, Patrick. 2013. *Stuck in Place: Urban Neighborhoods and the End of Progress Toward Racial Equality*. Chicago, IL: University of Chicago Press. www.press.uchicago.edu/ucp/books/book/chicago/S/bo14365260.html.
- Sharkey, Patrick. 2014. "Spatial Segmentation and the Black Middle Class." *American Journal of Sociology* 119(4):903-54. <https://doi.org/10.1086/674561>.
- Sharkey, Patrick and Jacob W. Faber. 2014. "Where, When, Why and For Whom Do Residential Contexts Matter? Moving Away From Dichotomous Understanding of Neighborhood Effects." *Annual Review of Sociology* 40:559-79. <https://doi.org/10.1146/annurev-soc-071913-043350>.
- Shonkoff, Jack P., and Deborah A. Phillips, editors. 2000. *From Neurons to Neighborhoods: The Science of Early Childhood Development*. Washington, DC: National Academies Press. <https://doi.org/10.17226/9824>.
- Silver, Eric, Edward P. Mulvey, and Jeffrey W. Swanson. 2002. "Neighborhood Structural Characteristics and Mental Disorder: Faris and Dunham Revisited." *Social Science and Medicine* 55(8):1457-70. [https://doi.org/10.1016/S0277-9536\(01\)00266-0](https://doi.org/10.1016/S0277-9536(01)00266-0).
- Small, Mario L., Scott W. Allard, Stefanie DeLuca, Philip M. E. Garboden, and Peter Rosenblatt. 2013. "Segregating Shelter: How Housing Policies Shape the Residential Locations of

- Low Income Minority Families.” *Annals of the American Academy of Political and Social Science*, 647:268-99. <https://doi.org/10.1177/0002716213479310>.
- Smith, Sandra S. 2007. *Lone Pursuit: Distrust and Defensive Individualism Among the Black Poor*. New York, NY: Russell Sage Foundation.
www.jstor.org/stable/10.7758/9781610445078.
- Thompson, Mark A. 1997. “The Impact of Spatial Mismatch on Female Labor Force Participation.” *Economic Development Quarterly* 11(2):138-45.
<https://doi.org/10.1177/089124249701100203>.
- Vartanian, Thomas P. 1997. “Neighborhood Effects on AFDC Exits: Examining the Social Isolation, Relative Deprivation, and Epidemic Theories.” *Social Service Review* 71(4):548-573. <https://doi.org/10.1086/604278>.
- Vermilyea, Todd, and James A. Wilcox. 2002. “Who Is Unbanked and Why: Results from a Large, New Survey of Low-and-Moderate Income Adults.” *Proceedings of the 38th Annual Conference on Bank Structure and Competition*. Chicago, IL: Federal Reserve Bank of Chicago.
- Votruba, Mark E., and Jeffrey R. Kling. 2009. “Effects of Neighborhood Characteristics on the Mortality of Black Male Youth: Evidence From Gautreaux, Chicago.” *Social Science and Medicine* 68(5):814-23. <https://doi.org/10.1016/j.socscimed.2008.12.018>.
- Waitzman, Norman J. and Ken R. Smith. 1998. “Phantom of the Area: Poverty-Area Residence and Mortality in the United States.” *American Journal of Public Health* 88(6):973-76.
<https://doi.org/10.2105/AJPH.88.6.973>.
- Watson, Tara. 2009. “Inequality and the Measurement of Residential Segregation By Income in American Neighborhoods.” *Review of Income and Wealth* 55(3):820-44.
<https://doi.org/10.1111/j.1475-4991.2009.00346.x>.
- Weinberg, Bruce A., Patricia B. Reagan, Jeffrey J. Yankow. 2004. “Do Neighborhoods Affect Hours Worked? Evidence from Longitudinal Data.” *Journal of Labor Economics* 22(4):891-924. <https://doi.org/10.1086/423158>.
- Wilde, Elizabeth T., and Robinson Hollister. 2007. “How Close Is Close Enough? Evaluating Propensity Score Matching Using Data From a Class Size Reduction Experiment.” *Journal of Policy Analysis and Management* 26(3):455–77.
<https://doi.org/10.1002/pam.20262>.
- Wilson, William J. 1987. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. Chicago, IL: University of Chicago Press.
www.press.uchicago.edu/ucp/books/book/chicago/T/bo13375722.html.
- Wilson, William J. 1996. *When Work Disappears*. New York, NY: Knopf.
<https://scholar.harvard.edu/wwilson/publications/when-work-disappears-world-new-urban-poor>.

- Wong, Vivian C., Peter M. Steiner, and Kylie L. Anglin. 2018. "What Can Be Learned From Empirical Evaluations of Nonexperimental Methods?" *Evaluation Review* 42(2):147-75. <https://doi.org/10.1177/0193841X18776870>.
- Yen, Irene H., and George A. Kaplan 1999a. "Neighborhood Social Environment and Risk of Death: Multilevel Evidence From the Alameda County Study." *American Journal of Epidemiology* 149(10):898-907. <https://doi.org/10.1093/oxfordjournals.aje.a009733>.
- Yen, Irene H., and George A. Kaplan. 1999b. "Poverty Area Residence and Changes in Depression and Perceived Health Status: Evidence From the Alameda County Study." *International Journal of Epidemiology* 28(1):90-4. <https://doi.org/10.1093/ije/28.1.90>.

Table 1. Baseline Sample Characteristics and Neighborhood Poverty Exposures

| | MTO Females | | | PSID Females | | | | | | | | |
|---|---------------------------------|---------------------------------|------------------|--------------|---------------------------------------|---------------------------------------|---------------------------|---------------------------------------|---------------------------------------|------------------------------|---------------------------------------|---------------------------------------|
| | Low-Poverty Voucher Group | Traditional Voucher Group | Control Group | No Weights | | | PSID Longitudinal Weights | | | PSID Cross-Sectional Weights | | |
| | | | | All | Neigh- borhood Poverty < 25% | Neigh- borhood Poverty ≥ 25% | All | Neigh- borhood Poverty < 25% | Neigh- borhood Poverty ≥ 25% | All | Neigh- borhood Poverty < 25% | Neigh- borhood Poverty ≥ 25% |
| Sample Size | 1433 | 660 | 1114 | 4299 | 3722 | 577 | 4299 | 3722 | 577 | 4299 | 3722 | 577 |
| Baseline Characteristics | | | | | | | | | | | | |
| Female head of household | 0.930 | 0.957 | 0.940 | 0.324 | 0.290 | 0.551 | 0.345 | 0.332 | 0.492 | 0.271 | 0.259 | 0.441 |
| Race & ethnicity | | | | | | | | | | | | |
| African-American Non-Hispanic | 0.635 | 0.620 | 0.641 | 0.309 | 0.244 | 0.731 | 0.130 | 0.100 | 0.473 | 0.107 | 0.082 | 0.450 |
| Hispanic | 0.309 | 0.327 | 0.299 | 0.067 | 0.054 | 0.149 | 0.080 | 0.062 | 0.288 | 0.071 | 0.056 | 0.270 |
| White | 0.025 | 0.023 | 0.026 | 0.586 | 0.661 | 0.097 | 0.751 | 0.798 | 0.207 | 0.781 | 0.820 | 0.248 |
| Age (mean in years) | | | | | | | | | | | | |
| Age at baseline | 32.8 | 33.0 | 32.7 | 41.2 | 41.3 | 40.0 | 44.2 | 44.4 | 42.5 | 44.3 | 44.5 | 42.8 |
| Age in 2009 | 45.5 | 46.0 | 45.5 | 53.3 | 53.5 | 52.1 | 56.3 | 56.5 | 54.6 | 56.5 | 56.6 | 54.9 |
| Number of children (mean) | 2.42 | 2.525 | 2.482 | 1.175 | 1.087 | 1.744 | 0.883 | 0.834 | 1.446 | 0.883 | 0.845 | 1.407 |
| Graduated high school | 0.382 | 0.349 | 0.362 | 0.792 | 0.827 | 0.563 | 0.812 | 0.837 | 0.529 | 0.829 | 0.849 | 0.550 |
| Household Income | | | | | | | | | | | | |
| < 100% of Poverty Line | 0.776 | 0.805 | 0.807 | 0.188 | 0.147 | 0.454 | 0.153 | 0.130 | 0.412 | 0.126 | 0.105 | 0.409 |
| 100-200% of Poverty Line | 0.189 | 0.186 | 0.190 | 0.159 | 0.149 | 0.229 | 0.135 | 0.132 | 0.175 | 0.125 | 0.120 | 0.185 |
| ≥ 200% or Poverty Line | 0.035 | 0.010 | 0.003 | 0.653 | 0.704 | 0.317 | 0.712 | 0.738 | 0.413 | 0.749 | 0.775 | 0.406 |
| Working | 0.261 | 0.255 | 0.232 | 0.659 | 0.680 | 0.525 | 0.657 | 0.670 | 0.504 | 0.653 | 0.664 | 0.498 |
| Receiving AFDC/TANF | 0.766 | 0.743 | 0.770 | 0.053 | 0.034 | 0.177 | 0.031 | 0.023 | 0.125 | 0.022 | 0.016 | 0.107 |
| Region | | | | | | | | | | | | |
| North Central (Chicago) | 0.204 | 0.213 | 0.207 | 0.253 | 0.260 | 0.215 | 0.266 | 0.272 | 0.199 | 0.233 | 0.238 | 0.175 |
| Northeast (Boston/NYC) | 0.430 | 0.443 | 0.435 | 0.152 | 0.158 | 0.113 | 0.202 | 0.205 | 0.169 | 0.206 | 0.209 | 0.171 |
| South (Baltimore) | 0.135 | 0.143 | 0.137 | 0.414 | 0.399 | 0.513 | 0.320 | 0.316 | 0.366 | 0.342 | 0.339 | 0.393 |
| Baseline census tract | | | | | | | | | | | | |
| Share poor | 0.494 | 0.497 | 0.502 | 0.143 | 0.115 | 0.325 | 0.120 | 0.103 | 0.319 | 0.114 | 0.100 | 0.312 |
| Share African-American | 0.556 | 0.568 | 0.563 | 0.225 | 0.173 | 0.563 | 0.122 | 0.100 | 0.375 | 0.111 | 0.093 | 0.357 |
| Neighborhood Poverty Exposure 1998/1999 to 2008/2009 | | | | | | | | | | | | |
| Mean | 0.296 | 0.317 | 0.378 | 0.133 | 0.102 | 0.337 | 0.114 | 0.094 | 0.339 | 0.109 | 0.092 | 0.337 |
| Median | 0.281 | 0.311 | 0.378 | 0.105 | 0.091 | 0.317 | 0.090 | 0.083 | 0.318 | 0.086 | 0.081 | 0.318 |
| 25th Percentile | 0.177 | 0.213 | 0.275 | 0.058 | 0.052 | 0.278 | 0.051 | 0.048 | 0.283 | 0.049 | 0.046 | 0.282 |
| 75th Percentile | 0.403 | 0.409 | 0.477 | 0.183 | 0.145 | 0.379 | 0.152 | 0.131 | 0.383 | 0.144 | 0.129 | 0.377 |

Notes: The regression models controlled for age in 2009 as opposed to age at baseline (and also included a squared age term).

Table 2. MTO Effects on the Characteristics of Participants' Residential Census Tracts

| | Control Mean | Experimental vs. Control | |
|--|--------------|--------------------------|-------------------|
| | | ITT | TOT |
| <i>Duration-weighted tract characteristics (over the 10- to 15-year follow-up period)</i> | | | |
| Tract share poor | 0.400 | -0.092 (0.006) | -0.188 (0.012) |
| Tract share African-American | 0.543 | -0.034 (0.008) | -0.070 (0.017) |
| Tract share minority | 0.881 | -0.062 (0.007) | -0.127 (0.015) |
| Tract share receiving public assistance | 0.187 | -0.048 (0.004) | -0.099 (0.008) |
| Tract share female-headed households | 0.530 | -0.074 (0.006) | -0.151 (0.011) |
| Tract share unemployed | 0.196 | -0.039 (0.003) | -0.080 (0.007) |
| Tract share residents under age 18 | 0.340 | -0.024 (0.003) | -0.050 (0.005) |
| Concentrated disadvantage index | 1.855 | -0.269 (0.021) | -0.551 (0.042) |
| <i>Tract share poor</i> | | | |
| 1 year post-random assignment | 0.482 | -0.161 (0.008) | -0.330 (0.015) |
| 5 years post-random assignment | 0.408 | -0.102 (0.008) | -0.209 (0.016) |
| 10 years post-random assignment | 0.339 | -0.053 (0.007) | -0.108 (0.015) |
| <i>Number of years (over the 10- to 15-year follow-up period) in tracts with poverty rate less than...</i> | | | |
| 10% | 0.515 | 1.329 (0.103) | 2.727 (0.211) |
| 25% | 2.623 | 2.535 (0.169) | 5.200 (0.347) |

Notes: This table presents the estimated effect of the MTO experimental (low-poverty voucher) treatment on a variety of neighborhood characteristics for female adults (N=2,550 with valid Census tract characteristics data). The concentrated disadvantage index is a weighted combination of census tract percent [i] in poverty, [ii] African-American, [iii] on welfare, [iv] unemployed, [v] female-headed family households, and [vi] under age 18, with loading factors developed using 2000 Census tracts in Chicago by Sampson, Sharkey, and Raudenbush (2008). Census tract characteristics are based on Census 2000 data. The tables includes both the intent-to-treat (ITT) and treatment-on-treated (TOT) estimates from an ordinary least squares regression where the outcome is regressed on the experimental (low-poverty voucher) flag as well as the standard set of MTO covariates and where the standard MTO weight is applied. Robust standard errors are in parentheses, and all effects are significant at the p<.01 level.

Table 3. Comparing experimental and nonexperimental estimates of the effects of low poverty neighborhoods on an economic index

| | Census Tract Poverty Rates | | | Economic Index for High-Poverty | Estimated Effect of Low-Poverty Nbhd on Economic Index | | N |
|--|----------------------------|-------------------------|---------------------|---------------------------------|--|---------|------|
| | Low-Poverty Group Mean | High-Poverty Group Mean | Pov-erty Rate Diff. | | Coeff. (SE) | P-Value | |
| I. Prior to trimming and reweighting | | | | | | | |
| A. Experimental estimate | | | | | | | |
| (A1) MTO exp TOT (std wghts) | 0.199 | 0.368 | -0.169 | -0.229 | -0.012 (0.067) | 0.863 | 2543 |
| B. Nonexperimental estimates (unweighted) | | | | | | | |
| (B1) PSID nonexp estimate (< 25% poverty) | 0.102 | 0.337 | -0.235 | -0.262 | 0.101 (0.044) | 0.021 | 4299 |
| (B2) MTO nonexp est (< 25% poverty) | 0.177 | 0.413 | -0.236 | -0.369 | 0.096 (0.046) | 0.036 | 1770 |
| C. Nonexp estimate matching gap (unweighted) | | | | | | | |
| (C1) PSID nonexp same poverty gap | 0.148 | 0.319 | -0.171 | -0.250 | 0.107 (0.045) | 0.018 | 2512 |
| (C2) Within study (MTO) nonexp same pov gap | 0.199 | 0.370 | -0.171 | -0.368 | 0.053 (0.050) | 0.295 | 1393 |
| II. Trimming and reweighting to make samples more demographically similar | | | | | | | |
| D. Reweighted quasi-experimental estimate | | | | | | | |
| (D1) MTO exp TOT (rewgtd) | 0.200 | 0.369 | -0.169 | -0.111 | -0.131 (0.120) | 0.276 | 2532 |
| E. Reweighted non-experimental estimate matching poverty gap/level | | | | | | | |
| (E1) PSID nonexp same poverty gap/level (rewgtd) (< 25% poverty) | 0.171 | 0.338 | -0.167 | -0.253 | 0.112 (0.068) | 0.103 | 850 |
| (E2) MTO nonexp same poverty gap/level (rewgtd) (< 25% poverty) | 0.183 | 0.353 | -0.170 | -0.256 | 0.069 (0.069) | 0.316 | 1738 |
| F. Reweighted non-experimental matching poverty gap/level & disadv gap | | | | | | | |
| (F1) PSID nonexp same pov & disadv. gap/level | 0.171 | 0.338 | -0.166 | -0.280 | 0.121 (0.073) | 0.095 | 850 |
| (F2) MTO nonexp same pov & disadv. gap/level | 0.183 | 0.353 | -0.170 | -0.249 | 0.091 (0.068) | 0.188 | 1738 |
| G. Reweighted non-experimental matching poverty gap/level & disadv/years gap | | | | | | | |
| (G1) PSID nonexp same poverty, disadv., and years | 0.171 | 0.336 | -0.165 | -0.244 | 0.083 (0.065) | 0.200 | 850 |
| (G2) MTO nonexp same poverty, disadv., and years | 0.183 | 0.354 | -0.171 | -0.332 | 0.120 (0.077) | 0.120 | 1738 |
| H. Reweighted non-experimental matching poverty gap/level & disadv/# of moves | | | | | | | |
| (H1) PSID nonexp same poverty, disadv., and moves | 0.171 | 0.337 | -0.166 | -0.247 | 0.079 (0.065) | 0.222 | 850 |
| (H2) MTO nonexp same poverty, disadv., and moves | 0.183 | 0.366 | -0.183 | -0.314 | 0.093 (0.058) | 0.112 | 1738 |

Table 3. (continued)

Notes: This table presents the estimated effects of living in low-poverty neighborhoods on an index of economic outcomes for female adults. The economic index is an average z-score of employment and earnings measures (standardized against the PSID female sample). The MTO TOT (row A1) compares the experimental (low poverty voucher) and control groups using a regression (with the standard MTO weights) and scaling up the intent-to-treat estimate by the inverse of the compliance rate. This table's measure of neighborhood poverty exposure is constructed for the MTO sample to be parallel to the PSID so it uses only biannual addresses during the present study's follow-up time. In Row A1, the Low-Poverty Group Mean is the experimental complier mean, the Poverty Rate Diff is the TOT for the effect on the mean neighborhood poverty rate during follow-up, and the High-Poverty Group Mean is the implied experimental control complier mean (the experimental complier mean minus the TOT). The non-experimental treatment estimates are generated by first creating high- and low- neighborhood poverty subsamples. Rows B1 and B2 divide the sample into high and low groups at 25% poverty. In rows C1 and C2, the high and low poverty groups have been constructed such that the difference between them in average poverty rate is similar to the size of the MTO experimental TOT of about 17 percentage points. The nonexperimental analyses regress the outcome on an indicator for the low-poverty group as well as the best-available controls for each sample. The MTO non-experimental treatment estimates use only the Section 8 (traditional voucher) and control groups and the model includes an indicator for the Section 8 group. In Panel I, the nonexperimental estimates are unweighted. In Panel II, the samples have been made more similar to each other on basic demographics by trimming and reweighting the data.

Here is some additional detail about how we construct the high and low nonexperimental groups for rows C1 and C2 in Panel I. The sample is initially split at 25% average neighborhood poverty. The MTO average neighborhood poverty (for the combined controls and traditional voucher group) is about 17.7% among those with average neighborhood poverty rates below 25 percent. To produce a contrast comparable to the MTO TOT, we create a "high poverty" group starting with respondents near 25 percent neighborhood poverty and incrementally expanding (by 0.5 percentage points) to encompass higher neighborhood poverty rates until we achieve a contrast of about 17 percentage points between the high and low poverty groups.

Table 4. Comparison of MTO experimental TOT and PSID non-experimental effects on self-sufficiency index, by subgroup

| | MTO Sample (N = 2543) | | | PSID Sample (N = 7520) | | |
|--|---------------------------|-------------------|--------|---------------------------------|---|----------|
| | Exper Complier Mean | Exp TOT | N | Low Poverty Group Mean | Estimated Effect of Low- Poverty Nbhd on Economic Index | N |
| <u>Working at baseline</u> | | | | | | |
| Working | 0.135 | -0.117 (0.126) | 633 | 0.317 | 0.241 (.034) | *** 2845 |
| Not working | -0.362 | 0.041 (0.080) | 1910 | -0.494 | .046 (.027) | * 1957 |
| <u>African-American</u> | | | | | | |
| African-American | -0.246 | 0.021 (0.081) | 1710 | -0.025 | .124 (.028) | *** 1576 |
| Not African-American | -0.192 | -0.044 (0.123) | 833 | -0.008 | .117 (.038) | *** 3226 |
| <u>High school graduate at baseline</u> | | | | | | |
| High school graduate | -0.287 | 0.219 (0.112) | ** 977 | 0.111 | .173 (.032) | *** 3542 |
| Not a high school graduate | -0.197 | -0.135 (0.084) | 1566 | -0.363 | .073 (.029) | ** 1260 |
| <u>Child under age 5 at baseline</u> | | | | | | |
| Child under age 5 | -0.196 | 0.033 (0.093) | 1322 | | | |
| No children under age 5 | -0.265 | -0.038 (0.099) | 1221 | | | |
| <u>Interacting baseline education and work</u> | | | | | | |
| Not h.s. grad and not working | -0.338 | -0.069 (0.100) | 1245 | -0.557 | .058 (.033) | * 801 |
| Not h.s. grad and working | 0.221 | -0.322 (0.158) | ** 321 | -0.025 | .168 (.055) | *** 459 |
| High school grad but not working | -0.408 | 0.235 (0.132) | * 665 | -0.449 | .076 (.043) | * 1156 |
| High school grad and working | 0.016 | 0.153 (0.203) | 312 | 0.382 | .236 (.043) | *** 2386 |
| <u>Interacting baseline education and race</u> | | | | | | |
| Not h.s. grad and not African-American | -0.172 | -0.101 (0.150) | 569 | -0.388 | .072 (.045) | 676 |
| Not h.s. grad and African-American | -0.212 | -0.153 (0.101) | 997 | -0.335 | .077 (.038) | ** 584 |

| | | | | | | | |
|---|--------|-------------------|--------|--------|----------------|-----|------|
| High school grad and not African-American | -0.231 | 0.068 (0.210) | 264 | 0.093 | .172 (.057) | *** | 2550 |
| High school grad and African-American | -0.309 | 0.280 (0.132) | ** 713 | 0.158 | .170 (.038) | *** | 992 |
| <u>Interacting baseline work and race</u> | | | | | | | |
| Not working and not Afr. American | -0.383 | 0.076 (0.152) | 644 | -0.553 | .046 (.043) | | 1218 |
| Not working and African-American | -0.354 | 0.026 (0.094) | 1266 | -0.395 | .058 (.036) | | 739 |
| Working and not African-American | 0.244 | -0.330 (0.204) | 189 | 0.323 | .231 (.059) | *** | 2008 |
| Working and African-American | 0.073 | 0.012 (0.159) | 444 | 0.302 | .232 (.043) | *** | 837 |
| <u>Interacting baseline kids under 5 and work</u> | | | | | | | |
| No kids under 5 and not working | -0.396 | -0.044 (0.121) | 858 | | | | |
| No kids under 5 and working | 0.014 | -0.029 (0.171) | 363 | | | | |
| Kids under 5 and not working | -0.336 | 0.108 (0.107) | 1052 | | | | |
| Kids under 5 and working | 0.284 | -0.224 (0.185) | 270 | | | | |

Note: MTO ITT estimates are from an OLS model predicting the economic index on an indicator for treatment group assignment interacted with the different subgroups and controlling for baseline covariates. The MTO TOT was calculated by dividing the ITT by the compliance rate. The PSID estimates are from an OLS model interacting an indicator for average poverty below 25% with the subgroups and controlling for baseline covariates. *** p < .01; ** p < .05; * p < .1

Appendix Table A1. Correlations of Average Tract Share Poor vs. Other Components of the Concentrated Disadvantage Index

| Average Census Tract Characteristic | MTO (N=3207) | PSID | | |
|-------------------------------------|-----------------|-----------------|---------------------------|-------------------------|
| | | All (N=4299) | Low-Poverty (N = 3722) | High-Poverty (N=577) |
| Share African-American | 0.227 | 0.593 | 0.459 | 0.146 |
| Share under age 18 | 0.664 | 0.369 | 0.092 | 0.350 |
| Share on welfare | 0.851 | 0.791 | 0.685 | 0.506 |
| Share unemployed | 0.831 | 0.795 | 0.659 | 0.469 |
| Share female-headed households | 0.733 | 0.770 | 0.632 | 0.403 |

Notes: This table presents correlations between census tract poverty and five other census tract characteristics in the MTO female and PSID female overall samples as well as the PSID female sample split into high ($\geq 25\%$) and low ($< 25\%$) average poverty groups. Poverty and the other tract characteristics are the components of a concentrated disadvantage index developed by Sampson, Sharkey, and Raudenbush (2008). Census tract characteristics are based Census 2000 data.

Appendix Table A2. Baseline Sample Characteristics for Common and Sample-Specific Covariate Models

| | MTO Females | | | | | | PSID Females | | |
|---|-----------------------------------|--------------------------------------|---------------------------------------|----------------------------------|----------------------------------|----------------------|---------------------|------------------------------|---------------------------|
| | In Common Covariate Models | In MTO Sample-Specific Models | In PSID Sample-Specific Models | Low-Poverty Voucher Group | Traditional Voucher Group | Control Group | All | Nghd Poverty < 25% | Nghd Poverty ≥ 25% |
| Sample Size | | | | 1433 | 660 | 1114 | 4299 | 3722 | 577 |
| Baseline Characteristics | | | | | | | | | |
| Female head of household | X | X | X | 0.930 | 0.957 | 0.940 | 0.324 | 0.289 | 0.551 |
| Marital status | | | | | | | | | |
| Married | | X | | 0.096 | 0.082 | 0.094 | | | |
| Separated/widowed | | X | | 0.168 | 0.177 | 0.163 | | | |
| Divorced | | X | | 0.100 | 0.095 | 0.089 | | | |
| Race & ethnicity | | | | | | | | | |
| African-American Non-Hispanic | X | X | X | 0.635 | 0.620 | 0.641 | 0.308 | 0.244 | 0.731 |
| Hispanic | X | X | X | 0.309 | 0.327 | 0.299 | 0.067 | 0.054 | 0.149 |
| White | X | X | X | 0.025 | 0.023 | 0.026 | 0.586 | 0.661 | 0.097 |
| Age in 2009 (mean in years) | X | X | X | 45.5 | 46.0 | 45.5 | 53.3 | 53.5 | 52.1 |
| Number of children (mean) | X | X | X | 2.420 | 2.525 | 2.482 | 1.175 | 1.087 | 1.744 |
| Education | | | | | | | | | |
| Graduated high school | X | X | X | 0.382 | 0.349 | 0.362 | 0.792 | 0.827 | 0.563 |
| Received certificate of General Educational Development | | X | | 0.161 | 0.185 | 0.199 | | | |
| Number of years completed | | | X | | | | 12.829 | 13.069 | 11.270 |
| Household Income | | | | | | | | | |
| < 100% of Poverty Line | X | | | 0.776 | 0.805 | 0.807 | 0.188 | 0.147 | 0.454 |
| 100-200% of Poverty Line | X | | | 0.189 | 0.186 | 0.190 | 0.159 | 0.148 | 0.229 |
| ≥ 200% or Poverty Line (omitted) | | | | 0.035 | 0.010 | 0.003 | 0.654 | 0.705 | 0.317 |
| Income in dollars | | X | X | \$12,665 | \$12,640 | \$12,390 | \$63,433 | \$67,859 | \$34,295 |
| Individual earnings in dollars | | | X | | | | \$19,288 | \$20,401 | \$11,954 |
| Government assistance | | | | | | | | | |
| Receiving AFDC/TANF | X | X | X | 0.766 | 0.743 | 0.770 | 0.053 | 0.034 | 0.177 |
| Receiving Supplemental Security Income | | X | | 0.160 | 0.169 | 0.162 | | | |

Appendix Table A2. (continued)

| | In Common Covariate Models | In MTO Sample- Specific Models | In PSID Sample- Specific Models | MTO Females | | | | | |
|---|-------------------------------------|---|--|-----------------------------|---------------------------------|------------------|-------------|--------------------------|--------------------------|
| | | | | PSID Females | | | Low- All | Nghd Poverty < 25% | Nghd Poverty ≥ 25% |
| | | | | Poverty Voucher Group | Traditional Voucher Group | Control Group | | | |
| Employment | | | | | | | | | |
| Working (full- or part-time) | X | | X | 0.261 | 0.255 | 0.232 | 0.659 | 0.680 | 0.525 |
| Working full-time | | X | | 0.139 | 0.146 | 0.137 | | | |
| Working part-time | | X | | 0.122 | 0.109 | 0.095 | | | |
| Number of months worked in past year | | X | | 3.430 | 3.331 | 3.042 | | | |
| Fair/poor self-rated health | | | X | | | | 0.117 | 0.105 | 0.185 |
| Region | | | | | | | | | |
| North Central (Chicago) | X | X | X | 0.205 | 0.213 | 0.207 | 0.253 | 0.260 | 0.215 |
| Northeast (Boston/NYC) | X | | X | 0.430 | 0.443 | 0.435 | 0.152 | 0.158 | 0.113 |
| South (Baltimore) | X | X | X | 0.135 | 0.143 | 0.137 | 0.413 | 0.399 | 0.513 |
| Boston | | X | | 0.201 | 0.209 | 0.202 | | | |
| Los Angeles | | reference | | 0.231 | 0.200 | 0.221 | 0.182 | 0.183 | 0.159 |
| Census tract characteristics | | | | | | | | | |
| Share poor | X | X | X | 0.494 | 0.497 | 0.502 | 0.143 | 0.115 | 0.325 |
| Share African-American Other | X | X | X | 0.556 | 0.568 | 0.563 | 0.225 | 0.173 | 0.563 |
| household characteristics | | | | | | | | | |
| Owns car | | X | | 0.185 | 0.179 | 0.163 | | | |
| Disabled household member | | X | | 0.146 | 0.168 | 0.148 | | | |
| No teens in household | | X | | 0.609 | 0.613 | 0.646 | | | |
| Renter (vs. home owner) | | | X | | | | 0.323 | 0.291 | 0.530 |
| Home value in dollars | | | X | | | | \$92,867 | \$101,861 | \$33,058 |
| Neighborhood characteristics | | | | | | | | | |
| Victimized in past 6 months | | X | | 0.431 | 0.415 | 0.414 | | | |
| Streets very unsafe at night | | X | | 0.494 | 0.522 | 0.513 | | | |
| Very dissatisfied with neighborhood | | X | | 0.480 | 0.481 | 0.468 | | | |
| No family in neighborhood | | X | | 0.642 | 0.610 | 0.637 | | | |
| No friends in neighborhood | | X | | 0.398 | 0.399 | 0.408 | | | |
| Chatted with neighbors at least once per week | | X | | 0.525 | 0.489 | 0.546 | | | |
| Very likely to report neighbor's child getting into trouble | | X | | 0.557 | 0.521 | 0.556 | | | |

Appendix Table A2. (continued)

| In Common Covariate Models | In MTO Sample-Specific Models | In PSID Sample-Specific Models | MTO Females | | | PSID Females | |
|--|-------------------------------|--------------------------------|---------------------------|---------------------------|---------------|--------------|--------------------|
| | | | Low-Poverty Voucher Group | Traditional Voucher Group | Control Group | All | Nghd Poverty < 25% |
| Very confident about finding new apartment | X | | 0.479 | 0.505 | 0.456 | | |
| Applied for Section 8 previously | X | | 0.402 | 0.381 | 0.429 | | |
| Neighborhood mobility | | | | | | | |
| Lived in neighborhood 5 or more years | X | | 0.598 | 0.617 | 0.606 | | |
| Moved more than three times in 5 years prior to baseline | X | | 0.093 | 0.091 | 0.109 | | |
| Reason for wanting to move via MTO | | | | | | | |
| To get away from drugs and gangs | X | | 0.785 | 0.754 | 0.778 | | |
| Better schools for the children | | | 0.489 | 0.551 | 0.480 | | |
| MTO long-term survey release 1 flag | X | | 0.501 | 0.000 | 0.500 | | |

Notes: Household income, individual earnings and home value are in 2009 dollars. For PSID, household income, individual earnings, and home value were top-coded at the 95th percentile value for PSID females and then standardized using the PSID female sample mean and standard deviation (the squared terms were squared after top-coding and then similarly standardized). While the z-scores were included in the regression models, we present the raw values in this table. The common covariate models also included a squared term for age. In addition, the MTO sample-specific models included a missing data flag for education level and a squared term for household income. The PSID sample-specific models included squared terms for number of years of education completed, household income, individual earnings, and home value.

Appendix Table A3. Balance Across Neighborhood Poverty Groups PRIOR TO Trimming and Propensity Score Reweighting

| | Low Poverty Groups | | | | High Poverty Groups | | | |
|---|------------------------|----------------------|-----------------------|-------------------------|---------------------|----------------------|-----------------------|-------------------------|
| | MTO Experimental Group | MTO Non-experimental | PSID Non-Experimental | Standardized difference | MTO Control | MTO Non-experimental | PSID Non-Experimental | Standardized difference |
| I. Sample Size | 1433 | 342 | 1992 | 2334 | 1110 | 1051 | 520 | 1571 |
| II. Baseline Characteristics Used for Propensity Score Model | | | | | | | | |
| Female head of household | 0.930 | 0.951 | 0.359 | 1.593 | 0.940 | 0.949 | 0.542 | 1.054 |
| Race & ethnicity | | | | | | | | |
| African-American Non-Hispanic | 0.635 | 0.569 | 0.359 | 0.434 | 0.641 | 0.612 | 0.733 | -0.26 |
| Hispanic | 0.309 | 0.353 | 0.071 | 0.731 | 0.299 | 0.337 | 0.148 | 0.451 |
| White | 0.025 | 0.023 | 0.532 | -1.380 | 0.026 | 0.029 | 0.102 | -0.3 |
| Age (mean) | 45.5 | 45.1 | 52.7 | -0.667 | 45.5 | 45.7 | 51.6 | -0.54 |
| Number of Children (mean) | 2.417 | 2.358 | 1.161 | 0.964 | 2.482 | 2.514 | 1.744 | 0.546 |
| Household Income < 100% of Poverty Line | 0.776 | 0.795 | 0.218 | 1.413 | 0.807 | 0.822 | 0.446 | 0.848 |
| Graduated high school | 0.381 | 0.413 | 0.753 | -0.734 | 0.362 | 0.334 | 0.567 | -0.482 |
| Working | 0.261 | 0.295 | 0.655 | -0.774 | 0.232 | 0.228 | 0.527 | -0.647 |
| Receiving AFDC/TANF | 0.766 | 0.723 | 0.056 | 1.871 | 0.770 | 0.779 | 0.175 | 1.519 |
| III. Additional Baseline Characteristics | | | | | | | | |
| Household Income 100-200% of Poverty Line | 0.189 | 0.205 | 0.195 | 0.026 | 0.190 | 0.170 | 0.235 | -0.161 |
| Region | | | | | | | | |
| North Central (Chicago) | 0.204 | 0.110 | 0.217 | -0.293 | 0.207 | 0.167 | 0.219 | -0.131 |
| Northeast (Boston/NYC) | 0.430 | 0.500 | 0.105 | 0.951 | 0.435 | 0.454 | 0.096 | 0.874 |
| South (Baltimore) | 0.135 | 0.214 | 0.498 | -0.620 | 0.137 | 0.120 | 0.527 | -0.966 |
| Census Tract: | | | | | | | | |
| Share poor | 0.494 | 0.434 | 0.161 | 2.319 | 0.502 | 0.481 | 0.311 | 1.371 |
| Share African-American | 0.556 | 0.519 | 0.245 | 0.851 | 0.563 | 0.518 | 0.558 | -0.114 |

Notes: This table shows sample sizes and group means for the high and low poverty groups shown in Table 3's rows A1, C1, and C2. Panel I shows group sample sizes. Panel II shows means on the basic baseline demographics. Panel III presents the means for additional baseline information including neighborhood characteristics. Panel IV shows the predicted economic index value based on the coefficients from a regression model that regresses the index on the characteristics from Panel I and Panel II combined and that uses only data from the low poverty group in the model.

Appendix Table A4. Balance Across Neighborhood Poverty Groups AFTER Trimming and Propensity Score Reweighting

| | Trimmed and Reweighted Lower Poverty Groups | | | | Trimmed and Reweighted Higher Poverty Groups | | | |
|---|---|----------------------|-----------------------|-------------------------|--|----------------------|-----------------------|-------------------------|
| | MTO Experimental Group | MTO Non-experimental | PSID Non-Experimental | Standardized difference | MTO Control | MTO Non-experimental | PSID Non-Experimental | Standardized difference |
| Sample Size | 1429 | 413 | 531 | 944 | 1103 | 1325 | 319 | 1644 |
| Sum of Weights | 404 | 361 | 410 | | 384 | 163 | 231 | |
| Baseline Characteristics Used for Propensity Score Model | | | | | | | | |
| Female head of household | 0.782 | 0.825 | 0.755 | 0.172 | 0.819 | 0.955 | 0.757 | 0.357 |
| Race & ethnicity | | | | | | | | |
| African-American Non-Hispanic | 0.735 | 0.724 | 0.723 | 0.002 | 0.739 | 0.698 | 0.739 | -0.066 |
| Hispanic | 0.211 | 0.202 | 0.224 | -0.055 | 0.213 | 0.225 | 0.221 | 0.012 |
| White | 0.018 | 0.014 | 0.028 | -0.098 | 0.018 | 0.041 | 0.022 | 0.139 |
| Age (mean) | 48.4 | 47.9 | 47.4 | 0.058 | 48.2 | 44.1 | 48.0 | -0.113 |
| Number of Children (mean) | 2.272 | 2.327 | 2.269 | 0.040 | 2.361 | 2.489 | 2.283 | 0.110 |
| Household Income < 100% of Poverty Li | 0.637 | 0.640 | 0.618 | 0.045 | 0.627 | 0.763 | 0.618 | 0.273 |
| Graduated high school | 0.504 | 0.505 | 0.535 | -0.060 | 0.480 | 0.391 | 0.509 | -0.236 |
| Working | 0.447 | 0.388 | 0.482 | -0.191 | 0.430 | 0.280 | 0.456 | -0.390 |
| Receiving AFDC/TANF | 0.347 | 0.394 | 0.313 | 0.170 | 0.367 | 0.736 | 0.330 | 1.043 |
| Additional Baseline Characteristics | | | | | | | | |
| Household Income 100-200% of Poverty Region | 0.303 | 0.353 | 0.204 | 0.331 | 0.368 | 0.232 | 0.219 | 0.040 |
| North Central (Chicago) | 0.201 | 0.139 | 0.167 | -0.049 | 0.180 | 0.347 | 0.227 | 0.322 |
| Northeast (Boston/NYC) | 0.437 | 0.440 | 0.084 | 0.885 | 0.448 | 0.348 | 0.090 | 0.755 |
| South (Baltimore) | 0.152 | 0.279 | 0.575 | -0.627 | 0.141 | 0.103 | 0.476 | -0.934 |
| Census Tract: | | | | | | | | |
| Share poor | 0.490 | 0.446 | 0.212 | 1.817 | 0.497 | 0.552 | 0.328 | 0.872 |
| Share African-American | 0.570 | 0.577 | 0.444 | 0.391 | 0.560 | 0.627 | 0.570 | 0.120 |

Notes: This table shows sample sizes and group means for the reweighted and trimmed high and low poverty groups shown in Table 3 in rows D1, G1, and G2. Panel I shows group sample sizes. Panel II shows means on the basic baseline demographics used in reweighting the data and Panel III additional baseline characteristics not used in reweighting. Panel IV shows the predicted economic index value based on the coefficients from a regression model that regresses the index on the characteristics from Panel II and Panel III combined and that uses only data from the low poverty group in the model.

Appendix Table A5. Neighborhood Averages for High and Low Poverty Groups

| | Census Tract Poverty Rates | | | National Concentrated Disadvantage Index | | | Number of Years in Low-Poverty Areas | | | Number of Census Tract Changes | | |
|--|----------------------------|--------------------|---------------|--|--------------------|---------------|--------------------------------------|--------------------|---------------|--------------------------------|--------------------|---------------|
| | Low-Poverty Group | High-Poverty Group | Diff in means | Low-Poverty Group | High-Poverty Group | Diff in means | Low-Poverty Group | High-Poverty Group | Diff in means | Low-Poverty Group | High-Poverty Group | Diff in means |
| A. Experimental estimate | | | | | | | | | | | | |
| (A1) MTO exp TOT (std wghts) | 0.199 | 0.368 | -0.169 | 1.044 | 1.526 | -0.482 | 6.392 | 2.375 | 4.017 | 1.617 | 1.377 | 0.240 |
| B. Nonexperimental estimates (unweighted) | | | | | | | | | | | | |
| (B1) PSID nonexp (< 25% poverty) | 0.102 | 0.337 | -0.235 | 0.503 | 1.395 | -0.892 | 9.811 | 0.582 | 9.229 | 0.994 | 1.069 | -0.076 |
| (B2) MTO nonexp (< 25% poverty) | 0.177 | 0.413 | -0.236 | 1.010 | 1.646 | -0.636 | 7.124 | 0.712 | 6.413 | 1.545 | 1.239 | 0.307 |
| C. Nonexp estimate matching gap (unweighted) | | | | | | | | | | | | |
| (C1) PSID same pov gap | 0.148 | 0.319 | -0.171 | 0.657 | 1.350 | -0.693 | 8.877 | 0.638 | 8.239 | 1.154 | 1.140 | 0.014 |
| (C2) MTO same pov gap | 0.199 | 0.370 | -0.171 | 1.098 | 1.514 | -0.415 | 6.205 | 0.880 | 5.325 | 1.645 | 1.370 | 0.274 |
| Trimming and reweighting to make samples more demographically similar | | | | | | | | | | | | |
| D. Reweighted quasi-experimental estimate | | | | | | | | | | | | |
| (D1) MTO exp TOT (rewgtd) | 0.200 | 0.369 | -0.169 | 1.052 | 1.496 | -0.444 | 6.240 | 2.762 | 3.478 | 1.514 | 1.294 | 0.220 |
| E. Reweighted non-experimental estimate matching poverty gap/level | | | | | | | | | | | | |
| (E1) PSID same pov gap/level (rewgtd) | 0.171 | 0.338 | -0.167 | 0.878 | 1.445 | -0.567 | 7.105 | 0.580 | 6.525 | 1.668 | 1.371 | 0.296 |
| (E2) MTO same pov gap/level (rewgtd) | 0.183 | 0.353 | -0.170 | 1.096 | 1.505 | -0.409 | 7.109 | 0.930 | 6.179 | 1.443 | 1.414 | 0.030 |
| F. Reweighted non-experimental matching poverty gap/level & disadv gap | | | | | | | | | | | | |
| (F1) PSID nonexp same pov & disadv. gap/level (rewgtd) | 0.171 | 0.338 | -0.166 | 0.878 | 1.341 | -0.462 | 7.105 | 0.627 | 6.478 | 1.668 | 1.332 | 0.335 |
| (F2) MTO nonexp same pov & disadv. gap/level (rewgtd) | 0.183 | 0.353 | -0.170 | 1.096 | 1.545 | -0.449 | 7.109 | 0.937 | 6.172 | 1.443 | 1.438 | 0.005 |

G. Reweighted non-experimental matching poverty gap/level & disadv/years gap

| | | | | | | | | | | | | |
|--|-------|-------|--------|-------|-------|--------|-------|-------|-------|-------|-------|--------|
| (G1) PSID nonexp similar poverty, disadv., and years | 0.171 | 0.336 | -0.165 | 0.878 | 1.428 | -0.550 | 7.105 | 0.818 | 6.287 | 1.668 | 1.409 | 0.258 |
| (G2) MTO nonexp similar poverty, disadv., and years | 0.183 | 0.354 | -0.171 | 1.096 | 1.560 | -0.464 | 7.109 | 3.396 | 3.712 | 1.443 | 1.791 | -0.348 |

H. Reweighted non-experimental matching poverty gap/level & disadv/moves

| | | | | | | | | | | | | |
|---|-------|-------|--------|-------|-------|--------|-------|-------|-------|-------|-------|--------|
| (H1) PSID nonexp same poverty, disadv., and moves | 0.171 | 0.337 | -0.166 | 0.878 | 1.439 | -0.561 | 7.105 | 0.691 | 6.414 | 1.668 | 1.445 | 0.223 |
| (H2) MTO nonexp same poverty, disadv., and moves | 0.183 | 0.366 | -0.183 | 1.096 | 1.515 | -0.419 | 7.109 | 1.029 | 6.079 | 1.443 | 1.530 | -0.087 |

Notes: This table presents mean neighborhood characteristics for the high and low poverty groups for each model (row) shown in Table 3. Means are presented for: 1) average neighborhood poverty, 2) concentrated disadvantage index (national), and 3) years in low poverty neighborhoods (poverty < 25%). See notes to Table 3 for further information on the construction of high and low poverty groups.

Appendix Table A6. Neighborhood Characteristic Percentiles for High and Low Poverty Groups

| | Census Tract Poverty Rates | | | | National Concentrated Disadvantage Index | | | | Number of Years in Low-Poverty Areas | | | | Number of Census Tract Changes | | | |
|--|----------------------------|-----------|--------------------|-----------|--|-----------|--------------------|-----------|--------------------------------------|-----------|--------------------|-----------|--------------------------------|-----------|--------------------|-----------|
| | Low-Poverty Group | | High-Poverty Group | | Low-Poverty Group | | High-Poverty Group | | Low-Poverty Group | | High-Poverty Group | | Low-Poverty Group | | High-Poverty Group | |
| | 25th %ile | 75th %ile | 25th %ile | 75th %ile | 25th %ile | 75th %ile | 25th %ile | 75th %ile | 25th %ile | 75th %ile | 25th %ile | 75th %ile | 25th %ile | 75th %ile | 25th %ile | 75th %ile |
| I. Prior to trimming and reweighting | | | | | | | | | | | | | | | | |
| A. Experimental estimate | | | | | | | | | | | | | | | | |
| (A1) MTO exp TOT | 0.124 | 0.256 | | | 0.660 | 1.370 | | | 3 | 11 | | | 1 | 2 | | |
| B. Nonexperimental estimates (unweighted) | | | | | | | | | | | | | | | | |
| (B1) PSID (< 25% poverty) | 0.052 | 0.145 | 0.278 | 0.379 | 0.292 | 0.603 | 1.034 | 1.755 | 10 | 11 | 0 | 0 | 0 | 2 | 0 | 2 |
| (B2) MTO (< 25% poverty) | 0.141 | 0.218 | 0.323 | 0.481 | 0.707 | 1.314 | 1.321 | 1.967 | 5 | 11 | 0 | 0 | 1 | 2 | 0 | 2 |
| C. Nonexp estimate matching gap (unweighted) | | | | | | | | | | | | | | | | |
| (C1) PSID same pov gap | 0.110 | 0.183 | 0.275 | 0.352 | 0.404 | 0.830 | 1.006 | 1.712 | 8 | 11 | 0 | 1 | 0 | 2 | 0 | 2 |
| (C2) MTO same pov gap | 0.176 | 0.225 | 0.307 | 0.430 | 0.791 | 1.370 | 1.274 | 1.763 | 4 | 9 | 0 | 1 | 1 | 2 | 0 | 2 |
| Trimming and reweighting to make samples more demographically similar | | | | | | | | | | | | | | | | |
| D. Reweighted quasi-experimental estimate | | | | | | | | | | | | | | | | |
| (D1) MTO TOT (rewgtd) | 0.128 | 0.250 | | | 0.650 | 1.370 | | | 2 | 11 | | | 1 | 2 | | |
| E. Reweighted non-experimental estimate matching poverty gap/level | | | | | | | | | | | | | | | | |
| (E1) PSID nonexp | 0.130 | 0.213 | 0.299 | 0.377 | 0.623 | 1.074 | 1.136 | 1.776 | 4 | 11 | 0 | 1 | 0 | 3 | 0 | 2 |
| (E2) MTO nonexp | 0.152 | 0.219 | 0.311 | 0.385 | 0.776 | 1.379 | 1.275 | 1.757 | 4 | 11 | 0 | 1 | 1 | 2 | 1 | 2 |
| F. Reweighted non-experimental matching poverty gap/level & | | | | | | | | | | | | | | | | |
| (F1) PSID nonexp | 0.130 | 0.213 | 0.299 | 0.375 | 0.623 | 1.074 | 1.067 | 1.580 | 4 | 11 | 0 | 1 | 0 | 3 | 0 | 2 |
| (F2) MTO nonexp | 0.152 | 0.219 | 0.308 | 0.388 | 0.776 | 1.379 | 1.280 | 1.830 | 4 | 11 | 0 | 1 | 1 | 2 | 1 | 2 |
| G. Reweighted non-experimental matching poverty gap/level & disadv/years gap | | | | | | | | | | | | | | | | |
| (G1) PSID nonexp | 0.130 | 0.213 | 0.281 | 0.372 | 0.623 | 1.074 | 1.069 | 1.762 | 4 | 11 | 0 | 1 | 0 | 3 | 0 | 2 |
| (G2) MTO nonexp | 0.152 | 0.219 | 0.291 | 0.388 | 0.776 | 1.379 | 1.269 | 1.827 | 4 | 11 | 1 | 5 | 1 | 2 | 1 | 2 |

Appendix Table A6. (continued)

H. Reweighted non-experimental matching poverty gap/level & disadv/moves

| | | | | | | | | | | | | | | | | |
|-------------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|---|----|---|---|---|---|---|---|
| (H1) PSID nonexp same poverty, d | 0.130 | 0.213 | 0.283 | 0.373 | 0.623 | 1.074 | 1.070 | 1.780 | 4 | 11 | 0 | 1 | 0 | 3 | 0 | 2 |
| (H2) MTO nonexp same poverty, d | 0.152 | 0.219 | 0.305 | 0.401 | 0.776 | 1.379 | 1.240 | 1.776 | 4 | 11 | 0 | 1 | 1 | 2 | 1 | 2 |

Notes: This table presents 25th and 75th percentile for neighborhood characteristics for the high and low poverty groups for each model (row) shown in Table 3. Means are presented for: 1) average neighborhood poverty, 2) concentrated disadvantage index (national), and 3) years in low poverty neighborhoods (poverty < 25%). In the TOT rows (A1 and D1) the low poverty group consists of the experimental compliers. [The high poverty percentiles for the TOT rows are blank because we have not estimated the control compliers percentiles.] See notes to Table 3 for further information on the construction of high and low poverty groups.

Appendix Table A7. Effects of Lower Poverty Neighborhoods on Economic Outcomes Index: Raw Difference and Common Covariate Models

| | <u>Census Tract Poverty Rates</u> | | | Mean Economic Index for High-Poverty Group | <u>Estimated Effect of Low-Poverty Nbhd on Economic Index</u> | | | | N |
|---|-----------------------------------|-----------------------------------|-------------------------------|--|---|---------|------------------------------|---------|------|
| | Low- Poverty Group Mean | High- Poverty Group Mean | Poverty Rate Difference | | <u>Raw Difference (No Covariates)</u> | | <u>Common Covariates</u> | | |
| | | | | | Coeff. (SE) | P-Value | Coeff. (SE) | P-Value | |
| A. MTO experimental TOT vs. PSID & MTO nonexperimental estimates | | | | | | | | | |
| (A1) MTO exp TOT (std wghts) | 0.199 | 0.368 | -0.169 | -0.229 | 0.000 (0.073) | 0.995 | -0.034 (0.068) | 0.613 | 2543 |
| B. Nonexperimental estimates (unweighted) | | | | | | | | | |
| (B1) PSID nonexp estimate (< 20% poverty) | 0.102 | 0.337 | -0.235 | -0.262 | 0.302 (0.041) | 0.000 | 0.088 (0.045) | 0.051 | 4299 |
| (B2) MTO nonexp est (< 20% poverty) | 0.177 | 0.413 | -0.236 | -0.369 | 0.204 (0.046) | 0.000 | 0.115 (0.045) | 0.011 | 1770 |
| C. Nonexp estimate matching gap (unweighted) | | | | | | | | | |
| (C1) PSID nonexp same poverty gap | 0.148 | 0.319 | -0.171 | -0.250 | 0.182 (0.044) | 0.000 | 0.108 (0.047) | 0.021 | 2512 |
| (C2) W/in study (MTO) nonexp same pov gap | 0.199 | 0.370 | -0.171 | -0.368 | 0.128 (0.052) | 0.014 | 0.057 (0.050) | 0.255 | 1393 |
| II. Trimming and reweighting to make samples more demographically similar | | | | | | | | | |
| D. Reweighted quasi-experimental estimate | | | | | | | | | |
| (D1) MTO exp TOT (rewgtd) | 0.200 | 0.369 | -0.169 | -0.111 | -0.087 (0.147) | 0.554 | -0.104 (0.126) | 0.410 | 2532 |
| E. Reweighted non-experimental estimate matching poverty gap/level | | | | | | | | | |
| (E1) PSID nonexp same poverty gap/level (rewgtd) | 0.171 | 0.338 | -0.167 | -0.253 | 0.108 (0.075) | 0.151 | 0.111 (0.070) | 0.111 | 850 |
| (E2) MTO nonexp same poverty gap/level (rewgtd) | 0.183 | 0.353 | -0.170 | -0.256 | 0.122 (0.081) | 0.133 | 0.091 (0.074) | 0.220 | 1738 |
| F. Reweighted non-experimental matching poverty gap/level & disadv gap | | | | | | | | | |
| (F1) PSID nonexp same pov & disadv gap/level | 0.171 | 0.338 | -0.166 | -0.280 | 0.135 (0.082) | 0.100 | 0.124 (0.074) | 0.094 | 850 |
| (F2) MTO nonexp same pov & disadv gap/level | 0.183 | 0.353 | -0.170 | -0.249 | 0.115 (0.080) | 0.154 | 0.116 (0.072) | 0.108 | 1738 |

Appendix Table A7. (continued)

| | | | | | | | | | | |
|--|-------|-------|--------|--------|---------|-------|---------|-------|------|--|
| G. Reweighted non-experimental matching poverty gap/level & disadv./years gap | | | | | | | | | | |
| (G1) PSID nonexp same poverty, disadv., and y | 0.171 | 0.336 | -0.165 | -0.244 | 0.099 | 0.142 | 0.088 | 0.177 | 850 | |
| | | | | | (0.067) | | (0.065) | | | |
| (G2) MTO nonexp same poverty, disadv., and y | 0.183 | 0.354 | -0.171 | -0.332 | 0.202 | 0.023 | 0.152 | 0.053 | 1738 | |
| | | | | | (0.089) | | (0.079) | | | |
| H. Reweighted non-experimental matching poverty gap/level & disadv./# of moves | | | | | | | | | | |
| (H1) PSID nonexp same poverty, disadv., and y | 0.171 | 0.337 | -0.166 | -0.247 | 0.102 | 0.125 | 0.084 | 0.195 | 850 | |
| | | | | | (0.067) | | (0.065) | | | |
| (H2) MTO nonexp same poverty, disadv., and y | 0.183 | 0.366 | -0.183 | -0.314 | 0.179 | 0.016 | 0.119 | 0.046 | 1738 | |
| | | | | | (0.074) | | (0.060) | | | |

Notes: This table presents the estimated effect of living in low-poverty neighborhoods on an index of economic outcomes for female adults using two alternative regression models: a) a model that does not include covariates, and b) a model that includes only a limited set of covariates common to both MTO and PSID. See notes for Table 3 for additional information about the construction of the economic index and the models in each row.

Appendix Table A8. Comparing experimental and nonexperimental estimates of the effects of low poverty neighborhoods on earnings

| | Census Tract Poverty Rates | | | Earnings for High-Poverty | Estimated Effect of Low-Poverty Nbhd on Earnings (Dollars) | | |
|--|----------------------------|-------------------------|--------------------|---------------------------|--|---------|------|
| | Low-Poverty Group Mean | High-Poverty Group Mean | Poverty Rate Diff. | | Coeff. (SE) | P-Value | N |
| I. Prior to trimming and reweighting | | | | | | | |
| A. Experimental estimate | | | | | | | |
| (A1) MTO exp TOT (std wgts) | 0.199 | 0.368 | -0.169 | 12654 | 447 (1181) | 0.705 | 2523 |
| B. Nonexperimental estimates (unweighted) | | | | | | | |
| (B1) PSID nonexp estimate (< 25% poverty) | 0.102 | 0.337 | -0.235 | 14062 | 3124 (916) | 0.001 | 4121 |
| (B2) MTO nonexp est (< 25% poverty) | 0.177 | 0.413 | -0.236 | 11196 | 2818 (840) | 0.001 | 1755 |
| C. Nonexp estimate matching gap (unweighted) | | | | | | | |
| (C1) PSID nonexp same poverty gap | 0.148 | 0.319 | -0.171 | 14137 | 2916 (921) | 0.002 | 2397 |
| (C2) Within study (MTO) nonexp same pov gap | 0.199 | 0.370 | -0.171 | 11484 | 2290 (914) | 0.012 | 1381 |
| II. Trimming and reweighting to make samples more demographically similar | | | | | | | |
| D. Reweighted quasi-experimental estimate | | | | | | | |
| (D1) MTO exp TOT (rewgtd) | 0.200 | 0.369 | -0.169 | 16471 | -2700 (2143) | 0.206 | 2512 |
| E. Reweighted non-experimental estimate matching poverty gap/level | | | | | | | |
| (E1) PSID nonexp same poverty gap/level (rewgtd) (< 25% poverty) | 0.171 | 0.338 | -0.167 | 13780 | 2377 (1248) | 0.057 | 798 |
| (E2) MTO nonexp same poverty gap/level (rewgtd) (< 25% poverty) | 0.183 | 0.353 | -0.170 | 14044 | 1732 (1240) | 0.163 | 1723 |
| F. Reweighted non-experimental matching poverty gap/level & disadv gap | | | | | | | |
| (F1) PSID nonexp same pov & disadv. gap/level | 0.171 | 0.338 | -0.166 | 13780 | 2350 (1344) | 0.081 | 798 |
| (F2) MTO nonexp same pov & disadv. gap/level | 0.183 | 0.353 | -0.170 | 14100 | 2039 (1275) | 0.110 | 1723 |
| G. Reweighted non-experimental matching poverty gap/level & disadv/years gap | | | | | | | |
| (G1) PSID nonexp same poverty, disadv., and years | 0.171 | 0.336 | -0.165 | 14383 | 1550 (1227) | 0.207 | 798 |
| (G2) MTO nonexp same poverty, disadv., and years | 0.183 | 0.354 | -0.171 | 11342 | 2936 (1375) | 0.033 | 1723 |
| H. Reweighted non-experimental matching poverty gap/level & disadv/# of moves | | | | | | | |
| (H1) PSID nonexp same poverty, disadv., and moves | 0.171 | 0.337 | -0.166 | 14282 | 1474 (1221) | 0.228 | 798 |
| (H2) MTO nonexp same poverty, disadv., and moves | 0.183 | 0.366 | -0.183 | 12391 | 2440 (1064) | 0.022 | 1723 |

Notes: This table presents the estimated effect of living in low-poverty neighborhoods on the earnings for female adults for models that parallel Table 3. See notes for Table 3 for additional information about the models in each row.

Appendix Table A9. Comparing experimental and nonexperimental estimates of the effects of low poverty neighborhoods on employment

| | Census Tract Poverty Rates | | | Employment for High-Poverty | Estimated Effect of Low-Poverty Nbhd on Employment | | |
|--|----------------------------|-------------------------|--------------------|-----------------------------|--|---------|------|
| | Low-Poverty Group Mean | High-Poverty Group Mean | Poverty Rate Diff. | | Coeff. (SE) | P-Value | N |
| I. Prior to trimming and reweighting | | | | | | | |
| A. Experimental estimate | | | | | | | |
| (A1) MTO exp TOT (std wgts) | 0.199 | 0.368 | -0.169 | 0.580 | -0.020 (0.042) | 0.638 | 2547 |
| B. Nonexperimental estimates (unweighted) | | | | | | | |
| (B1) PSID nonexp estimate (< 25% poverty) | 0.102 | 0.337 | -0.235 | 0.511 | 0.031 (0.026) | 0.223 | 4298 |
| (B2) MTO nonexp est (< 25% poverty) | 0.177 | 0.413 | -0.236 | 0.487 | 0.025 (0.028) | 0.366 | 1774 |
| C. Nonexp estimate matching gap (unweighted) | | | | | | | |
| (C1) PSID nonexp same poverty gap | 0.148 | 0.319 | -0.171 | 0.521 | 0.039 (0.027) | 0.143 | 2511 |
| (C2) Within study (MTO) nonexp same pov gap | 0.199 | 0.370 | -0.171 | 0.482 | -0.001 (0.031) | 0.965 | 1397 |
| II. Trimming and reweighting to make samples more demographically similar | | | | | | | |
| D. Reweighted quasi-experimental estimate | | | | | | | |
| (D1) MTO exp TOT (rewgtd) | 0.200 | 0.369 | -0.169 | 0.608 | -0.058 (0.074) | 0.432 | 2536 |
| E. Reweighted non-experimental estimate matching poverty gap/level | | | | | | | |
| (E1) PSID nonexp same poverty gap/level (rewgtd) (< 25% poverty) | 0.171 | 0.338 | -0.167 | 0.526 | 0.049 (0.043) | 0.253 | 849 |
| (E2) MTO nonexp same poverty gap/level (rewgtd) (< 25% poverty) | 0.183 | 0.353 | -0.170 | 0.529 | 0.022 (0.041) | 0.592 | 1742 |
| F. Reweighted non-experimental matching poverty gap/level & disadv gap | | | | | | | |
| (F1) PSID nonexp same pov & disadv. gap/level | 0.171 | 0.338 | -0.166 | 0.504 | 0.060 (0.044) | 0.174 | 849 |
| (F2) MTO nonexp same pov & disadv. gap/level | 0.183 | 0.353 | -0.170 | 0.534 | 0.034 (0.040) | 0.391 | 1742 |
| G. Reweighted non-experimental matching poverty gap/level & disadv/years gap | | | | | | | |
| (G1) PSID nonexp same poverty, disadv., and years | 0.171 | 0.336 | -0.165 | 0.522 | 0.044 (0.040) | 0.280 | 849 |
| (G2) MTO nonexp same poverty, disadv., and years | 0.183 | 0.354 | -0.171 | 0.516 | 0.042 (0.048) | 0.381 | 1742 |
| H. Reweighted non-experimental matching poverty gap/level & disadv/# of moves | | | | | | | |
| (H1) PSID nonexp same poverty, disadv., and moves | 0.171 | 0.337 | -0.166 | 0.521 | 0.041 (0.041) | 0.315 | 849 |
| (H2) MTO nonexp same poverty, disadv., and moves | 0.183 | 0.366 | -0.183 | 0.511 | 0.027 (0.035) | 0.449 | 1742 |

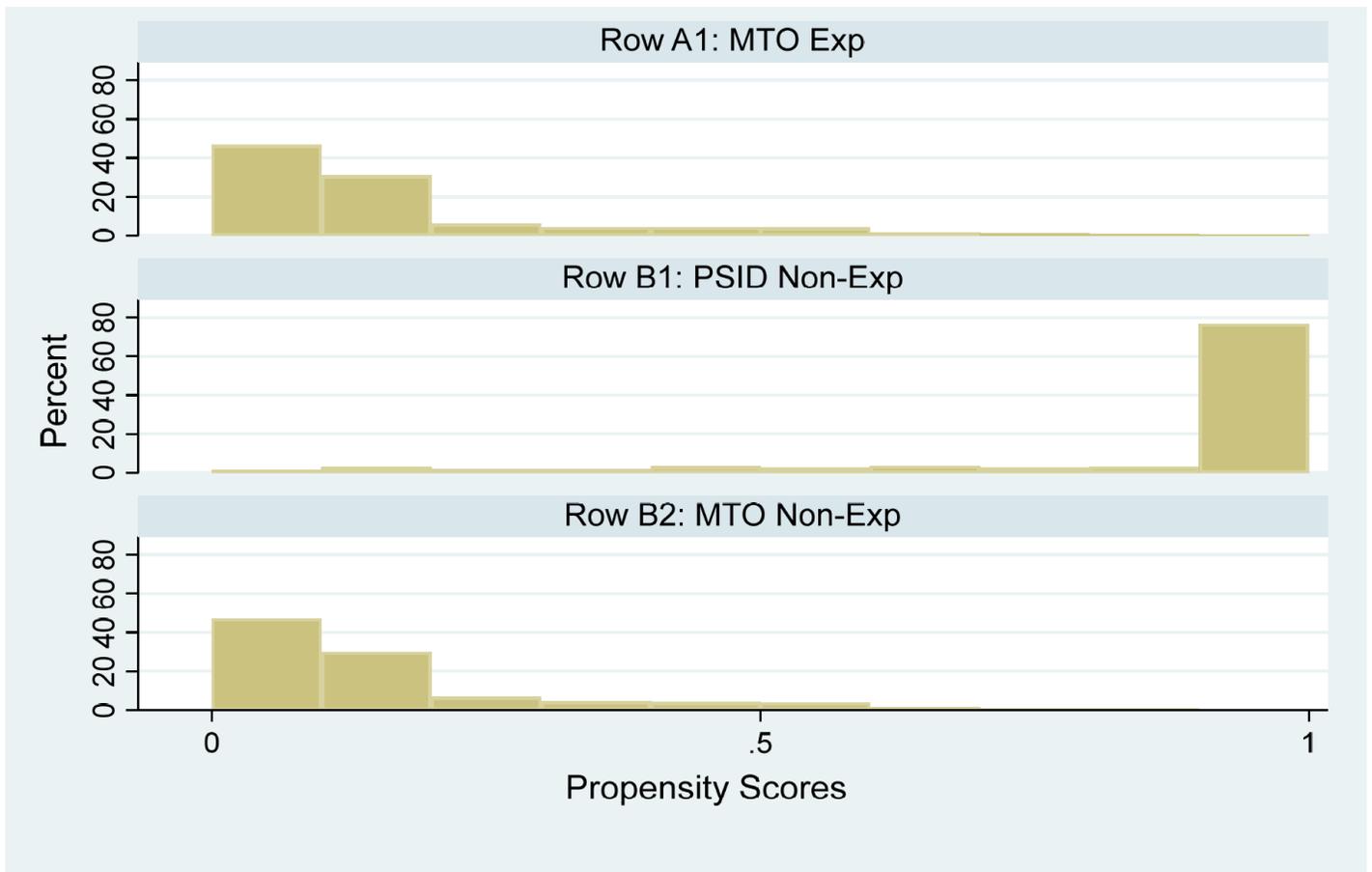
Notes: This table presents the estimated effect of living in low-poverty neighborhoods on the employment for female adults for models that parallel Table 3. See notes for Table 3 for additional information about the models in each row.

Appendix Table A10. MTO Effects on the National Percentile Ranks of Participants' Residential Census Tracts

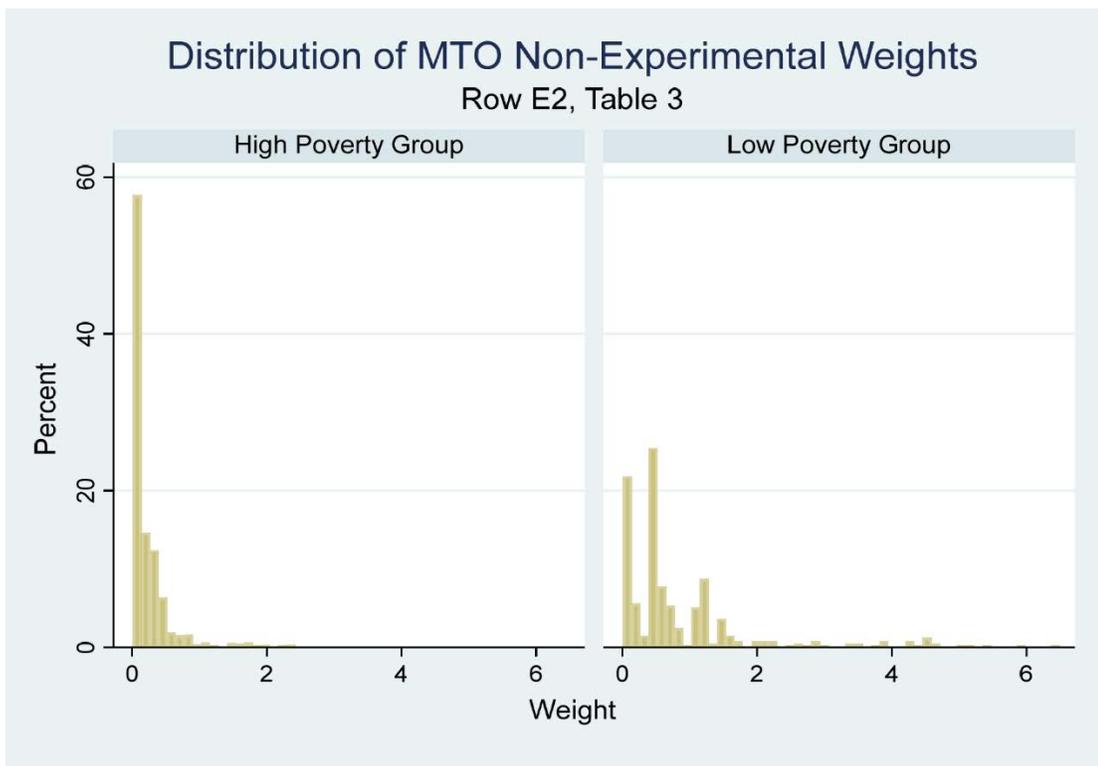
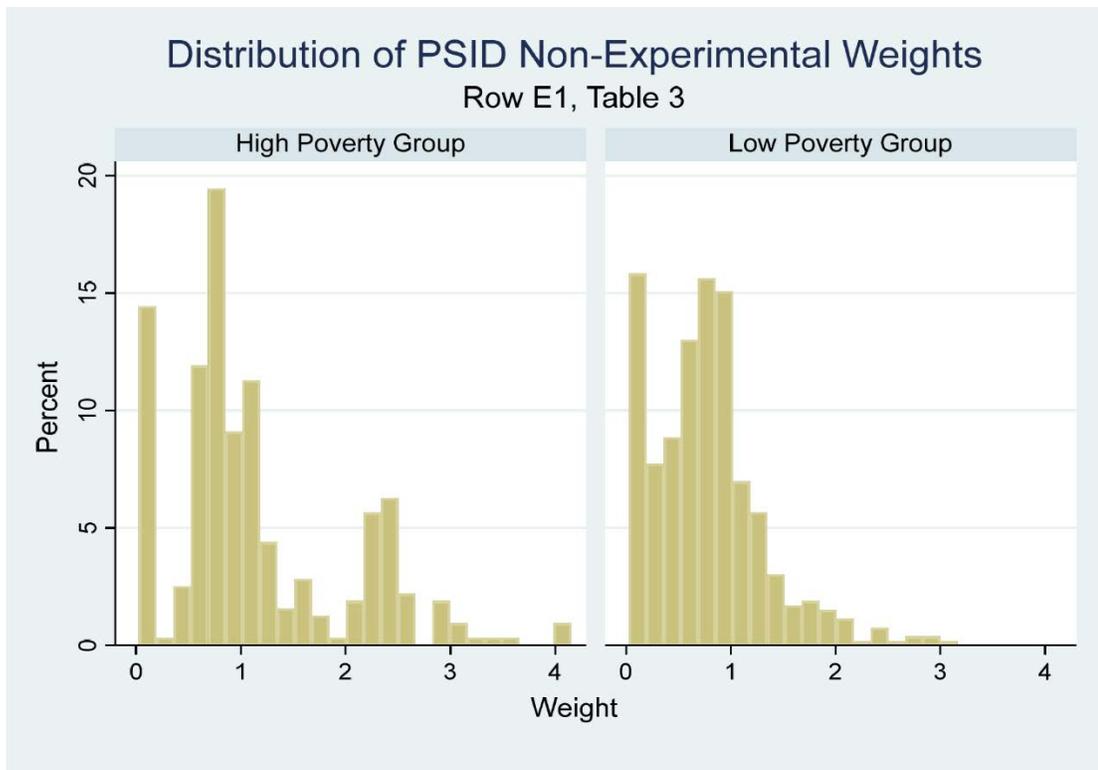
| | Control Mean | Experimental vs. Control | |
|--|-----------------|-----------------------------|-------------------|
| | | ITT | TOT |
| <i>Duration-weighted national percentile ranks of tract characteristics (over the 10- to 15-year follow-up period)</i> | | | |
| Tract share poor percentile rank | 0.887 | -0.103 (0.006) | -0.211 (0.013) |
| Tract share African-American percentile rank | 0.865 | -0.029 (0.005) | -0.059 (0.011) |
| Tract share minority percentile rank | 0.913 | -0.038 (0.004) | -0.079 (0.009) |
| Tract share receiving public assistance percentile rank | 0.896 | -0.077 (0.006) | -0.159 (0.013) |
| Tract share female-headed households percentile rank | 0.924 | -0.049 (0.005) | -0.101 (0.009) |
| Tract share unemployed percentile rank | 0.886 | -0.078 (0.006) | -0.160 (0.012) |
| Tract share residents under age 18 percentile rank | 0.793 | -0.063 (0.009) | -0.129 (0.018) |
| Concentrated disadvantage index percentile rank | 0.921 | -0.056 (0.005) | -0.116 (0.010) |
| <i>Tract share poor percentile rank</i> | | | |
| 1 year post-random assignment | 0.943 | -0.176 (0.008) | -0.361 (0.016) |
| 5 years post-random assignment | 0.897 | -0.114 (0.008) | -0.233 (0.017) |
| 10 years post-random assignment | 0.840 | -0.062 (0.009) | -0.128 (0.020) |

Notes: This table presents the estimated effect of the MTO experimental (low-poverty voucher) treatment on a variety of neighborhood characteristics for female adults (N=2,550 with valid Census tract characteristics data). The concentrated disadvantage index is a weighted combination of census tract percent [i] in poverty, [ii] African-American, [iii] on welfare, [iv] unemployed, [v] female-headed family households, and [vi] under age 18, with loading factors developed using 2000 Census tracts in Chicago by Sampson, Sharkey, and Raudenbush (2008). Census tract characteristics are based on Census 2000 data. The tables includes both the intent-to-treat (ITT) and treatment-on-treated (TOT) estimates from an ordinary least squares regression where the outcome is regressed on the experimental (low-poverty voucher) flag as well as the standard set of MTO covariates and where the standard MTO weight is applied. Robust standard errors are in parentheses, and all effects are significant at the p<.01 level.

Appendix Figure A1: Propensity Score Distributions Pre-Trimming



Appendix Figure A2: Distributions of Propensity Score Weights Post-Trimming



Appendix Table B1. Comparing experimental and nonexperimental estimates of the effects of low poverty neighborhoods on the K6 index (higher is better)

| | <u>Census Tract Poverty Rates</u> | | | Mean K6 Index for High- Poverty Group | <u>Estimated Effect of Low-Poverty Nbhd on K6 z- score</u> | | |
|--|-----------------------------------|-----------------------------------|-------------------------------|---|--|---------|------|
| | Low- Poverty Group Mean | High- Poverty Group Mean | Pov- erty Rate Diff. | | Coeff. (SE) | P-Value | N |
| I. Prior to trimming and reweighting | | | | | | | |
| A. Experimental estimate | | | | | | | |
| (A1) MTO exp TOT (std wgts) | 0.199 | 0.368 | -0.169 | -0.691 | 0.231 (0.099) | 0.019 | 2547 |
| B. Nonexperimental estimates (unweighted) | | | | | | | |
| (B1) PSID nonexp estimate (< 25% poverty) | 0.102 | 0.337 | -0.235 | -0.052 | 0.024 (0.079) | 0.759 | 1516 |
| (B2) MTO nonexp est (< 25% poverty) | 0.177 | 0.413 | -0.236 | -0.655 | 0.144 (0.064) | 0.025 | 1774 |
| C. Nonexp estimate matching gap (unweighted) | | | | | | | |
| (C1) PSID nonexp same poverty gap (< 25% poverty) | 0.148 | 0.319 | -0.171 | -0.055 | 0.042 (0.083) | 0.617 | 1065 |
| (C2) Within study (MTO) nonexp same pov gap (< 25% poverty) | 0.199 | 0.370 | -0.171 | -0.643 | 0.134 (0.072) | 0.061 | 1397 |
| II. Trimming and reweighting to make samples more demographically similar | | | | | | | |
| D. Reweighted quasi-experimental estimate | | | | | | | |
| (D1) MTO exp TOT (rewgtd) | 0.200 | 0.369 | -0.169 | -0.672 | 0.310 (0.194) | 0.110 | 2536 |
| E. Reweighted non-experimental estimate matching poverty gap/level | | | | | | | |
| (E1) PSID nonexp same poverty gap/level (rewgtd) (< 25% poverty) | 0.170 | 0.339 | -0.169 | -0.036 | -0.079 (0.126) | 0.531 | 523 |
| (E2) MTO nonexp same poverty gap/level (rewgtd) (< 25% poverty) | 0.183 | 0.353 | -0.170 | -0.543 | 0.194 (0.083) | 0.020 | 1742 |
| F. Reweighted non-experimental matching poverty gap/level & disadv gap | | | | | | | |
| (F1) PSID nonexp same pov & disadv. gap/level | 0.170 | 0.337 | -0.167 | -0.110 | -0.045 (0.135) | 0.742 | 523 |
| (F2) MTO nonexp same pov & disadv. gap/level | 0.183 | 0.353 | -0.170 | -0.539 | 0.187 (0.137) | 0.024 | 1742 |
| G. Reweighted non-experimental matching poverty gap/level & disadv/years gap | | | | | | | |
| (G1) PSID nonexp same poverty, disadv., and years | 0.170 | 0.334 | -0.164 | -0.108 | -0.071 (0.124) | 0.568 | 523 |
| (G2) MTO nonexp same poverty, disadv., and years | 0.183 | 0.354 | -0.171 | -0.681 | 0.230 (0.100) | 0.021 | 1742 |
| H. Reweighted non-experimental matching poverty gap/level & disadv/# of moves | | | | | | | |
| (H1) PSID nonexp same poverty, disadv., and moves | 0.170 | 0.331 | -0.161 | -0.101 | -0.069 (0.128) | 0.591 | 523 |
| (H2) MTO nonexp same poverty, disadv., and moves | 0.183 | 0.366 | -0.183 | -0.629 | 0.223 (0.078) | 0.004 | 1742 |

Notes: This table presents the estimated effect of living in low-poverty neighborhoods on psychological distress for female adults. The psychological distress measure is a short-term psychological distress index known as K6 (standardized against the PSID female sample). See notes for Table 3 for additional information about the models in each row. See notes to Table 3 for further information on the construction of high and low poverty groups and the meanings of each column.

Appendix Table B2. Comparing experimental and nonexperimental estimates of the effects of low poverty neighborhoods on a physical health index

| | <u>Census Tract Poverty Rates</u> | | | <u>Mean Phys Health Index for High-Poverty Group</u> | <u>Estimated Effect of Low-Poverty Nbhd on Physical Health Index</u> | | |
|--|-----------------------------------|--------------------------------|----------------------------|--|--|----------------|----------|
| | <u>Low-Poverty Group Mean</u> | <u>High-Poverty Group Mean</u> | <u>Pov-erty Rate Diff.</u> | | <u>Coeff. (SE)</u> | <u>P-Value</u> | <u>N</u> |
| I. Prior to trimming and reweighting | | | | | | | |
| A. Experimental estimate | | | | | | | |
| (A1) MTO exp TOT (std wgt) | 0.199 | 0.368 | -0.169 | -0.818 | 0.062 (0.104) | 0.550 | 2546 |
| B. Nonexperimental estimates (unweighted) | | | | | | | |
| (B1) PSID nonexp estimate (< 25% poverty) | 0.102 | 0.337 | -0.235 | -0.337 | 0.062 (0.057) | 0.279 | 4291 |
| (B2) MTO nonexp est (< 25% poverty) | 0.177 | 0.413 | -0.236 | -0.857 | 0.027 (0.068) | 0.690 | 1773 |
| C. Nonexp estimate matching gap (unweighted) | | | | | | | |
| (C1) PSID nonexp same poverty gap (< 25% poverty) | 0.148 | 0.319 | -0.171 | -0.350 | 0.053 (0.061) | 0.382 | 2508 |
| (C2) Within study (MTO) nonexp same pov gap (< 25% poverty) | 0.199 | 0.370 | -0.171 | -0.850 | 0.072 (0.074) | 0.336 | 1396 |
| II. Trimming and reweighting to make samples more demographically similar | | | | | | | |
| D. Reweighted quasi-experimental estimate | | | | | | | |
| (D1) MTO exp TOT (rewgtd) | 0.200 | 0.369 | -0.169 | -0.803 | 0.062 (0.184) | 0.736 | 2535 |
| E. Reweighted non-experimental estimate matching poverty gap/level | | | | | | | |
| (E1) PSID nonexp same poverty gap/level (rewgtd) (< 25% poverty) | 0.171 | 0.338 | -0.167 | -0.322 | 0.097 (0.088) | 0.272 | 848 |
| (E2) MTO nonexp same poverty gap/level (rewgtd) (< 25% poverty) | 0.183 | 0.353 | -0.170 | -0.794 | 0.023 (0.095) | 0.808 | 1741 |
| F. Reweighted non-experimental matching poverty gap/level & disadv gap | | | | | | | |
| (F1) PSID nonexp same pov & disadv. gap/level | 0.171 | 0.338 | -0.166 | -0.351 | 0.091 (0.088) | 0.302 | 848 |
| (F2) MTO nonexp same pov & disadv. gap/level | 0.183 | 0.353 | -0.170 | -0.759 | -0.026 (0.094) | 0.782 | 1741 |
| G. Reweighted non-experimental matching poverty gap/level & disadv/years gap | | | | | | | |
| (G1) PSID nonexp same poverty, disadv., and years | 0.171 | 0.336 | -0.165 | -0.314 | 0.081 (0.087) | 0.353 | 848 |
| (G2) MTO nonexp same poverty, disadv., and years | 0.183 | 0.354 | -0.171 | -0.848 | 0.147 (0.108) | 0.174 | 1741 |
| H. Reweighted non-experimental matching poverty gap/level & disadv/# of moves | | | | | | | |
| (H1) PSID nonexp same poverty, disadv., and moves | 0.171 | 0.337 | -0.166 | -0.311 | 0.076 (0.085) | 0.372 | 848 |

| | | | | | | | |
|--|-------|-------|--------|--------|------------------|-------|------|
| (H2) MTO nonexp same poverty, disadv., and moves | 0.183 | 0.366 | -0.183 | -0.827 | 0.051 (0.086) | 0.552 | 1741 |
|--|-------|-------|--------|--------|------------------|-------|------|

Notes: This table presents the estimated effect of living in low-poverty neighborhoods on an index of physical health for female adults. The physical health index is an average z-score of fair/poor self-rated health, obesity, and asthma measures (standardized against the PSID female sample). See notes for Table 3 for additional information about the construction of the models in each row. See notes to Table 3 for further information on the construction of high and low poverty groups and the meanings of each column.