www.sciencemag.org/cgi/content/full/337/6101/1505/DC1

Supplementary Material for

Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults

Jens Ludwig,* Greg J. Duncan, Lisa A. Gennetian, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling, Lisa Sanbonmatsu

*To whom correspondence should be addressed. E-mail: jludwig@uchicago.edu

Published 21 September 2012, *Science* **337**, 1505 (2012)
DOI: 10.1126/science.1224648

This PDF file includes:

1. Moving to Opportunity (MTO) demonstration design and study sample

The U.S. Department of Housing and Urban Development's (HUD) Moving to Opportunity (MTO) demonstration was authorized by the U.S. Congress in the Housing and Community Development Act of 1992 (for more background on MTO, see *(34, 35)*).

MTO enrolled families between 1994 and 1998 in five cities: Baltimore, Boston, Chicago, Los Angeles, and New York. To be eligible, families had to have at least one child under age 18 and live in public housing developments or project-based assisted housing in high-poverty areas, defined as a census tract in which more than 40 percent of the population was living in poverty in 1990. (Census tracts are geographic areas defined by the U.S. Census Bureau that typically contain 2,500 to 8,000 residents, with boundaries that were originally drawn to be "homogenous with respect to population characteristics, economic status, and living conditions." *(36)*) The Public Housing Authorities (PHAs) in each city conducted outreach to all eligible households through fliers, tenant associations, and other means, and all those interested received the opportunity to apply for this special program. At orientation meetings, families were told they would be randomly assigned to one of three groups if they applied. Those heads of households who remained interested after the briefing were screened for Section 8 housing voucher eligibility, completed the MTO baseline survey, signed an enrollment agreement, and then were randomly assigned to one of the three MTO program groups.

The context in which the MTO demonstration was carried out is worth keeping in mind since it is relevant for thinking about to what samples and settings these results may generalize. MTO was carried out during a time in which concentrated poverty and crime rates were declining, and HUD's HOPE VI program was demolishing many public housing projects across the country. Labor market conditions also varied considerably over the study period: Unemployment rates were relatively low during the 1990s in the period immediately following MTO enrollments. But starting in 2001 the unemployment rate increased for several years, then declined, and then has surged in recent years since the financial crisis of 2007.

A total of 4,604 eligible households enrolled in MTO, representing around one-quarter of the population of MTO-eligible families *(34, 35)*. Eligible applicants were randomly assigned to one of three groups:[*]

1. The *MTO Low-Poverty Voucher (LPV)* group received Section 8 rental assistance certificates or vouchers that they could use only in census tracts with 1990 poverty rates below 10 percent. In each city, a nonprofit organization under contract to the PHA provided mobility counseling to help low-poverty group families locate and lease suitable housing in a low-poverty area. Families who stayed in their new neighborhoods less than a year did not receive a new voucher. After one year, families were able to use their voucher to relocate without any special MTO-imposed constraints on their moves. Aside from this requirement, families assigned to the low-poverty voucher group were required to abide by all of the regular rules and requirements of the Section 8 certificate and voucher programs, including having a limited amount of time to search for housing and lease-up before they lost the rights to their subsidy, being required to contribute 30 percent of their adjusted income toward rent (the same rent requirement as in public housing), and prohibitions on rental assistance to households engaging in certain types of criminal activity.

2. The *MTO Traditional Voucher (TRV)* group received regular Section 8 certificates or vouchers that were not subject to any special location restrictions under the MTO program. These families received no special mobility counseling in MTO beyond what is usually offered by local housing authorities to housing-voucher recipients.

3. The *MTO control group* received no certificates or vouchers through MTO, but continued to be eligible for project-based assistance and whatever other social programs and services to which the families would otherwise be entitled.

Assignment rates within sites were adjusted during implementation of MTO to compensate for the fact that the lease-up rate for the MTO voucher groups turned out to be higher than had been anticipated. The sample weights used in the quantitative analyses presented in the text and below adjust for differences among sites and over time in the random assignment ratio. The weights that we use also account for the fact that for budgetary reasons we could afford to survey just a randomly-selected two-thirds of adults in the *TRV* group, and for the fact that the two-phase survey sampling approach used by our survey subcontractor randomly selected a sub-sample of remaining hard-to-reach cases for intensive follow-up during phase two after a certain target number of completed surveys had been met during phase one (see below and *(37)* for details).

In the main paper itself we focus on presenting estimates that pool together the two randomized treatment groups (the *LPV* and *TRV* groups) to improve statistical power, and because the effects of each treatment on the housing and neighborhood characteristics of MTO families converge over time. In these supplementary materials we present more detailed results for the two treatment groups pooled together, and estimates for each treatment group separately as well.

2. Methods and materials

2.1 Data sources

The HUD-sponsored evaluation included a baseline survey conducted just prior to randomization and an "interim MTO study," which gathered uniform data across all five sites and examined outcomes for MTO adults and youth at 4-7 years after random assignment *(7, 37–40)*.

More than a decade after randomization and the baseline survey, our research team was engaged by HUD to follow up with MTO families to assess a variety of outcomes. These data were collected for our research team by the Survey Research Center (SRC) of University of Michigan's Institute of Social Research from June 2008 to April 2010, on average 12.7 years after randomization (range 10.0 to 15.3). The sample frame included one adult from each family in the *LPV* and control groups, as well as youth who were living in the baseline households and were ages 10-20 at the end of 2007. For budgetary reasons, we randomly sampled two-thirds of adults in the *TRV* group, who were also interviewed a few months later, on average, than the other groups.

Our focus here is primarily on MTO adults in part because our goal is to learn more about neighborhood effects on long-term well-being, which may not be evident for adolescents.

Another reason we focus on adults is because we examine an outcome measure – self-reported subjective well-being (SWB) – that provides a comprehensive picture of how people's lives are affected and we believe is new to the neighborhood effects literature, and more is currently known about how to measure SWB for adults than for youth *(19)*.

The data collection plan for our long-term follow-up study of MTO families was reviewed and approved by the federal Office of Management and Budget and the Institutional Review Boards at HUD, the National Bureau of Economic Research, the University of Chicago, the University of Michigan, and Northwestern University.

Target respondents were traced and, when contacted, offered $50 to complete a survey about health, economic conditions, residential history, and other outcomes, drawing mostly on questions from existing national studies. (The full set of survey instruments is available at www.mtoresearch.org). Respondents were offered an additional $25 to provide physical and biological measures at the end of the survey. Written informed consent was obtained before beginning interviews.

Trained interviewers using Computer-Assisted Personal Interviewing on laptop computers administered the survey primarily in the respondent's homes, with the session scheduled at the respondent's convenience. Interviewers were blinded to MTO group assignments. After 75-80% of the sample was interviewed in the initial phase of fieldwork, a probability subsample of 35% of remaining hard-to-reach cases were selected for further recruitment efforts *(41)*. The latter interviews were up-weighted to adjust for the sub-sampling of hard-to-reach cases. A total of 3,273 adults were successfully interviewed.

To account for two-phase sampling, we calculated effective response rates *(41)*. Response rates were calculated using American Association of Public Opinion Research definition RR1w *(42)*. Specifically, the response rate calculations account for the change over time in the MTO random assignment ratios as well as the two-phase survey sampling design of the long-term evaluation. The weights equal the product of the random assignment ratio weight and the sampling weight (equal to 1 for families interviewed in Phase 1, equal to 1/0.35 for families who were randomly selected for the Phase 2 survey sample, and equal to 0 for families who were not randomly selected for the Phase 2 survey sample). The "effective response rate" (ERR) is equal to the weighted number of interviews divided by the weighted survey sample frame total minus the weighted number of decedents.

The ERR for our long-term survey of MTO adults overall was 89.6%. The ERR equaled 90.8% for the *LPV* group, 86.6% for the *TRV* group, and 90.0% for the control group. We cannot reject the null hypothesis that the ERR for the *LPV* and control groups are the same, but the *TRV* group's ERR statistically differs from that of the control group.

2.2 Measures

This section describes the key dependent and independent variables that we analyze. We begin by discussing our measures of neighborhood conditions. We then discuss the key dependent variables from our long-term surveys that go into our outcome indices for MTO adults in the

domains of mental health, physical health, and economic self-sufficiency. The elements of these outcome indices were pre-specified to the current study, based on what was constructed for the interim (5-year) MTO follow-up *(7)*. Each variable within these outcome indices is first re-scaled so that higher values equal "better" outcomes, then converted to Z-scores by subtracting the control group mean and dividing by the control group standard deviation (so each variable has a mean of zero and a standard deviation of one), then averaged across all individual outcomes within the domain, and then re-scaled again so that the index itself has a standard deviation of one. For people missing data on any element of the index, we impute the group average value of that variable, which yields estimates that are the equivalent of the average of the coefficients from estimating effects on each individual Z-scored element of the index using just those observations with non-missing values on the individual outcome variable *(7)*. Aggregating outcomes improves statistical power to detect impacts in that domain among outcomes that move in the same direction, and helps reduce the risk of "false positives" from examining large numbers of individual outcomes *(7)*. Lastly we discuss our measure of SWB.

A. Neighborhood socio-demographic composition

To measure neighborhood socio-demographic composition, HUD has tracked MTO respondents from baseline through the time of our long-term survey. Our long-term survey instrument itself also included a series of questions that reported back to respondents the data HUD collected about their residential histories and asked them to make any additions, deletions, or corrections that might be necessary. We geo-coded the address histories of MTO program participants over the 10-15 year study period, linked them to tract-level data from the 1990 and 2000 decennial censuses and the 2005-09 American Community Survey, and interpolated tract attributes for the years that fall between these Census Bureau data collections to measure tract characteristics at the time each family was living at the given address. We calculated tract characteristics for the addresses at which participants were living at baseline, 1, 5 and 10-15 years after random assignment in May, 2008 (basically the start of our long-term survey fieldwork). We also calculated duration-weighted average tract characteristics for each participant's post-baseline address history, where the tract characteristics for each address are weighted by the share of the follow-up study period the family spent at each address.

We examine census tract poverty rates, the neighborhood measure that MTO was explicitly designed to change for program participants, as well as the share of tract residents who are members of racial or ethnic minority groups. We focus primarily on duration-weighted average tract characteristics, which allow for the possibility that lagged as well as current neighborhood environments that people experience might be relevant for their outcomes *(16)*. Sensitivity analyses that use information on the characteristics of the tracts in which families reside at the time the long-term survey fieldwork was launched (May 2008) yield qualitatively similar results, as discussed in more detail below.

B. Economic self-sufficiency

Our economic self-sufficiency index is composed of the following variables: an indicator for whether the respondent is currently employed and not on Temporary Assistance to Needy Families (TANF), that is, cash welfare; an indicator for currently employed; total annual

earnings; an indicator for currently on TANF; and total annual income from government programs. As noted above, we first convert each measure so that higher values represent better economic outcomes (in this case, reversing the sign for currently on TANF and total annual income from government programs), then convert them to Z-scores and average them together, and then re-scale the average so that the overall index has a standard deviation of one. Note that including the measure for "currently employed and not on TANF" is intended to capture the possible interaction effect of absence of welfare receipt and employment above and beyond the main effects of these two measures, and follows exactly the way this index was constructed for the interim MTO study by Kling, Liebman and Katz *(7)*.

Our measures of employment status and annual earnings come from long-term MTO adult survey questions that are taken from the Current Population Survey, which are used by the Bureau of Labor Statistics to measure aggregate changes in labor market conditions. The MTO surveys also included questions about public assistance receipt and total household income for the most recent calendar year preceding the survey – 2007, 2008, or 2009, depending on their survey interview date. The MTO survey used an innovative technique for collecting retrospective reports of total household income by embedding an income calculator. Respondents were first asked about specific sources of income in the most recent calendar year for earnings, government sources, and then income from other sources. If a respondent did not know the total value of these income sources, the interviewer asked a series of yes-no questions about different income amounts to determine the income category. These income sources are then added up to calculate for the respondent a total household income level. The respondent then had the opportunity to agree with or to adjust the total household income.

C. Physical health

Our physical health index consists of the following individual measures: self-reported health is fair or poor; the respondent had an asthma attack the past year; obesity (body mass index (BMI), equal to weight in kilograms divided by height in meters squared, of 30 or higher); hypertension; and trouble carrying groceries or climbing stairs. As with the economic outcome index, each element of the index is scored so that a more positive value corresponds to better health, then converted to a Z-score using the control group's distribution. The index is the average of these Z-scores, re-scaled so that it has a standard deviation of one.

We measured self-reported health using a standard question from the health studies literature that has been used on a variety of other national surveys: "In general, how is your health: excellent, very good, good, fair, or poor?" Previous research shows that self-reported health is strongly related to life expectancy among adults *(43, 44)*. Some researchers have questioned whether self-rated health categories represent the same objective levels of health among respondents of different socio-economic status *(45, 46)*. Thus, one potential challenge in comparing self-reported health across MTO groups is that the adults living in more affluent neighborhoods may judge their own health by higher standards (for example, more negatively, related to others), and this might lead to impact estimates that understate any benefits of MTO moves on health status.

We also asked adult respondents about a variety of specific health problems they might have experienced, including whether they had an asthma or a wheezing attack the past year. Another

measure of a specific health outcome that we asked about came from national survey questions that have been used to measure activities of daily living (ADLs) to assess an individual's functional status and quality of life *(47)*. We focused on ADLs that were likely to be relevant for our sample of largely middle-aged adults and asked them if their health limited them "in climbing several flights of stairs or lifting or carrying groceries," and considered them to have limitations if they indicated they were limited "a lot" or "a little" versus "not limited at all."

Our measure for obesity (BMI≥30) comes from direct physical health measurements carried out by our survey interviewers. We used protocols developed for the Health and Retirement Survey to measure adult height and weight *(48)*. For height, respondents removed their shoes and stood on a smooth surface with their heels and shoulders against a wall. Interviewers placed a rafter square on the respondent's head, marked the height on a wall, and then measured it in inches to the nearest quarter inch using a tape measure. For weight, respondents removed their shoes and heavy objects from their pockets, as well as any heavy outer layers of clothing before stepping on the scale. Interviewers used a digital floor scale to measure weight to the nearest half pound. If weight or height could not be measured or if quality checks revealed an unusual value, we obtained self-reports from the respondents.

Interviewers also took respondents' blood pressure using a large-sized automated blood pressure cuff (Omron automated sphygmomanometer model HEM-711DLX). Respondents sat at a table with both feet flat on the floor and their arm resting, palm up, on the table. The cuff was placed on the respondent's upper left arm, about half an inch above the elbow. Interviewers tried to collect two separate readings from each adult respondent. If two valid diastolic and systolic readings were obtained, we used the average of each measure. If only one valid reading was obtained, we used the single-reading values. We used the definitions suggested by the National Institutes of Health, National Heart, Lung and Blood Institute, and Obesity Education Initiative for hypertension (systolic pressure of 140 mmHg or higher, or diastolic pressure of 90mmHg or higher; see also *(49)*). Because our measure of hypertension does not distinguish between having hypertension and being treated for hypertension, a limitation of our study is our inability to distinguish between people who are being effectively treated for hypertension versus those with no hypertension.

D. Mental health

Our index of adult mental health outcomes consists of a psychological distress index score for the past month, lifetime depression, lifetime Generalized Anxiety Disorder, calm and peaceful during the past month, and normal sleep last night. As with the economic and physical health indices, we score each variable so that a more positive value represents better mental health, and then create the index by taking the average of these variables in Z-score form.

We measure psychological distress using the Kessler 6 (K6) scale, which consists of six questions and is the most widely used scale of nonspecific psychological distress in the literature *(50, 51)*. The K6 consists of six questions about feelings over the past month of sadness, nervousness, restlessness, hopelessness, feeling that everything is an effort, and worthlessness *(51)*. Response options for the six questions were all, most, some, a little, or none of the time. The raw scores from the K6 can range from 0 (no distress) to 24 (highest level of distress).

We measure depression and Generalized Anxiety Disorder by asking adults to complete sections of the World Health Organization's (WHO) Composite International Diagnostic Interview (CIDI) *(52)* and the National Co-Morbidity Survey Replication *(53)*, the most widely used epidemiological interview for mental disorders in the world, which are designed to generate diagnoses of specific disorders defined by the Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV). We also replicated the question from the MTO interim evaluation that asked adults how much of the time in the past 30 days they felt calm and peaceful. Response categories match those for the K6 item. Finally we asked adults to report on how much sleep they received last night, by asking them when they went to bed and when they woke up. Normal sleep is between seven and eight hours per night.

E. Self-reported subjective well-being (SWB)

Our primary measure of subjective well-being (SWB) is based on responses to the question: "Taken all together, how would you say things are these days – would you say that you are very happy, pretty happy, or not too happy?" This question has a long history in the social sciences, having been included, for example, in the General Social Survey (GSS) since the 1970s. We employ the same three-point response scale as in the GSS to facilitate comparisons between the responses of our MTO sample with those of nationally representative samples.

The question of how many response points are optimal is a long-standing one, given the tradeoff between additional information that comes from more response points versus potential difficulties respondents might have in dealing with more response options. Some studies suggest there are diminishing returns to additional response options *(54)*, and that 3-point response scales for questions about feelings may capture 80-90% of the variation captured by 7-point scales *(55)* and are fine when the focus is on group averages *(56, 57)* as is the case with our study of MTO. We assume responses can be represented on an equal-interval linear scale. Our results are similar when we relax this functional form assumption, consistent with previous studies *(58)*.

Previous studies find that self-reports about SWB are reasonably stable over time. For example Krueger and Schkade find that the correlation of life satisfaction reports taken two weeks apart is 0.59 *(59, p. 1838)*. Using an average of multiple life satisfaction measures, Lucas, Diener and Suh find a correlation of 0.77 for measures taken four weeks apart *(60, p. 620)*.

Answers to the GSS question or similar SWB questions have also been shown to be correlated in expected ways with people's objective circumstances. Responses to questions about how happy people are at a point in time relate in the expected direction with what people are doing at that moment, such as engaging in leisure versus working or commuting. More global assessments of life satisfaction are correlated with self-reported positive and negative life events, including unemployment and income. Self-reports on different types of well-being questions have also been found to correlate with physical and mental health outcomes, spousal reports of the person's happiness, duration of genuine (or "Duchenne") smiles, blood pressure, and sleep quality *(29, 61–70)*. Average SWB reports across states correlate highly with objective indicators of quality of life *(30)*. Previous research shows a correlation of 0.3 between self-reports of life satisfaction and left-right difference in brain activation, relevant because the left prefrontal

cortex helps process approach and pleasurable stimuli, while the right side processes avoidance and aversive stimuli *(29, 71)*.

Previous research also provides some evidence indicating that people have a common understanding of "happiness." For example, previous studies suggest that people are able to predict the satisfaction levels of others *(61, 72)*. Previous research also suggests that people seem to translate numerical happiness scales into similar verbal labels *(73)*.[†]

2.3 Analytic strategy

In this section we first describe how we estimate the intention-to-treat (ITT) effects presented in Table 2 and Figure 1 in the main text. We then discuss how we can move beyond the pure design of the randomized MTO experiment to identify the effects of actually moving through MTO, known as the effects of treatment on the treated (TOT). We conclude by discussing how we use random assignment to treatment in MTO to construct instrumental variables to use in estimating the relationship between specific neighborhood characteristics and our outcomes of interest.

A. Intention-to-treat (ITT) effects

We begin with simple comparisons of the average happiness of adults assigned to different MTO groups, known as the ITT effect, which identifies the causal effect of offering families the services made available through the *LPV* or *TRV* treatments. One advantage of the ITT estimate is that it fully capitalizes on the strength of MTO's randomized experimental design. The disadvantage of the ITT estimation is that it does not provide any information about the size of the effect on those who actually change neighborhoods, or about the relationship between specific neighborhood characteristics and people's life outcomes.

Let Y represent some outcome of interest. In the main paper itself we present the results from estimating a model using pooled data from all three MTO groups with Z consisting of a single indicator for assignment to either the *LPV* or *TRV* group. We initially pool the two randomized MTO treatment groups together to improve statistical power and because, as noted below, the effects of assignment to the *LPV* versus *TRV* groups on neighborhood environments that families experience converge over time. In these supplementary materials we also present additional results that show findings for each of the two randomized MTO treatment groups separately.

We calculate the ITT effect as π_{11} in equation (S1) using ordinary least squares (OLS), conditioning on a set of (pre-random assignment) baseline characteristics ($\mathbf{X}$). These include indicators for each person's MTO demonstration site, and survey measures of the socio-demographic characteristics of household members. Because the distribution of pre-program characteristics should be balanced across treatment groups with random assignment, conditioning on these variables serves mainly to improve the precision of the treatment effect estimates. All estimates in this paper are computed using the sample weights described above. In practice, the coefficients from applying ordinary least squares to dichotomous dependent variables tend to be quite similar to the average marginal effects that come from probit or logit models *(74)*, and indeed we find that average marginal effects calculated from probit models are quite similar to the OLS estimates reported here. Our categorical outcome measure for SWB

(reported on a 3-point scale) presents a different sort of estimation challenge. In the main text, we assume that responses can be represented on a linear scale. In this appendix, we also estimate a series of ordinal representations of the responses using ordered probit or logit maximum likelihood models and show that we obtain qualitatively similar results.

$$(S1) \quad Y = Z\pi_{11} + \mathbf{X}\pi_{12} + e_1$$

Unbiased estimation of the ITT effect requires several assumptions that we believe are likely to be met in the MTO application. The first assumption is that random assignment was carried out correctly, which we believe is the case based on a review of the randomization procedures employed by Abt Associates, which carried out random assignment on behalf of HUD, and given evidence presented in tables 1 and S1 that the distribution of baseline characteristics are very similar randomly assigned MTO groups. A second assumption is that there is no selective attrition in our measurement of follow-up outcomes across randomized groups. We believe this assumption is likely to be met because of the high overall effective response rate achieved by our survey subcontractor (around 90 percent), and because this effective response rate is generally similar across randomly assigned groups.

A third assumption for the standard interpretation of the ITT estimate is that the effect of MTO random assignment on a given family is independent of the treatment-assignment status of other families in the study sample, which Rubin called the "stable unit treatment value assumption" (SUTVA) and what has also been called the "no-interference" assumption *(75)*. Michael Sobel has raised concerns that SUTVA may not be met in the MTO application, as could occur if for example families assigned to the treatment group share information with controls that lead some control group families to also move to lower-poverty neighborhoods *(76)*.

As we have indicated elsewhere *(77)*, we think major violations of SUTVA are unlikely. Only around one-quarter of eligible public housing families applied to participate in MTO. Since around two-thirds of families that signed up for MTO were assigned to treatment, and fewer than three of five assigned to treatment moved with a MTO voucher, the share of public housing families who moved out of public housing through MTO is not more than 10% (that is, 25%*66%*60%). The actual share will be lower still given that not all public housing families were eligible for MTO (for example because they did not include children). Moreover the families that signed up for MTO seem to have been fairly socially isolated at baseline: 55 percent of household heads indicated on the baseline surveys when applying to MTO that they had no friends in the baseline neighborhood, and 65 percent reported that they had no family in the neighborhood. For this reason we suspect that social interactions among MTO families were probably limited. Moreover the MTO program administrators tried to limit the clustering of families assigned to the MTO experimental group in low-poverty neighborhoods, an effort that maps of MTO relocation outcomes suggest were successful; see for example *(37)*. Understanding more about the degree to which MTO families both within and across randomly assigned groups had important social interactions with one another remains a useful topic for future research.

B. Effects of treatment on the treated (TOT)

It is also possible to use data from the MTO experiment to estimate the effects of MTO moves on those who actually move through MTO, known as the effect of treatment on the treated (TOT) and calculated as the ITT effect divided by the treatment take-up rate *(78)*. The standard error for the TOT effect is calculated the same way, by dividing the ITT standard error by the treatment take-up rate, so that the p-value for the ITT and TOT estimates will be the same under this method. The TOT estimates derived this way are very similar to what comes from using two-stage least squares to estimate the effects of relocating through either the *LPV* or *TRV* group, using indicators of random assignment to the *LPV* and *TRV* group as instrumental variables *(79)*. Since 48% of the *LPV* group relocated with a MTO voucher, the *LPV* TOT effect will be (1/.48) = 2.08 times as large as the *LPV* ITT effect. Similarly, because 63% of the *TRV* group relocated with a MTO voucher the TOT effect for the *TRV* group is (1/.63) = 1.59 times as large as the ITT effect. While we do not present the TOT estimates in our tables, readers can calculate the TOT point estimates and standard errors from our ITT estimates using these scale-up factors.

C. Estimating the relationship between outcomes and specific neighborhood conditions

Also of interest is to understand the relationship between specific neighborhood attributes (*W*) and outcomes *Y*. Let **W** represent one or more measures of candidate mediating mechanisms through which MTO might influence happiness, such as the poverty rate for the census tracts in which MTO families are residing, while ***X*** represents the baseline control variables discussed above. The relationship between the candidate mediator(s) and happiness is summarized by the parameter(s) π_{21} in outcome equation (S2).

$$(S2)\quad Y = \mathbf{W}\,\pi_{21} + \mathbf{X}\,\pi_{22} + e_2$$

For purposes of estimation of equation (S2), we view any single variable used as an element of **W** to be a summary measure of neighborhood economic disadvantage. For example, when *W* is a scalar equal to the tract poverty rate, we interpret π_{21} as the effect of moving to a neighborhood with a lower poverty rate and other aspects of neighborhood economic disadvantage that co-vary with tract poverty rates. We provide a similar interpretation with the other key mediating measure that we examine, namely the share of the census tract population that is minority.

OLS estimation of (S2) may be biased by endogenous residential choices. Families that wind up living in lower-poverty tracts may be systematically different from those who live in high-poverty areas in ways that are difficult to measure in a social science dataset and may directly affect people's outcomes. This type of selection bias (or omitted variables bias) manifests itself as a correlation between **W** and e_2 in equation (S2), and would lead ordinary least squares estimates to mistakenly attribute to **W** the effects of unobserved measures in e_2.

Rather than use ordinary least squares to estimate (S2), we use the random assignment of families to treatment and control conditions in MTO as an instrumental variable for **W** and estimate (S2) using two-stage least squares (2SLS) and related instrumental variables (IV) estimators. One possible way to do this would be to use a single indicator for assignment to

either of the two MTO treatment conditions (*Z*) as an instrument for a single candidate mediating measure, *W*. The first stage equation (S3) is used to generate a predicted value of the mediating measure that is then substituted for the actual measure in the second stage. The second-stage equation (S2) estimates the relationship between neighborhood conditions and the outcomes isolating the experimentally-induced variation in the mediator.

$$\text{(S3)}\quad W = Z\pi_{31} + \mathbf{X}\pi_{32} + e_3$$

One potential drawback of this type of one-instrument, one-mediator ("just-identified") model is precision. In a model without baseline covariates, 2SLS basically collapses the data into just two points, one for treatment-group families and one for controls. The 2SLS estimate would be equivalent to a "visual instrumental variables" or VIV *(80)* graph that fits a line between the average outcome and mediator values for adults assigned to treatment and the average outcome and mediator values for adults assigned to control.

A different approach is to interact treatment assignment with different baseline characteristics of sample members to increase the number of instrumental variables. If the effect of treatment assignment on the candidate mediator of interest varies by baseline sub-groups, then interacting treatment assignment with baseline sub-group indicators can improve the precision of the second-stage 2SLS estimates by increasing the explanatory power of the first-stage equation *(80)*. This approach requires the assumption that the only reason that some sub-groups would experience more pronounced changes in outcomes in response to treatment assignment is because those sub-groups experience relatively larger changes in the mediator in response to treatment assignment. Another potential cost of this approach arises in applications where the instruments are weak (low explanatory power in the first stage equation), since the bias associated with instrumental variables estimation with weak instruments generally increases as the number of instruments increases.

The results presented in the main paper follow the approach of Kling, Liebman, and Katz *(7)*, who used as instrumental variables interactions between two indicators for assignment to the *LPV* or *TRV* group (*Z*) and five site indicators (*S*), controlling for the main demonstration-site effects in the baseline covariates, ***X***. They showed that there is indeed substantial variation across the five MTO demonstration sites in the degree to which treatment assignment affects neighborhood poverty and other candidate mediators (see also *(81)*).

In a model in which the only covariates in **X** are the five MTO demonstration site indicators, the 2SLS estimates are equivalent to the slope in a VIV graph of the line that is fit between the 15 data points for the average mediator and outcome values for the cells defined by the three MTO random assignment groups (*LPV*, *TRV*, and control) and five MTO demonstration sites. The model essentially estimates a "dose-response" relationship, and asks whether those groups that experience relatively larger changes in some candidate mediator as a result of treatment also experience larger changes in the outcome.

We begin by presenting instrumental variables estimates that are calculated using 2SLS. One limitation of 2SLS is the assumption that responses to the GSS happiness question fall on a linear scale. As a sensitivity analysis we also relax this assumption and re-estimate equations (S2) and

(S3) using instrumental variables ordered probit, following the control-function approach from Rivers and Vuong *(82)*, and obtain qualitatively similar results.

A different limitation of 2SLS is that if the instruments are weak, that is, have low explanatory power in the first stage equation, the standard errors from 2SLS estimation may be too small and hence lead us to over-reject the null hypothesis that mediator **W** has no effect on the outcome *Y* (see for example *(83)*). In response to this "weak-instrument" concern we also re-estimate equations (S2) and (S3) using limited information maximum likelihood (LIML), which requires the assumption that the error terms in the first and second stage equations are jointly normal but can have better finite-sample properties than 2SLS in applications where the instruments are weak. As another sensitivity test we also estimate our model using a modified version of LIML suggested by Fuller *(84)* designed to decrease the variability of LIML estimators in small samples and has better performance with weak instruments. The Fuller version of LIML with Fuller adjustment factor 1 has a higher order mean bias of zero *(85, 86)*, although we also present results with adjustment factors of 2 and 4 that tend to have lower mean-squared error. However we demonstrate below that our instruments seem to have good first-stage explanatory power for most candidate mediators. Even in cases where the first-stage explanatory power for some mediators is relatively lower, our results do not seem to suffer from a weak-instruments problem.

3. Supplementary results

3.1 Descriptive characteristics of the MTO Sample

Table S1 is an expanded version of table 1 in the main text, and displays descriptive characteristics for our study sample of MTO adults separately for the control group, for the two treatment groups (*TRV* and *LPV*) pooled together, and for the two treatment groups separately as well. Almost all of the households that signed up for MTO were female-headed, nearly two-thirds were African-American, and most of the rest were Hispanic. Three-quarters of household heads were on welfare at baseline, and less than 40% had completed high school.

More than 40% of households that applied had a household member victimized by a crime during the previous six months. Three-quarters of MTO families reported that getting away from gangs and drugs—that is, crime—was the first or second most important reason for enrolling in the program, far more than any other reason. More than half of the households said the first or second most important reason for signing up for MTO was so that that their children could attend a better school, and about 45% indicated that getting a bigger or better apartment was their first or second reason for enrolling in MTO. Only a small fraction of families (6%) indicated getting a job was one of their top reasons for signing up for MTO.

Table S1 also confirms that random assignment appears to have been correctly carried out in MTO, given the balance across randomized MTO groups in the distributions of the observed baseline characteristics. An omnibus F-test fails to reject the null hypothesis that the set of baseline characteristics shown in the table are similar for the *LPV* versus the control group (*P*=.473) or *TRV* group versus controls (*P*=.268) or for the *LPV* and *TRV* groups pooled together versus controls (*P*=.309).[‡]

Some 48% of the adults assigned to the MTO *LPV* group and 63% of those assigned to the *TRV* group were able to lease-up and relocate using an MTO voucher (the MTO "compliance rate"). The MTO compliance rate is less than 100% for a variety of reasons including that families were given only a limited amount of time to search for a new unit, many housing units were not affordable under voucher program rules, and some landlords may discriminate against voucher holders. The voucher use rate in MTO is in line with what other studies have found—equal to 65% in the Experimental Housing Allowance Program *(87, p. 146)*, and around 20% in the Gautreaux mobility program in Chicago *(88, p. 67)*. The compliance rate is higher for the *TRV* group presumably in part because their vouchers had no special MTO-imposed geographic restriction, unlike the vouchers that were offered to adults in the *LPV* group. The compliers within the MTO treatment groups are younger, more dissatisfied with their original neighborhoods, and have fewer children than the noncompliers (for details see *(89, 90)*).

Table S2 presents descriptive statistics on the distribution of happiness reports by MTO adults in the long-term follow-up survey, compared with adults interviewed in the 2000-2006 waves of the GSS.[§] MTO control group adults are less likely than the average American adult to be very happy (23% versus 34%) and are over twice as likely to say they are not too happy (28% versus 11%). On the other hand, MTO control group adults are more economically disadvantaged than the average American, and the final column of table S2 shows that on average adults in the MTO control group are slightly happier than other Americans with similar socio-demographic characteristics.[**]

3.2 MTO effects on neighborhood and housing conditions

Table S3 displays estimates of MTO effects on housing and neighborhood conditions that could be candidate mediators for the effect of MTO moves on adult outcomes.

The top panel of table S3 shows the average baseline neighborhood characteristics of the control group, the ITT effect of the pooled voucher treatment (*LPV* and *TRV* groups together), and then the ITT effects of the *LPV* and *TRV* groups separately. At baseline the average control group family is living in a census tract that is about 53% poor, or about 3.2 standard deviations above the national average as calculated from the national census-tract-poverty distribution from the 2000 decennial census. There are no differences across randomly assigned groups in baseline characteristics, as we would expect with random assignment.

The second panel of table S3 shows that MTO moves accomplished their goal of helping families move into lower-poverty neighborhoods. One year after random assignment, the average control group family lived in a census tract with a 50% poverty rate, which was equal to about 33% for families assigned to the *LPV* group and 37% for families assigned to the *TRV* group ($P<.01$ in both cases). These changes are from 1.1 to 1.3 standard deviations within the national tract-poverty distribution, or about 0.8 to 1.0 standard deviations within the control group's tract-poverty distribution. While table S3 focuses on presenting ITT effects for parsimony, as discussed above the TOT effects will be about 2.1 times as large as the ITT effects for the *LPV* group and about 1.6 times as large as the ITT effect for the *TRV* group.

The difference across MTO groups in census tract poverty rates narrowed over time, due largely to declines over time in the tract poverty rates experienced by the control group – which went from 53% at baseline to 33% 10 years later. The control group trend is due more to control families moving into lower-poverty neighborhoods over time on their own, as opposed to control families living in neighborhoods that are gentrifying around them.[††]

Ten to fifteen years after baseline at the time we went into field to collect the long-term surveys (around May 2008), the ITT effects of the *LPV* and *TRV* groups on tract poverty rates equaled about 4 and 2 percentage points, respectively (P<.05 in both cases). These impacts equal .30 and .17 standard deviations, respectively, in the national census-tract poverty distribution in the 2000 census (or about .22 and .13 standard deviations in the MTO control group distribution). The duration-weighted average tract poverty rate for all addresses between random assignment and 10-15 years later (May 2008) was around 40% for the control group, with an ITT effect of about 9 percentage points for the *LPV* group and 6 percentage points for the *TRV* group (*P*<.01 in both cases). We also find that using a duration-weighted measure of a broader set of indicators for concentrated disadvantage based on Sampson, Sharkey, and Raudenbush yields results that are qualitatively similar to the results for duration-weighted tract poverty *(20)*.

MTO moves also led treatment-group families to live in census tracts that had slightly lower minority shares compared with controls, even if they were still mostly minority. One year after random assignment the average tract share minority was 91 percent for control families, with ITT effects of 11 percentage points for the *LPV* group and 3 percentage points for the *TRV* group (*P*<.01 in both cases). As a way to think about the magnitude of these estimates, the *LPV* ITT effect is equal to about .36 standard deviations in the national tract distribution, or about .57 standard deviations in the control group's tract-minority distribution. The ITT effects on duration-weighted tract minority share was 6 percentage points for the *LPV* group (*P*<.01, equal to .19 standard deviations in the national tract distribution, or .37 standard deviations of the MTO control group's distribution) and 1 point for the *TRV* group.

Table S3 shows that MTO families assigned to treatment wound up making about one-half extra move over the course of the study period compared to the control group, which moved on average 2.2 times over the 10-15 year period. Table S3 also shows that MTO moves led to gains in housing-unit quality.

A growing body of research suggests that more detailed measures of neighborhood social process may be better predictors of neighborhood influences on behavior than are measures of neighborhood socio-demographic composition *(5)*. Table S3 shows that MTO moves helped families move into communities where neighbors were more likely to intervene to prevent youth from spraying graffiti (a measure of "collective efficacy" *(3)*), increased the chances that MTO movers had at least one close friend who had a college degree, and reduced the likelihood that MTO movers felt unsafe in their neighborhoods during the day.

The fact that MTO changed so many aspects of the social and physical environment of families at once will present a challenge for isolating which specific mediators are most important for affecting people's outcomes. For our IV estimates below we interpret neighborhood poverty rates and neighborhood racial segregation as broad markers for a collection of attributes of the

neighborhood environment that are relatively more strongly correlated with concentrated poverty than with neighborhood racial segregation. It is worth noting that different candidate mediating measures such as housing-unit quality or safety tend to be more strongly correlated with tract poverty than with tract minority share (see also *(7)*).

3.3 MTO effects on adult outcomes

Table S4 presents more details about the impacts of MTO on our adult outcome measures. The first row shows that MTO treatment assignment overall has no statistically significant effect on our index of economic outcomes. The estimated ITT effect is if anything negative and marginally significant for the *TRV* group, although as discussed elsewhere we believe this is mostly an artifact of interviewing the *TRV* group on average slightly later in calendar time than controls, when the recent recession would have had more pronounced effects on outcomes of the *TRV* group when they were interviewed *(23)*. The next two rows show that the effects of *LPV* and *TRV* assignment are quite similar to one another in magnitude for both our physical health and mental health outcome indices. Pooling the two treatment groups together provides just enough gain in precision to be able to reject the null hypothesis that the MTO treatment effect on mental health is zero at $P<.10$.

The second panel of table S4 shows our results for subjective well-being (SWB) as our outcome of interest. About 23% of the control group reported being "very happy." The ITT estimates in the first row of this panel imply that those who are assigned to treatment are nearly 3 percentage points more likely to report being very happy, with ITT effects of 1 point (not significant) for the *LPV* group and 5 points ($P<.05$) for the *TRV* group. Around 73% of controls report being very happy or pretty happy; assignment to treatment increases that likelihood by 3 to 5 percentage points. Table S4 shows that average marginal effects from probit and logit models for whether families are above the very happy threshold or the very or pretty happy threshold are nearly identical to the OLS coefficients.

The third section of the second panel of table S4 show the results of combining information about both of the happiness cut points (very happy versus pretty happy or not too happy, and pretty happy or very happy versus not too happy) into a single measure, assuming cardinality using a three-point scale in the third section of the panel. The average control adult is about at the "pretty happy" point on the scale (1.95). The pooled treatment groups are about .07 higher on this scale ($P<.05$), with ITT effects of .06 for the *LPV* group ($P<.10$) and .08 for the *TRV* group ($P<.10$). These results imply that the MTO restriction to initially move into a low-poverty census tract that was imposed on the *LPV* group (combined with any effect of mobility counseling) but not the *TRV* group had little effect on the long-term effects of offering families the chance to move to a less-distressed area on SWB, since the *LPV* and *TRV* ITT effects are not statistically different from one another ($P=.45$). The next two rows show the coefficients from ordered probit and logit models; the marginal effects implied by these ordered probit and logit coefficients are nearly identical to those from OLS (available upon request).

Although the GSS "happiness" question is the only measure on the MTO long-term surveys that was explicitly designed to provide a global self-assessment of SWB, the remainder of table S4 shows that our findings of improved SWB among treatment group adults are corroborated by

other measures that we would expect to be related to global SWB. For example, we find evidence of beneficial MTO impacts on a measure of whether respondents met clinical criteria for major depression, from a fully structured assessment that was included on the MTO long-term surveys designed to measure mental health disorders as defined by the Diagnostic and Statistical Manual, 4th edition *(53)*. Responses on the long-term MTO surveys imply that around 20% of control group adults have ever had major depression; the figure is about 4 percentage points lower for adults assigned to either treatment group ($P<.05$), with ITT effects of about 3 percentage points for the *LPV* group and nearly 5 points for the *TRV* group. Our long-term survey also included a validated, six-question screening scale of psychological distress called the K6 *(51)*. Assignment to treatment in MTO improved mental health on this K6 scale by about .08 to .10 standard deviations.

3.4 Neighborhood effects on subjective well-being

The estimates shown in table 2 and Figure 1 in the main paper and tables S3 and S4 make clear that the MTO program generated sizable changes for participating families in the characteristics of their neighborhoods and in their SWB. The next step in our analysis is to estimate the size of the relationship between specific neighborhood or housing measures with SWB by relating the size of MTO's impacts on these mediators to the size of MTO's impacts on SWB.

These results are presented in table S5. We present results from estimating equations (S2) and (S3) above using interactions of MTO random assignment outcomes and demonstration site as instrumental variables for different candidate mediating measures. Each panel in the table presents results for a different outcome measure used as the dependent variable in our second-stage equation (S2) – economic self-sufficiency in the first panel, followed by physical health, then mental health, and finally SWB. Within each panel we present the coefficient and standard error for the relationship between the outcome measure and a candidate mediating measure that is included one-at-a-time as endogenous explanatory variables in equations (S2) and (S3), focusing on duration-weighted tract poverty rates and tract minority share. The columns show results from different estimation approaches, so that each entry in the table is the second-stage coefficient for the relationship between the candidate mediator and outcome listed at left, calculated using the estimation procedure described by the column heading.

The results shown in the first row, first column of table S5 indicate that a 1 standard deviation reduction in neighborhood economic disadvantage (as represented by tract poverty rate) reduces the value of the economic self-sufficiency index by .043 standard deviations, which is not statistically significant at conventional thresholds. This is calculated by applying 2SLS to equations (S2) and (S3). By way of reference, in the MTO data a 1 standard deviation change in tract poverty rates is equal to about 13 percentage points, compared to a control mean of about 40% tract poverty over the entire follow-up study period. The final column shows that the F test statistic for the explanatory power of our instruments in the first-stage equation is 29.9, which is significant and well above the usual rule-of-thumb for weak instrument problems of F=10 *(91)*. Given that the instrumental variables have strong explanatory power in the first stage for tract poverty rates, it is not surprising then that the other estimation procedures shown in subsequent columns in SE5 that are designed to deal with the weak instrument problem all yield estimates that are quite similar to those from 2SLS.

The second row of this top panel shows that neighborhood minority composition does not have a statistically significant relationship with economic outcomes, either.

The second and third panels of table S5 show that neighborhood economic disadvantage as represented by tract poverty rate has a statistically significant relationship with our physical and mental health outcomes, while tract minority share does not. A 1 standard deviation decrease in tract poverty is associated with an improvement in the physical health index of about .11 standard deviations, with a similarly-sized gain in mental health as well ($P<.10$ in both cases, consistent across estimation methods).

Some readers might wonder whether the statistically insignificant relationship between tract minority share and physical and mental health outcomes is simply a reflection of limited variation in tract minority share in the MTO data. It is true that the overall difference between the treatment and control groups pooled across all five sites in tract minority share (that is the overall ITT effect) is fairly modest. For example table S3 showed that the difference between the two treatment groups pooled together and the control group in duration-weighted tract minority share was about 4.6 percentage points. However this is not the source of variation that we use to identify our IV estimates. Our IV model instead relies on the variation *across treatment groups across sites* in tract minority share. The range of variation in the explanatory variable in our second-stage IV model is about 10 percentage points for tract minority share, about twice the ITT effect presented in table S3. By way of comparison the range of variation in duration-weighted tract poverty for our IV estimates is on the order of 13 percentage points.[‡‡]

The results shown in the first row, first column of the final panel of table S5 indicate that a 1 standard deviation reduction in neighborhood economic disadvantage (as represented by tract poverty rate) is associated with a relatively larger beneficial increase on SWB, equal to .14 standard deviations ($P<.01$). The last row of table S5 also shows that tract minority share has no statistically significant relationship with SWB among MTO adults. Even though the first-stage F-test statistics are lower in models in which we use tract minority share as the endogenous explanatory variable compared to those in which tract poverty is the endogenous explanatory variable, our estimates for the association between tract minority share and SWB are quite similar whether we use 2SLS or the other estimation approaches like LIML designed to adjust for any problems with weak instruments.

Note also that the "dose-response" relationship implied by these results (and shown graphically in Figure 2A) helps rule out concerns that our estimates for the MTO effects on SWB are simply due to a "Hawthorne effect." In principle one might be concerned that MTO respondents assigned to either the *LPV* or *TRV* group provide higher SWB values simply because they "won the lottery," or because they think the interviewer expects them to be happier. But winning the MTO lottery is something that is common across all five MTO demonstration sites. If our results were simply due to a Hawthorne effect, we would not expect to see the largest gains in SWB in places where the "treatment-dose" on some neighborhood condition like poverty rate is most strongly affected by MTO moves. We can further refine this exercise by excluding data from the control group, and just compare levels of SWB of families assigned to the *LPV* and *TRV* groups

in different sites. The estimates become less precise than our main results, but are qualitatively similar – which also helps rule out a generic moving effect as an explanation for our findings.

Previous theorizing about "neighborhood effects" has emphasized the possibility that people's life outcomes may depend not just on their current neighborhood environments, but also on people's accumulated exposure to different neighborhood environments in the past *(16)*. Current neighborhood environments may affect people's outcomes by shaping the immediate environment in which different decisions and behaviors occur, or what Robert Sampson calls "situational" influences *(16)*. But neighborhoods may also shape the way people's preferences develop, or their capacities to engage in different behaviors – what Sampson calls "developmental effects." To allow for both types of effects, the main results we present in the paper use duration-weighted neighborhood environments that are averaged over each MTO participant's history of residential addresses from randomization to the start of our long-term follow-up survey period in May 2008. But we can also estimate what is in some sense a more restrictive model that assumes that only current neighborhood environments matter, that is, neighborhood environments measured at the start of the follow-up survey period. Table S6 shows that magnitudes of the estimated coefficients tend to increase when we use tract characteristics measured as of the time of the long-term survey rather than duration-weighted tract characteristics, and the statistical significance of the estimated relationship between tract poverty and mental health index becomes a bit more sensitive to our choice of specific estimation approach (2SLS versus LIML or some version of the Fuller-adjusted LIML). But the overall pattern of results shown in table S6 using tract characteristics measured at the time of the long-term survey is qualitatively similar to what we see in table S5 when we use duration-weighted tract characteristics.

Because SWB or "happiness" has no natural metric, one way to interpret the size of MTO's impacts is to compare them in size to disparities in SWB that have been of social science or policy concern. Table S5 shows that the effects of a 1 standard deviation decrease in neighborhood economic disadvantage as represented by duration-weighted tract poverty rates is to increase SWB by .14 standard deviations, which is equal to about .10 units on a three-point scale assuming cardinality. Data from the nationally representative GSS surveys carried out from 2000 to 2006 show that the gap between blacks and whites on the same SWB scale is equal to 0.15. The implication is that the effect of MTO moves on SWB is about two-thirds as large as the overall gap in SWB between blacks and whites in the U.S.[§§] Another comparison is to the gap in SWB between married and single people in the GSS, which equals 0.32 units. So the effect of moving through MTO on SWB is equal to about one-third of the gap between married and single adults.

A different benchmark comes from the gradient in SWB found among people with different annual incomes. Previous research has found that the relationship between self-reported SWB and income seems to be linear in the natural log of income *(69, 70)*. Because survey reports about annual income are subject to a great deal of measurement error, which may attenuate the estimated relationship with SWB, we instrument for family income using educational attainment as an instrumental variable, following Stevenson and Wolfers *(69)*. The coefficients in the second column of table S7 from this IV estimation suggest that the effect of a 1 standard deviation decline in neighborhood economic disadvantage as represented by tract poverty rates is

about equivalent to a 0.5 log point change in family income, or about the difference in SWB between families with permanent incomes that differ by $13,000 per year.

One might reasonably worry about whether educational attainment meets the exclusion restriction for a valid instrument, since schooling could potentially directly influence happiness other than through its effect on annual earnings. However, these IV estimates are more conservative for our MTO application than the OLS estimates, in the sense that the larger coefficient for the relationship between income and SWB from the IV estimates compared to the OLS estimates in table S7 implies that our estimated neighborhood effects on SWB are equivalent to relatively more modest changes in annual income. For that reason, in the main body of the text we focus on these smaller income differentials as our points of comparison for the size of the MTO effects.

We can also use our estimates for the implied dollar equivalent for the relationship between a 1 standard deviation change in tract poverty rates and SWB to consider the change in SWB over time for families in the bottom of the income distribution due to increased poverty concentration over time. We calculated changes over time in the average tract poverty rate of families in the bottom quintile and bottom two quintiles (i.e. bottom 40%) of the income distribution using data from the 1970 decennial census *(92)* and the 2005-09 waves of the American Community Survey (ACS; *(93)*), which for convenience we refer to as "2007 data." The average family income of MTO families in the long-term follow-up survey (about $20,000) is somewhat higher than that of the bottom quintile in 1969 ($15,336) and somewhat lower than that of the bottom two quintiles averaged together in 1969 ($24,603) *(32)*. Given how disadvantaged MTO participants are on so many other dimensions besides income, it may be that MTO results are more likely to generalize to those families in the bottom quintile of the overall U.S. income distribution than to those in the bottom 40% of the distribution. For completeness we report results for both groups in these supplementary materials.

One challenge in comparing tract poverty rates from 1970 to 2007 is that 1970 census tract data are limited to urban areas and cover only about 73% of all U.S. families whereas the 2007 tracts cover all families. For comparability, we focus our main comparisons on the set of counties with tract level data for both 1970 and 2007. For completeness, we also show all available 1970 tracts compared to all 2007 tracts.

To calculate the average tract poverty rate for the bottom quintiles, we average the tract poverty rate weighted by the number of people in each income bin below the cutoff *(94)*. For the 1970 data, the 1969 quintile cutoffs align with the family income bin; this is not the case for the ACS and so we linearly interpolate the number of people in the income bin (that spans the cutoff) that are below the cutoff value.

The top panel of table S8 shows average tract poverty rates for families in the bottom income quintile and the bottom panel shows families in the bottom two income quintiles together. The columns compare poverty rates for 1970 and 2007 using all the census tracts that are available in each year, and then again just for tracts in those counties for which we have data at both points in time. Tract poverty rates for the bottom quintile increased by about 2.4 percentage points over the past 40 years (using data from matched counties as described above). Using our estimated

relationship between poverty and happiness, this would suggest an average decrease on the happiness scale of about .02. This decrease in happiness is similar to the predicted change in happiness that would be associated with a decrease of $1,400 in income. This is slightly larger than the gain in total real income that families in the bottom quintile of the overall U.S. income distribution experienced from 1969 to 2007 (from $15,336 to $16,622, for a change of $1,286). For the bottom two-quintiles, our estimates would imply that the increase in neighborhood poverty experienced by this group over the past 40 years (about 1.5 percentage points) would decrease the SWB of families by nearly $1,500, which is nearly half the growth in real income that families in the bottom 40% of the income distribution experienced from 1969 to 2007 ($24,603 to $28,123).[***]

3.5 Mechanisms through which MTO moves affect happiness

Table S9 shows that duration-weighted tract poverty is more important for SWB than is duration-weighted tract minority share when we "horse race" them against each other by including both mediators at a time in the same model described by equations (S2) and (S3) above. As with the instrumental variables estimates presented in table S5, we use interactions of MTO treatment assignment and demonstration site as instruments for our candidate mediators. Table S10 presents similar results using tract characteristics measured at the start of the long-term survey period (May 2008), rather than duration-weighted tract characteristics, which imposes the assumption that only current and not lagged neighborhoods matter for SWB. The results in tables S9 and S10 are qualitatively similar but less precisely estimated using just current neighborhood conditions.

The top three panels of table S9 show that we simply do not have adequate statistical power in the MTO data to be able to distinguish the effects of neighborhood economic disadvantage (as reflected by tract poverty) and tract minority share using economic self-sufficiency, physical health, or mental health as our outcomes of interest. However the figures shown in Figure S1 and S2 point at least suggestively in the direction of tract poverty being more important than is racial segregation for physical health or mental health. Panel C of both Figures S1 and S2 shows that a decrease in tract poverty rates (holding tract minority share constant) seems to improve physical or mental health. On the other hand, reductions in tract minority share (Panel D) if anything seem to lead to reductions in physical or mental health.

The final panel of table S9 shows that we do have adequate statistical power to show that neighborhood economic disadvantage as reflected by tract poverty is more important for SWB than is tract minority share when both are included in the same 2SLS model (equations S2 and S3 above). When both mediators are included in the same model, we find that a 1 standard deviation reduction in tract poverty is associated with increased levels of SWB equal to about .26 standard deviations (P=.005). In contrast, a 1 standard deviation reduction in tract minority share is associated with a *reduction* in SWB for adults, by about .28 standard deviations (P=.098). While a t-test just barely rejects the null hypothesis that the slope of the relationship between SWB and duration-weighted tract minority share is different from zero at P<.10, an F-test lets us soundly reject the null hypothesis that the coefficients on tract poverty and tract minority share are equal to one another (P=.030 with 2SLS estimation, and with similar P-values using our other estimation approaches). This is the key result behind our conclusion that a decline in

neighborhood economic disadvantage is more beneficial for SWB than is a decline in neighborhood minority composition.

Endnotes

[*] In previous discussions of the MTO results the three randomized groups have also been referred to as the experimental group, the Section 8 only-group, and the control group.

[†] Even if MTO respondents differ in the thresholds that they use to map experienced utility into happiness reports on the GSS question, this would not pose any problems for our analysis so long as the MTO treatment itself did not affect the happiness thresholds, because in that case the distribution of happiness thresholds would be similar across the three MTO groups by virtue of random assignment. The methodological challenge for our analysis comes from the possibility that the MTO treatment itself shifts the happiness thresholds people use for reporting, which we discuss further below.

[‡] We conduct an omnibus F-test of the differences between the treatment and control groups by estimating a seemingly unrelated regression where all of the characteristics listed in table S1 are stacked as Y (outcome) variables and the only X variable is an indicator for treatment group status and a constant. This approach follows Jacob, Ludwig, and Miller *(95)*.

[§] The version of the GSS data we use was downloaded from Justin Wolfers' website, used for Stevenson and Wolfers *(70)*. http://bpp.wharton.upenn.edu/jwolfers/data.shtml

[**] We re-weighted the GSS sample to have similar socio-demographic characteristics to the MTO control group, specifically with respect to race/ethnicity, age, gender, marital status, family income, urban residence (versus suburban or rural residence), and educational attainment. Similarly, if we use the GSS data to regress happiness against these socio-demographic characteristics, and then use the coefficients from this GSS regression to predict the happiness levels that MTO control group families should have based on their background factors, the predicted values are fairly similar to observed values.

[††] We test this by reproducing the estimates shown in table 2 in the main text and table S3 in these supplementary materials measuring the share poor in each tract using just data from the 2000 decennial census, rather than interpolating each census tract's poverty rate at the time the MTO family was actually living in the tract. The estimates using 2000 tract poverty rates are fairly similar to those shown in tables 2 and S3, suggesting that most of the change in the control group's tract poverty rate over time occurs because control families are moving into lower-poverty areas, rather than because the control group is living in census tracts that are becoming less poor around them.

[‡‡] Figure 1B shows that the span of support for duration-weighted tract minority share (that is, the horizontal distance between the left-most and right-most data point on the x-axis) is equal to about .6 standard deviations. The results in table S3 suggest that 1 standard deviation in the MTO control group's distribution for duration-weighted tract minority share is about 16 percentage points, so that the range of variation in the explanatory variable in our second-stage IV estimates is around (.6*16)=10 percentage points, or about twice the ITT effect presented in table S3. Under the same logic, Figure 1A suggests that the span of the support in our second-stage model when we use duration-weighted tract poverty rates as the explanatory variable of interest is about

1.05 standard deviations. Table S3 suggests that a 1.05 standard deviation change in the MTO control group's distribution in duration-weighted tract poverty rates would be about 13 percentage points.

[§§] Stevenson and Wolfers show that the black-white gap in SWB has declined by about 40 percent from 1972 to 2008; in standard deviation units, from about 0.45 standard deviations to about 0.27 standard deviations *(28)*. Despite this substantial convergence the black-white gap in SWB is still sizable.

[***] We note that a smaller tract poverty change for families in the bottom 40% of the income distribution has about the same dollar-equivalent association with SWB as a slightly larger change in tract poverty for families in the bottom quintile because of the concavity between the relationship between family income and SWB.

Table S1. Full list of baseline characteristics (1994-98) controlled for in the main analysis *** = $P<.01$, ** = $P<.05$, * = $P<.10$ on pair wise t-test of the difference between the control group and (a) the two MTO treatment groups (the low-poverty voucher and traditional voucher group pooled together), (b) the low-poverty voucher group, or (c) the tradition voucher group. An omnibus F-test fails to reject the null hypothesis that the set of baseline characteristics presented below is the same for both the control group and the randomly assigned housing voucher treatment groups (p-value for the control vs. pooled treatment groups comparison is $P=0.309$; p-value for the low-poverty housing voucher vs. control comparison is $P=0.473$; and p-value for the traditional housing voucher vs. control comparison is $P=0.268$). All values represent shares (except income). Values are calculated using sample weights to account for changes in random assignment ratios across randomization cohorts, for survey sample selection, and for two-phase interviewing. Missing values (except for income) were imputed based on randomization site and whether randomized through 1997 or in 1998. The baseline head of household reported on the neighborhood characteristics listed here. Data source and sample: Baseline survey. All adults interviewed as part of the long-term survey (N=3,273).

	Control	MTO treatment (voucher) groups	MTO low-poverty voucher group	MTO traditional voucher group
	N=1139	N=2134	N=1456	N=678
Female	0.978	0.984	0.988 *	0.978
Age as of December 31, 2007				
≤ 35	0.143	0.140	0.145	0.132
36-40	0.229	0.222	0.212	0.236
41-45	0.234	0.230	0.236	0.223
46-50	0.175	0.192	0.184	0.203
> 50	0.249	0.246	0.251	0.240
Race and ethnicity				
African-American (any ethnicity)	0.660	0.640	0.648	0.629
Other non-white (any ethnicity)	0.270	0.283	0.283	0.283
Hispanic ethnicity (any race)	0.304	0.325	0.314	0.340
Other demographic characteristics				
Never married	0.637	0.623	0.623	0.624
Parent before age 18	0.246	0.261	0.249	0.277
Working	0.245	0.270	0.271	0.269
Enrolled in school	0.167	0.167	0.161	0.174
High school diploma	0.361	0.367	0.381	0.347
GED	0.199	0.169 *	0.159 **	0.183
Receiving Aid to Families with Dependent Children (AFDC)	0.763	0.752	0.763	0.736
Household characteristics				
Household income (2009 dollars)	$12,438.64	$12,833.64	$12,865.83	$12,788.32
Own car	0.170	0.190	0.190	0.190
Disabled household member	0.148	0.154	0.145	0.168
No teens in household	0.646	0.609 *	0.608 *	0.610
Household size				
Two	0.194	0.218	0.223	0.210
Three	0.330	0.297 *	0.302	0.291
Four or more	0.221	0.235	0.233	0.238

Table S1. (continued)

	Control	MTO treatment (voucher) groups	MTO low-poverty voucher group	MTO traditional voucher group
Site				
Baltimore	0.135	0.136	0.134	0.140
Boston	0.205	0.203	0.201	0.207
Chicago	0.205	0.206	0.205	0.209
Los Angeles	0.226	0.225	0.233	0.214
New York	0.229	0.229	0.227	0.231
Neighborhood characteristics				
Household member was crime victim in last six months	0.416	0.425	0.434	0.414
Streets unsafe at night	0.512	0.503	0.493	0.517
Very dissatisfied with neighborhood	0.467	0.478	0.478	0.477
Lived in neighborhood 5+ years	0.606	0.606	0.599	0.616
Moved more than 3 times in past 5 years	0.108	0.092	0.093	0.090
No family in neighborhood	0.639	0.628	0.640	0.611
No friends in neighborhood	0.409	0.398	0.396	0.400
Chatted with neighbors at least once per week	0.549	0.508 **	0.524	0.486 **
Very likely to tell neighbor about child getting into trouble	0.555	0.541	0.556	0.521
Confident about finding a new apartment	0.456	0.486	0.477	0.499
Had Section 8 voucher before	0.426	0.391 *	0.400	0.379 *
Primary or secondary reason for wanting to move				
To get away from gangs and drugs	0.779	0.770	0.786	0.749
Better schools for children	0.481	0.516 *	0.491	0.553 ***
To get a bigger or better apartment	0.457	0.440	0.441	0.438
To get a job	0.069	0.058	0.063	0.050

Table S2. Comparison of MTO subjective well-being responses on long-term follow-up survey with adults in General Social Survey (GSS). Author calculations from MTO data, and from the GSS, adjusting for sampling weights. GSS results come from the 2000, 2002, 2004, and 2006 waves. The last column represents the GSS sample when it is reweighted to look like the MTO sample with respect to race/ethnicity, age, gender, marital status, family income, urban residence, and educational attainment. The categories on the 3-point happiness scale are: 1=not too happy, 2=pretty happy, 3=very happy.

	MTO				GSS Adults	
	Control group	**Treatment (voucher) groups**	**Low-poverty voucher group**	**Traditional voucher group**	**Full sample**	**Reweighted to match socio-demographics of the MTO sample**
	N=1138	N=2128	N=1455	N=673	N=8311	N=6276
Distribution of responses						
Very happy	0.228	0.252	0.237	0.273	0.336	0.179
Pretty happy	0.497	0.513	0.531	0.488	0.559	0.552
Not too happy	0.275	0.235	0.232	0.239	0.106	0.269
Mean						
Happiness on 3-point scale	1.953	2.017	2.005	2.034	2.230	1.910

Table S3. MTO effects on expanded set of housing and neighborhood condition measures. *** = $P<.01$, ** = $P<.05$, * = $P<.10$ on two-tailed t-test. Robust standard errors shown in parentheses. ITT = Intent-to-Treat or estimated impact of being offered an MTO housing voucher. The control mean is unadjusted. The MTO treatment (voucher) group impacts pooled the low-poverty voucher and traditional voucher groups, while the low-poverty voucher vs. control and traditional voucher vs. control impacts were estimated in separate regressions. All impacts were estimated using a weighted ordinary least squares (OLS) regression model controlling for baseline covariates and field release. Final-evaluation Census tract characteristics are determined by their address on May 31, 2008 (just prior to the start of the long-term survey fielding period). The concentrated disadvantage index is a weighted combination of census tract percent [i] poverty, [ii] on welfare, [iii] unemployed, [iv] female-headed family households, and [v] under age 18, with loading factors developed using 2000 Census tracts in Chicago by Sampson, Sharkey, and Raudenbush *(20)*, but does not include percent African-American. The safety measure reflects whether the respondent felt unsafe or very unsafe (vs. safe or very safe) in the neighborhood during the day. Housing problems include peeling paint, broken plumbing, rats, roaches, broken locks, broken windows, and broken heating system. Data source and sample: Self-reported measures come from the adult long-term survey. Census tract characteristics are interpolated data from the 2000 decennial censuses as well as the 2005-09 American Community Survey. The sample is all adults interviewed.

	Control mean	ITT of MTO treatment (voucher) groups vs. control	ITT of MTO low-poverty voucher vs. control	ITT of MTO traditional voucher vs. control	N
Tract share poor					
At baseline					
Share poor	0.531	-0.004	-0.004	-0.003	3227
		(0.004)	(0.005)	(0.006)	
Share poor, z-score on U.S. tracts	3.172	-0.034	-0.036	-0.021	3227
		(0.034)	(0.037)	(0.049)	
Share poor, z-score on MTO controls	0.000	-0.028	-0.030	-0.018	3227
		(0.029)	(0.031)	(0.041)	
1 year post-random assignment					
Share poor	0.499	-0.160 ***	-0.169 ***	-0.134 ***	3224
		(0.007)	(0.008)	(0.009)	
Share poor, z-score on U.S. tracts	2.916	-1.294 ***	-1.372 ***	-1.085 ***	3224
		(0.054)	(0.062)	(0.073)	
Share poor, z-score on MTO controls	0.000	-0.984 ***	-1.043 ***	-0.825 ***	3224
		(0.041)	(0.047)	(0.056)	
5 years post-random assignment					
Share poor	0.399	-0.089 ***	-0.098 ***	-0.065 ***	3208
		(0.007)	(0.007)	(0.010)	
Share poor, z-score on U.S. tracts	2.109	-0.724 ***	-0.793 ***	-0.526 ***	3208
		(0.056)	(0.060)	(0.083)	
Share poor, z-score on MTO controls	0.000	-0.543 ***	-0.594 ***	-0.394 ***	3208
		(0.042)	(0.045)	(0.062)	
10-15 years post-random assignment (May 2008)					
Share poor	0.311	-0.034 ***	-0.037 ***	-0.021 **	3206
		(0.007)	(0.007)	(0.010)	
Share poor, z-score on U.S. tracts	1.396	-0.275 ***	-0.298 ***	-0.171 **	3206
		(0.053)	(0.057)	(0.080)	
Share poor, z-score on MTO controls	0.000	-0.203 ***	-0.220 ***	-0.126 **	3206
		(0.039)	(0.042)	(0.059)	

Table S3. (continued)

	Control mean	ITT of MTO treatment (voucher) groups vs. control	ITT of MTO low-poverty voucher vs. control	ITT of MTO traditional voucher vs. control	N
Tract share poor (continued)					
Duration-weighted					
Share poor	0.396	-0.082 ***	-0.088 ***	-0.062 ***	3270
		(0.005)	(0.006)	(0.007)	
Share poor, z-score on U.S. tracts	2.082	-0.666 ***	-0.716 ***	-0.501 ***	3270
		(0.041)	(0.046)	(0.058)	
Share poor, z-score on MTO controls	0.000	-0.653 ***	-0.702 ***	-0.491 ***	3270
		(0.040)	(0.045)	(0.057)	
Duration-weighted poverty rate is…					
Less than 20%	0.054	0.196 ***	0.233 ***	0.104 ***	3270
		(0.013)	(0.015)	(0.019)	
Less than 30%	0.242	0.237 ***	0.268 ***	0.148 ***	3270
		(0.018)	(0.019)	(0.027)	
Less than 40%	0.512	0.206 ***	0.199 ***	0.207 ***	3270
		(0.018)	(0.020)	(0.028)	
Tract share minority					
At baseline					
Share minority	0.912	0.003	0.001	0.007	3227
		(0.006)	(0.007)	(0.010)	
Share minority, z-score on U.S. tracts	1.898	0.011	0.005	0.023	3227
		(0.021)	(0.021)	(0.032)	
Share minority, z-score on MTO controls	0.000	0.018	0.008	0.037	3227
		(0.034)	(0.035)	(0.052)	
1 year post-random assignment					
Share minority	0.904	-0.087 ***	-0.111 ***	-0.031 ***	3224
		(0.008)	(0.009)	(0.011)	
Share minority, z-score on U.S. tracts	1.875	-0.279 ***	-0.356 ***	-0.098 ***	3224
		(0.025)	(0.028)	(0.036)	
Share minority, z-score on MTO controls	0.000	-0.450 ***	-0.574 ***	-0.158 ***	3224
		(0.041)	(0.045)	(0.058)	
5 years post-random assignment					
Share minority	0.886	-0.046 ***	-0.056 ***	-0.014	3208
		(0.008)	(0.009)	(0.012)	
Share minority, z-score on U.S. tracts	1.815	-0.147 ***	-0.181 ***	-0.046	3208
		(0.026)	(0.028)	(0.038)	
Share minority, z-score on MTO controls	0.000	-0.231 ***	-0.285 ***	-0.072	3208
		(0.041)	(0.043)	(0.060)	
10-15 years post-random assignment (May 2008)					
Share minority	0.844	-0.024 **	-0.036 ***	0.004	3206
		(0.009)	(0.010)	(0.015)	
Share minority, z-score on U.S. tracts	1.681	-0.076 **	-0.115 ***	0.013	3206
		(0.030)	(0.032)	(0.048)	
Share minority, z-score on MTO controls	0.000	-0.103 **	-0.157 ***	0.018	3206
		(0.041)	(0.043)	(0.065)	

Table S3. (continued)

	Control mean	ITT of MTO treatment (voucher) groups vs. control	ITT of MTO low-poverty voucher vs. control	ITT of MTO traditional voucher vs. control	N
Tract share minority (continued)					
Duration-weighted					
Share minority	0.880	-0.046 ***	-0.060 ***	-0.010	3270
		(0.006)	(0.007)	(0.010)	
Share minority, z-score on U.S. tracts	1.798	-0.148 ***	-0.191 ***	-0.033	3270
		(0.020)	(0.022)	(0.031)	
Share minority, z-score on MTO controls	0.000	-0.285 ***	-0.368 ***	-0.063	3270
		(0.039)	(0.042)	(0.059)	
Other tract characteristics					
10-15 years post-random assignment (May 2008)					
Concentrated disadvantage index	1.128	-0.093 ***	-0.104 ***	-0.053 **	3206
		(0.017)	(0.018)	(0.025)	
Concentrated disadvantage index, z-score on MTO controls	0.000	-0.220 ***	-0.245 ***	-0.125 **	3206
		(0.039)	(0.042)	(0.058)	
Share college graduates	0.220	0.016 ***	0.021 ***	0.003	3206
		(0.006)	(0.006)	(0.009)	
Duration-weighted					
Concentrated disadvantage index	1.389	-0.221 ***	-0.235 ***	-0.171 ***	3270
		(0.014)	(0.016)	(0.020)	
Concentrated disadvantage index, z-score on MTO controls	0.000	-0.599 ***	-0.637 ***	-0.462 ***	3270
		(0.038)	(0.042)	(0.053)	
Share college graduates	0.161	0.034 ***	0.042 ***	0.014 **	3270
		(0.004)	(0.004)	(0.005)	
Residential mobility					
Number of moves after random assignment	2.165	0.584 ***	0.555 ***	0.588 ***	3273
		(0.068)	(0.073)	(0.103)	
Safety, housing and neighborhood problems, and social networks					
Feel unsafe during day	0.196	-0.039 **	-0.036 **	-0.047 **	3262
		(0.015)	(0.016)	(0.023)	
Number of housing problems (0-7)	2.051	-0.380 ***	-0.359 ***	-0.395 ***	3267
		(0.076)	(0.080)	(0.115)	
Likely or very likely to report kids spraying graffiti (collective efficacy)	0.589	0.064 ***	0.078 ***	0.018	3255
		(0.020)	(0.021)	(0.030)	
One or more friends with college degree	0.532	0.049 **	0.071 ***	-0.018	3203
		(0.020)	(0.021)	(0.031)	

Table S4. MTO effects on outcome indices, subjective well-being (SWB), and other mental health measures. *** = $P<.01$, ** = $P<.05$, * = $P<.10$ on two-tailed t-test. Robust standard errors shown in parentheses. ITT = Intent-to-Treat or estimated impact of being offered an MTO housing voucher. The control mean is unadjusted. The MTO treatment (voucher) group impacts pooled the low-poverty voucher and traditional voucher groups, while the low-poverty voucher vs. control and traditional voucher vs. control impacts were estimated in separate regressions. All impacts are weighted estimates controlling for baseline covariates and field release. The ordered probit and ordered logit model rows present raw coefficients. The index components are as follows: economic self-sufficiency – currently self-sufficient (employed and not on TANF), currently employed, 2009 earnings, currently on TANF, and 2009 government income; physical health – self-reported health fair/poor, asthma attack past year, obesity, hypertension, trouble carrying/climbing; mental health – psychological distress index score for the past month, lifetime depression, lifetime Generalized Anxiety Disorder, calm and peaceful during the past month, and normal hours of sleep last night. The K6 index is a sum of five-point likert responses asking how much of the time in the past 30 days respondents felt: so sad that nothing could cheer them up, nervous, restless, hopeless, worthless, and that everything was an effort. For the components of the three indices and the K6, we flipped the coding so that a higher score indicates better outcomes and then z-scored the measure by standardizing against the control group mean and sd. Each index is the average of its components, restandardized using the control mean and sd after averaging. Data source and sample: Adult long-term survey. All interviewed adults.

	Control mean	ITT of MTO treatment (voucher) groups vs. control	ITT of MTO low-poverty voucher vs. control	ITT of MTO traditional voucher vs. control	N
Outcome indices					
Economic self-sufficiency index	0.000	-0.061	-0.029	-0.112 *	3271
		(0.038)	(0.040)	(0.059)	
Physical health index	0.000	0.063	0.056	0.067	3273
		(0.039)	(0.042)	(0.058)	
Mental health index	0.000	0.070 *	0.069	0.063	3273
		(0.041)	(0.042)	(0.062)	
Subjective well-being measures					
Very happy					
OLS model	0.228	0.026	0.010	0.050 *	3266
		(0.017)	(0.018)	(0.027)	
Probit model, average marginal effects	0.228	0.026	0.011	0.050 *	3266
		(0.017)	(0.018)	(0.027)	
Logit model, average marginal effects	0.228	0.026	0.011	0.053 *	3266
		(0.018)	(0.018)	(0.027)	
Very happy or pretty happy					
OLS model	0.725	0.044 **	0.045 **	0.034	3266
		(0.017)	(0.018)	(0.027)	
Probit model, average marginal effects	0.725	0.043 ***	0.045 **	0.034	3266
		(0.017)	(0.018)	(0.026)	
Logit model, average marginal effects	0.725	0.042 **	0.045 **	0.032	3266
		(0.017)	(0.018)	(0.026)	
Happiness 3-point scale					
OLS model	1.953	0.069 **	0.056 *	0.084 *	3266
		(0.028)	(0.029)	(0.043)	
Ordered probit model		0.113 **	0.092 *	0.137 *	3266
		(0.046)	(0.048)	(0.070)	
Ordered logit model		0.192 **	0.160 *	0.230 *	3266
		(0.078)	(0.084)	(0.120)	
Major depression with hierarchy					
Ever had major depression	0.203	-0.037 **	-0.032 *	-0.045 *	3269
		(0.016)	(0.017)	(0.023)	
Absence of psychological distress					
K6 index, reversed and z-scored	0.000	0.104 ***	0.106 **	0.081	3273
		(0.039)	(0.042)	(0.060)	

Table S5. Instrumental variables (IV) estimates of the relationship between MTO adult outcomes from the long-term survey and duration-weighted tract poverty rate or tract share minority, entered one at a time in the IV model. *** = $P<.01$, ** = $P<.05$, * = $P<.10$ on two-tailed t-test. Robust standard errors shown in parentheses. This table reports coefficient estimates for the various IV regressions shown using site and treatment group interactions as instruments. Each regression also controlled for baseline covariates and field release and was weighted. Columns labeled 2SLS report results for two-stage least squares, LIML is an unmodified limited information maximum likelihood (LIML) model, and those labeled Fuller present Fuller-modified LIML models with constants 1, 2 and 4, respectively. All measures are standardized on the control group mean and sd. See Table S4 for a description of the indices. Subjective well-being (SWB) scale refers to the 3-point happiness scale (1=not too happy, 2=pretty happy, 3=very happy). Share poor is the fraction of census tract residents living below the poverty threshold, and share minority is the fraction census tract minority residents. Both share poor and share minority are average measures weighted by the amount of time respondents lived at each of their addresses between random assignment and May 31, 2008 (just prior to the start of the long-term survey fielding period). Data source and sample: SWB and the index components were self-reported on the MTO long-term survey. Share poor and share minority come from interpolated data from the 1990 and 2000 decennial census as well as the 2005-09 American Community Survey. The sample is all adults interviewed as part of the long-term survey (N=3,273).

	Model					First Stage Statistics	
Outcome and Single Mediator Included in Model	**2SLS**	**LIML**	**Fuller (c=1)**	**Fuller (c=2)**	**Fuller (c=4)**	**Partial R-Sq.**	**Angrist-Pischke F-stat**
Outcome=Economic self-sufficiency index							
Share poor (duration-weighted)	0.043	0.048	0.047	0.047	0.046	0.097	29.827
	(0.054)	(0.056)	(0.056)	(0.055)	(0.055)		
Share minority (duration-weighted)	0.028	0.033	0.032	0.032	0.031	0.035	10.493
	(0.095)	(0.108)	(0.107)	(0.106)	(0.104)		
Outcome=Physical health index							
Share poor (duration-weighted)	-0.106 *	-0.112 *	-0.111 *	-0.111 *	-0.110 *	0.096	29.648
	(0.054)	(0.058)	(0.058)	(0.058)	(0.057)		
Share minority (duration-weighted)	-0.087	-0.106	-0.105	-0.104	-0.102	0.035	10.509
	(0.096)	(0.120)	(0.119)	(0.118)	(0.115)		
Outcome=Mental health index							
Share poor (duration-weighted)	-0.104 *	-0.106 *	-0.105 *	-0.105 *	-0.105 *	0.096	29.648
	(0.057)	(0.058)	(0.058)	(0.058)	(0.058)		
Share minority (duration-weighted)	-0.151	-0.161	-0.160	-0.159	-0.156	0.035	10.509
	(0.101)	(0.110)	(0.109)	(0.108)	(0.106)		
Outcome=Subjective well-being scale							
Share poor (duration-weighted)	-0.141 ***	-0.143 **	-0.143 **	-0.143 **	-0.142 ***	0.098	30.265
	(0.054)	(0.056)	(0.056)	(0.055)	(0.055)		
Share minority (duration-weighted)	-0.069	-0.073	-0.073	-0.073	-0.072	0.035	10.697
	(0.098)	(0.115)	(0.114)	(0.113)	(0.111)		

Table S6. Instrumental variables (IV) estimates of the relationship between MTO adult outcomes from the long-term survey and tract poverty rate or tract share minority at the start of the long-term survey fielding period (May 2008), entered one at time in the IV model. *** = $P<.01$, ** = $P<.05$, * = $P<.10$ on two-tailed t-test. Robust standard errors shown in parentheses. This table reports coefficient estimates for the various IV regressions shown using site and treatment group interactions as instruments. Each regression also controlled for baseline covariates and field release and was weighted. Columns labeled 2SLS report results for two-stage least squares, LIML is an unmodified limited information maximum likelihood (LIML) model, and those labeled Fuller present Fuller-modified LIML models with constants 1, 2 and 4, respectively. All measures are standardized on the control group mean and sd. See Table S4 for a description of the indices. Subjective well-being (SWB) scale refers to the 3-point happiness scale (1=not too happy, 2=pretty happy, 3=very happy). Share poor is the fraction of census tract residents living below the poverty threshold, and share minority is the fraction census tract minority residents. Both share poor and share minority represent the address where the respondent was living on May 31, 2008 (just prior to the start of the long-term survey fielding period). Data source and sample: SWB and the index components were self-reported on the MTO long-term survey. Share poor and share minority come from interpolated data from the 2000 decennial census as well as the 2005-09 American Community Survey. The sample is all adults interviewed as part of the long-term survey (N=3,273).

	Model					First Stage Statistics	
Outcome and Single Mediator Included in Model	**2SLS**	**LIML**	**Fuller (c=1)**	**Fuller (c=2)**	**Fuller (c=4)**	**Partial R-Sq.**	**Angrist-Pischke F-stat**
Outcome=Economic self-sufficiency index							
Share poor (May 2008)	0.043	0.086	0.081	0.076	0.067	0.015	4.013
	(0.149)	(0.191)	(0.186)	(0.181)	(0.172)		
Share poor (May 2008)	0.082	0.132	0.125	0.119	0.108	0.011	2.747
	(0.171)	(0.242)	(0.233)	(0.224)	(0.208)		
Outcome=Physical health index							
Share poor (May 2008)	-0.382 **	-0.521 **	-0.508 **	-0.495 **	-0.473 **	0.015	3.985
	(0.159)	(0.241)	(0.233)	(0.225)	(0.212)		
Share poor (May 2008)	-0.216	-0.451	-0.426	-0.404	-0.366	0.011	2.787
	(0.174)	(0.431)	(0.401)	(0.375)	(0.332)		
Outcome=Mental health index							
Share poor (May 2008)	-0.281 *	-0.336	-0.329	-0.323	-0.310 *	0.015	3.985
	(0.164)	(0.208)	(0.203)	(0.197)	(0.188)		
Share poor (May 2008)	-0.232	-0.317	-0.306	-0.296	-0.279	0.011	2.787
	(0.187)	(0.278)	(0.267)	(0.256)	(0.237)		
Outcome=Subjective well-being scale							
Share poor (May 2008)	-0.422 ***	-0.503 **	-0.493 **	-0.483 **	-0.465 **	0.016	4.116
	(0.159)	(0.204)	(0.198)	(0.193)	(0.183)		
Share poor (May 2008)	-0.107	-0.153	-0.149	-0.144	-0.137	0.011	2.765
	(0.179)	(0.385)	(0.363)	(0.343)	(0.310)		

Table S7. Subjective well-being model including natural log of income, General Social Survey (GSS) data years 2000-06. *** = $P<.01$, ** = $P<.05$, * = $P<.10$ on two-tailed t-test. Robust standard errors shown in parentheses. Weighted ordinary least squares. The first column regresses the natural log of income (in 2009 dollars) on the happiness scale measure, controlling for the demographic characteristics of the GSS survey respondent presented above. The instrumental variable (IV) regression in the second column uses education level groups as the instruments for the natural log of income. The education categories are less than high school, high school (omitted category), associate degree, bachelor's degree, and master's degree. The categories on the 3-point happiness scale are: 1=not too happy, 2=pretty happy, 3=very happy.

	Dependent variable = 3-point happiness scale	**Dependent variable = 3-point happiness scale**
	Estimation = Ordinary least squares	**Estimation = Instrumental variables (educational attainment as instruments for log of income)**
Income		
Family Income, thousands (natural log)	0.091 *** (0.009)	0.196 *** (0.028)
Additional covariates		
Gender and age		
Female	0.032 ** (0.016)	0.044 *** (0.017)
Age	-0.013 *** (0.003)	-0.018 *** (0.003)
Age-squared	0.0001 *** (0.00003)	0.0002 *** (0.00003)
Work status		
Temporarily not working	-0.105 * (0.055)	-0.109 ** (0.055)
Unemployed	-0.122 ** (0.051)	-0.081 (0.053)
Retired	0.057 * (0.033)	0.090 *** (0.034)
Currently in school	0.056 (0.048)	0.106 ** (0.052)
Keeping house	-0.038 (0.027)	0.003 (0.030)
Marital status		
Widowed	-0.311 *** (0.034)	-0.248 *** (0.038)
Divorced	-0.270 *** (0.023)	-0.200 *** (0.029)
Separated	-0.374 *** (0.047)	-0.278 *** (0.053)
Never married	-0.252 *** (0.024)	-0.196 *** (0.028)
Race		
Black	-0.051 ** (0.026)	-0.007 (0.029)
Other	-0.081 *** (0.031)	-0.063 * (0.033)
Constant	2.272 *** (0.089)	1.872 *** (0.128)
Year dummies included?	YES	YES
Regional dummies included?	YES	YES
R^2	0.096	0.074
N	7248	7238

Table S8. Change in neighborhood poverty from 1970 to 2007 for the bottom of the income distribution and its implied impact on happiness. Data source and sample: Average tract poverty rates are calculated using data from the 1970 decennial census and 2005-09 American Community Survey. The income cutoffs for the first quintile are $25,862 and $28,825 for 1969 and 2007 and the cutoffs for the second quintile are $41,379 and $51,217 for 1969 and 2007 (*(94)*, converted to 2009 dollars). The average income of families in the bottom quintile changed from $15,336 to $16,622 between 1969 and 2007 (*(32)*, converted to 2009 dollars). For families in the bottom two quintiles, average income changed from $24,603 in 1969 to $28,123 in 2007. "All tracts" includes all tracts with available data in 1970 (urban areas only) and all U.S. tracts in 2007. "Tracts in Matching Counties" restricts the 2007 tracts to the 615 counties with tract-level data from the 1970 census.

	All Available Tracts	**Tracts in Matching Counties**
Families in Bottom Quintile of Income (Below 20th Percentile)		
1970 Average Tract Poverty Rate	17.59%	17.59%
2007 Average Tract Poverty Rate	20.55%	20.00%
Change in Poverty Rate Between 1970 and 2007	2.96%	2.41%
Implied impact on happiness of change in poverty	-0.0234	-0.0191
Income equivalent of impact on happiness	-$1,725	-$1,422
Families in Bottom Two Quintiles of Income (Below 40th Percentile)		
1970 average tract poverty rate	15.49%	15.49%
2007 average tract poverty rate	17.65%	17.05%
Change in poverty rate between 1970 and 2007	2.16%	1.56%
Implied impact on happiness of change in poverty	-0.0171	-0.0123
Income equivalent of impact on happiness	-$2,053	-$1,499
Proportion of U.S. Families Included		
1970 Tracts (available only for urban areas)	73%	73%
2007 Tracts	100%	71%

Table S9. Instrumental variables (IV) estimates of the relationship between MTO adult outcomes from the long-term survey and duration-weighted tract poverty rate and tract share minority, entered simultaneously in the IV model. *** = $P<.01$, ** = $P<.05$, * = $P<.10$ on two-tailed t-test. Robust standard errors shown in parentheses. This table reports coefficient estimates for the various IV regressions shown using site and treatment group interactions as instruments. Each regression presents coefficients for the respective neighborhood measure controlling for the other mediator listed. Each regression also controlled for baseline covariates and field release and was weighted. Columns labeled 2SLS report results for two-stage least squares, LIML is an unmodified limited information maximum likelihood (LIML) model, and those labeled Fuller present Fuller-modified LIML models with constants 1, 2 and 4, respectively. All measures are standardized on the control group mean and standard deviation. See Table S4 for a description of the indices. Subjective well-being (SWB) scale refers to the 3-point happiness scale (1=not too happy, 2=pretty happy, 3=very happy). Share poor is the fraction of census tract residents living below the poverty threshold, and share minority is the fraction census tract minority residents. Both share poor and share minority are average measures weighted by the amount of time respondents lived at each of their addresses between random assignment and May 31, 2008 (just prior to the start of the long-term survey fielding period). Data source and sample: SWB and the index components were self-reported on the MTO long-term survey. Share poor and share minority come from interpolated data from the 1990 and 2000 decennial census as well as the 2005-09 American Community Survey. The sample is all adults interviewed as part of the long-term survey (N=3,273).

	Model					First Stage Statistics		
Outcome and Both Mediators Included in Model	**2SLS**	**LIML**	**Fuller (c=1)**	**Fuller (c=2)**	**Fuller (c=4)**	**Partial R-Sq.**	**Angrist-Pischke F-stat**	**Cragg-Donald F-stat**
Outcome=Economic self-sufficiency index								
Share poor, controlling for share minority (duration-weighted)	0.073	0.088	0.086	0.085	0.082	0.052	14.126	6.132
	(0.087)	(0.103)	(0.101)	(0.100)	(0.097)			
Share minority, controlling for share poor (duration-weighted)	-0.068	-0.093	-0.091	-0.088	-0.084	0.019	4.484	
	(0.155)	(0.196)	(0.192)	(0.188)	(0.181)			
P-value of test that coefficients are equal	0.539	0.530	0.530	0.531	0.532			
Outcome=Physical health index								
Share poor, controlling for share minority (duration-weighted)	-0.157 *	-0.186	-0.184	-0.182	-0.178	0.053	14.210	6.220
	(0.089)	(0.118)	(0.115)	(0.113)	(0.110)			
Share minority, controlling for share poor (duration-weighted)	0.119	0.173	0.169	0.166	0.158	0.019	4.546	
	(0.159)	(0.233)	(0.227)	(0.222)	(0.212)			
P-value of test that coefficients are equal	0.242	0.290	0.287	0.284	0.278			
Outcome=Mental health index								
Share poor, controlling for share minority (duration-weighted)	-0.089	-0.090	-0.090	-0.090	-0.090	0.053	14.210	6.220
	(0.091)	(0.100)	(0.098)	(0.097)	(0.095)			
Share minority, controlling for share poor (duration-weighted)	-0.034	-0.036	-0.036	-0.036	-0.035	0.019	4.546	
	(0.160)	(0.183)	(0.179)	(0.176)	(0.170)			
P-value of test that coefficients are equal	0.817	0.842	0.838	0.835	0.829			
Outcome=Subjective well-being scale								
Share poor, controlling for share minority (duration-weighted)	-0.261 ***	-0.279 ***	-0.276 ***	-0.273 ***	-0.268 ***	0.052	14.246	6.077
	(0.093)	(0.102)	(0.100)	(0.099)	(0.096)			
Share minority, controlling for share poor (duration-weighted)	0.279 *	0.316 *	0.310 *	0.304 *	0.293 *	0.019	4.552	
	(0.169)	(0.191)	(0.187)	(0.184)	(0.177)			
P-value of test that coefficients are equal	0.030	0.035	0.034	0.033	0.032			

Table S10. Instrumental variables (IV) estimates of the relationship between outcomes and tract poverty rate and tract share minority at the start of the long-term survey fielding period (May 2008), entered simultaneously in the IV model. *** = $P<.01$, ** = $P<.05$, * = $P<.10$ on two-tailed t-test. Robust standard errors shown in parentheses. This table reports coefficient estimates for the various IV regressions shown using site and treatment group interactions as instruments. Each regression presents coefficients for the respective neighborhood measure controlling for the other mediator listed. Each regression also controlled for baseline covariates and field release and was weighted. Columns labeled 2SLS report results for two-stage least squares, LIML is an unmodified limited information maximum likelihood (LIML) model, and those labeled Fuller present Fuller-modified LIML models with constants 1, 2 and 4, respectively. All measures are standardized on the control group mean and standard deviation. See Table S4 for a description of the indices. Subjective well-being (SWB) scale refers to the 3-point happiness scale (1=not too happy, 2=pretty happy, 3=very happy). Share poor is the fraction of census tract residents living below the poverty threshold, and share minority is the fraction census tract minority residents. Both share poor and share minority represent the address where the respondent was living on May 31, 2008 (just prior to the start of the long-term survey fielding period). Data source and sample: SWB and the index components were self-reported on the MTO long-term survey. Share poor and share minority come from interpolated data from the 2000 decennial census as well as the 2005-09 American Community Survey. The sample is all adults interviewed as part of the long-term survey (N=3,273).

	Model					First Stage Statistics		
Outcome and Both Mediators Included in Model	**2SLS**	**LIML**	**Fuller (c=1)**	**Fuller (c=2)**	**Fuller (c=4)**	**Partial R-Sq.**	**Angrist-Pischke F-stat**	**Cragg-Donald F-stat**
Outcome=Economic self-sufficiency index								
Share poor, controlling for share minority (May 2008)	0.014	0.045	0.041	0.038	0.031	0.015	3.511	3.473
	(0.166)	(0.213)	(0.207)	(0.202)	(0.192)			
Share minority, controlling for share poor (May 2008)	0.075	0.108	0.103	0.099	0.092	0.011	2.288	
	(0.191)	(0.272)	(0.261)	(0.251)	(0.233)			
P-value of test that coefficients are equal	0.841	0.879	0.877	0.874	0.867			
Outcome=Physical health index								
Share poor, controlling for share minority (May 2008)	-0.368 **	-0.492 *	-0.480 *	-0.469 *	-0.449 *	0.015	3.492	3.536
	(0.181)	(0.277)	(0.267)	(0.258)	(0.242)			
Share minority, controlling for share poor (May 2008)	-0.037	-0.076	-0.071	-0.067	-0.060	0.011	2.330	
	(0.200)	(0.356)	(0.338)	(0.322)	(0.294)			
P-value of test that coefficients are equal	0.313	0.449	0.435	0.423	0.400			
Outcome=Mental health index								
Share poor, controlling for share minority (May 2008)	-0.236	-0.274	-0.269	-0.264	-0.256	0.015	3.492	3.536
	(0.184)	(0.231)	(0.225)	(0.219)	(0.208)			
Share minority, controlling for share poor (May 2008)	-0.117	-0.156	-0.150	-0.145	-0.136	0.011	2.330	
	(0.207)	(0.283)	(0.272)	(0.262)	(0.244)			
P-value of test that coefficients are equal	0.722	0.788	0.779	0.771	0.756			
Outcome=Subjective well-being scale								
Share poor, controlling for share minority (May 2008)	-0.473 **	-0.565 **	-0.553 **	-0.542 **	-0.520 **	0.015	3.574	3.455
	(0.184)	(0.233)	(0.227)	(0.220)	(0.209)			
Share minority, controlling for share poor (May 2008)	0.134	0.183	0.177	0.170	0.158	0.011	2.274	
	(0.218)	(0.304)	(0.291)	(0.280)	(0.259)			
P-value of test that coefficients are equal	0.080	0.108	0.104	0.100	0.094			

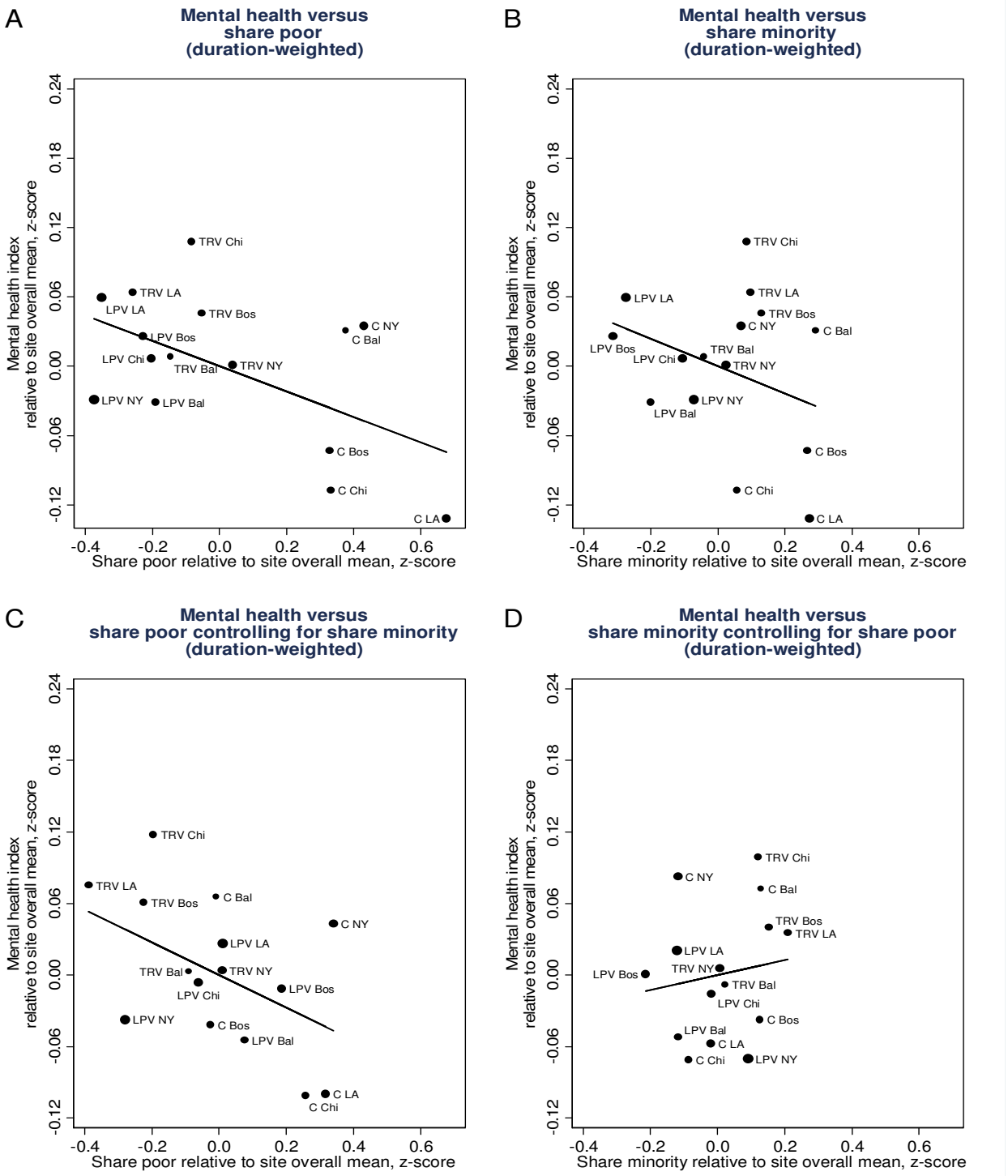

Fig. S1. Instrumental variable estimation of the relationship between mental health (higher values indicate better mental health) and average (duration-weighted) tract poverty rate (panel A), tract share minority (panel B), tract poverty controlling for minority share (panel C), and tract minority share controlling for tract poverty (panel D).

Fig. S1. (continued)
The y-axis is the mental health index expressed in standard deviation units relative to the control group standard deviation. The index has the following components: psychological distress index score for the past month, lifetime depression, lifetime Generalized Anxiety Disorder, calm and peaceful during the past month, and normal hours of sleep last night. Each component was standardized using the mean and standard deviation for the control group, and negative items were flipped such that higher values indicate better outcomes. The index is the average of its components, restandardized using the control mean and standard deviation after averaging. Share poor (shown on the x-axis of Panels A and C) is the fraction of census tract residents living below the poverty threshold. Share minority (shown on the x-axis of panels B and D) represents the fraction of census tract residents who are members of racial or ethnic minority groups. Tract shares are linearly interpolated from the 1990 and 2000 decennial census and 2005-09 American Community Survey and are weighted by the time respondents lived at each of their addresses from random assignment through May 2008. Both measures are expressed as z-scores standardized by the control group mean and standard deviation. The points represent the site (Bal = Baltimore, Bos = Boston, Chi = Chicago, LA = Los Angeles, NY = New York City) and treatment group (LPV = low-poverty voucher, TRV = traditional voucher, C = control group). The line through the data points is equivalent to a two-stage least-squares estimate of the relationship between subjective well-being and the mediator, using site-group interactions as instruments for the mediator (conditional on site main effects). The size of each point is proportional to the sum of the weights for that group and, correspondingly, to the weight that the point receives in the two-stage least squares regression. The estimated impact of 1 standard deviation (sd) decrease in poverty (Panel A) is 0.110sd better mental health (N=3270, SE=0.057, P=0.052), and the estimated impact of 1sd decrease in minority share (Panel B) is 0.119sd better mental health (N=3270, SE=0.103, P=0.250). The estimated impact of 1sd decrease in poverty controlling for minority share (Panel C) is 0.137sd better mental health (N=3270, SE=0.091, P=0.134), and the estimated impact of 1sd decrease in minority share controlling for poverty (Panel D) is 0.062sd worse mental health (N=3270, SE=0.164, P=0.704). The p-value from an F test of whether the coefficients on poverty and minority share are the same (that is, whether the slope in panel C equals the slope in panel D) is 0.411.

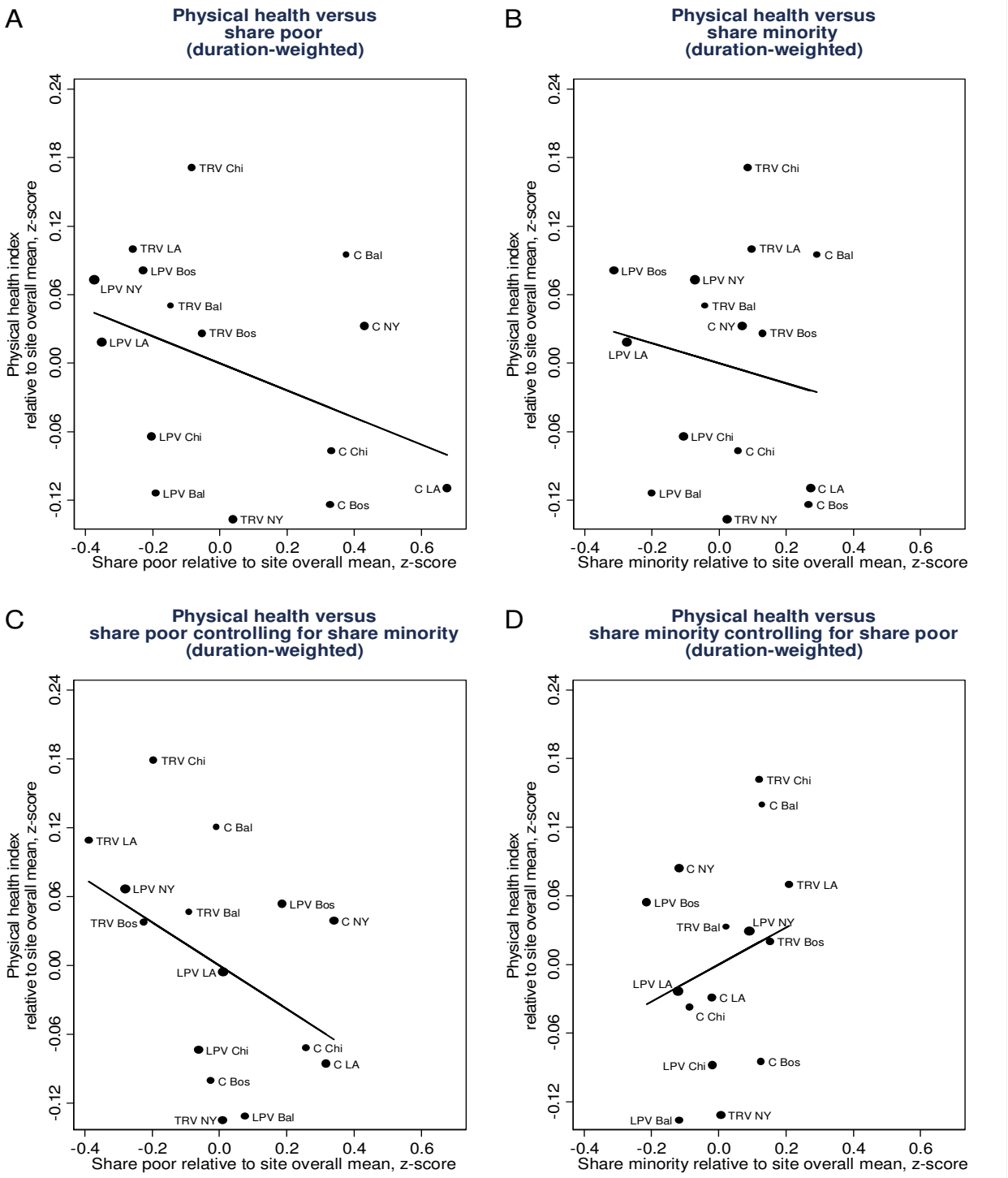

Fig. S2. Instrumental variable estimation of the relationship between physical health (higher values indicate better physical health) and average (duration-weighted) tract poverty rate (panel A), tract share minority (panel B), tract poverty controlling for minority share (panel C), and tract minority share controlling for tract poverty (panel D).

Fig. S2. (continued)
The y-axis is the physical health index expressed in standard deviation units relative to the control group standard deviation. The index has the following components: self-reported health fair/poor, asthma attack past year, obesity, hypertension, trouble carrying/climbing. Each component was standardized using the mean and standard deviation for the control group, and negative items were flipped such that higher values indicate better outcomes. The index is the average of its components, restandardized using the control mean and standard deviation after averaging. Share poor (shown on the x-axis of Panels A and C) is the fraction of census tract residents living below the poverty threshold. Share minority (shown on the x-axis of panels B and D) represents the fraction of census tract residents who are members of racial or ethnic minority groups. Tract shares are linearly interpolated from the 1990 and 2000 decennial census and 2005-09 American Community Survey and are weighted by the time respondents lived at each of their addresses from random assignment through May 2008. Both measures are expressed as z-scores standardized by the control group mean and standard deviation. The points represent the site (Bal = Baltimore, Bos = Boston, Chi = Chicago, LA = Los Angeles, NY = New York City) and treatment group (LPV = low-poverty voucher, TRV = traditional voucher, C = control group). The line through the data points is equivalent to a two-stage least-squares estimate of the relationship between subjective well-being and the mediator, using site-group interactions as instruments for the mediator (conditional on site main effects). The size of each point is proportional to the sum of the weights for that group and, correspondingly, to the weight that the point receives in the two-stage least squares regression. The estimated impact of 1 standard deviation (sd) decrease in poverty (Panel A) is 0.119sd better physical health (N=3270, SE=0.056, P=0.035), and the estimated impact of 1sd decrease in minority share (Panel B) is 0.088sd better physical health (N=3270, SE=0.102, P=0.388). The estimated impact of 1sd decrease in poverty controlling for minority share (Panel C) is 0.188sd better physical health (N=3270, SE=0.095, P=0.047), and the estimated impact of 1sd decrease in minority share controlling for neighborhood poverty (Panel D) is 0.162sd worse physical health (N=3270, SE=0.174, P=0.354). The p-value from an F test of whether the coefficients on poverty and minority share are the same (that is, whether the slope in panel C equals the slope in panel D) is 0.172.

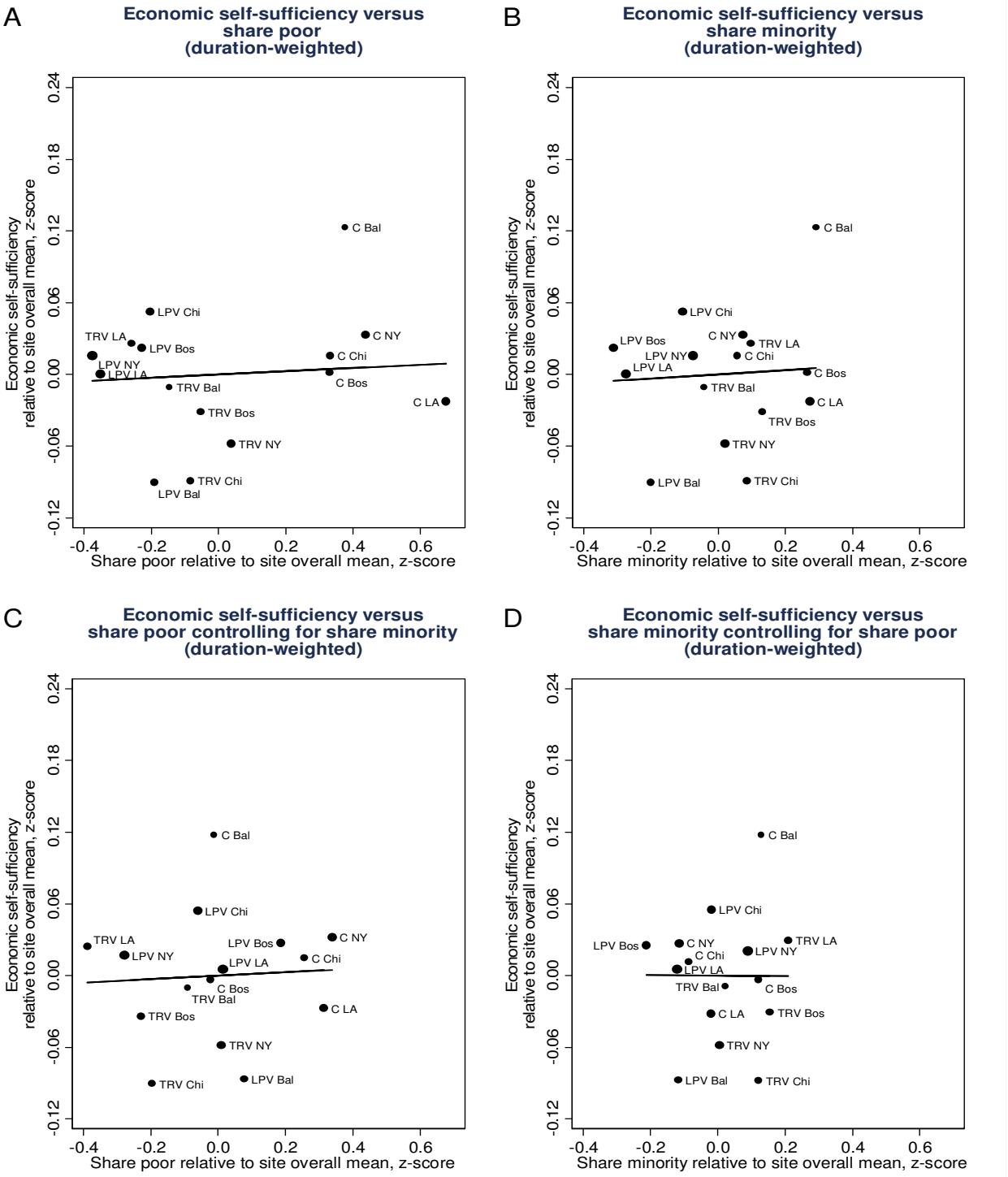

Fig. S3. Instrumental variable estimation of the relationship between economic self-sufficiency (higher values indicate greater economic self-sufficiency) and average (duration-weighted) tract poverty rate (panel A), tract share minority (panel B), tract poverty controlling for minority share (panel C), and tract minority share controlling for tract poverty (panel D).

Fig. S3. (continued)
The y-axis is the economic self-sufficiency index expressed in standard deviation units relative to the control group standard deviation. The index has the following components: currently self-sufficient (employed and not on TANF), currently employed, 2009 earnings, currently on TANF, and 2009 government income. Each component was standardized using the mean and standard deviation for the control group, and negative items were flipped such that higher values indicate better outcomes. The index is the average of its components, restandardized using the control mean and standard deviation after averaging. Share poor (shown on the x-axis of Panels A and C) is the fraction of census tract residents living below the poverty threshold. Share minority (shown on the x-axis of panels B and D) represents the fraction of census tract residents who are members of racial or ethnic minority groups. Tract shares are linearly interpolated from the 1990 and 2000 decennial census and 2005-09 American Community Survey and are weighted by the time respondents lived at each of their addresses from random assignment through May 2008. Both measures are expressed as z-scores standardized by the control group mean and standard deviation. The points represent the site (Bal = Baltimore, Bos = Boston, Chi = Chicago, LA = Los Angeles, NY = New York City) and treatment group (LPV = low-poverty voucher, TRV = traditional voucher, C = control group). The line through the data points is equivalent to a two-stage least-squares estimate of the relationship between subjective well-being and the mediator, using site-group interactions as instruments for the mediator (conditional on site main effects). The size of each point is proportional to the sum of the weights for that group and, correspondingly, to the weight that the point receives in the two-stage least squares regression. The estimated impact of 1 standard deviation (sd) decrease in poverty (Panel A) is 0.018sd lower economic self-sufficiency (N=3268, SE=0.056, P=0.746), and the estimated impact of 1sd decrease in minority share (Panel B) is 0.023sd lower economic self-sufficiency (N=3268, SE=0.102, P=0.822). The estimated impact of 1sd decrease in poverty controlling for minority share (Panel C) is 0.019sd lower economic self-sufficiency (N=3268, SE=0.095, P=0.838), and the estimated impact of 1sd decrease in minority share controlling for neighborhood poverty (Panel D) is 0.003sd higher economic self-sufficiency (N=3268, SE=0.171, P=0.987). The p-value from an F test of whether the coefficients on poverty and minority share are the same (that is, whether the slope in panel C equals the slope in panel D) is 0.930.

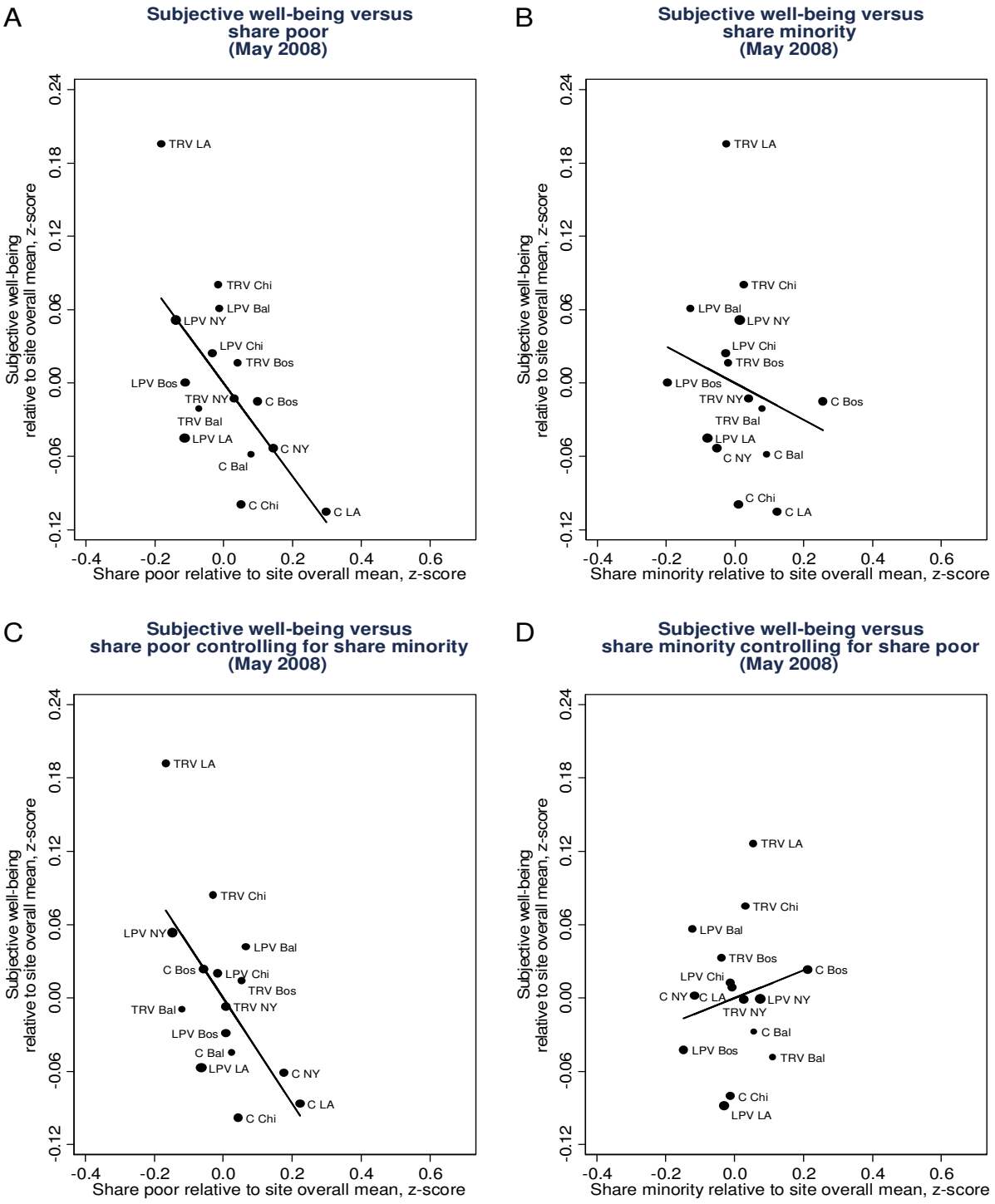

Fig. S4. Instrumental variable estimation of the relationship between subjective well-being (SWB) and May 2008 tract poverty rate (panel A), tract share minority (panel B), tract poverty controlling for minority share (panel C), and tract minority share controlling for tract poverty (panel D).

Fig. S4. (continued)
The y-axis is a 3-point happiness scale (1=not too happy, 2=pretty happy, 3=very happy) expressed in standard deviation (sd) units relative to the control group. Share poor is the fraction of census tract residents living below the poverty threshold. Share minority is the fraction of census tract residents who are members of racial or ethnic minority groups. Tract shares are linearly interpolated from the 2000 decennial census and 2005-09 American Community Survey and represent the address where the respondent was living on May 31, 2008 (just prior to the start of the long-term survey fielding period). Share poor and minority are z-scores, standardized by the control group mean and sd. The points represent the site (Bal = Baltimore, Bos = Boston, Chi = Chicago, LA = Los Angeles, NY = New York City) and treatment group (LPV = low-poverty voucher, TRV = traditional voucher, C = control group). The slope of the line is equivalent to a 2SLS estimate of the relationship between SWB and the mediator, using site-group interactions as instruments for the mediator (conditional on site main effects). The estimated impact of 1sd decrease in poverty (Panel A) is a 0.381sd increase in SWB (N=3199, SE=0.154, P=0.014), and the estimated impact of 1sd decrease in minority share (Panel B) is a 0.150sd increase in SWB (N=3199, SE=0.177, P=0.398). The estimated impact of 1sd decrease in poverty controlling for minority share (Panel C) is a 0.431sd increase in SWB (N=3199, SE=0.188, P=0.022), and the estimated impact of 1sd decrease in minority share controlling for poverty (Panel D) is a 0.113sd decrease in SWB (N=3199, SE=0.224, P=0.613). The p-value from an F test of whether the coefficients on poverty and minority share are the same (that is, whether the slope in panel C equals the slope in panel D) is 0.135.

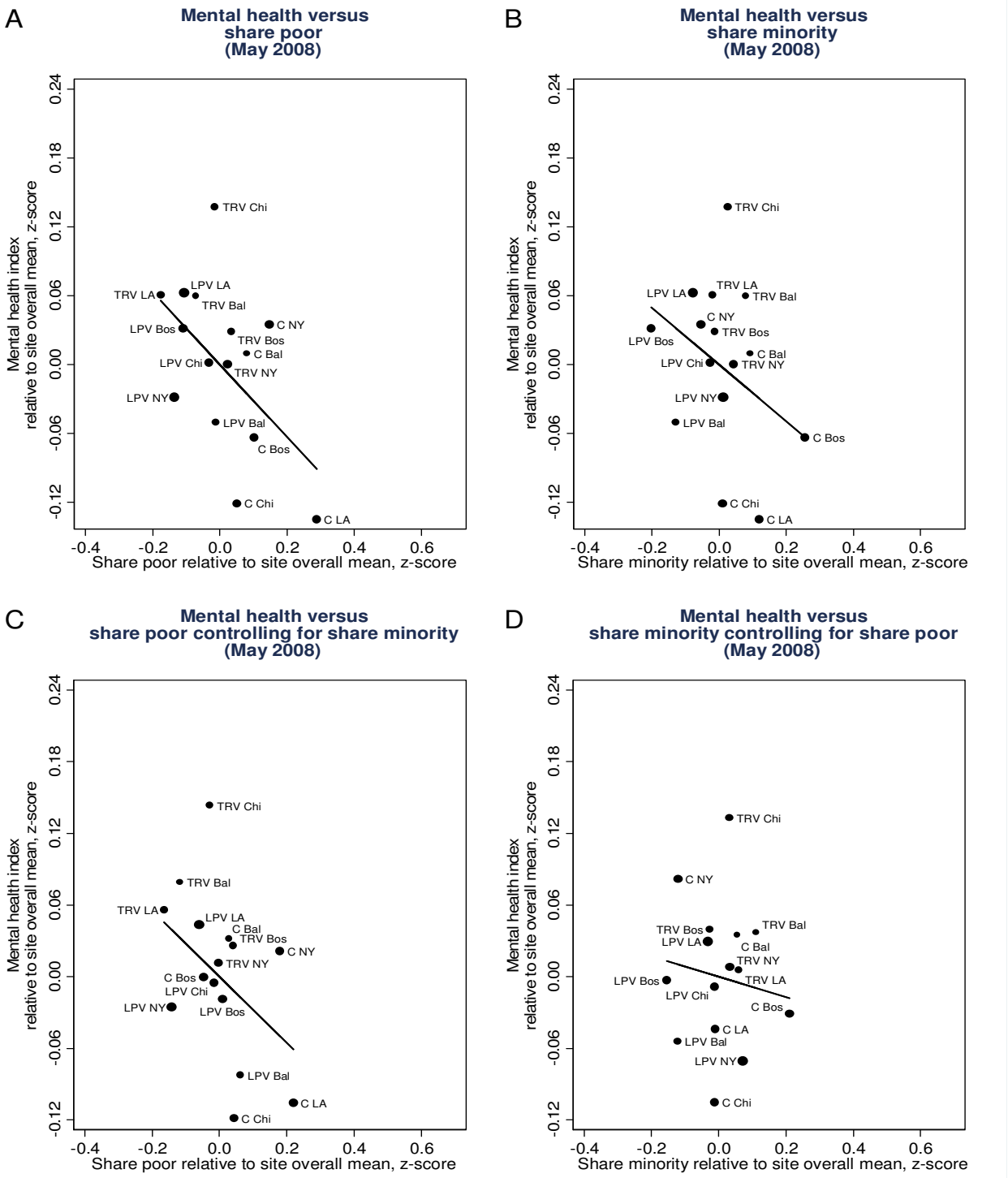

Fig. S5. Instrumental variable estimation of the relationship between mental health (higher values indicate better mental health) and May 2008 tract poverty rate (panel A), tract share minority (panel B), tract poverty controlling for minority share (panel C), and tract minority share controlling for tract poverty (panel D).

Fig. S5. (continued)
The y-axis is the mental health index expressed in standard deviation units relative to the control group standard deviation. The index has the following components: psychological distress index score for the past month, lifetime depression, lifetime Generalized Anxiety Disorder, calm and peaceful during the past month, and normal hours of sleep last night. Each component was standardized using the mean and standard deviation for the control group, and negative items were flipped such that higher values indicate better outcomes. The index is the average of its components, restandardized using the control mean and standard deviation after averaging. Share poor (shown on the x-axis of Panels A and C) is the fraction of census tract residents living below the poverty threshold. Share minority (shown on the x-axis of panels B and D) represents the fraction of census tract residents who are members of racial or ethnic minority groups. Share poor and minority are linearly interpolated from the 2000 decennial census and 2005-09 American Community Survey and represent the address where the respondent was living on May 31, 2008 (just prior to the start of the long-term survey fielding period). Both measures are expressed as z-scores standardized by the control group mean and standard deviation. The points represent the site (Bal = Baltimore, Bos = Boston, Chi = Chicago, LA = Los Angeles, NY = New York City) and treatment group (LPV = low-poverty voucher, TRV = traditional voucher, C = control group). The line through the data points is equivalent to a two-stage least-squares estimate of the relationship between subjective well-being and the mediator, using site-group interactions as instruments for the mediator (conditional on site main effects). The size of each point is proportional to the sum of the weights for that group and, correspondingly, to the weight that the point receives in the two-stage least squares regression. The estimated impact of 1 standard deviation (sd) decrease in poverty (Panel A) is 0.315sd better mental health (N=3206, SE=0.160, P=0.048), and the estimated impact of 1sd decrease in minority share (Panel B) is 0.247sd better mental health (N=3206, SE=0.189, P=0.191). The estimated impact of 1sd decrease in poverty controlling for minority share (Panel C) is 0.277sd better mental health (N=3206, SE=0.186, P=0.135), and the estimated impact of 1sd decrease in minority share controlling for poverty (Panel D) is 0.085sd better mental health (N=3206, SE=0.217, P=0.694). The p-value from an F test of whether the coefficients on poverty and minority share are the same (that is, whether the slope in panel C equals the slope in panel D) is 0.584.

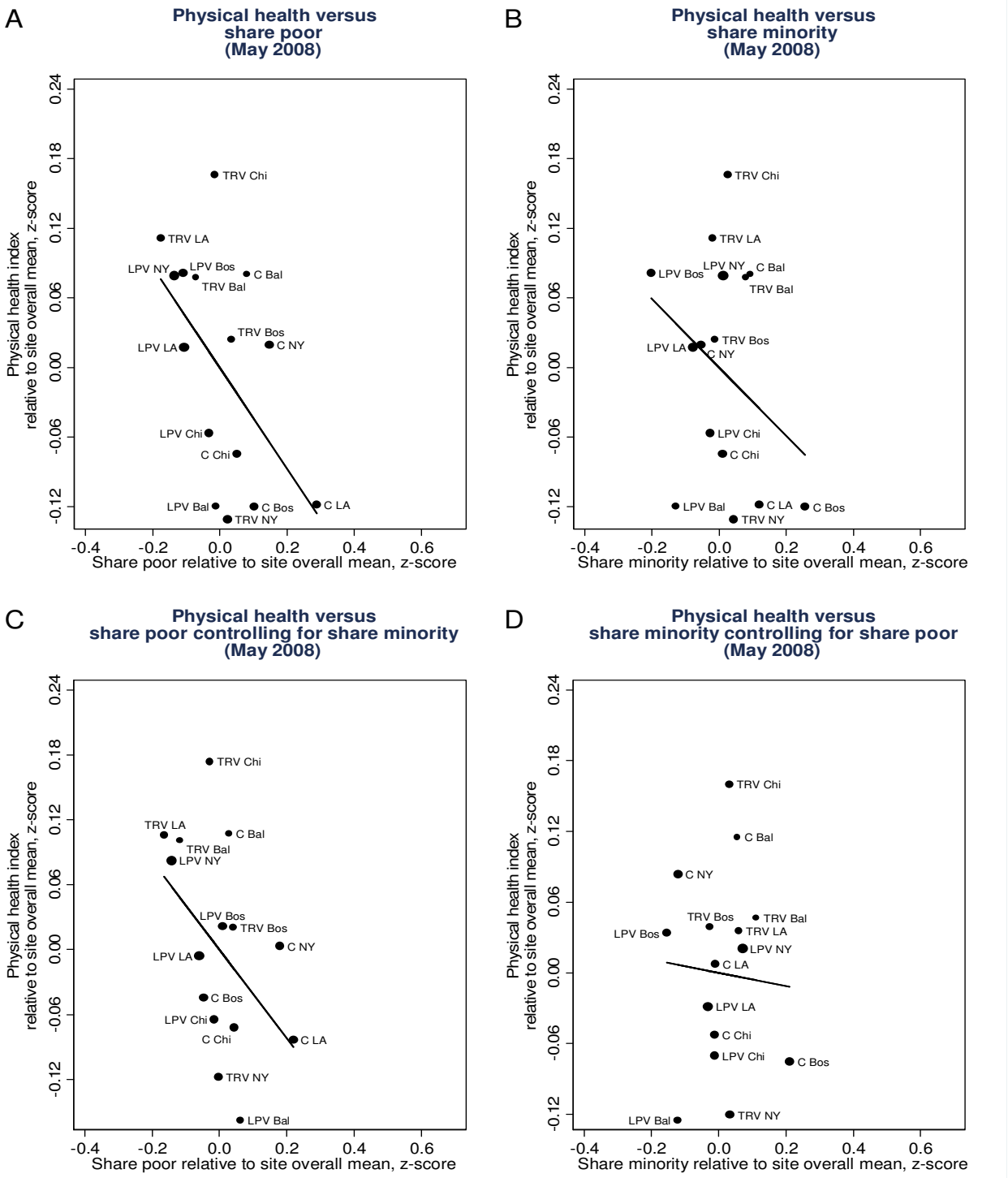

Fig. S6. Instrumental variable estimation of the relationship between physical health (higher values indicate better physical health) and May 2008 tract poverty rate (panel A), tract share minority (panel B), tract poverty controlling for minority share (panel C), and tract minority share controlling for tract poverty (panel D).

Fig. S6. (continued)
The y-axis is the physical health index expressed in standard deviation units relative to the control group standard deviation. The index has the following components: self-reported health fair/poor, asthma attack past year, obesity, hypertension, trouble carrying/climbing. Each component was standardized using the mean and standard deviation for the control group, and negative items were flipped such that higher values indicate better outcomes. The index is the average of its components, restandardized using the control mean and standard deviation after averaging. Share poor (shown on the x-axis of Panels A and C) is the fraction of census tract residents living below the poverty threshold. Share minority (shown on the x-axis of panels B and D) represents the fraction of census tract residents who are members of racial or ethnic minority groups. Share poor and minority are linearly interpolated from the 2000 decennial census and 2005-09 American Community Survey and are determined by the address where the respondent was living on May 31, 2008 (just prior to the start of the long-term survey fielding period). Both measures are expressed as z-scores standardized by the control group mean and standard deviation. The points represent the site (Bal = Baltimore, Bos = Boston, Chi = Chicago, LA = Los Angeles, NY = New York City) and treatment group (LPV = low-poverty voucher, TRV = traditional voucher, C = control group). The line through the data points is equivalent to a two-stage least-squares estimate of the relationship between subjective well-being and the mediator, using site-group interactions as instruments for the mediator (conditional on site main effects). The size of each point is proportional to the sum of the weights for that group and, correspondingly, to the weight that the point receives in the two-stage least squares regression. The estimated impact of 1 standard deviation (sd) decrease in poverty (Panel A) is 0.434sd better physical health (N=3206, SE=0.159, P=0.006), and the estimated impact of 1sd decrease in minority share (Panel B) is 0.294sd better physical health (N=3206, SE=0.182, P=0.106). The estimated impact of 1sd decrease in poverty controlling for minority share (Panel C) is 0.409sd better physical health (N=3206, SE=0.190, P=0.031), and the estimated impact of 1sd decrease in minority share controlling for neighborhood poverty (Panel D) is 0.056sd better physical health (N=3206, SE=0.216, P=0.797). The p-value from an F test of whether the coefficients on poverty and minority share are the same (that is, whether the slope in panel C equals the slope in panel D) is 0.322.

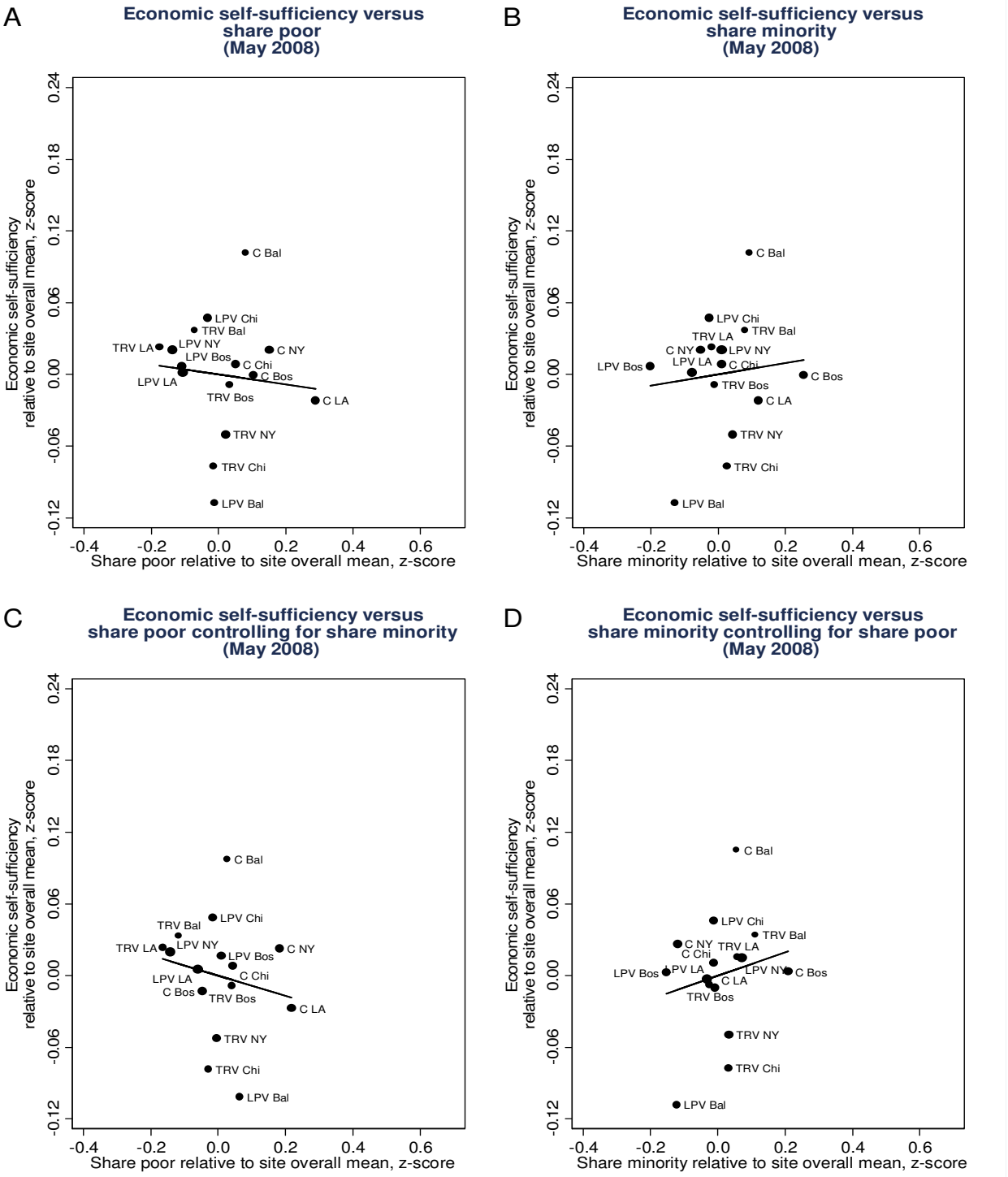

Fig. S7. Instrumental variable estimation of the relationship between economic self-sufficiency (higher values indicate greater economic self-sufficiency) and May 2008 tract poverty rate (panel A), tract share minority (panel B), tract poverty controlling for minority share (panel C), and tract minority share controlling for tract poverty (panel D).

Fig. S7. (continued)
The y-axis is the economic self-sufficiency index expressed in standard deviation units relative to the control group standard deviation. The index has the following components: currently self-sufficient (employed and not on TANF), currently employed, 2009 earnings, currently on TANF, and 2009 government income. Each component was standardized using the mean and standard deviation for the control group, and negative items were flipped such that higher values indicate better outcomes. The index is the average of its components, restandardized using the control mean and standard deviation after averaging. Share poor (shown on the x-axis of Panels A and C) is the fraction of census tract residents living below the poverty threshold. Share minority (shown on the x-axis of panels B and D) represents the fraction of census tract residents who are members of racial or ethnic minority groups. Share poor and minority are linearly interpolated from the 2000 decennial census and 2005-09 American Community Survey and are determined by the address where the respondent was living on May 31, 2008 (just prior to the start of the long-term survey fielding period). Both measures are expressed as z-scores standardized by the control group mean and standard deviation. The points represent the site (Bal = Baltimore, Bos = Boston, Chi = Chicago, LA = Los Angeles, NY = New York City) and treatment group (LPV = low-poverty voucher, TRV = traditional voucher, C = control group). The line through the data points is equivalent to a two-stage least-squares estimate of the relationship between subjective well-being and the mediator, using site-group interactions as instruments for the mediator (conditional on site main effects). The size of each point is proportional to the sum of the weights for that group and, correspondingly, to the weight that the point receives in the two-stage least squares regression. The estimated impact of 1 standard deviation (sd) decrease in poverty (Panel A) is 0.055sd higher economic self-sufficiency (N=3204, SE=0.153, P=0.721), and the estimated impact of 1sd decrease in minority share (Panel B) is 0.063sd lower economic self-sufficiency (N=3204, SE=0.183, P=0.731). The estimated impact of 1sd decrease in poverty controlling for minority share (Panel C) is 0.112sd higher economic self-sufficiency (N=3204, SE=0.180, P=0.536), and the estimated impact of 1sd decrease in minority share controlling for neighborhood poverty (Panel D) is 0.129sd lower economic self-sufficiency (N=3204, SE=0.213, P=0.546). The p-value from an F test of whether the coefficients on poverty and minority share are the same (that is, whether the slope in panel C equals the slope in panel D) is 0.484.

References and Notes

1. E. Kneebone, C. Nadeau, A. Berube, *The Re-Emergence of Concentrated Poverty: Metropolitan Trends in the 2000s* (2011); available at www.brookings.edu/~/media/Files/rc/papers/2011/1103_poverty_kneebone_nadeau_berube/1103_poverty_kneebone_nadeau_berube.pdf.

2. U.S. Census Bureau, *Poverty Thresholds for 2011 by Size of Family and Number of Related Children Under 18 Years* (2012); available at www.census.gov/hhes/www/poverty/data/threshld/thresh11.xls (accessed July 12 2012).

3. R. J. Sampson, S. W. Raudenbush, F. Earls, Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science* **277**, 918 (1997). doi:10.1126/science.277.5328.918 Medline

4. S. A. Macintyre, A. Ellaway, in *Neighborhoods and Health*, I. Kawachi, L. F. Berkman, Eds. (Oxford Univ. Press, New York, 2003), pp. 20–42.

5. R. J. Sampson, J. D. Morenoff, T. Gannon-Rowley, Assessing "neighborhood effects": Social processes and new directions in research. *Annu. Rev. Sociol.* **28**, 443 (2002). doi:10.1146/annurev.soc.28.110601.141114

6. I. Kawachi, L. F. Berkman, Eds., *Neighborhoods and Health* (Oxford Univ. Press, New York, 2003).

7. J. R. Kling, J. B. Liebman, L. F. Katz, Experimental Analysis of Neighborhood Effects. *Econometrica* **75**, 83 (2007). doi:10.1111/j.1468-0262.2007.00733.x

8. G. Myrdal, *An American Dilemma: The Negro Problem and Modern Democracy, Volume 1* (Harper Brothers, New York, 1944).

9. U.S. National Advisory Commission on Civil Disorders, *The Kerner Report* (1968; www.eisenhowerfoundation.org/docs/kerner.pdf).

10. D. S. Massey, N. Denton, *American Apartheid: Segregation and the Making of the Underclass* (Harvard Univ. Press, Cambridge, MA, 1993).

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

12. E. L. Glaeser, J. L. Vigdor, *The End of the Segregated Century: Racial Separation in America's Neighborhoods, 1980–2010* (Manhattan Institute Civic Report No. 66, New York, 2012); available at www.manhattan-institute.org/pdf/cr_66.pdf.

13. T. Watson, Inequality and the measurement of residential segregation by income in American neighborhoods. *Rev. Income Wealth* **55**, 820 (2009). doi:10.1111/j.1475-4991.2009.00346.x

14. S. F. Reardon, K. Bischoff, *More Unequal and More Separate: Growth in the Residential Segregation of Families by Income, 1970-2009* (Report prepared for the Russell Sage Foundation US2010 project, New York, 2011); available at www.s4.brown.edu/us2010/Data/Report/report111111.pdf.

15. J. Ludwig *et al.*, Neighborhoods, obesity, and diabetes—A randomized social experiment. *N. Engl. J. Med.* **365**, 1509 (2011). doi:10.1056/NEJMsa1103216 Medline

16. R. J. Sampson, *Great American City: Chicago and the Enduring Neighborhood Effect* (Univ. Chicago Press, Chicago, 2012).

17. T. W. Smith, P. Marsden, M. Hout, J. Kim, General social surveys, 1972-2010 [machine-readable data file] Sponsored by National Science Foundation. NORC ed. Chicago: National Opinion Research Center [producer]; Storrs, CT: The Roper Center for Public Opinion Research, University of Connecticut (2011).

18. P. Dolan, R. Layard, R. Metcalfe, Measuring Subjective wellbeing for public policy: Recommendations on measures (Special Paper No. 23) (2011); available at http://cep.lse.ac.uk/pubs/download/special/cepsp23.pdf.

19. E. S. Huebner, Research on assessment of life satisfaction of children and adolescents. *Soc. Indic. Res.* **66**, 3 (2004). doi:10.1023/B:SOCI.0000007497.57754.e3

20. R. J. Sampson, P. Sharkey, S. W. Raudenbush, Durable effects of concentrated disadvantage on verbal ability among African-American children. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 845 (2008). doi:10.1073/pnas.0710189104 Medline

21. J. Ludwig, Guest editor's introduction: Special issue on MTO. *Cityscape* **14**, 1.

22. R. J. Sampson, Moving to inequality: Neighborhood effects and experiments meet social structure. *Am. J. Sociol.* **114**, 189 (2008). doi:10.1086/589843

23. L. Sanbonmatsu *et al*., *Moving to Opportunity for Fair Housing Demonstration Program: Final Impacts Evaluation* (U.S. Department of Housing and Urban Development, Office of Policy Development and Research, Washington, DC 2011); available at www.huduser.org/publications/pdf/MTOFHD_fullreport_v2.pdf.

24. B. A. Jacob, Public housing, housing vouchers and student achievement: Evidence from public housing demolitions in Chicago. *Am. Econ. Rev.* **94**, 233 (2004). doi:10.1257/000282804322970788

25. P. Oreopoulos, The long-run consequences of living in a poor neighborhood. *Q. J. Econ.* **118**, 1533 (2003). doi:10.1162/003355303322552865

26. S. DeLuca, G. J. Duncan, M. Keels, R. Mendenhall, Gautreaux mothers and their children: An update. *Housing Policy Debate* **20**, 7 (2010). doi:10.1080/10511481003599829

27. R. Ellickson, in *Property Rights and Land Policies*, G. Ingram, Y.-H. Hong, Eds. (Lincoln Institute of Land Policy, Cambridge, MA, 2009), pp. 418–452.

28. B. Stevenson, J. Wolfers, Subjective and objective indicators of racial progress. Working Paper, University of Pennsylvania Wharton School (2010); available at http://bpp.wharton.upenn.edu/jwolfers/Papers/Happiness_Race.pdf.

29. D. Kahneman, A. B. Krueger, Developments in the measurement of subjective well-being. *J. Econ. Perspect.* **20**, 3 (2006). doi:10.1257/089533006776526030

30. A. J. Oswald, S. Wu, Objective confirmation of subjective measures of human well-being: evidence from the U.S.A. *Science* **327**, 576 (2010). doi:10.1126/science.1180606 Medline

31. S. F. Reardon, K. Bischoff, Income inequality and income segregation. *Am. J. Sociol.* **116**, 1092 (2011). doi:10.1086/657114 Medline

32. U.S. Census Bureau, *Table F-3. Mean Income Received by Each Fifth and Top 5 Percent of Families, All Races: 1966 to 2010* (2010); available at www.census.gov/hhes/www/income/data/historical/families/2010/F03AR_2010.xls (accessed May 03 2012).

33. Y. Ben-Shalom, R. A. Moffitt, J. K. Scholz, An Assessment of the Effectiveness of Anti-Poverty Programs in the United States, Working Paper 17042 (National Bureau of Economic Research, Cambridge, MA), (2011).

34. J. Goering *et al*., *Moving to Opportunity for Fair Housing Demonstration Program: Current Status and Initial Findings* (U.S. Department of Housing and Urban Development, Washington, DC, 1999).

35. J. Goering, J. D. Feins, T. Richardson, in *What Have We Learned About Housing Mobility and Poverty Deconcentration? In Choosing a Better Life? Evaluating the Moving to Opportunity Social Experiment*, J. Goering, J. Feins, Eds. (The Urban Institute Press, 2003), pp. 3–36.

36. U.S. Census Bureau, Geography Division, *Census Tracts and Block Numbering Areas* (2000); available at www.census.gov/geo/www/cen_tract.html (accessed December 1, 2011).

37. L. Orr *et al*., *Moving to Opportunity for Fair Housing Demonstration Program: Interim Impacts Evaluation* (U.S. Department of Housing and Urban Development, Office of Policy Development and Research, Washington, DC, 2003); available at www.huduser.org/Publications/pdf/MTOFullReport.pdf.

38. J. G. Fortson, L. Sanbonmatsu, Child health and neighborhood conditions: Results from a randomized housing voucher experiment. *J. Hum. Resour.* **45**, 840 (2010). doi:10.1353/jhr.2010.0025

39. J. R. Kling, J. Ludwig, L. F. Katz, Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment. *Q. J. Econ.* **120**, 87 (2005).

40. L. Sanbonmatsu, J. R. Kling, G. J. Duncan, J. Brooks-Gunn, Neighborhoods and academic achievement: Results from the Moving to Opportunity experiment. *J. Hum. Resour.* **41**, 649 (2006).

41. R. M. Groves, F. J. Fowler, M. P. Couper, J. M. Lepkowski, E. Singer, *Survey Methodology* (Wiley, New York, ed. 7, 2004).

42. The American Association of Public Opinion Research, *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys, 7th Edition* (2011).

43. E. L. Idler, S. V. Kasl, Self-ratings of health: Do they also predict change in functional ability? *J. Gerontol. B Psychol. Sci. Soc. Sci.* **50B**, S344 (1995). doi:10.1093/geronb/50B.6.S344 Medline

44. E. L. Idler, Y. Benyamini, Self-rated health and mortality: A review of twenty-seven community studies. *J. Health Soc. Behav.* **38**, 21 (1997). doi:10.2307/2955359 Medline

45. C. Delpierre, V. Lauwers-Cances, G. D. Datta, T. Lang, L. Berkman, Using self-rated health for analysing social inequalities in health: A risk for underestimating the gap between

socioeconomic groups? *J. Epidemiol. Community Health* **63**, 426 (2009). doi:10.1136/jech.2008.080085 Medline

46. J. B. Dowd, A. Zajacova, Does self-rated health mean the same thing across socioeconomic groups? Evidence from biomarker data. *Ann. Epidemiol.* **20**, 743 (2010). doi:10.1016/j.annepidem.2010.06.007 Medline

47. J. M. Wiener, R. J. Hanley, R. Clark, J. F. Van Nostrand, Measuring the activities of daily living: Comparisons across national surveys. *J. Gerontol.* **45**, S229 (1990). Medline

48. Institute for Social Research, Health and retirement study: Physical measures and biomarkers (2008); available at http://hrsonline.isr.umich.edu/modules/meta/2008/core/qnaire/online/2008PhysicalMeasuresBiomarkers.pdf.

49. A. V. Chobanian *et al*., National High Blood Pressure Education Program Coordinating Committee, Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *Hypertension* **42**, 1206 (2003). doi:10.1161/01.HYP.0000107251.49515.c2 Medline

50. T. A. Furukawa, R. C. Kessler, T. Slade, G. Andrews, The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of Mental Health and Well-Being. *Psychol. Med.* **33**, 357 (2003). doi:10.1017/S0033291702006700 Medline

51. R. C. Kessler *et al*., Screening for serious mental illness in the general population. *Arch. Gen. Psychiatry* **60**, 184 (2003). doi:10.1001/archpsyc.60.2.184 Medline

52. R. C. Kessler, T. B. Ustün, The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *Int. J. Methods Psychiatr. Res.* **13**, 93 (2004). doi:10.1002/mpr.168 Medline

53. R. C. Kessler, K. R. Merikangas, The National Comorbidity Survey Replication (NCS-R): Background and aims. *Int. J. Methods Psychiatr. Res.* **13**, 60 (2004). doi:10.1002/mpr.166 Medline

54. D. F. Alwin, Information transmission in the survey interview: Number of categories and the reliability of attitude measurement. *Sociol. Methodol.* **22**, 83 (1992). doi:10.2307/270993

55. F. M. Andrews, S. B. Withey, *Social Indicators of Well-Being: Americans' Perceptions of Life Quality* (Plenum, New York, 1976).

56. J. Jacoby, M. S. Matell, Three-point Likert scales are good enough. *J. Mark. Res.* **8**, 495 (1971). doi:10.2307/3150242

57. D. R. Lehmann, H. James, Are three-point scales always good enough? *J. Mark. Res.* **9**, 444 (1972). doi:10.2307/3149313

58. A. Ferrer-i-Carbonell, P. Frijters, How important is methodology for the estimates of the determinants of happiness? *Econ. J.* **114**, 641 (2004). doi:10.1111/j.1468-0297.2004.00235.x

59. A. B. Krueger, D. A. Schkade, The Reliability of subjective well-being measures. *J. Public Econ.* **92**, 1833 (2008). doi:10.1016/j.jpubeco.2007.12.015 Medline

60. R. E. Lucas, E. Diener, E. M. Suh, Discriminant validity of well-being measures. *J. Pers. Soc. Psychol.* **71**, 616 (1996). doi:10.1037/0022-3514.71.3.616 Medline

61. E. Sandvik, E. Diener, L. Seidlitz, Subjective well-being: The convergence and stability of self-report and non-self-report measures. *J. Pers.* **61**, 317 (1993). doi:10.1111/j.1467-6494.1993.tb00283.x

62. E. Diener, E. M. Suh, in *Well-Being: the Foundations of Hedonic Psychology*, D. Kahneman, E. Diener, N. Schwartz, Eds. (1999), pp. 434–450.

63. E. Diener, E. M. Suh, R. E. Lucas, H. L. Smith, Subjective well-being: Three decades of progress. *Psychol. Bull.* **125**, 276 (1999). doi:10.1037/0033-2909.125.2.276

64. B. S. Frey, A. Stutzer, *Happiness and Economics* (Princeton Univ. Press, Princeton, NJ, 2002).

65. D. G. Blanchflower, A. J. Oswald, Well-being over time in Britain and the USA. *J. Public Econ.* **88**, 1359 (2002). doi:10.1016/S0047-2727(02)00168-8

66. R. Layard, *Happiness: Lessons from a New Science* (Penguin, London, 2005).

67. A. Steptoe, J. Wardle, M. Marmot, Positive affect and health-related neuroendocrine, cardiovascular, and inflammatory processes. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6508 (2005). doi:10.1073/pnas.0409174102 Medline

68. R. Di Tella, R. MacCulloch, Some uses of happiness data in economics. *J. Econ. Perspect.* **20**, 25 (2006). doi:10.1257/089533006776526111

69. B. Stevenson, J. Wolfers, Happiness inequality in the United States. *J. Legal Stud.* **37**, (S2), S33 (2008). doi:10.1086/592004

70. B. Stevenson, J. Wolfers, in *Brookings Papers on Economic Activity, Spring 2008*, D. W. Elmendorf, N. Gregory, Mankiw, L. H. Summers, Eds. (Brookings Institution Press, Washington, DC, 2008), vol. 39, pp. 1–102.

71. H. L. Urry *et al*., Making a life worth living: neural correlates of well-being. *Psychol. Sci.* **15**, 367 (2004). doi:10.1111/j.0956-7976.2004.00686.x Medline

72. E. Diener, R. E. Lucas, in *Well-being: The Foundations of Hedonic Psychology*, D. Kahneman, E. Diener, N. Schwarz, Eds. (Guilford Press, 1999), vol. 37, pp. 213–229.

73. B. M. S. van Praag, Ordinal and cardinal utility: An integration of the two dimensions of the welfare concept. *J. Econom.* **50**, 69 (1991). doi:10.1016/0304-4076(91)90090-Z

74. J. D. Angrist, Estimation of limited dependent variable models with dummy endogenous regressors. *J. Bus. Econ. Stat.* **19**, 2 (2001). doi:10.1198/07350010152472571

75. D. B. Rubin, Comment on "Randomization analysis of experimental data: The Fisher randomization test," by D. Basu. *J. Am. Stat. Assoc.* **75**, 591 (1980). doi:10.2307/2287653

76. M. E. Sobel, What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference. *J. Am. Stat. Assoc.* **101**, 1398 (2006). doi:10.1198/016214506000000636

77. J. Ludwig *et al*., What can we learn about neighborhood effects from the Moving to Opportunity Experiment? *Am. J. Sociol.* **114**, 144 (2008). doi:10.1086/588741

78. H. S. Bloom, Accounting for no-shows in experimental evaluation designs. *Eval. Rev.* **8**, 225 (1984). doi:10.1177/0193841X8400800205

79. J. D. Angrist, G. W. Imbens, D. B. Rubin, Identification of causal effects using instrumental variables. *J. Am. Stat. Assoc.* **91**, 444 (1996). doi:10.1080/01621459.1996.10476902

80. J. D. Angrist, J.-S. Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton Univ. Press, Princeton, NJ, 2009).

81. J. Ludwig, J. R. Kling, Is crime contagious? *J. Law Econ.* **50**, 491 (2007). doi:10.1086/519807

82. D. Rivers, Q. Vuong, Limited information estimators and exogeneity tests for simultaneous probit models. *J. Econom.* **39**, 347 (1988). doi:10.1016/0304-4076(88)90063-2

83. M. P. Murray, Avoiding invalid instruments and coping with weak instruments. *J. Econ. Perspect.* **20**, 111 (2006). doi:10.1257/jep.20.4.111

84. W. A. Fuller, Some properties of a modification of the limited information estimator. *Econometrica* **45**, 939 (1977). doi:10.2307/1912683

85. J. Stock, M. Yogo, in *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*, D. W. K. Andrews, J. H. Stock, Eds. (Cambridge Univ. Press, 2005), pp. 80–108.

86. J. Hahn, J. A. Hausman, G. Kuersteiner, Estimation with weak instruments: Accuracy of higher order bias and MSE approximations. *Econometrics J.* **7**, 272 (2004). doi:10.1111/j.1368-423X.2004.00131.x

87. S. D. Kennedy, M. L. Leger, *Final Comprehensive Report of the Freestanding Housing Voucher Demonstration* (U.S. Department of Housing and Urban Development, Office of Policy Development and Research. Washington, DC, 1990).

88. L. S. Rubinowitz, J. E. Rosenbaum, *Crossing the Class and Color Lines: From Public Housing to White Suburbia* (Univ. Chicago Press, Chicago, 2000).

89. J. D. Feins, M. Shroder, Moving to Opportunity: The demonstration's design and its effects on mobility. *Urban Stud.* **42**, 1275 (2005). doi:10.1080/00420980500150599

90. M. Shroder, Locational constraint, housing counseling, and successful lease-up in a randomized housing voucher experiment. *J. Urban Econ.* **51**, 315 (2002). doi:10.1006/juec.2001.2247

91. D. Staiger, J. Stock, Instrumental variables regression with weak instruments. *Econometrica* **65**, 557 (1997). doi:10.2307/2171753

92. U.S. Census Bureau, *1970 Census Data*. Prepared by Social Explorer.

93. U.S. Census Bureau, *American Community Survey 2005–2009*. Prepared by Social Explorer.

94. U.S. Census Bureau, *Table F-1. Income Limits for Each Fifth and Top 5 Percent of Families (All Races): 1947 to 2010* (2010); available at www.census.gov/hhes/www/income/data/historical/families/2010/F01AR_2010.xls (accessed May 03 2012).

95. B. A. Jacob, J. Ludwig, The effects of housing assistance on labor supply: Evidence from a voucher lottery. *American Economic Review* **102**, 272 (2012).