Econometrica, Vol. 75, No. 1 (January, 2007), 83–119

EXPERIMENTAL ANALYSIS OF NEIGHBORHOOD EFFECTS

BY JEFFREY R. KLING, JEFFREY B. LIEBMAN, AND LAWRENCE F. KATZ[1]

Families, primarily female-headed minority households with children, living in high-poverty public housing projects in five U.S. cities were offered housing vouchers by lottery in the Moving to Opportunity program. Four to seven years after random assignment, families offered vouchers lived in safer neighborhoods that had lower poverty rates than those of the control group not offered vouchers. We find no significant overall effects of this intervention on adult economic self-sufficiency or physical health. Mental health benefits of the voucher offers for adults and for female youth were substantial. Beneficial effects for female youth on education, risky behavior, and physical health were offset by adverse effects for male youth. For outcomes that exhibit significant treatment effects, we find, using variation in treatment intensity across voucher types and cities, that the relationship between neighborhood poverty rate and outcomes is approximately linear.

KEYWORDS: Social experiment, housing vouchers, neighborhoods, mental health, urban poverty.

THE RESIDENTS OF DISADVANTAGED NEIGHBORHOODS fare substantially worse on a wide range of socioeconomic and health outcomes than do those with more affluent neighbors. Economic models of residential sorting—partially motivated by these observed associations between neighborhood characteristics and individual outcomes—suggest that inefficient equilibria can arise when individual outcomes are influenced by neighbors and individuals do not take their external effects on neighbors into account in their location decisions (e.g., Benabou (1993)).

[1]This paper integrates material previously circulated in Kling and Liebman (2004), Kling, Liebman, Katz, and Sanbonmatsu (2004), and Liebman, Katz, and Kling (2004). Support for this research was provided by grants from the U.S. Department of Housing and Urban Development (HUD), the National Institute of Child Health and Development (NICHD), and the National Institute of Mental Health (R01-HD40404 and R01-HD40444), the National Science Foundation (9876337 and 0091854), the Robert Wood Johnson Foundation, the Russell Sage Foundation, the Smith Richardson Foundation, the MacArthur Foundation, the W. T. Grant Foundation, and the Spencer Foundation. Additional support was provided by grants to Princeton University from the Robert Wood Johnson Foundation and from NICHD (5P30-HD32030 for the Office of Population Research), by the Princeton Industrial Relations Section, the Bendheim–Thomas Center for Research on Child Wellbeing, the Princeton Center for Health and Wellbeing, and the National Bureau of Economic Research. We are grateful to Todd Richardson and Mark Shroder of HUD; to Eric Beecroft, Judie Feins, Barbara Goodson, Robin Jacob, Stephen Kennedy, Larry Orr, and Rhiannon Patterson of Abt Associates; to our collaborators Jeanne Brooks-Gunn, Alessandra Del Conte Dickovick, Greg Duncan, Tama Leventhal, Jens Ludwig, Bruce Psaty, Lisa Sanbonmatsu, and Robert Whitaker; to our research assistants Ken Fortson, Jane Garrison, Erin Metcalf, and Josh Meltzer; and to numerous colleagues and three anonymous referees for valuable suggestions. Any findings or conclusions expressed are those of the authors.

It is hard to judge from theory alone whether the externalities from having neighbors of higher socioeconomic status are predominantly beneficial (from social connections, positive role models, reduced exposure to violence, and more community resources), inconsequential (only family influences, genetic endowments, individual human capital investments, and the broader nonneighborhood social environment matter), or adverse (from competition with advantaged peers and discrimination).[2] Empirical assessment of the importance of such externalities has also proven difficult using nonexperimental data because individuals sort across neighborhoods for reasons that are likely to be correlated with the underlying determinants of their outcomes.

In this paper, we avoid the problem of endogenous neighborhood selection by using data from a randomized experiment in which some families living in high-poverty U.S. housing projects were offered Section 8 housing vouchers to enable them to move to lower poverty neighborhoods while others were not offered vouchers.[3] Thus our analysis provides direct evidence on the existence, direction, and magnitude of neighborhood effects for important socioeconomic and health outcomes in both adult and youth populations. The findings also bear on key housing policy decisions such as whether it is better to provide housing subsidies tied to public housing projects or housing vouchers that can be used in the private-sector rental market.

The research design used in this paper is based on comparisons of three groups to which households were randomly assigned in the Moving to Opportunity (MTO) social experiment operated in five cities—Baltimore, Boston, Chicago, Los Angeles, and New York—by the U.S. Department of Housing and Urban Development. A control group received no new assistance, but continued to be eligible for public housing. A Section 8 group received a traditional Section 8 voucher, without geographic restriction. An experimental group received a Section 8 voucher restricted for 1 year to a census tract with a poverty rate of less than 10 percent and also received mobility counseling. Our sample consists of 4,248 households assigned from 1994–1997 at the five sites.

In 2002, extensive data were collected on outcomes from five key domains: economic self-sufficiency, mental health, physical health, risky behavior, and education. This paper provides the main results from MTO for adults and for youth ages 15–20 at all five sites an average of 5 years after random assignment,

[2]See, for example, Jencks and Mayer (1990), Becker and Murphy (2000), Brock and Durlauf (2001), and Kawachi and Berkman (2003).

[3]Low-income U.S. families receive federal housing assistance through subsidies tied to residence in public housing projects or through vouchers (now called Housing Choice Vouchers, but commonly known as Section 8) that subsidize rents for private sector units. These programs are not entitlements and have wait lists that can be substantial. Tenants in U.S. public housing and those whose use Section 8 housing vouchers both pay approximately 30 percent of their incomes in rent. The value of a voucher is the difference between 30 percent of income and the city's fair market rent, set at the 40th percentile of area rents. See Olsen (2003) for an overview of U.S. housing programs for the poor.

providing the most comprehensive experimental analysis to date of neighborhood effects.[4]

1. DATA AND DESCRIPTIVE STATISTICS

The data for this study come from a baseline survey, administrative data, and an impact evaluation survey conducted in 2002 of one adult and up to two randomly selected children in each MTO household. The baseline survey was administered to household heads prior to random assignment. Administrative data on earnings and welfare benefits were obtained from state and county agencies in California, Illinois, Maryland, Massachusetts, and New York.[5] The 2002 survey had an effective response rate of 90 percent for adults and female youth and 86 percent for male youth.[6] All statistical estimates in this paper use weights.[7] The baseline covariates included in our reported regressions are displayed in Appendix A, Tables A1 and A2 for adults and youth, respectively. Details of the outcome variables are provided in Appendix A.

The Boston, Los Angeles, and New York families in our sample are mainly black or Hispanic; those in Baltimore and Chicago are nearly all black. Overall,

[4] Additional results are available from Sanbonmatsu, Kling, Duncan, and Brooks-Gunn (2006), who analyzed reading and math test scores for children ages 6–20, and from Kling, Ludwig, and Katz (2005), who analyzed arrest records for youth ages 15–25. The earlier single-site pilot studies of MTO are collected in Goering and Feins (2003). Detailed background information on the MTO demonstration and the interim evaluation survey is contained in Orr et al. (2003).

[5] Four states provided individual-level earnings information on each MTO sample member who matched to the Unemployment Insurance (UI) records. Massachusetts provided only data aggregated to groups that consisted of at least 10 MTO individuals. Agency data were linked by Social Security number using information only from the state in which the sample member resided at the time of random assignment. Earnings and welfare amounts were inflation adjusted to 2001 dollars using the Consumer Price Index for all urban consumers.

[6] An initial phase from January–June 2002 resulted in an 80% response rate for adults. At that point, we drew a 3-in-10 subsample of remaining cases and located 48% of them. The purpose of the subsampling was to concentrate resources on finding hard-to-locate families so as to minimize the potential for nonresponse bias. We calculate the effective response rate for adults as $80 + (1 - 0.8) \times 48 = 89.6$. The effective response rate for youths is calculated using this same approach.

[7] The weights have three components (Orr et al. (2003)). First, subsample members receive greater weight because, in addition to themselves, they represent individuals who we did not attempt to contact during the subsampling phase. Second, youth from large families receive greater weight because we randomly sampled two children per household, implying that youth from large families are representative of a larger fraction of the study population; this component does not apply to adults. Third, all individuals are weighted by the inverse of their probability of assignment to their experimental group to account for changes in the random assignment ratios over time. The ratio of individuals randomly assigned to treatment groups was changed during the course of the demonstration to minimize the minimum detectable effects after take-up of the vouchers turned out to be different than had been projected. This third component of the weights prevents time or cohort effects from confounding the results. Our weights imply that each random assignment period is weighted in proportion to the number of people randomly assigned in that period. Analyses of administrative data use only the third component of the weights.

85 percent of the households are female-headed and either African–American or Hispanic. Ninety-eight percent of the sample adults are female, and 93 percent were ages 25–54 as of December 31, 2001. At the time of random assignment, one quarter of sample adults were employed, three-quarters were receiving Aid for Families with Dependent Children (AFDC), more than half had never married, fewer than half had graduated from high school, and a quarter had been teenage parents. In a baseline survey, a majority said they wanted to move out of public housing "to get away from drugs and gangs."

Participants volunteered for this study, presumably because they were interested in moving out of their original high-poverty neighborhoods. Although this may be the most relevant population when considering incremental expansion of the use of housing vouchers to replace public housing, care should be taken in applying these results to populations with different characteristics. The experiment did not result in large clusters of moves to the same new neighborhoods; therefore, it is unlikely to have had large effects on receiving neighborhoods or via "social multiplier" or "general equilibrium" effects (Manski (1993), Heckman (2001)).

Families in the treatment groups had 4–6 months to find qualified housing and move using a MTO voucher. The fraction of treatment group families who used a MTO voucher to move, which we refer to as the compliance rate, was 47 percent for the experimental group and 60 percent for the Section 8 group. Compared to noncompliers (those in the treatment groups who did not use a MTO voucher), compliers (those who moved using a MTO voucher) were younger and more likely to have had no teenage children at baseline, to have reported that their neighborhood was very unsafe at night, to have said that they were very dissatisfied with their apartment, to have been enrolled in school, and to have forecast that they would be "very likely" to find a new apartment if offered a voucher. Compliance rates differed substantially by site from a low of 32 percent in the Chicago experimental group to a high of 77 percent in the Los Angeles Section 8 group.[8]

To characterize the neighborhoods in which families lived and the differences in residential location for those who used a MTO voucher versus those who did not, Figure 1 shows several densities of neighborhood (census tract) poverty rates. The poverty rates are duration-weighted averages over locations lived at since random assignment and they use linear interpolation for poverty rates between the Census years of 1990 and 2000. Figure 1 indicates that experimental compliers lived in neighborhoods with significantly lower poverty rates than did controls, with nearly 60 percent living in neighborhoods below 20 percent poverty; Section 8 compliers also lived in lower poverty neighborhoods,

[8]The intensity of housing search assistance provided to the experimental group by the nonprofits responsible for the counseling varied considerably across sites, as did the tightness of local housing markets.

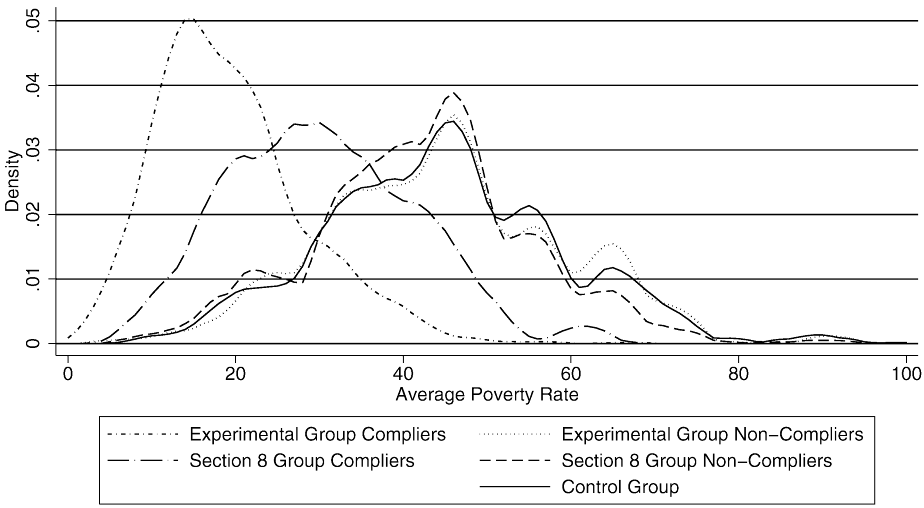

FIGURE 1.—Densities of average poverty rate, by group. Average poverty rate is a duration-weighted average of tract locations from random assignment through 12/31/2001. Poverty rate is based on linear interpolation of 1990 and 2000 Censuses. Density estimates used an Epanechnikov kernel with a half-width of 2.

but their density is shifted by a more modest amount. The densities for experimental noncompliers, Section 8 noncompliers, and controls are quite similar to each other.[9]

Additional descriptive statistics of the residential locations are shown in Table I. The experimental and Section 8 groups are both substantially less likely to live in very poor areas with visible drug activity, and somewhat more likely to live in areas with greater adult employment and a lower share of minority residents. Members of the treatment groups feel safer and are less likely to report a household member having been victimized by crime in the previous 6 months. The 0.82 average share minority for experimental group tracts indicates that although families moved to lower poverty census tracts, these families did not move to distant white suburban areas. In the experimental group, only 16 percent moved 10 miles or more, and only 12 percent had an average tract share minority less than half (Appendix, Table F2 in the supplement to this article (Kling, Liebman, and Katz (2007)), hereafter referenced as Web Appendix).

[9]This implies that there was little selection of the type typically hypothesized, where compliers would have been more likely to have moved to lower poverty neighborhoods even if they had not been offered a voucher (and the poverty distribution for controls would therefore exhibit greater density at lower neighborhood-poverty rates than would the density for noncompliers).

TABLE I

DESCRIPTIVE STATISTICS OF NEIGHBORHOOD CHARACTERISTICS[a]

	Experimental (i)	Section 8 (ii)	Control (iii)
Average census tract poverty rate	0.33	0.35	0.45
Average census tract poverty rate above 30%	0.52	0.62	0.87
Respondent saw illicit drugs being sold or used in neighborhood during past 30 days	0.33	0.34	0.46
Streets are safe or very safe at night	0.70	0.65	0.56
Member of household victimized by crime during past 6 months	0.17	0.16	0.21
Average census tract share on public assistance	0.16	0.17	0.23
Average census tract share of adults employed	0.83	0.83	0.78
Average census tract share workers in professional and managerial occupations	0.26	0.23	0.21
Average census tract share minority	0.82	0.87	0.90

[a]Census tract characteristics are the average for an individual's addresses from randomization through 2001 weighted by duration. Except for "professional and managerial occupations" (for which only 2000 Census data were used due to differences in the occupation classification used for the 1990 Census and 2000 Census), values for intercensus years are interpolated. "Saw illicit drugs," "streets are safe," and "victimized by crime" are based on adult report in the 2002 survey. All experimental − control and Section 8 − control differences have p-values < 0.05.

2. ANALYSIS

We focus on 15 primary outcomes for adults and 15 primary outcomes for youth. Prior to examining the data, we decided to examine youth results pooled by gender and separately for females and males—both because the prevalence for some outcomes differs greatly by gender and results in different statistical power to detect effects and because there had been some evidence of more beneficial effects for boys in earlier MTO research (Katz, Kling, and Liebman (2001), Ludwig, Duncan, and Hirschfield (2001)) and in welfare reform research (Bos et al. (1999)). With 15 outcomes, four population groups (adults, all youth, female youth, and male youth), and two treatment groups, there is a total of 120 treatment effect estimates in this set.

To draw general conclusions about the experiment's results, we first present findings for summary indices that aggregate information over multiple treatment effect estimates (later we present estimates for specific outcomes). For example, we create an index of economic self-sufficiency that averages together five measures of employment, earnings, and public assistance receipt. The aggregation improves statistical power to detect effects that go in the same direction within a domain.[10] The summary index Y is defined to be the equally weighted average of z-scores of its components, with the sign of each measure oriented (as indicated in the notes to Table II) so that more beneficial outcomes have higher scores. The z-scores are calculated by subtracting the control group mean and dividing by the control group standard deviation.[11] Thus, each component of the index has mean 0 and standard deviation 1 for the control group.

We begin by estimating intent-to-treat (ITT) effects—differences between treatment and control group means. Estimation of the ITT effect π_1 is from Equation (1). Let Z be an indicator for treatment group assignment and let X be a matrix of baseline covariates:

$$Y = Z\pi_1 + X\beta_1 + \varepsilon_1. \tag{1}$$

[10] O'Brien (1984) constructed a global test statistic for multiple outcomes with maximum power against the alternative that all effects have the same sign and effect size. An adaptation of the O'Brien approach is discussed in Appendix B of the Web Appendix and implemented in Kling and Liebman (2004), which uses seemingly unrelated regression effects for specific outcomes to estimate the covariance of the effects and calculates the mean effect size for groups of estimates in a second step. The average z-score index used in this paper is much simpler to work with, particularly for our results that relate neighborhood poverty rates to outcomes. The two approaches yield identical treatment effects when there is no item nonresponse and no regression adjustment.

[11] If an individual has a valid response to at least one component measure of an index, then any missing values for other component measures are imputed at the random assignment group mean. This results in differences between treatment and control means of an index being the same as the average of treatment and control means of the components of that index (when the components are divided by their control group standard deviation and have no missing value imputation), so that the index can be interpreted as the average of results for separate measures scaled to standard deviation units.

TABLE II

MEAN EFFECT SIZES FOR SUMMARY MEASURES OF OUTCOMES[a]

	All Adults		All Youth		Female Youth		Male Youth		M – F Youth	
	E – C (i)	S – C (ii)	E – C (iii)	S – C (iv)	E – C (v)	S – C (vi)	E – C (vii)	S – C (viii)	E – C (ix)	S–C (x)
Economic self-sufficiency	0.017 (0.031)	0.037 (0.033)								
Absence of physical health problems	0.012 (0.024)	0.019 (0.026)	−0.038 (0.038)	−0.020 (0.040)	0.025 (0.053)	0.077 (0.055)	−0.112* (0.053)	−0.114 (0.061)	−0.138 (0.076)	−0.192* (0.084)
Absence of mental health problems	0.079* (0.030)	0.029 (0.033)	0.102 (0.053)	0.138* (0.056)	0.267* (0.062)	0.192* (0.067)	−0.052 (0.080)	0.054 (0.092)	−0.319* (0.101)	−0.138 (0.113)
Absence of risky behavior			−0.023 (0.043)	−0.039 (0.050)	0.142* (0.053)	0.129* (0.059)	−0.181* (0.062)	−0.208* (0.071)	−0.323* (0.080)	−0.337* (0.092)
Education			0.050 (0.041)	0.028 (0.047)	0.138* (0.065)	0.056 (0.068)	−0.053 (0.047)	−0.001 (0.060)	−0.191* (0.080)	−0.057 (0.090)
Overall	0.036 (0.020)	0.028 (0.022)	0.018 (0.025)	0.018 (0.026)	0.136* (0.034)	0.109* (0.034)	−0.099* (0.031)	−0.078* (0.037)	−0.235* (0.047)	−0.187* (0.051)

[a]E – C denotes experimental – control; S – C denotes Section 8 – control. Estimates are the intent-to-treat mean effect sizes, from Equation (1), fully interacted with gender in columns (v)–(x) as described in the text. The estimated equations all include site indicators and the baseline covariates listed in Appendix A with those in Table A1 included for adults and those in Tables A1 and A2 included for youth. M – F Youth is male – female difference. Adult economic self-sufficiency: + adult not employed and not on TANF + employed + 2001 earnings – on TANF – 2001 government income. Adult mental health: – distress index – depression symptoms – worrying + calmness + sleep. Adult physical health: – self-reported health fair/poor – asthma attack past year – obesity – hypertension – trouble carrying/climbing. Adult overall includes 15 measures in self-sufficiency, physical health, and mental health. Youth physical health: – self-reported health fair/poor – asthma attack past year – obesity – nonsports injury past year. Youth mental health: – distress index – depression symptoms – anxiety symptoms. Youth risky behavior: – marijuana past 30 days – smoking past 30 days – alcohol past 30 days – ever pregnant or gotten someone pregnant. Youth education: + graduated high school or still in school + in school or working + WJ-R broad reading score + WJ-R broad math score. Youth overall includes 15 measures in physical health, mental health, risky behavior, and education. Sample sizes in the E, S, and C groups are 1,453, 993, and 1,080 for adults and 749, 510, and 548 for youth ages 15–20 on 12/31/2001. Robust standard errors adjusted for household clustering are in parentheses; * = p-value < 0.05.

The covariates (X) are included to improve estimation precision and to account for chance differences between groups in the distribution of pre-random-assignment characteristics. Table II shows ITT results by domain and population group for indices, with the measures included in each index indicated in the notes to Table II and the details on each measure presented in Appendix A. The absolute magnitudes of the indices are in units akin to standardized test scores: the ITT estimate shows where the mean of the treatment group is in the distribution of the control group in terms of standard deviation units.

For adults, the direction of effects is positive for both the experimental and Section 8 groups relative to the control group for all three domains: economic self-sufficiency, physical health, and mental health. The effect on mental health for the experimental group is much larger in magnitude than the others and is the only adult estimate that is statistically significant at the 5 percent level. For results that pool all youth, the direction of effects is positive for mental health and education, and negative for physical health and risky behavior. Again, the effects on mental health (for both the experimental group and Section 8 groups) are much larger and have p-values below 0.06. For the overall index that averages together all fifteen outcomes, the results in columns (i)–(iv) for adults and for all youth are positive in sign, but the magnitudes are not large enough to reject a null hypothesis of no effect with 95 percent confidence. Thus, for adults and for all youth, the evidence of effects from relocation to lower poverty neighborhoods is strongest for the domain of mental health.

The overall results for youth average together estimates that differ substantially for female and male youth. Columns (v)–(viii) of Table II show large positive effects on mental health and risky behavior for female youth, and large negative effects on physical health and risky behavior for male youth.[12] This gender pattern in results was the opposite of what we expected.[13] Yet, as shown in columns (ix) and (x), the medium-term effects for females are more beneficial than for males for all four domains and in both treatment groups relative to the control group.

As a complement to the summary indices, we also examined results for each specific outcome that was a component of an index. Because the magnitudes of these separate outcomes are often easier to interpret than those of the summary indices, we show in Table III all outcomes with ITT effects significant at the 5 percent level. In addition to ITT effects, we also report the effect of treatment-on-treated (TOT) for these measures. We estimate this effect using

[12]These results are based on estimation of $Y = (1 - G)(X\beta_{10} + Z\pi_{10}) + G(X\beta_{11} + Z\pi_{11}) + v$, where G is an indicator for gender, and X includes household and individual-specific characteristics.

[13]Existing nonexperimental papers find larger beneficial effects for boys from living in advantaged neighborhoods than for girls (examples include Entwisle, Alexander, and Olson (1994), Ramirez-Valles, Zimmerman, and Juarez (2002), and Crane (1991)). The predominant mechanism proposed in these studies is that boys spend more time hanging out in the neighborhood and, therefore, are influenced more heavily by the neighborhood.

TABLE III

SPECIFIC OUTCOMES WITH EFFECTS SIGNIFICANT AT 5 PERCENT LEVEL[a]

	E/S (i)	CM (ii)	ITT (iii)	TOT (iv)	CCM (v)
A. Adult outcomes					
Obese, BMI ≥ 30	E – C	0.468	−0.048	−0.103	0.502
			(0.022)	(0.047)	
Calm and peaceful	E – C	0.466	0.061	0.131	0.443
			(0.022)	(0.047)	
Psychological distress, K6 z-score	E – C	0.050	−0.092	−0.196	0.150
			(0.046)	(0.099)	
B. Youth (female and male) outcomes					
Ever had generalized anxiety symptoms	E – C	0.089	−0.044	−0.099	0.164
			(0.019)	(0.042)	
	S – C	0.089	−0.063	−0.114	0.147
			(0.019)	(0.035)	
Ever had depression symptoms	S – C	0.121	−0.039	−0.069	0.134
			(0.019)	(0.035)	
C. Female youth outcomes					
Psychological distress, K6 scale z-score	E – C	0.268	−0.289	−0.586	0.634
			(0.094)	(0.197)	
Ever had generalized anxiety symptoms	E – C	0.121	−0.069	−0.138	0.207
			(0.027)	(0.055)	
	S – C	0.121	−0.075	−0.131	0.168
			(0.029)	(0.051)	
Used marijuana in the past 30 days	E – C	0.131	−0.065	−0.130	0.202
			(0.029)	(0.059)	
	S – C	0.131	−0.072	−0.124	0.209
			(0.032)	(0.056)	
Used alcohol in past 30 days	S – C	0.206	−0.091	−0.155	0.306
			(0.038)	(0.056)	
D. Male youth outcomes					
Serious nonsports accident or injury in past year	E – C	0.062	0.087	0.215	0
			(0.026)	(0.064)	
	S – C	0.062	0.080	0.157	0
			(0.028)	(0.058)	
Ever had generalized anxiety symptoms	S – C	0.055	−0.049	−0.098	0.126
			(0.024)	(0.047)	
Smoked in past 30 days	E – C	0.125	0.103	0.257	0
			(0.032)	(0.084)	
	S – C	0.125	0.151	0.293	0.014
			(0.037)	(0.073)	

[a]E/S: indicates whether the row is experimental − control (E – C) or Section 8 − control (S – C). CM, control mean; ITT, intent-to-treat, from Equation (1); TOT, treatment-on-treated, from Equation (2); CCM, control complier mean. Robust standard errors adjusted for household clustering are in parentheses. The estimated equations all include site indicators and the baseline covariates listed in Appendix A with those in Table A1 included for adults and those in Tables A1 and A2 for youth. Rows shown in the table to illustrate magnitudes were selected based on ITT p-values < 0.05 and are 17 of 120 from the set of specific contrasts (E − C, S − C), based on the outcomes (15 for adults and 15 for youth) and subgroups—adults, youth (female and male), female youth, and male youth—described in the notes to Table II.

the offer of a MTO voucher as an instrumental variable for MTO voucher use, so Z is the excluded instrument for an indicator D of compliance in the two stage least squares (2SLS) estimation:

$$(2) \qquad Y = D\gamma_2 + X\beta_2 + \varepsilon_2.$$

The TOT parameter γ_2 is equal to the ITT parameter divided by the regression-adjusted compliance rate. We interpret these 2SLS estimates as treatment-on-treated estimates rather than local average treatment effect estimates (Imbens and Angrist (1994)) because the endogenous variable is use of a voucher offered by the MTO program, and MTO vouchers are never offered to the control group; there are no always-takers in the terminology of Angrist, Imbens, and Rubin (1996).[14] This TOT approach relies on the assumption that there was no average effect of being offered a MTO voucher on those who did not use a MTO voucher, which we believe is a reasonable approximation, but not strictly true.[15] Under this assumption, we can assess the average magnitude of the effect of the voucher offer for those who complied and used a MTO voucher to move to a lower poverty neighborhood.

As an example, results for the outcome of adult obesity, using a standard body mass index cut point (BMI $\geq$ 30 kg/m^2), are shown in the first row of Table III. The results in column (i) indicate that this row compares the experimental and control groups (E – C). Column (iii) reveals that the ITT effect on obesity is a 5 percentage point reduction in obesity for the experimental group relative to the control group. Assuming no effect on those who did not use a MTO voucher to move, the TOT effect was just over 10 percentage points.

Because outcomes are directly observed for treatment group compliers and we have a TOT estimate, we can estimate the mean level of each outcome for those in the control group who would have complied if they had been offered a voucher—which we refer to as the control complier mean (CCM; Katz, Kling, and Liebman (2001))—based on the relationship that CCM $= E[Y|Z = 1,$

[14]See, also, Heckman (1990). Over time, some control group members do receive housing vouchers from other sources, but they tend to receive them significantly later than treatment group members do. Therefore, in our TOT estimates, we do not define control group voucher recipients as having been treated. We do interpret both our ITT and TOT estimates as averages of heterogeneous treatment effects across individuals.

[15]For the experimental group, this assumption implies that the later outcomes of households who met with a housing mobility counselor were not affected by the counselor if that household did not make a subsidized move through the MTO program. For both treatment groups, this assumption implies that the experience of housing search induced by assignment to a treatment group did not affect later outcomes if that household did not make a subsidized program move. For noncompliers, we believe that the effects of mobility counselors (who mainly provided housing advice and not general social services) on self-sufficiency and health outcomes are likely to be orders of magnitude smaller than the effects of moving to a new residential location. The TOT approach also requires that the control group was not affected by the experience of losing the voucher lottery, something we view as a reasonable approximation.

$D = 1] - \gamma$.[16] For the fraction obese, the CCM in column (v) is estimated to be 0.502.[17] Thus a 10 percentage point change would be a decline in relative risk among compliers of 21 percent and a relative odds ratio of 0.66 for obesity for experimental relative to control compliers.[18]

These effects on adult obesity are of substantial magnitude, yet they are the smallest in relative risk and relative odds among the 13 binary outcomes in Table III. The two continuous outcomes in Table III (the z-scores from the K6 distress index) have TOT effect sizes of 0.2 and 0.5 standard deviations. Thus, each of the TOT effects in Table III appears to be of a substantively important magnitude.

Another metric in which the magnitude of the results can be assessed is in terms of the association between the neighborhood poverty rate (W) and the outcome (Y).[19] This relationship is summarized by the parameter γ_3, the coefficient on the neighborhood poverty rate, in the outcome equation

$$Y = W\gamma_3 + X\beta_3 + \varepsilon_3. \tag{3}$$

For the purposes of the estimation of (3), we view the neighborhood poverty rate as a summary measure of neighborhood quality. Thus, γ_3 should be interpreted as the effect of moving to a neighborhood with a lower poverty rate and the bundle of associated differences in neighborhood characteristics, and not as the effect of changing the poverty rate while holding other characteristics of the neighborhood constant. To be precise, the assumptions that underlie this regression are that there is a scalar index of neighborhood quality (Q say) that is linearly related to the outcome and that the poverty rate is proportional to this index ($W = Q\alpha$ for some scalar α).

Ordinary least squares (OLS) estimation of (3) may be biased by endogenous residential location choices, but treatment group assignment can be used

[16]For binary outcomes, sampling variation can produce negative estimates of the CCM. Our method assumes that there would have been the same fraction of noncompliers in the control group as were observed in the treatment group and that these noncompliers had exactly the same outcome prevalence as treatment noncompliers. In any one sample, this method may produce a negative estimate of the CCM if a particular realization of the treatment noncomplier mean is higher than the realization of the control noncomplier mean, even though the method is unbiased in repeated sampling. We report a CCM of zero when the CCM estimate for a binary outcome is negative.

[17]Body mass index is 30 for a woman five feet four inches tall and 175 pounds. Nationally for ages 18–44, 33 percent of black women and 22 percent of Hispanic women have BMI ≥ 30 (Lucas, Schiller, and Benson (2004)).

[18]The odds of obesity for control compliers are $0.502/(1 - 0.502) = 1.01$. The odds for experimental compliers are $0.399/(1 - 0.399) = 0.664$. The relative odds are $0.664/1.01 = 0.659$.

[19]This approach permits direct comparison with the large nonexperimental literature. In addition, sociological threshold models (Granovetter (1978)) and economic models of sorting across neighborhoods, schools, and classrooms (Arnott and Rowse (1987), de Bartolome (1990), Fernandez and Rogerson (1996), Benabou (1993)) hinge on the exact form of the relationship between neighborhood or peer group characteristics and individual outcomes.

to form instrumental variables. Our 2SLS estimates of (3) use all sample members, regardless of MTO group, and include a full set of site-by-treatment interactions as the excluded instruments for the neighborhood poverty rate (while controlling for site main effects).[20] For our instrumental variables estimation in (3) to be consistent, we must further assume that the entire effect of the instruments works through the neighborhood characteristics that are indexed by Q and not through some other omitted variable. We caution that the effects of neighborhood characteristics on individual outcomes may be heterogeneous; therefore, the external validity of these findings to other contexts may depend on the degree of similarity to the MTO population and their moves.

If X contains only site indicators, then the 2SLS estimate of γ_3 using the site-by-treatment interactions instruments is the slope of the line fit through a scatterplot of the 15 outcome and poverty rate means for the three random assignment groups in each of five sites, normalized so that each site has mean 0. This approach is depicted graphically in Figure 2, which has four panels for four summary indices (adult mental health, female youth mental health, female youth overall, and male youth overall). The plots show that there is a consistent pattern across the sites and groups that larger differences in poverty rates (relative to the site mean) are associated with differences of larger magnitude in outcomes. The relationship between poverty rate and outcomes appears fairly linear and, in each case, a test of the overidentifying restrictions (Davidson and MacKinnon (1993)) has a p-value of less than 0.30, indicating that the data are consistent with a linear model. We interpret the evidence in Figure 2 as supportive of a dose–response relationship. In sites with larger differences in neighborhood poverty rates between the treatment and control groups (larger doses), there are larger treatment effects on outcomes (larger responses).

Estimates based on Equation (3) are given in Table IV for selected outcomes. The first column shows OLS estimates using data for the control group only—results illustrative of the approach taken in the nonexperimental literature on neighborhood effects (e.g., Brooks-Gunn, Duncan, and Aber (1997)) that could have been applied in analysis of this population without the random assignment of vouchers.[21] The second column shows 2SLS estimates of γ_3

[20]In the first stage of 2SLS, the p-value on the excluded instruments is less than 0.0001 for all samples: adults, youth, female youth, and male youth.

[21]Our control group sample differs in important ways from the OLS samples often used in the nonexperimental literature. In particular, our sample is more homogenous. Also, in typical observational data sets, current variation in neighborhoods is the result of mobility choices that have occurred over an extended period of time. In contrast, the variation in neighborhood quality in the MTO control group is, to a large extent, the result of recent moves. The form of selection bias in estimates using the MTO control group might, therefore, be different from that of estimates calculated using standard data sets. For example, recent transitory shocks may introduce a correlation structure between outcomes and neighborhood types in the MTO control group data that would not be as prominent in typical observational data sets.

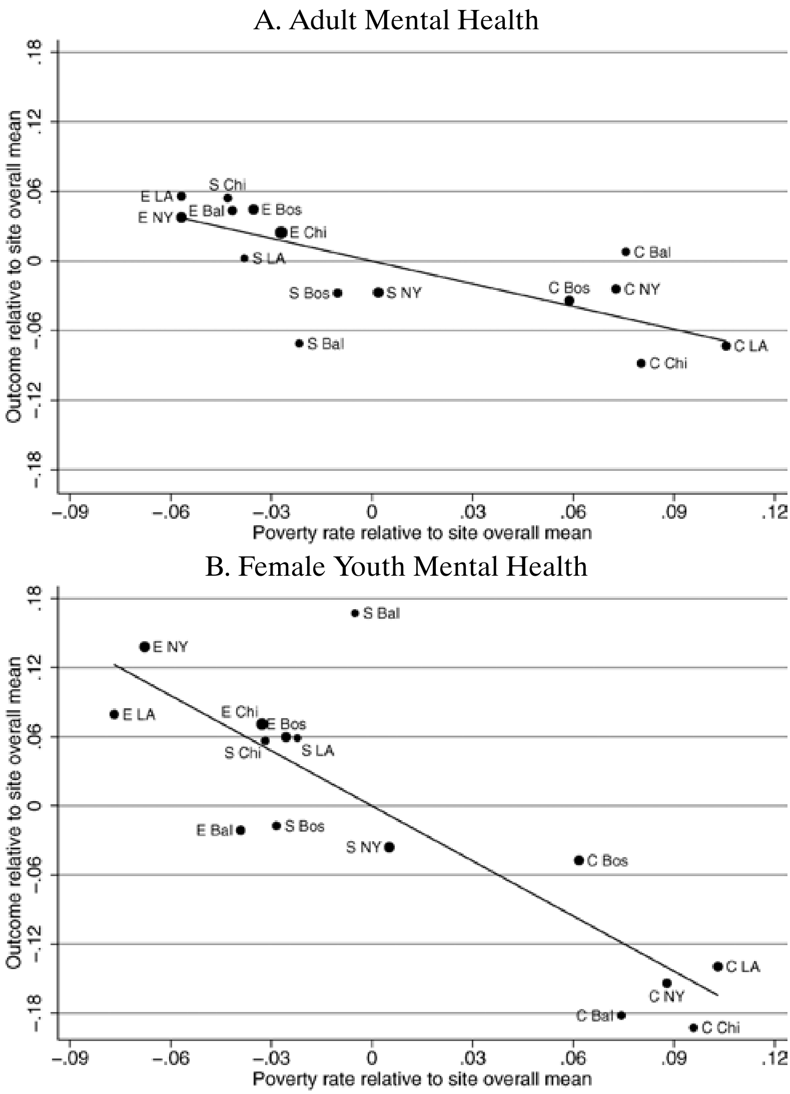

FIGURE 2.—Partial regression leverage plots. The index on the horizontal axis is expressed in standard deviation units relative to the control group overall standard deviation for each variable. The components of the overall and mental health indices are described in the notes to Table II. The poverty rate is an average across tracts since random assignment, weighted by residential duration, using linear interpolation between the 1990 and 2000 Censuses. The line passes through the origin with the slope from 2SLS estimation of Equation (3) of the outcome on poverty rate and site indicators, using group-by-site interactions as instrumental variables. The points are from a partial regression leverage plot of the group outcome means on the group poverty rate means, conditional on site main effects, as described in the text. The size of each point is proportional to the sample size of that group and, correspondingly, to the weight each point receives in the 2SLS regression.

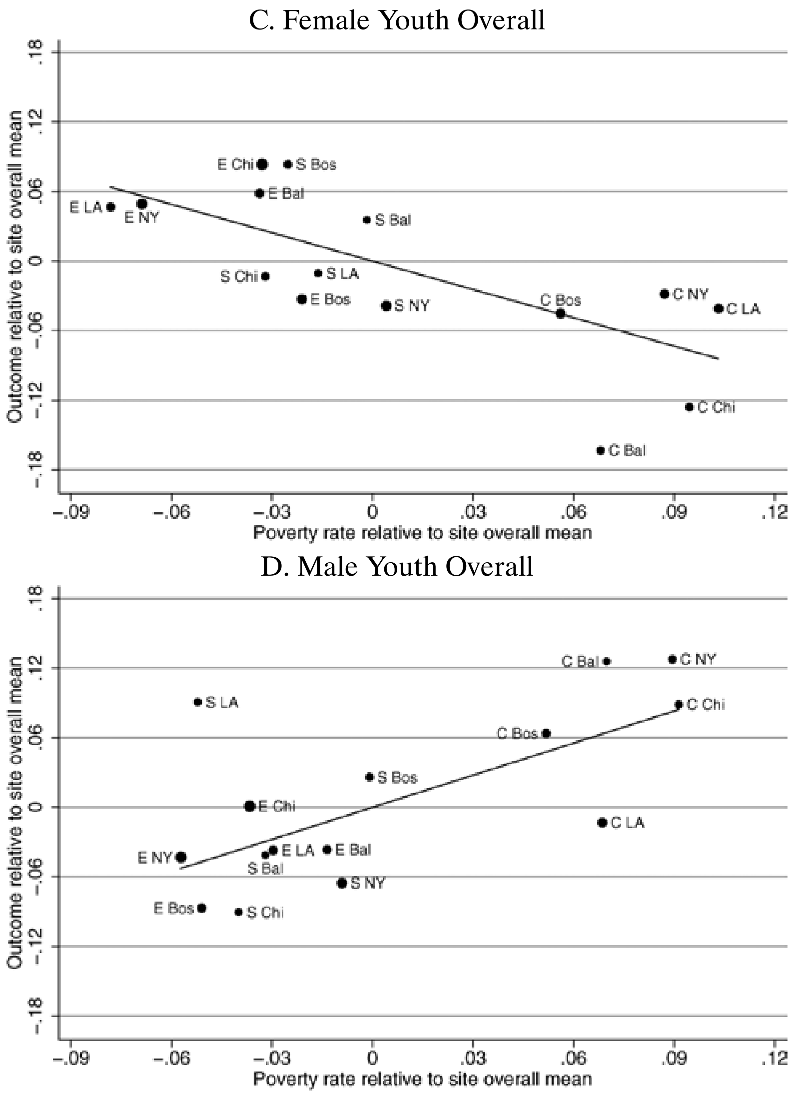

FIGURE 2 (*Continued*).

for the entire sample with site–group interactions as excluded instruments for the neighborhood poverty rate W and a full set of covariates in X. The 2SLS estimates bear little relationship to the OLS estimates, implying that endogeneity is a substantial issue for nonexperimental approaches to these data. Differences in neighborhood poverty rates of 10 percentage points (roughly the treatment–control difference in poverty rates in Table I) are associated

TABLE IV

EFFECTS OF NEIGHBORHOOD POVERTY RATES ON SELECTED OUTCOMES[a]

		Models			
		OLS	2SLS	2SLS	
Variables	Group	Poverty (i)	Poverty (ii)	Poverty (iii)	Compliance (iv)
Mental health	Adult	0.13	−0.62*	−1.35*	−0.17
		(0.17)	(0.24)	(0.60)	(0.13)
	Youth (female and male)	0.57	−0.97*	−0.18	0.20
		(0.34)	(0.41)	(0.87)	(0.21)
	Female youth	0.99	−1.84*	−1.88	−0.01
		(0.61)	(0.50)	(1.09)	(0.25)
Risky behavior	Female youth	−0.61	−0.94*	−1.03	−0.02
		(0.42)	(0.39)	(0.85)	(0.19)
Overall	Female youth	−0.03	−0.90*	−1.03	−0.03
		(0.28)	(0.26)	(0.56)	(0.12)
Physical health	Male youth	−0.84*	1.07*	1.77	0.18
		(0.35)	(0.49)	(1.09)	(0.26)
Risky behavior	Male youth	−0.06	1.46*	0.94	−0.13
		(0.42)	(0.54)	(1.29)	(0.31)
Overall	Male youth	−0.13	0.80*	1.47*	0.17
		(0.23)	(0.28)	(0.68)	(0.16)

[a]The OLS model is from Equation (3) with no excluded instruments, using the control group only; the 2SLS is from Equation (3) with 10 site-by-treatment interactions as excluded instruments, using the entire sample. Columns (i) and (ii) are each based on separate estimation of Equation (3), with W including poverty rate. Each row in columns (iii) and (iv) contains coefficients from one estimate of Equation (3) with W including poverty rate and an indicator for treatment compliance as endogenous variables. Units of summary indices are standard deviations of control group outcomes. The estimated equations all include site indicators and a full set of covariates that combine baseline variables about adults listed in Table A1 and those about youth listed in Table A2 (for youth outcomes only): age, gender, race, marital status, employment, education, mobility history, attitudes about neighborhood, special classes for youth, and behavioral or emotional problems of youth. Poverty rate is averaged over tracts since random assignment, weighted by duration, using linear interpolation between 1990 and 2000 Censuses. Standard errors are in parentheses, adjusted for correlation between same-sex siblings. * = p-value < 0.05. Rows shown in the table to illustrate magnitudes were selected based on 2SLS column (ii) p-value < 0.05 and are 8 of 19 from set of four adult, five youth (female and male), five female youth, and five male youth summary indices shown in Table II.

with outcome effect sizes similar in magnitude to those for the ITT effects in Table II.

To test the hypothesis that differences in poverty rates had the primary effects on outcomes as opposed to simply using a MTO voucher to move out of public housing, we also enriched W in Equation (3) to include both the poverty rate and an indicator for compliance (D); the results are reported in columns (iii) and (iv) of Table IV. Comparing columns (ii) and (iii), results without and with controls for compliance are quite similar for female youth and are within sampling error for adult mental health and for male youth (accounting for the covariance of the estimates). For the more precisely estimated models (adult mental health, female youth overall, and male youth overall), the coefficients on poverty rates are large both in absolute magnitude and relative to their standard errors, while the coefficients on compliance have the wrong sign and are small relative to their standard errors, providing some evidence that the poverty rate effect was more important than the "move per se" effect. We have also examined other models with two endogenous variables, such as poverty and poverty-squared, intercept shifts in poverty rates, and kink points in poverty rates; although there is no evidence of nonlinearities in these models, the research design has little power to identify these effects.

3. DISCUSSION

Adult Economic Self-Sufficiency

The idea that residence in a distressed community can limit an individual's economic prospects was advanced by Wilson (1987), and the related hypothesis that proximity to employment is important has its roots in the spatial mismatch hypothesis of Kain (1968). However, we found no significant evidence of treatment effects on earnings, welfare participation, or amount of government assistance after an average of 5 years since random assignment. As shown in Figure 3, the fraction of MTO adults with positive quarterly UI earnings increased from less than 25 percent in early 1995 to more than 50 percent in 2001. The time patterns are, however, similar for the three randomly assigned groups.[22]

We have examined possible reasons for the lack of effects. There do not appear to be important differences across the three MTO groups in job accessibility, to the extent that aggregate employment growth in establishments at the zip code level reflects available job vacancies (Web Appendix, Table F14).[23] The MTO intervention also had only small impacts on job-related social networks. Although the intervention modestly increased the fraction of sample members who "had a friend who graduated from college or earned more than $30,000 a year," only about 8 percent of the sample "found a job through someone living in their neighborhood such as a friend, relative, or acquaintance," and this

[22]The strong U.S. labor market, welfare reform, and declining share of sample members with preschool-age children are the most likely explanations for this upward trend. In analyses shown in Web Appendix, Tables F3 and F4, we found similar results based on survey and administrative data, suggesting little bias in self-reports.

[23]It may be surprising that the MTO intervention—which assisted families to move out of some of the most concentrated pockets of U.S. poverty—had no discernable overall effects on employment, given a recent survey (Ihlandfeldt and Sjoquist (1998)) that concluded the evidence supports the spatial mismatch hypothesis that inner-city low-skilled minority workers have weak access to jobs because job opportunities are disproportionately in suburban areas and housing market discrimination plus commuting costs prevent minorities from reaching suburban jobs. However, the MTO experiment provides only a weak test of this view of spatial mismatch because the effects on distance moved and on local area job growth are small. If spatial mismatch is broadly construed to encompass how residence in distressed, crime-ridden communities may inhibit job access, the results indicate that moving to communities that are much safer does not have detectable effects on employment.

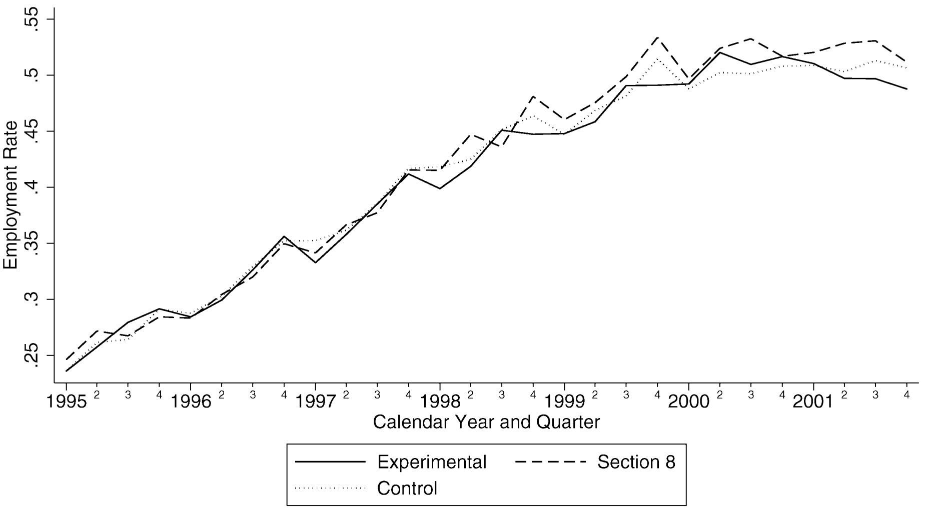

FIGURE 3.—Employment rates over time. Employment is the fraction with positive earnings per quarter from Unemployment Insurance records in California, Illinois, Maryland, Massachusetts, and New York.

proportion did not vary across MTO groups (Web Appendix, Table F10). In subsequent follow-up work based on 67 semistructured open-ended interviews with MTO adults, Turney et al. (2006) found that transportation difficulties and disrupted social networks were additional barriers to employment in the experimental group.

We also explored whether effects differed by baseline characteristics: in general, results do not differ appreciably by these characteristics.[24] However, in results shown in the Web Appendix, Tables F7 and F8, we find suggestive evidence using administrative UI data of interesting dynamics in the treatment effects on employment and earnings for younger adults (those younger than 33 years at random assignment): initial negative treatment effects in the first 2 years after random assignment fade away over time for the Section 8 group and turn positive and substantial in the fourth and fifth years after random assignment for the experimental group.[25]

[24]Economic self-sufficiency results by age are given in the Web Appendix, Tables F6–F8. Results for other household head characteristics and by gender composition of the household are available from the authors.

[25]Examining effects by age was an exploratory, not a confirmatory, exercise. This type of searching for significant effects in subgroups raises the chance of concluding that there are statistically significant results even when the null hypothesis of no effects is true.

Adult Physical Health

Our early work at the Boston site (Katz, Kling, and Liebman (2001)) suggested that the MTO intervention may have had important health impacts and led us to expand the outcomes studied beyond the economic and housing outcomes that were the original focus of the MTO demonstration. In the much more extensive health data gathered in the current study, there was no broad pattern of physical health improvements for the treatment groups. The intervention did not have statistically significant effects on self-reported overall health, hypertension, asthma, or trouble carrying groceries or climbing stairs, and there was not a statistically significant effect on the physical health index in Table II.[26]

There was a large and statistically significant effect on obesity (see Table III), possibly related to the reduced psychological distress and increase in exercise and nutrition that we also observed for the treatment groups. The interpretation of the obesity results depends largely on one's reason for focusing on obesity. If a researcher was searching for social experiments about the effects of neighborhoods on obesity (and this is the only one), focusing on the t-statistic of 2.2 and the associated per-comparison p-value of 0.03 is highly suggestive of an important effect of residential location on obesity. If a researcher is searching the MTO results for significant effects, however, we suggest some caution in interpreting the obesity results. When using a method that focuses on particular results from among many outcomes of a family because of their t-statistics, there is considerable likelihood that the obesity results could have been observed by chance under a joint null hypothesis of no effects for all estimates in that family.[27]

Adult Mental Health

In contrast to the results for physical health, the adult mental health results were quite consistent across specific measures (distress, depression, anxiety, calmness, and sleep) in finding beneficial effects for the experimental group

[26]Analyses by age show a positive and significant impact of the MTO experimental treatment on our summary measure of physical health outcomes for the younger adults and no significant overall impact for older adults. These health impacts come from aggregating five consistently signed estimates with small magnitudes rather than from a large effect on any one measure (Web Appendix, Table F6). This result, along with the suggestive evidence of employment gains among younger adults, leads us to speculate that the habits and behaviors of younger adults may be more malleable and, therefore, more responsive to a change in residential environment.

[27]The probability that the second largest t-statistic among 30 adult estimates is 2.2 or higher under the joint null hypothesis of no effect is a familywise adjusted p-value of 0.80, which is indicative of the fact that there is little power to reject the joint null hypothesis for specific outcomes. The adjusted p-values throughout this paper are based on a bootstrap procedure that accounts for covariance among estimates, using a method adapted from Westfall and Young (1993) as described in Appendix B of the Web Appendix.

relative to the control group. This consistency led to the large mean (ITT) effect size estimate of 0.08 standard deviations for the adult mental health summary measure in Table II. The confidence level that the results are not due to chance is quite high under a method where the focus on mental health is determined exogenously (leading to per-comparison inference) or endogenously from the high t-statistic (leading to familywise inference).[28] The magnitude of the mental health results—for example, a 45 percent reduction in relative risk among compliers of scoring above the K6 screening cut point for serious mental illness (Kessler et al. (2003))—is comparable to that found in some of the most effective clinical and pharmacologic mental health interventions.[29]

The overall pattern of adult results—with the agreement of estimated effects based on self-reports and administrative records of economic outcomes, with the effects concentrated in the single domain of mental health, and with the mental health effect sizes systematically related to changes in neighborhood poverty rates in Figure 2A and in Table IV—indicates that there are beneficial impacts on mental health of moving to less distressed neighborhoods. In addition, this pattern is contrary to a model in which the unrestricted choice of the Section 8 group should have led to better outcomes than the restricted choice of the experimental group. Based in part on evidence from the extensive qualitative interviews that have been done with MTO participants and the strong associations shown in the MTO quantitative research, we believe that the leading hypothesis for the mechanism that produces the mental health improvements involves the reduction in stress that occurred when families moved away from dangerous neighborhoods in which the fear of random violence influenced all aspects of their lives (Kling, Liebman, and Katz (2005), Popkin, Harris, and Cunningham (2002)).

[28]We focus our familywise inference on the summary indices, thereby reducing the dimension of the inference problem. For adults, we focus on eight ITT estimates for summary indices in Table II. The familywise adjusted p-value for the mental health index t-statistic being 2.8 or higher (i.e., the largest t-statistic on estimates for eight adult indices) under the joint null hypothesis of no effect on any of the eight adult indices in Table II is 0.06.

[29]In a study often cited as an exemplar of an effective clinical intervention, Wells et al. (2000) analyzed outcomes of depressive patients randomized to obtain usual care or improved quality care (better training of medical staff and better follow-up with patients). Twelve months later, the fraction with depressive symptoms in the quality improvement group was 0.42, while the fraction was 0.51 in the usual care group—a reduction in relative risk of 18 percent. A metanalysis of clinical trials of medications for major depressive disorder found that on average 50 percent of patients who received active medication showed improvement compared with 29 percent who received a placebo—a reduction in relative risk of 30 percent (Walsh et al. (2002)). The reduction in relative risk for major depressive episode among MTO experimental compliers was also 30 percent (Web Appendix, Table F5). Improved mental health for the treatment groups relative to the control group was a mechanism that we had hypothesized might increase employment and earnings. Although the effects on mental health reported here are large, Kling, Liebman, Katz, and Sanbonmatsu (2004) calculated that these are unlikely to translate into effects on earnings large enough to detect.

Female Youth Outcomes

Teenage youth are often seen as the population most at risk from the adverse effects of high-poverty neighborhoods. In this study, 15 specific outcomes were assessed for youth within the four domains of physical health, mental health, risky behavior, and education. A summary index of all 15 outcomes shows large benefits for female youth in the experimental and Section 8 groups relative to the control group. The pattern of beneficial effects is quite consistent across outcomes: 13 of 15 outcomes have the sign of a treatment effect in a beneficial direction for both treatments. The magnitudes of the effects are largest for mental health, still substantial for education and risky behavior, and small for physical health. For example, the experimental compliers have a relative risk of serious generalized anxiety symptoms 70 percent lower than the control complier mean.

To assess statistical significance, we adopt a framework similar to that used for adult outcomes. If there is ex ante interest by a researcher in a particular estimate, then the per-comparison p-value for that estimate is appropriate. When considering the many estimates for youth simultaneously, there is a high probability of observing a few large estimates due to sampling variability, even if there was no true effect. To account for these multiple comparisons while restricting the set to a manageable size, we considered inference for three youth subgroups (all, female, and male), two treatments (experimental and Section 8), and five domains (physical health, mental health, risky behavior, education, and overall), which correspond to the 30 estimates in columns (iii)–(viii) of Table II. We calculated familywise adjusted p-values, which are similar to Bonferroni corrections, but adjusted for the ordering of the tests and the covariance of the estimates as described in Appendix B of the Web Appendix. The estimate in this set with the largest t-statistic was the overall summary index for the experimental group, and the probability of observing an effect this large or larger as the maximum of the 30 estimates was less than 0.001 under the joint null hypothesis of no effects for the 30 estimates. The familywise adjusted p-value was 0.003 for the experimental group mental health index for female youth. Based on these calculations, we conclude that the overall pattern and the mental health result for the experimental group were quite unlikely to have occurred by chance, even if the focus was on these results because of their large t-statistics.

Male Youth Outcomes

The results for the overall summary index of male youth outcomes are of almost exactly the same magnitude as for female youth but with the opposite sign, implying more adverse effects in the treatment groups than in the control group. Among specific outcomes, the effects are largest for injuries and for substance use, as shown in Table III, leading to large effects for the male physical health and risky behavior summary indices in Table II. These summary index measures for males have highly significant per-comparison p-values, although familywise adjusted p-values were greater than 0.05.

There are a number of issues that complicate interpretation of the results for male youth. First, males in the treatment groups exhibited more behavior and other problems at baseline than did those in the control group (there were no such baseline differences among females).[30] This imbalance appears largely to be due to the random sampling that occurred when we subsampled children for our interviews, rather than to survey attrition or imbalance in the original random assignment.[31] The key question for our analysis is whether our regression controls for baseline covariates are sufficient to adjust for the imbalance. We suspect that our regression adjustment is sufficient to remove most of the potential bias. Most importantly, in analysis of administrative arrest data using the full set of MTO youth (with little imbalance in covariates and no survey attrition), Kling, Ludwig, and Katz (2005) found adverse treatment effects for males and beneficial effects for females, which supports our conclusion that gender differences in effects are substantial and not simply an artifact of sampling issues.

A second issue is that the rates of some adverse outcomes in the control group seem implausibly low. For example, the proportion of nonsports injuries for male youth in the control group is barely half as high as the injury rate for females, whereas our analysis of National Health Interview Survey data found nonsports injury rates for male youth over 30 percent higher than for female youth. Moreover, assuming that the injury rate among control noncompliers is the same as for treatment noncompliers implies a control complier mean of less than zero. Both of these facts are consistent with a low realization of injury rates for the particular sample and time period for which we have data; we speculate that the injury rate for males would be at least as high for females if we were to run the experiment again. The rates of substance use for

[30]This result comes from a summary index of baseline covariates constructed in the same manner as the outcomes indices (normalizing the control group for each gender to mean 0 and standard deviation 1). The index includes variables collected prior to random assignment for age, gifted classes, school suspension, problems at school, behavior problems, learning problems, physical activity problems, and other medical problems.

[31]For the following discussion, E − C and S − C refer to ITT differences in the baseline covariate index in standard deviation units for male youth (with standard errors). For all 1,604 male youth ages 15–20 in MTO households, the E − C difference was −0.022 (0.031) and the S − C difference was −0.019 (0.034). For the 923 male youth we attempted to survey (drawing two children per household and a 3-in-10 subsample of initial nonrespondents), the E − C difference was −0.133 (0.053) and the S − C difference was −0.108 (0.055). For the 879 youth with whom we completed surveys, the E − C difference was −0.158 (0.054) and the S − C difference was −0.129 (0.056). Because there was less than half a standard error difference, respectively, between the estimates for all male youth for whom surveys were attempted and for those completed but large imbalance between treatment groups in baseline covariates for those attempted, we conclude that the imbalance was largely driven by random sampling. For comparison with the 928 female youth surveyed, the E − C difference was −0.021 (0.044) and the S − C difference was −0.047 (0.055).

males in the control group are also low relative to demographically similar individuals in national data, whereas there are smaller differences between MTO controls and national data for females.[32] Our interpretation of the results is that issues such as random covariate imbalance, survey attrition, and surprisingly low prevalence of adverse outcomes among controls may exaggerate the magnitude of the adverse effects for males, but that these factors are not large enough to account for the differences in effects between males and females.

Understanding Gender Differences

We collected extensive data on mediating factors that could potentially help to determine the mechanisms responsible for the gender differences in the results for youth. In brief, we find that there were large effects of the treatment on neighborhood characteristics and smaller effects on school characteristics, but no significant differences by gender in the treatment effects on neighborhoods or schools (see the Web Appendix, Tables G3–G7 for details).[33] We conclude that female youth and male youth in the treatment groups responded to similar new neighborhood environments in different ways.

Several mechanisms could potentially explain the gender differences in neighborhood effects. There has been a broad trend over the past two decades of gains in education and employment for minority women—gains that have not been shared by minority men (Altonji and Blank (1999)). Moves through MTO may remove barriers to benefiting from these gains for female youth, whereas the male youth have poor prospects even in lower poverty neighborhoods. It is also likely that girls suffer disproportionately from domestic violence and sexual abuse (Popkin, Harris, and Cunningham (2002)), and the MTO intervention may have reduced their exposure to such events, providing benefits from the moves for girls that were not nearly as relevant for boys. We did find some statistically significant survey evidence that female youth are more likely to have three or more adult role models to whom they are comfortable talking about their problems (Web Appendix, Table G5) and that the mean effect size on a summary index of adult contact measures was significantly higher for female youth than for male youth. Although the probability is

[32]In Appendix D of the Web Appendix, we describe our method for producing results for the National Longitudinal Survey of Youth 1997 (NLSY97) that are adjusted to match the demographics of the MTO sample. The following results in parentheses are for the MTO control group, the adjusted NLSY97, and the unadjusted NLSY97, respectively: marijuana (females: 0.13, 0.10, 0.16; males: 0.12, 0.23, 0.18), cigarettes (females: 0.19, 0.25, 0.33; males: 0.13, 0.33, 0.33), and alcohol (females: 0.21, 0.26, 0.44; males: 0.14, 0.32, 0.46). The deviations of the MTO controls from the adjusted NLSY97 across these three outcomes are substantially smaller for females than for males, consistent with a random draw of unusually low substance use among males in the control group.

[33]We also find that the MTO intervention had no significant impacts on parenting practices, peer characteristics, school engagement, or access to health care for either boys or girls (Web Appendix, Tables G4–G7).

quite high of observing at least one effect this large under a joint null hypothesis of no gender differences in the many mediating mechanisms we examined, our interpretation is that differences in adult contact are the most likely contributor to at least part of the gender difference in effects from among the mechanisms about which we have data to examine.[34]

To further develop ideas about why the gender differences in effects might have occurred, a qualitative research effort was designed to follow up on the results reported in this paper. Clampet-Lundquist, Edin, Kling, and Duncan (2006) examined data from semistructured open-ended interviews with 83 youth in Chicago and Baltimore randomly selected from the MTO experimental and control groups. Gender differences were found in effects for outcomes collected as part of this qualitative data, complementing similar results from the 2002 survey and from administrative arrest records. Before seeing the data, our leading hypothesis for why youth who moved to lower poverty neighborhoods might experience adverse outcomes was "relative deprivation"; for example, male youth in the experimental group could find themselves surrounded by relatively higher achieving peers and react negatively to being "further down the ladder." However, the qualitative interviews did not provide any evidence that relative deprivation was a salient experience for male youth in the experimental group.

Clampet-Lundquist et al. did find some evidence to support several other hypotheses. First, the nondominant cultural capital skills (e.g., use of language) that male youth learned in their high-poverty neighborhoods may have isolated them from the mainstream when they moved to lower poverty contexts. This cultural conflict sometimes manifested itself through police harassment or through fear and misunderstanding, and may have led to maladaptive behavior. Second, male youth may have experienced more negative peer effects. Male youth tended to spend their free time on neighborhood basketball courts and on street corners, in closer proximity to illegal activity than the female

[34]We also found that the beneficial effects on adult mental health outcomes overall (such as distress and calmness) were not evident for adults with male youth in their households, with the difference for male–female youth being significant. This finding could be the effect of adverse male outcomes rather than the cause. Male youth may have less effective coping strategies in stressful situations (Zaslow and Hayes (1986), Coleman and Hendry (1999), Kraemer (2000)) and the disruption of moving itself may have been greater for male youth. However, at least two strands of evidence run counter to this hypothesis. First, mobility rates were slightly higher among households with female youth than those with male youth (Web Appendix, Table G8). Second, the adverse effects for males did not manifest themselves right after the initial moves, as predicted by a simple mobility disruption model, but only after several years. In studies of single MTO sites 1–3 years after random assignment, there were either no gender differences in effects reported, or more beneficial effects for males than females (Katz, Kling, and Liebman (2001), Leventhal and Brooks-Gunn (2003a, 2003b)). In administrative arrest data, the adverse effects on male property crime are not found in the first two years after random assignment, but in the third and fourth years (Kling, Ludwig, and Katz (2005)).

youth, who tended to spend free time at home or in malls or other more supervised spaces. Many of the male youth in the control group developed strategies of "keeping to themselves" that helped them stay out of trouble, whereas fewer experimental group males used such strategies. Relatedly, experimental group males may have responded to peer pressure to signal that they had not abandoned their origin neighborhood culture by participating in deviant activities and returning to the origin neighborhood.[35] Third, male youth in the control group had particularly high rates of contact with father figures (stepfathers, mother's boyfriends, and uncles), likely contributing to the differential adult contact effects observed in the survey. In the qualitative study, rates of contact with biological fathers were similar for all groups and genders, and the prevalence of important contact with other adults outside of the family was also low for all.

Reconciling OLS and 2SLS Estimates

If low neighborhood poverty rates were beneficial and the neighborhood selection process operated such that people with unobserved characteristics associated with good outcomes tended to locate in lower poverty neighborhoods, these assumptions would predict that the OLS estimates of effects of higher poverty rates would be more adverse than 2SLS estimates using site–treatment group interactions as instruments for neighborhood poverty. In Table IV, however, we found OLS estimates (based on the control group only) that were often of opposite sign from 2SLS (for the full sample). The implied selection process is that adults and families with female teenagers likely to have adverse outcomes tended to move to low poverty neighborhoods, and families with male teenagers likely to have beneficial outcomes tended to move to low-poverty neighborhoods.[36] These selection patterns suggest that identifying the direction of bias in nonexperimental studies of neighborhood effects can be more complex than is typically assumed.

[35] According to our survey data, male experimental group youth did make more visits to their origin neighborhoods than did females, but the male–female difference is insignificant (Web Appendix, Table G5). For males, a two-audience signaling process (Austen-Smith and Fryer (2005)) could encourage them to avoid peer sanction through participation in deviant group activities rather than engage in more prosocial behaviors that are ultimately valued by employers—a process that could be more important when there is more uncertainty about social group affiliation due to greater racial, ethnic, or economic diversity among peers (as there would tend to be for youth in families using MTO vouchers). Fryer and Torelli (2005) find some evidence consistent with greater peer sanction against prosocial activities of black males than those of females.

[36] These implications about the pattern of residential sorting are borne out within the treatment groups as well. For nearly all outcomes, the compliers are more similar to the noncompliers than to the control group. For female youth and adults, this pattern can only be consistent with beneficial treatment effects if compliers otherwise would have had poor outcomes. For male youth, this pattern can only be consistent with adverse treatment effects if compliers otherwise would have had good outcomes.

Younger Children

Sanbonmatsu et al. (2006) hypothesized that greater effects of the MTO treatment would be found on children who are younger than the youth whose outcomes we study in this paper because the younger children (such as those ages 6–10) will have had less lifetime exposure to the high-poverty neighborhoods, but they found no statistically significant treatment effects for younger children on reading test scores, math test scores, or behavior problems. However, the main outcomes for which we found large treatment effects for teenage youth—mental health problems and risky behavior—have very low prevalence at younger ages, and it is therefore too early to tell whether the outcomes of the younger children will be different.

4. CONCLUSION

Using a housing voucher lottery that caused otherwise similar groups of families to reside in very different neighborhoods, we have investigated the effects of moving out of some of the highest poverty neighborhoods in the United States on outcomes for adults and teenage youth. Our findings—no discernable effects on adult economic self-sufficiency, improvements in adult mental health, beneficial outcomes for teenage girls, and adverse outcomes for teenage boys—have three important implications.

First, housing mobility by itself does not appear to be an effective antipoverty strategy—at least over a 5-year horizon. The MTO demonstration program was motivated by theories and nonexperimental empirical results that suggested that there would be large economic gains from moving to lower poverty neighborhoods. However, we found no consistent evidence of treatment effects on adult earnings or welfare participation. Whether economic gains begin to appear in the longer run, particularly among MTO children, remains to be seen.

Second, even in the absence of economic gains, policies that move families out of distressed public housing projects using rental vouchers are likely to have benefits that significantly exceed their costs. Because the MTO intervention produced large mental health improvements and because other research suggests that it is cheaper to provide a unit of subsidized housing with vouchers than in a public housing project (Olsen (2000)), an offer of a housing voucher is likely to pass the Kaldor–Hicks criterion—the gains to those who benefit would be large enough to hypothetically compensate those who experience adverse effects and still leave those who benefit better off. We note, however, that spillovers onto neighborhoods to which these families moved remain unknown. If there were large negative spillovers, this conclusion could be reversed. In addition, the largely offsetting male and female youth results complicate the welfare analysis.

Third, substantively important neighborhood effects do exist, but only for some outcomes. Teenagers—the population often thought to be most affected by neighborhood conditions—exhibited effects on the broadest range of outcomes. The evidence that the effects of housing vouchers appear to accrue from changes in neighborhood characteristics rather than from moves per se suggests that interventions that substantially improve distressed neighborhoods could have effects as least as large as those observed from moving to lower poverty neighborhoods.[37] Although numerous nonexperimental studies document strong associations between neighborhood characteristics and individual outcomes, these associations appear to be much weaker in the studies with the most credible identification strategies.[38] Because the current study used randomization to solve the selection problem, because it studied families who made very large moves as measured by changes in neighborhood poverty rates, and because it collected extensive data on teenagers, it provides us with the clearest answer so far to the threshold question of whether important neighborhood effects exist.

The Brookings Institution, 1775 Massachusetts Ave., NW, Washington, DC 20036, U.S.A.; jkling@brookings.edu, www.brookings.edu/scholars/jkling.htm,

John F. Kennedy School of Government, Harvard University, 79 JFK Street, Cambridge, MA 02138, U.S.A.; jeffrey_liebman@harvard.edu, www.jeffreyliebman.com,

and

Dept. of Economics, Harvard University, Littauer Center, Cambridge, MA 02138, U.S.A.; lkatz@harvard.edu, post.economics.harvard.edu/faculty/katz/katz.html.

Manuscript received May, 2004; final revision received June, 2006.

APPENDIX A: DESCRIPTION OF BASELINE COVARIATES AND OUTCOMES

The covariates (X) used in Equations (1)–(3) were drawn from data collected in a baseline survey conducted prior to random assignment. For analysis of adults, the covariates were those in Table A1 and six Legendre polynomials for adult date of birth. For analysis of youth, all the covariates in Table A1 were used as well as those in Table A2, six Legendre polynomials for youth date of birth, and five indicators for missing data on special class for gifted students or did advanced work, special school, class, or help for learning problem in past two years, special school, class, or help for behavioral or emotional problems in past two years, problems that made it difficult for him/her to get to school

[37]Bloom, Riccio, and Verma (2005) found substantial positive earnings effects in a community-based intervention.

[38]For example, Ellen and Turner (1997) report that "some recent studies that have done the most careful job of controlling for unobserved family characteristics … find no independent neighborhood effects, casting doubt on the robustness of results from other studies." Also, recent quasi-experimental studies (Jacob (2004), Oreopoulos (2003)) find little or no effect of living in high-poverty housing projects on child outcomes.

TABLE A1

ADULT BASELINE CHARACTERISTICS[a]

Variable	Control Mean (i)	Experimental Mean (ii)	Experimental CP Mean (iii)	Experimental NCP Mean (iv)	Experimental CP−NCP (v)	Section 8 Mean (vi)	Section 8 CP Mean (vii)	Section 8 NCP Mean (viii)	Section 8 CP−NCP (ix)
Demographics									
Age in years (as of December 2001)	39.6	39.7	38.1	41.2	−3.1*	40.1	38.3	42.6	−4.2*
Male	0.02	0.01	0.00	0.02	−0.01*	0.02	0.02	0.02	0.00
Baltimore site	0.15	0.15	0.17	0.13	0.04	0.15	0.18	0.10	0.09*
Boston site	0.21	0.22	0.21	0.22	−0.02	0.22	0.18	0.28	−0.10*
Chicago site	0.22	0.23	0.16	0.30	−0.14*	0.23	0.25	0.19	0.06
Los Angeles site	0.16	0.16	0.21	0.10	0.11*	0.15	0.19	0.09	0.11*
New York site	0.25	0.25	0.25	0.24	0.01	0.25	0.19	0.34	−0.15*
African–American	0.66	0.67	0.67	0.66	0.01	0.66	0.70	0.60	0.10*
Other race	0.27	0.26	0.23	0.29	−0.05*	0.26	0.22	0.31	−0.09*
Hispanic ethnicity, any race	0.29	0.29	0.28	0.29	−0.02	0.30	0.27	0.35	−0.08*
Never married	0.62	0.62	0.66	0.58	0.08*	0.62	0.65	0.58	0.07
Teen parent	0.24	0.25	0.26	0.24	0.02	0.26	0.30	0.21	0.08*
Economic and education									
Working	0.25	0.29	0.29	0.28	0.02	0.25	0.26	0.24	0.03
On AFDC	0.75	0.74	0.76	0.72	0.04	0.75	0.78	0.70	0.08*
In school	0.16	0.16	0.20	0.12	0.07*	0.16	0.18	0.12	0.05*
High school diploma	0.38	0.41	0.41	0.42	−0.01	0.41	0.41	0.40	0.01
General equivalency diploma	0.21	0.18	0.21	0.15	0.06*	0.19	0.20	0.18	0.01

Continues

TABLE A1—*Continued*

Variable	Control Mean (i)	Experimental Mean (ii)	Experimental CP Mean (iii)	Experimental NCP Mean (iv)	Experimental CP−NCP (v)	Section 8 Mean (vi)	Section 8 CP Mean (vii)	Section 8 NCP Mean (viii)	Section 8 CP−NCP (ix)
Household									
Had car	0.15	0.17	0.19	0.15	0.04	0.16	0.18	0.14	0.05
Household member with a disability	0.16	0.16	0.15	0.17	−0.02	0.17	0.14	0.20	−0.06*
Household member victimized by crime during past 6 months	0.41	0.42	0.46	0.39	0.07*	0.43	0.45	0.39	0.05
No teen children	0.62	0.59	0.66	0.53	0.13*	0.61	0.67	0.52	0.15*
Household of size 2	0.20	0.23	0.27	0.19	0.09*	0.21	0.23	0.18	0.05
Household of size 3	0.32	0.30	0.31	0.30	0.01	0.31	0.30	0.31	−0.01
Household of size 4	0.22	0.23	0.23	0.24	−0.01	0.23	0.23	0.22	0.00
Neighborhood and housing									
Lived in neighborhood 5 or more years	0.62	0.61	0.60	0.62	−0.02	0.63	0.57	0.72	−0.15*
Moved more than 3 times in past 5 years	0.11	0.08+	0.09	0.07	0.02	0.09	0.11	0.06	0.05*
Very dissatisfied with neighborhood	0.46	0.46	0.52	0.41	0.11*	0.47	0.52	0.39	0.13*
Streets very unsafe at night	0.49	0.48	0.52	0.45	0.07*	0.49	0.53	0.43	0.10*
Chats with neighbors at least once a week	0.55	0.52	0.49	0.55	−0.06*	0.50	0.50	0.50	0.00

Continues

TABLE A1—*Continued*

Variable	Control Mean (i)	Experimental Mean (ii)	Experimental CP Mean (iii)	Experimental NCP Mean (iv)	Experimental CP−NCP (v)	Section 8 Mean (vi)	Section 8 CP Mean (vii)	Section 8 NCP Mean (viii)	Section 8 CP−NCP (ix)
Neighborhood and housing (continued)									
Respondent very likely to tell neighbor if saw neighbor's child getting into trouble	0.56	0.53	0.50	0.57	−0.07*	0.55	0.56	0.53	0.03
No family living in neighborhood	0.65	0.65	0.66	0.64	0.02	0.62	0.63	0.60	0.03
No friends living in neighborhood	0.41	0.40	0.43	0.38	0.05	0.38	0.40	0.34	0.06
Very sure would find an apartment in another part of city	0.45	0.45	0.51	0.40	0.11*	0.48	0.54	0.40	0.14*
To get away from gangs or drugs was primary or secondary reason for moving	0.78	0.77	0.79	0.75	0.04	0.75	0.77	0.73	0.05
Better schools was primary or secondary reason for moving	0.48	0.47	0.50	0.46	0.04	0.52	0.53	0.49	0.05
Had applied for Section 8 voucher before	0.45	0.41	0.44	0.39	0.05	0.39+	0.38	0.40	−0.03
N	1,080	1,453	694	759		993	585	408	

[a]AFDC, Aid to Families with Dependent Children; CP, complier; NCP, noncomplier; *, difference between treatment compliers and noncompliers is statistically significant at the 5 percent level; +, difference between treatment and control mean is statistically significant at 5 percent level.

TABLE A2

YOUTH BASELINE CHARACTERISTICS[a]

	Female			Male		
	Exp (1)	Sec8 (2)	Con (3)	Exp (4)	Sec8 (5)	Con (6)
African–American	0.68	0.64	0.67	0.64	0.65	0.59
Special class for gifted students or did advanced work	0.15	0.17	0.17	0.17*	0.15*	0.27
Special school, class, or help for learning problem in past 2 years	0.13	0.13	0.12	0.29	0.25	0.30
Special school, class, or help for behavioral or emotional problems in past 2 years	0.07	0.08	0.05	0.18	0.17	0.11
Problems that made it difficult to get to school and/or to play active games	0.03	0.06	0.06	0.11*	0.08	0.05
Problems that required special medicine and/or equipment	0.05	0.07	0.05	0.13	0.14	0.09
School asked to talk about problems child having with schoolwork or behavior in past 2 years	0.19	0.23	0.19	0.41	0.37	0.33
Suspended or expelled from school in past 2 years	0.09	0.10	0.07	0.23	0.20	0.15

[a] Exp, experimental; Sec8, Section 8; Con, control; *, p-value < 0.05 on difference between experimental or Section 8 and control group. Baseline data were collected at random assignment, during 1994–1997. Surveys were completed in experimental, Section 8, and control groups with 749, 510, and 548 respondents, respectively, ages 15–20 on 12/31/2001 for a total sample size of 1,807.

and/or to play active games or sports, and suspended or expelled from school in past two years.

Descriptions of the 15 outcomes examined for adults and the 15 examined for youth follow:

Adult economic self-sufficiency. Our measure of employment is an indicator for whether the adult had worked for pay during the week prior to survey.

Welfare receipt is measured as being a beneficiary of Temporary Assistance for Needy Families (TANF) at the time of the survey.

Economic self-sufficiency is an indicator for working for pay during the previous week and not receiving TANF.

Earnings in 2001 is the amount self-reported to have been earned from all employers before taxes and deductions during 2001.

Government income is the amount received altogether in the form of TANF, Supplemental Security Income (SSI), unemployment benefits, Social Security, General Assistance, and related programs in 2001.

Adult physical health. To assess overall health, respondents were asked, "In general is your health excellent, very good, good, fair, or poor?" The analyses examined whether the respondent reported that he or she was in fair or poor health.

Our measure of physical limitation was whether the respondent reported having at least a little trouble "lifting or carrying groceries" or "climbing several flights of stairs."

Respondents were asked questions from the National Health Interview Survey sequence on asthma or wheezing attacks. As our dichotomous measure, we examined the fraction of respondents who had an attack during the past year.

Subjects self-reported their height and weight. We use the standard definition of obesity, BMI ≥ 30 kg/m^2 (National Institutes of Health (1998)).

Our measure of hypertension is based on the JNC7 stage 1 systolic and diastolic blood pressure cut-points: systolic $\geq$140 mm HG or diastolic $\geq$90 mm HG (Chobanian et al. (2003)). Systolic and diastolic blood pressures were measured from a single reading near the end of the survey from an Omron HEM-737. This device satisfied the American Association for the Advancement of Medical Instrumentation standards for accuracy (Anwar et al. (1998)). Subjects were seated and had been at rest for at least 30 minutes.

Adult mental health. Distress during the past 30 days was assessed using the K6 scale, developed by Kessler et al. (2002). This scale score can range from 0 to 24, which we normalize to a z-score by subtracting the mean of 5.8 and dividing by the standard deviation of 5.4.

Depression was assessed using the Composite International Diagnostic Interview—Short Form (CIDI-SF; Kessler et al. (1998)). If during a 2-week period in the past year the respondent reported dysphoric mood (feeling "sad, blue, or depressed") or anhedonia (having "lost interest in most things"), then he or she was assigned a probability of having had a major depressive episode (MDE) according to the number endorsed of seven possible symptoms that correspond to those used for *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition (DSM-IV) psychiatric diagnosis. The probability is based on a mapping between the CIDI-SF screening questions and more detailed assessments in the National Comorbity Survey (Walters, Kessler, Nelson Christopher, and Mroczek (2002)).

For worrying, respondents were asked the two initial screening items from the CIDI-SF sequence on generalized anxiety disorder, and we analyzed the fraction of the sample who answered "yes" to "felt worried, tense, or anxious" or "worried a lot more than most people would in your situation" (Kessler et al. (1998)).

Respondents were asked if they felt "calm and peaceful" at least most of the time during the past month—one of the items from the mental health inventory in the RAND Health Insurance Experiment and the Short Form-36 (Ware, Snow, Kosinski, and Gandek (1993)).

Sleep was measured as the amount of time that participant usually spent sleeping each night, and we analyzed the fraction that usually sleep at least 7 and less than 9 hours per night.

Youth education. Our measure of whether a youth had graduated from high school or was still in school is based on a parental report.

To assess idleness, we asked youth whether they were in school, on summer vacation from school, working during the past week, or none of the above.

Reading achievement was assessed using the Woodcock–Johnson Revised (WJ-R) Broad Reading assessment, which includes letter–word identification and passage comprehension subtests.

Math achievement was assessed using the Woodcock–Johnson Revised (WJ-R) Broad Math assessment, which includes applied problems and calculation subtests. As discussed in Sanbonmatsu et al. (2006), some systematic differences in scores on tests administered by particular interviewers were detected for reading and math tests, and the results presented in Table III are adjusted for interviewer effects after controlling for census tract fixed effects.

Youth physical health. Our analysis of youth health is based on self-reported information in response to questions drawn mainly from the National Health Interview Survey about general health status, injuries, asthma attacks, height, and weight. Case, Lubotsky, and Paxson (2002) and Currie and Stabile (2003) find that self-reported poor health correlates strongly with children's chronic conditions, bed days, and hospitalization episodes.

For asthma attacks, we follow the standard practice of combining attacks requiring medical attention with other episodes of wheezing or whistling in the chest (Pearce, Beasley, Burgess, and Crane (1998)).

We asked for details of any injuries, accidents, or poisonings that required medical attention or were serious enough to limit activities during the previous 12 months, and we focused our analysis on nonsports injuries.

To assess obesity, we collected self-reported height and weight and calculated the body mass index for each individual. Our measure of obesity is body mass index greater than the 95th percentile of the national norms for the youth's age and gender.

Youth mental health. Our distress measure, developed by Kessler et al. (2002) for the National Health Interview Survey, is commonly scored by summing the scale scores of the items, with the total ranging from 0 to 24. This measure is commonly known as the K6 and is based on a six-item Likert scale that measures how much of the time during the past 30 days the youth felt "so depressed nothing could cheer you up," "nervous," "restless or fidgety," "hopeless," "everything was an effort," or "worthless." Our results are reported as z-scores, scaling by the standard deviation. The sample mean is 5.0 and the standard deviation is 4.7.

Our measure of serious depression involved a series of screening questions about the duration and intensity of the feelings and the presence of related symptoms during the worst period in life. A youth is considered to have had a major depressive episode during his or her lifetime if he or she met the following five conditions:

A. The youth experienced a period in which for most of the day he or she felt one of the following: sad, empty or depressed, very discouraged or hopeless about how things were going in his or her life, or loss of interest and boredom with most things usually enjoyed like work, hobbies, and personal relationships.

B. Either felt this way most of the day almost every day for a period of 2 weeks or longer, or for a period of 3 days or longer and had a year or more in life when felt this way just about every month for several days or longer.

C. During times when mood was most severe and frequent, the feelings usually lasted not less than 3 hours a day.

D. These feelings were either more than mild, sometimes felt so bad that nothing could cheer him or her up, or sometimes felt so bad that he or she could not carry out daily activities.

E. These feelings were accompanied by changes in sleeping, eating, energy, his or her ability to keep mind on things, feeling badly about his or herself, or other problems.

Our measure of anxiety also involved a series of screening questions about the duration and intensity of the feelings and the presence of related symptoms during the worst period in life. A youth was considered to have had generalized anxiety disorder during his or her lifetime if the following four criteria were met:

A. The youth reported there was a period when he or she was either worrying a lot more about things than other people with the same problems, much more nervous or anxious than most people with the same problems, or anxious or worried most days.

B. The youth reported being worried about nothing in particular, everything, or more than one specific thing.

C. The youth sometimes or often either found it hard to stop the worries or anxiety or could not think about anything else no matter how hard he or she tried.

D. The period of being anxious, nervous, or worried lasted at least one month.

Youth risky behavior. Our measures of risky behavior drew upon survey self-reports using items from the National Longitudinal Survey of Youth 1997.

Alcohol use was measured as any use within the past 30 days.

Marijuana use was measured as any use within the past 30 days.

Smoking was measured as any cigarette smoking within the past 30 days.

Our measures of whether a female youth had borne a child or a male youth had fathered a child were also based on self-reported responses to survey questions.

REFERENCES

ALTONJI, J., AND R. BLANK (1999): "Race and Gender in the Labor Market," in *Handbook of Labor Economics*, Vol. 3C, ed. by O. Ashenfeleter and D. Card. Amsterdam: North-Holland, 3143–3260.

ANGRIST, J. D., G. W. IMBENS, AND D. B. RUBIN (1996): "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91, 444–472.

ANWAR, Y. A., S. GIACCO, E. J. MCCABE, B. E. TENDLER, AND W. B. WHITE (1998): "Evaluation of the Efficacy of the Omron HEM-737 Intellisense Device for Use on Adults According to the

Recommendations of the Association for Advancement of Medical Instrumentation," *Blood Pressure Monitoring*, 3, 261–265.
ARNOTT, R., AND J. ROWSE (1987): "Peer Group Effects and Educational Attainment," *Journal of Public Economics*, 32, 287–305.
AUSTEN-SMITH, D., AND R. FRYER (2005): "An Economic Analysis of Acting White," *Quarterly Journal of Economics*, 120, 551–583.
BECKER, G., AND K. M. MURPHY (2000): *Social Economics*. Chicago: University of Chicago Press.
BENABOU, R. (1993): "Workings of a City: Location, Education, and Production," *Quarterly Journal of Economics*, 108, 619–652.
BLOOM, H. S., J. A. RICCIO, AND N. VERMA (2005): *Promoting Work in Public Housing: The Effectiveness of Jobs-Plus.* New York: MDRC.
BOS, J., T. BROCK, G. DUNCAN, R. GRANGER, A. HUSTON, AND V. LLOYD (1999): *New Hope for People with Low Incomes: Two-Year Results of a Program to Reduce Poverty and Reform Welfare.* New York: MDRC.
BROCK, W. A., AND S. N. DURLAUF (2001): "Interactions-Based Models," in *Handbook of Econometrics*, Vol. 5, ed. by J. J. Heckman and E. Leamer. Amsterdam: North-Holland, 3297–3380.
BROOKS-GUNN, J., G. J. DUNCAN, AND L. ABER (1997): *Neighborhood Poverty: Context and Consequences for Children*, Vol. I. New York: Russell Sage Foundation.
CASE, A. C., D. LUBOTSKY, AND C. H. PAXSON (2002): "Economic Status and Health in Childhood: The Origins of the Gradient," *American Economic Review*, 92, 1308–1334.
CHOBANIAN, A. V., G. L. BAKRIS, H. R. BLACK, W. C. CUSHMAN, L. A. GREEN, J. L. IZZO JR., D. W. JONES, B. J. MATERSON, S. OPARIL, J. T. WRIGHT JR., AND E. J. ROCCELLA (2003): "The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: The JNC 7 Report," *Journal of the American Medical Association*, 289, 2560–2572.
CLAMPET-LUNDQUIST, S., K. EDIN, J. R. KLING, AND G. J. DUNCAN (2006): "Moving At-Risk Youth Out of High-Risk Neighborhoods: Why do Girls Fare Better than Boys?" Working Paper, Industrial Relations Section, Princeton University.
COLEMAN, J., AND L. B. HENDRY (1999): *The Nature of Adolescence.* London: Routledge.
CRANE, J. (1991): "The Epidemic Theory of Ghettos and Neighborhood Effects on Dropping out and Teenage Childbearing," *American Journal of Sociology*, 96, 1226–1259.
CURRIE, J., AND M. STABILE (2003): "Socioeconomic Status and Health: Why Is the Relationship Stronger for Older Children?" *American Economic Review*, 93, 1813–1823.
DAVIDSON, R., AND J. G. MACKINNON (1993): *Estimation and Inference in Econometrics*. New York: Oxford University Press.
DE BARTOLOME, C. (1990): "Equilibrium and Inefficiency in a Community Model with Peer Effects," *Journal of Political Economy*, 98, 100–113.
ELLEN, I. G., AND M. A. TURNER (1997): "Does Neighborhood Matter? Assessing the Recent Evidence," *Housing Policy Debate*, 8, 833–866.
ENTWISLE, D. R., K. L. ALEXANDER, AND L. S. OLSON (1994): "The Gender Gap in Math: Its Possible Origins in Neighborhood Effects," *Child Development*, 67, 2400–2416.
FERNANDEZ, R., AND R. ROGERSON (1996): "Income Distribution, Communities and the Quality of Public Education," *Quarterly Journal of Economics*, 111, 135–164.
FRYER, R., AND P. TORRELLI (2005): "An Empirical Analysis of Acting White," Unpublished Manuscript, Harvard University.
GOERING, J. M., AND J. D. FEINS (eds.) (2003): *Choosing a Better Life: Evaluating the Moving to Opportunity Social Experiment*. Washington, DC: Urban Institute Press.
GRANOVETTER, M. (1978): "Threshold Models of Collective Behavior," *American Journal of Sociology*, 83, 1420–1443.
HECKMAN, J. J. (1990): "Varieties of Selection Bias," *American Economic Review*, 80, 313–318.
——— (2001): "Accounting for Heterogeneity, Diversity and General Equilibrium in Evaluating Social Programs," *Economic Journal*, 111, 654–699.

IHLANFELDT, K. R., AND D. J. SJOQUIST (1998): "The Spatial Mismatch Hypothesis: A Review of Recent Studies and Their Implications for Welfare Reform," *Housing Policy Debate*, 9, 849–892.

IMBENS, G. W., AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475.

JACOB, B. A. (2004): "Public Housing, Housing Vouchers, and Student Achievement: Evidence from Public Housing Demolitions in Chicago," *American Economic Review*, 94, 233–258.

JENCKS, C., AND S. E. MAYER (1990): "The Social Consequences of Growing up in a Poor Neighborhood," in *Inner City Poverty in the United States*, ed. by L. Lynn Jr. and M. McGeary. Washington, DC: National Academy Press, 111–186.

KAIN, J. F. (1968): "Housing Segregation, Negro Employment and Metropolitan Decentralization," *Quarterly Journal of Economics*, 82, 175–197.

KATZ, L. F., J. R. KLING, AND J. B. LIEBMAN (2001): "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment," *Quarterly Journal of Economics*, 116, 607–654.

KAWACHI, I., AND L. F. BERKMAN (2003): "Introduction," in *Neighborhoods and Health*, ed. by I. Kawachi and L. F. Berkman. Oxford: Oxford University Press, 1–19.

KESSLER, R. C., G. ANDREWS, L. J. COLPE, E. HIRIPI, D. K. MROCZEK, S. T. NORMAND, E. E. WALTERS, AND A. M. ZASLAVSKY (2002): "Short Screening Scales to Monitor Population Prevalences and Trends in Non-Specific Psychological Distress," *Psychological Medicine*, 32, 959–976.

KESSLER, R. C., G. ANDREWS, D. K. MROCZEK, T. B. USTUN, AND H. U. WITTCHEN (1998): "The World Health Organization Composite International Diagnostic Interview Short-Form (CIDI-SF)," *International Journal of Methods in Psychiatric Research*, 7, 171–185.

KESSLER, R. C., P. R. BARKER, L. J. COLPE, J. F. EPSTEIN, J. C. GFROERER, E. HIRIPI, M. J. HOWES, S. T. NORMAND, R. W. MANDERSHEID, E. E. WALTERS, AND A. M. ZASLAVSKY (2003): "Screening for Serious Mental Illness in the General Population," *Archives of General Psychiatry*, 60, 184–189.

KLING, J. R., AND J. B. LIEBMAN (2004): "Experimental Analysis of Neighborhood Effects on Youth," Working Paper 483, Industrial Relations Section, Princeton University.

KLING, J. R., J. B. LIEBMAN, AND L. F. KATZ (2005): "Bullets Don't Got No Name: Consequences of Fear in the Ghetto," in *Discovering Successful Pathways in Children's Development: New Methods in the Study of Childhood and Family Life*, ed. by T. S. Weisner. Chicago: University of Chicago Press, 243–281.

——— (2007): "Supplement to 'Experimental Analysis of Neighborhood Effects'," *Econometrica Supplementary Material*, 75, http://www.econometricsociety.org/ecta/supmat/tables.pdf.

KLING, J. R., J. B. LIEBMAN, L. F. KATZ, AND L. SANBONMATSU (2004): "Moving to Opportunity and Tranquility: Neighborhood Effects on Adult Economic Self-Sufficiency and Health from a Randomized Housing Voucher Experiment," Working Paper 481, Industrial Relations Section, Princeton University.

KLING, J. R., J. LUDWIG, AND L. F. KATZ (2005): "Neighborhood Effects on Crime for Female and Male Youth: Evidence from a Randomized Housing Voucher Experiment," *Quarterly Journal of Economics*, 120, 87–130.

KRAEMER, S. (2000): "The Fragile Male," *British Medical Journal*, 321, 23–30.

LEVENTHAL, T., AND J. BROOKS-GUNN (2003a): "Moving to Opportunity: An Experimental Study of Neighborhood Effects on Mental Health," *American Journal of Public Health*, 93, 1576–1582.

——— (2003b): "New York City Site Findings: The Early Impacts of Moving to Opportunity on Children and Youth," in *Choosing a Better Life: Evaluating the Moving to Opportunity Social Experiment*, ed. by J. M. Goering and J. D. Feins. Washington, DC: The Urban Institute Press, 213–244.

LIEBMAN, J. B., L. F. KATZ, AND J. R. KLING (2004): "Beyond Treatment Effects: Estimating the Relationship Between Neighborhood Poverty and Individual Outcomes in the MTO Experiment," Working Paper 493, Industrial Relations Section, Princeton University.

LUCAS, J. W., J. S. SCHILLER, AND V. BENSON (2004): *Vital and Health Statistics, Series 10, Number 218.* Washington, DC: National Center for Health Statistics.

LUDWIG, J., G. J. DUNCAN, AND P. HIRSCHFIELD (2001): "Urban Poverty and Juvenile Crime: Evidence from a Randomized Housing-Mobility Experiment," *Quarterly Journal of Economics*, 116, 655–679.

MANSKI, C. F. (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60, 531–542.

NATIONAL INSTITUTES OF HEALTH (1998): "Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults—The Evidence Report," *Obesity Research*, 6, 51S–209S.

O'BRIEN, P. C. (1984): "Procedures for Comparing Samples with Multiple Endpoints," *Biometrics*, 40, 1079–1087.

OLSEN, E. O. (2000): "The Cost-Effectiveness of Alternative Methods of Delivering Housing Subsidies," Unpublished Manuscript, University of Virginia.

——— (2003): "Housing Programs for Low-Income Households," in *Means-Tested Transfer Programs in the United States*, ed. by R. Moffitt. Chicago: University of Chicago Press and NBER, 365–441.

OREOPOULOS, P. (2003): "The Long-Run Consequences of Growing up in a Poor Neighborhood," *Quarterly Journal of Economics*, 118, 1533–1575.

ORR, L., J. D. FEINS, R. JACOB, E. BEECROFT, L. SANBONMATSU, L. F. KATZ, J. B. LIEBMAN, AND J. R. KLING (2003): *Moving to Opportunity: Interim Impacts Evaluation.* Washington, DC: U.S. Department of Housing and Urban Development.

PEARCE, N., R. BEASLEY, C. BURGESS, AND J. CRANE (1998): *Asthma Epidemiology: Principles and Methods*. Oxford: Oxford University Press.

POPKIN, S. J., L. E. HARRIS, AND M. K. CUNNINGHAM (2002): *Families in Transition: A Qualitative Analysis of the MTO Experience*. Washington, DC: Urban Institute.

RAMIREZ-VALLES, J., M. A. ZIMMERMAN, AND L. JUAREZ (2002): "Gender Differences of Neighborhood and Social Control Process: A Study of the Timing of First Intercourse among Low-Achieving, Urban, African American Youth," *Youth and Society*, 33, 418–442.

SANBONMATSU, L., J. R. KLING, G. J. DUNCAN, AND J. BROOKS-GUNN (2006): "Neighborhoods and Academic Achievement: Results from the MTO Experiment," *Journal of Human Resources*, forthcoming.

TURNEY, K., S. CLAMPET-LUNDQUIST, K. EDIN, J. R. KLING, AND G. J. DUNCAN (2006): "Neighborhood Effects on Barriers to Employment: Results from a Randomized Housing Mobility Experiment," *Brookings–Wharton Papers on Urban Affairs 2006*, forthcoming.

WALSH, B., S. N. SEIDMAN, R. SYSKO, AND M. GOULD (2002): "Placebo Response in Studies of Major Depression: Variable, Substantial, and Growing," *Journal of the American Medical Association*, 287, 1840–1847.

WALTERS, E. E., R. C. KESSLER, C. B. NELSON CHRISTOPHER, AND D. M. MROCZEK (2002): "Scoring the World Health Organization's Composite International Diagnostic Interview Short Form (CIDI-SF)," http://www3.who.int/cidi/CIDISFScoringMemo12-03-02.pdf.

WARE, J. E., K. K. SNOW, M. KOSINSKI, AND B. GANDEK (1993): *SF-36 Health Survey: Manual and Interpretation Guide.* Boston: The Health Institute, New England Medical Center.

WELLS, K. B., C. SHERBOURNE, M. SCHOENBAUM, N. DUAN, L. MEREDITH, J. UNÜTZER, J. MIRANDA, M. F. CARNEY, AND L. V. RUBENSTEIN (2000): "Impact of Disseminating Quality Improvement Programs for Depression in Managed Primary Care: A Randomized Controlled Trial," *Journal of the American Medical Association*, 283, 212–220.

WESTFALL, P. H., AND S. S. YOUNG (1993): *Resampling-Based Multiple Testing: Examples and Methods for P-Value Adjustment.* New York: Wiley.

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

ZASLOW, M. J., AND C. D. HAYES (1986): "Sex Differences in Child's Response to Psychosocial Stress: Toward a Cross-Context Analysis," in *Advances in Developmental Psychology*, Vol. 4, ed. by M. Lamb, A. Brown, and B. Rogoff. Hillsdale, NJ: Lawrence Erlbaum.