

MDRC Working Papers on Research Methodology

**Exploring the Feasibility and Quality
of Matched Neighborhood
Research Designs**

**David C. Seith
Nandita Verma
Howard S. Bloom
George C. Galster**



June 2003

The Urban Change Neighborhood Indicators Database was supported by: the Cleveland, Ford, Charles Stewart Mott, W.K. Kellog, Robert Wood Johnson, John S. and James L. Knight, Joyce, George Gund, William Penn, James Irvine, California Wellness, and Edna McConnell Clark Foundations as well as the Pew Charitable Trusts and the U.S. Department of Health and Human Services (including interagency funds from the U.S. Department of Agriculture). The Impact Methods Initiative is supported by the Pew Charitable Trusts. The Neighborhood Work Advancement and Support Center Funding Partners include the Rockefeller and Annie E. Casey Foundations.

Dissemination of MDRC publications is also supported by the following foundations that help finance MDRC's public policy outreach and expanding efforts to communicate the results and implications of MDRC's work to policymakers, practitioners, and others: The Atlantic Philanthropies; the Alcoa, Ambrose Monell, Bristol-Myers Squibb, Fannie Mae, Ford, Grable, and Starr Foundations; and the Open Society Institute.

The findings and conclusions presented in this paper do not necessarily represent the official positions or policies of the funders.

Author Affiliations and Contact Information

David C. Seith, Nandita Verma, Howard S. Bloom (MDRC); George C. Galster (Wayne State University). Please send questions or comments to howard.bloom@mdrc.org.

For information about MDRC[®] and copies of our publications, see our Web site: www.mdrc.org.
Copyright © 2003 by MDRC. All rights reserved.

Abstract

Many evaluations of neighborhood-level interventions have relied on neighborhood matching strategies, but to the authors' knowledge the validity of this methodology has not been tested. This paper responds to the demand for rigorous evidence on the use of quasi-experimental neighborhood matches for assessing the effectiveness of community-wide interventions. Using neighborhood-level data in Cleveland, Ohio and Philadelphia, Pennsylvania, to match potential target neighborhoods to virtual comparison neighborhoods, this paper evaluates the number of target neighborhoods that can be matched and how well they stay matched over time. It identifies a range of matching variables and constraints that appear to strike the best balance between matchability and match quality. The results compare favorably to those generated by two less restrictive alternatives. The paper ends with suggestions for replication in other sites, with other outcomes, and in other time periods. It tentatively concludes that the neighborhood-matching algorithm described in this analysis is both operationally feasible and offers respectable accuracy in detecting the magnitude of impacts that might be expected from neighborhood-based employment interventions.

Contents

Abstract	iii
List of Tables and Figures	vii
Introduction	1
Approach	2
Discussion of Measurement Error Requirements	11
Findings	13
Next Steps	29
Applying the Model to Other Cities	29
Conclusions	30
Appendix	31
References	37

List of Tables and Figures

Table

1	Matching Constraints	9
2	Distribution of the Absolute Pre-Post Measurement Error	11
3	Matchability Diagnostics	14
4	Matching Constraints and Matchability: Cleveland	17
5	Matching Constraints and Match Quality: Cleveland	19
6	Matching Constraints and Matchability: Philadelphia	23
7	Matching Constraints and Match Quality: Philadelphia	25

Figure

1	Neighborhood Matching Exercises: St. Clair-Superior and Seven Comparison Tracts	7
2	A Portrait of the Median Measurement Error Surface Given Variations in Three Matching Constraints	28
A.1	1990 Poverty Rates of Potential Target Tracts Cleveland, Ohio	31
A.2	1990 Ethnic Composition of Potential Target Tracts Cleveland, Ohio	32
A.3	1990 Median Amount of Home Purchase Loans Relative to Average Among Target Tracts Cleveland, Ohio	33
A.4	1990 Poverty Rates of Potential Target Tracts Philadelphia, Pennsylvania	34
A.5	1990 Ethnic Composition of Potential Target Tracts Philadelphia, Pennsylvania	35
A.6	1990 Median Amount of Home Purchase Loans Relative to Average Among Target Tracts Philadelphia, Pennsylvania	36

Introduction

This paper responds to the demand for rigorous evidence on the efficacy of different methodologies for assessing the impact of neighborhood-level interventions. An increasing awareness of the neighborhood-level concentration of poverty has motivated a substantial investment of resources in neighborhood-level interventions to achieve a variety of individual and neighborhood impacts. But individual random assignment research designs, the strongest research designs for assessing program impacts, are simply not applicable to neighborhood-based interventions, which in principle and practice offer services to all eligible residents. Instead, many evaluations employ quasi-experimental comparisons of target neighborhoods and comparison neighborhoods matched in a pre-intervention period. A comprehensive review of these evaluations concludes that despite sophisticated matching methods, the change trajectories of target and comparison neighborhoods matched at baseline could diverge over the course of the intervention due to secular influences, increasing the risk of error in the assessment of program impacts.¹ The authors of that literature review call for rigorous assessments of the degree of error likely to arise in matched neighborhood designs.

This paper summarizes an empirical exercise conducted in response to that challenge. Specifically, this analysis:

- identifies an eligible pool of target neighborhoods that mimic those likely to be targeted for neighborhood-level interventions;
- matches them to “virtual comparison neighborhoods” (explained below) using easily accessible decennial census tract data and annual home purchase loan data;
- evaluates the “target neighborhood matchability,” which is the likelihood that any given target neighborhood can be matched in the 1992-1994 pre-intervention period;
- estimates the probability distribution of the measurement error for the program impact indicator, using the four-year change in annual, tract-level welfare receipt rates as an outcome measure;
- compares the results of these matching exercises with two less stringent alternatives;

¹Hollister and Hill, 1995.

- offers empirical guidance concerning: the practical challenges of identifying and defining eligible target and virtual comparison neighborhoods, the range of variables found to provide an acceptable matching surface, and the trade-offs between matchability and measurement error yielded by altering matching constraints within that range.

This paper begins with a discussion of the approach that was used. It describes the unique database that made the analysis possible and the definition of neighborhoods used for this analysis. It describes three sets of variables: those used to *screen* out ineligible neighborhoods, those used to *match* eligible target and virtual comparison neighborhoods, and the *outcome variable*. It also describes the *diagnostics* that were used to assess the two dimensions of matching performance — matchability and measurement error. The second section steps back from the mechanics just described to discuss the operational and research performance requirements of the matching exercise. The third section evaluates the findings and offers empirical guidance concerning the practical challenges mentioned above. The fourth section identifies next steps for replicating the findings of this exercise in other sites and with other outcomes. The fifth section describes operational lessons that could be learned by applying the matching algorithms developed in this paper to other cities. Finally, the paper concludes by discussing the implications of these findings for future measurement and evaluation research.

Approach

The first challenge implied by the research question is to identify: an *outcome* variable related to aggregate employment behavior in a neighborhood, a set of universally available neighborhood *matching* variables, and the *screening variables* that will identify neighborhoods eligible to be targeted for an intervention. In this simulation exercise, as in actual interventions, all screening and matching is performed using publicly available data in a pre-intervention period. The unique contribution of this simulation is the opportunity it affords to assess the measurement error for a potential outcome indicator in the absence of an actual intervention. In an actual intervention, it is practically impossible to rigorously assess the degree of target vs. virtual comparison neighborhood differences arising from influences completely exogenous to the program. In fact, to the extent that evaluators become aware of such differences at all it is usually in the worst case scenarios when the exogenous events impacting target and comparison tracts are so severe that they undermine the entire inferential integrity of the evaluation.

The database and outcome variable

Consider the MDRC Neighborhood Work Advancement and Support Center Initiative (NWASC), a proposed 4-6 year intervention that would offer work supports and employment services to all residents of an urban neighborhood, typically viewed as an area of 7,000-20,000

people. Obviously one of the most important performance outcomes for the NWASC initiative would be a neighborhood-level measure of employment. If eligible target and virtual comparison neighborhoods could be matched on baseline characteristics such that their employment trajectories for the following 4-6 years would be very similar in the absence of any interventions, the minimum research group comparability required for a quasi-experimental design would be secured. Note that it is not necessary that the target and virtual comparison neighborhoods have equivalent levels of baseline or follow-up employment, only that their employment trajectories are similar.

Unfortunately, it is practically impossible to obtain administrative records of employment and earnings outcomes for all residents of each of the potential target and virtual comparison neighborhoods. However, MDRC was fortunate to have access to a related indicator, the proportion of neighborhood residents receiving public assistance, specifically the annual average monthly welfare receipt rate. MDRC's Urban Change Neighborhood Indicators Database (UCNID) offers a complete, geo-coded 1992-2000 benefit history of all TANF, Food Stamps, and Medicaid recipients living in Cuyahoga and Philadelphia Counties. The UCNID contains monthly benefit receipt information, demographic characteristics, estimated benefit payment, and residential address for each case that ever received public assistance between 1992 and 2000. For the purposes of this exercise, it was possible to create the annual average monthly rate of all individual neighborhood residents receiving TANF. As welfare receipt rates are inversely correlated with employment rates, they provide a reasonable proxy for neighborhood employment.² Neighborhood welfare receipt rates were measured for a baseline period (1992-1994 monthly average) and a follow-up period (1996-1998 monthly average).

Identifying target neighborhoods

The challenge here is specifying target neighborhoods that would mimic those where an actual NWASC intervention might occur. The objective is to utilize to the extent possible "neighborhood" boundaries that have been delineated meaningfully by local authorities, yet are also consistent with the NWASC service model and the research goal of accurate impact assessment. *Target neighborhoods* were operationalized as collections of poor census tracts within local planning areas, where tracts must meet several qualifying conditions and collectively contain 7,000 – 20,000 residents. The details of this operationalization follow.

First, this report takes advantage of two local definitions of a neighborhood. One definition is the census tract, which is a relatively permanent, county subdivision delineated by local

²For example, in Cleveland, average 1992-1994 welfare rates were positively correlated with the 1990 unemployment rate ($r = 0.61^{***}$) and negatively correlated with the 1990 total labor force participation rate ($r = -0.45^{***}$).

appointees of the US Census Bureau to be homogeneous with respect to population characteristics, economic status, and living conditions; in metropolitan areas tracts typically contain 4,000 residents. The other definition is a statistical planning area (SPA) — a geographic area bearing some corporate identity from the perspective of the locality and consisting of 1 to 30 contiguous census tracts, depending on the city.

Second, within each SPA constituent census tract(s) were evaluated to see if they met qualifying criteria consistent with the NWASC service model and research requirements of the proposed NWASC impact assessment, as follows.

- **Consistent with the service model.** Neighborhoods targeted by the NWASC service model are urban and residential, with moderate to high concentrations of poverty, and are not dominated by site-based public housing. Urban tracts lie within the municipal limits (in this cases, the cities of Cleveland or Philadelphia). Residential tracts have at least 100 residents in 1990 that are not zoned for non-residential land uses. Poverty rates of 20 percent or higher identify moderate- to high-poverty tracts, and tracts where 60 percent or more of residential properties were designated as site-based public housing were excluded from the analysis.³
- **Amenable to evaluation research.** Target neighborhoods considered amenable to evaluation research must not be missing information on the key matching and outcome variables.⁴ Moreover, because (1) observed neighborhood-wide program impacts are least likely to be observed in areas with high rates of mobility, and (2) the proposed NWASC initiative is designed to offer services to residents over a 4-6 year period, this analysis stipulates that tract level five-year mobility rates (observed during the prior census) must be less than or equal to 75 percent.⁵

³As the percentage of total tax assessed properties that were owned by the Philadelphia Housing Authority did not exceed 24 percent in any tract, none of the Philadelphia tracts were excluded for this reason.

⁴One otherwise eligible tract in Cleveland and sixteen otherwise eligible tracts in Philadelphia were excluded due to missing Home Mortgage Disclosure Act (HMDA) data.

⁵The decennial US Census tracks the one and five-year mobility rates at the census tract level, that is the proportion of tract residents who have lived in the same address for one and five years. Nationally, the average five-year residential mobility rate in tracts with poverty rates greater than 20 percent is 50 percent. Importantly, the residential mobility rate is most likely higher than the tract level mobility rate, because not all families who move from one address to another move out of the census tract. One of the operating assumptions of place-based initiatives is that residents remain in the neighborhood long enough to benefit from the intervention. What proportion of residents must remain and for how long are matters of judgment. This analysis has specified a 75 percent maximum five-year residential mobility rate. In Philadelphia, where the average five-year residential mobility rate ranges from 7 percent to 83 percent, 12 otherwise eligible tracts were excluded due to

(continued)

Third, census tracts qualifying through the above criteria were aggregated within their corresponding SPAs. Because the NWASC intervention is intended to serve neighborhoods with 7,000-20,000 residents, this requirement was imposed on the constituent tracts of the target neighborhoods. SPAs whose constituent qualifying census tracts totaled fewer than 7,000 residents were removed from the analysis, while those with more than 20,000 residents in qualifying tracts were subdivided into smaller tract clusters of appropriate populations to constitute target neighborhoods.⁶

The matching algorithm

Once target neighborhoods were selected, the matching process could begin. It is crucial at the outset to recognize that, although the ultimate objective is to match target neighborhoods, the match itself is executed at the census tract level. That is, for each constituent tract in a particular target neighborhood the algorithm searches for other tracts outside the neighborhood that match it on specified matching variables. More formally, this exercise employs a “one-to-many, matching with replacement” sequence. “One-to-many,” means that each target tract comprising a given neighborhood is matched to one or more comparison tracts.

For an illustration of the tract-to-tract matching algorithm, consider Figure 1. On the Northeastern shores of Cleveland, Ohio sits the L-shaped neighborhood of St. Clair-Superior, a locally defined neighborhood outlined in gray with 8 tracts illustrated in different patterns. Each of these constituent tracts within St. Clair-Superior qualified as part of the St.-Clair Superior target neighborhood according to the criteria above; they were residential, urban census tracts without missing data, where poverty rates were 20 percent or greater, five-year mobility rates were less than 75 percent, and site-based public housing occupied less than 60 percent of each tract. Each of these 8 tracts, the constituent building blocks of the St. Clair-Superior neighborhood, were independently matched to one or more comparison tracts from the City of Cleveland at large.

For example, note that the gray tract in the Southwest corner of the St. Clair-Superior neighborhood is matched to two, separate, light gray tracts in the Southwest sec-

higher than acceptable five-year mobility rates. In Cleveland, where the average five-year residential mobility rate ranges from 9 percent to 69 percent, no tracts were excluded due to higher than acceptable five-year mobility rates. Thus, site variations in the five-year residential mobility rate will affect the applicability of place-based interventions in general, and of this matching strategy in particular.

⁶Of the 30 eligible neighborhoods in Cleveland, 10 neighborhoods had fewer than 7,000 residents in 1990 and were removed from the analysis. Two neighborhoods had more than 20,000 residents in 1990. These were subdivided into 4 neighborhoods and added to the original 18 to yield a pool of 22 eligible neighborhoods. Of the 31 eligible neighborhoods in Philadelphia, 7 had fewer than 7,000 residents in 1990 and were removed from the analysis. Thirteen neighborhoods had more than 20,000 residents in 1990. These were subdivided into 34 neighborhoods and added to the original 11 to yield a pool of 45 eligible neighborhoods.

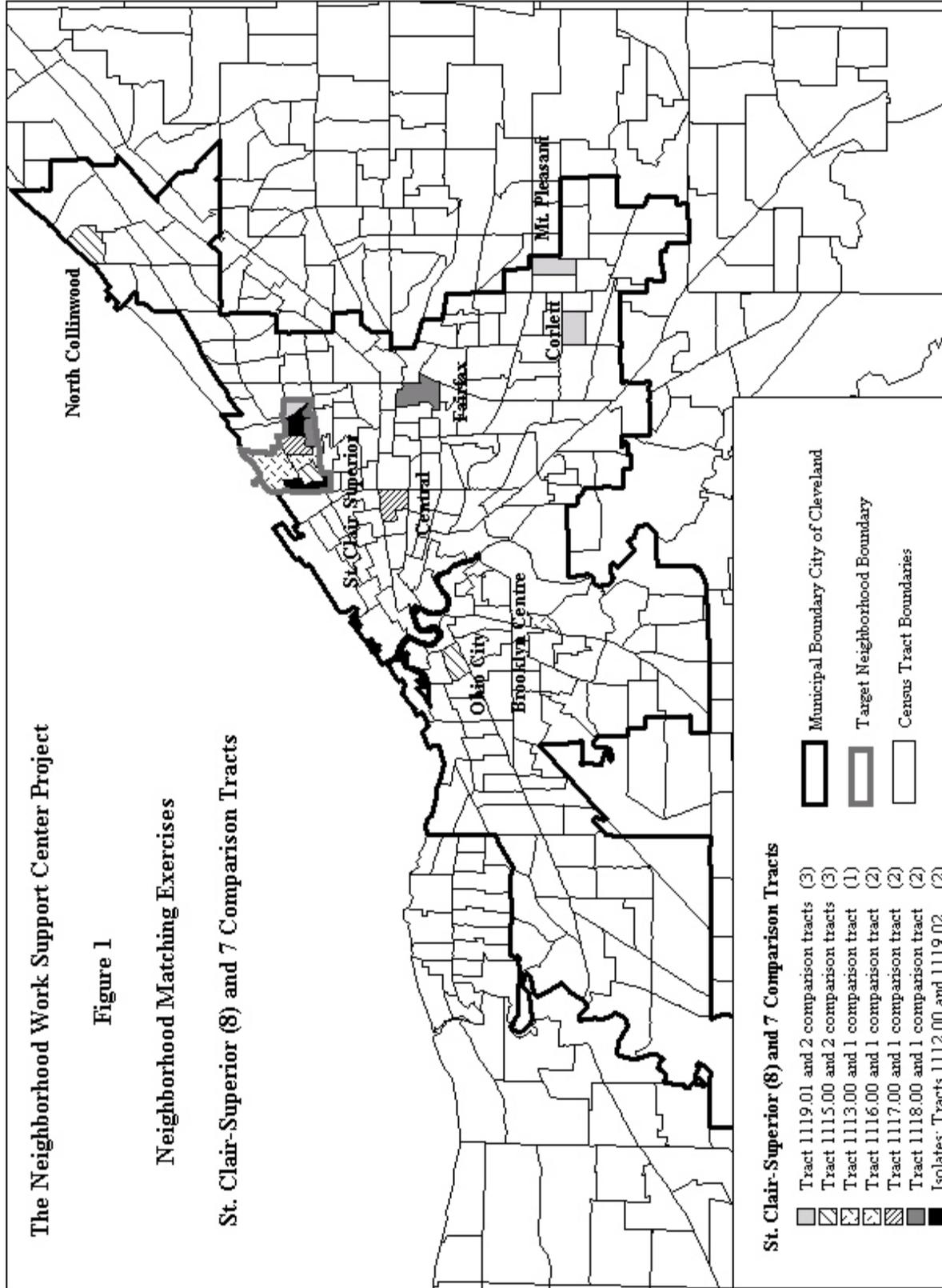
tion of Cleveland, one from the Corlett neighborhood and the other from the Mt. Pleasant neighborhood. “Matching with replacement,” means that each of the comparison tracts can be used as a match more than once. For example, the single comparison tract from the Brooklyn Centre neighborhood is independently matched to two separate tracts in St. Clair-Superior.

Note that not every target tract within every target neighborhood necessarily can be matched. Target tracts that are not matched are called “isolates.” As an illustration in the St. Clair-Superior neighborhood shown in Figure 1, census tracts that matched to others are shaded in patterns and isolates are shaded in black. Unfortunately, in a real experimental context the isolates might well be included in the demonstration, but could not be used in the impact assessment. Thus, it is preferable to minimize the number of isolates in a given intervention site but, as discussed below, this cannot be accomplished without a concomitant increase in measurement error.

Next, once each of the target tracts cycles through the matching algorithm, the outcome differences of each matched target tract are averaged at the neighborhood level. More simply, the change in welfare receipt from the baseline period (1992-1994) to the follow-up period (1996-1998) is calculated separately for each target tract. These tract level pre-post changes in welfare receipt are averaged at the neighborhood level for all successfully matched tracts of the target neighborhood. Tract level pre-post changes in welfare receipt are also calculated for each comparison tract. The weighted average of these tract-level pre-post changes in welfare receipt among each comparison tract successfully matched to one or more target tracts becomes the counterfactual.

The absolute difference between the average target neighborhood matched tracts’ pre-post change in welfare receipt and the weighted average of the pre-post change in welfare receipt for the comparison tracts becomes *the indicator of the measurement error* for the given target neighborhood. This measurement error is essentially the naturally occurring difference between the target neighborhood and its comparison tracts that would be inappropriately attributed to the intervention in a quasi-experimental context.

Before leaving the discussion of the matching algorithm, two final points must be made. Initially, note that a target neighborhood is not compared to an area consisting of a contiguous collection of census tracts, but rather a collection of perhaps widely scattered comparison census tracts, which this report calls a “virtual comparison neighborhood.” As an illustration, while the target tracts in Figure 1 all belong by definition to one target neighborhood, St. Clair-Superior, the comparison tracts are drawn from neighborhoods throughout the city — North Collinwood, Central, Fairfax, Mt. Pleasant, Corlett, Ohio City, and Brooklyn Centre. Thus,



although one may think of a target neighborhood and a comparison “neighborhood,” the latter is a *virtual neighborhood*, the average of similar, comparison tracts located throughout the city. Although this dispersion of comparison tracts initially may seem unnatural, it arguably provides a better “reading” of outcomes in comparable, “control” areas citywide, in the same way that the temperature of a swimming pool is better gauged by readings from several areas. (See comparison tract redundancy and virtual comparison neighborhood redundancy below.)

Finally, note how spatial autocorrelation comes into play in this analysis. Target neighborhoods may differ in the degree to which their constituent target tracts are contiguous and proximate, especially in larger SPAs. Given that one might expect inter-tract spillovers (greater spatial autocorrelation) to be greater in the case of target neighborhoods comprised of tightly clustered, contiguous target tracts, the measured local impacts of an intervention like NWASC might well be greater in such cases due to stronger spatial synergisms.⁷

The matching constraints

With the aforementioned points about the matching algorithm in mind, consider the matching constraints shown below in Table 1. The specification of these constraints was guided by the principles of *face validity* and *predictive validity*. Apart from their instrumental value, three of the four matching constraints — poverty, ethnicity, and the mean home purchase mortgage amount — were imposed to ensure that the match would incorporate dimensions of neighborhood identity and process that would be expected of any legitimate evaluation. Although it may have been possible to adequately match tracts using other indicators, each of these characteristics taps into something conventionally considered fundamental to neighborhood identity and arguably relevant to neighborhood employment outcomes. (For an illustration of residential patterns of poverty, ethnicity, and mean home purchase loan amounts in Cleveland and Philadelphia see Appendix Maps A.1 – A.6.)

While the face validity afforded by poverty and ethnicity will be readily appreciated, the third of these indicators — the median amount of home purchase loans — may merit a brief explanation. Responding to concerns about credit discrimination in the mortgage lending market, in 1975 the federal government passed the Home Mortgage Disclosure Act (HMDA), requiring lenders to report the approval rates and amounts for home purchase by applicant characteristics (such as race, sex), and by property location. HMDA loan amounts have been found to

⁷A target neighborhood likely has considerably stronger spatial autocorrelation among its constituent tracts than the perhaps widely scattered comparison tracts of its corresponding virtual comparison neighborhood. However, this probably does not bias the estimate of impact. Like each target tract, each comparison tract is subject to a welter of exogenous and endogenous spatially autocorrelated forces that cannot be directly observed, even if these forces are emanating from tracts that are not employed in this analysis. This does not obviate their ability to serve as proxies for the counterfactual of the target tracts.

be closely correlated with single-family home sales prices at the census tract level across multiple cities.⁸ Economists have demonstrated that a wide range of neighborhood amenities and disamenities are capitalized into the values of properties located nearby.⁹ Thus, mortgage loan data provide a broad summary measure of the market's evaluation of a whole range of neighborhood demographic, economic, social, public service, and physical characteristics.

The Neighborhood Work Support Center Project

Table 1

Matching Constraints

Constraint	Description	Source
Poverty rate in 1990	The proportion of individual tract residents whose income falls below the federal poverty threshold.	U.S. Census
Ethnic composition 1990	To be eligible for a match, the compositional proportion of each of three ethnic groups in the comparison tract — White, Black, and Hispanic — must match the respective proportion in the target tract within the range specified.	U.S. Census
Standard Deviation of Home Purchase Loan Amount 1993-1995	Based on Home Mortgage Disclosure Act of 1975 (HMDA) reports. Thresholds of $\pm\$9,669$ and $\pm\$4,835$ represent one standard deviation and 0.5 standard deviation of the distribution of home purchase loans among the 129 target tracts in Cleveland. Thresholds of $\pm\$23,629$ and $\pm\$11,815$ represent one standard deviation and 0.5 standard deviation of the distribution of home purchase loans among the 123 target tracts in Philadelphia.	Home Mortgage Disclosure Act (HMDA) Reports
Distance	The minimum required distance in miles between a target tract and a comparison tract.	Calculated ¹⁰

⁸Walker, et al., 2002.

⁹Polinsky and Shavell, 1976; Bartik, 1988; Grieson and White, 1989; Palmquist, 1992.

¹⁰Distances are measured from target tract to comparison tract centroids in degrees of latitude and longitude and then converted to miles.

The fourth matching constraint, the minimum required distance in miles between each target and comparison tract, was necessitated by concerns about contamination between target and comparison areas for a real NWASC demonstration. Proximity to the treatment site is often a consideration of neighborhood-level interventions, where the risks of contamination increase with proximity.¹¹ In this case, the closer the target and virtual comparison neighborhoods, the greater the chance that residents of the virtual comparison neighborhoods may receive treatments intended only for residents of the target neighborhood, and treated residents of the target neighborhood may move into virtual comparison neighborhoods during the treatment period.

Measurement Error Diagnostics

The second dimension of assessment is measurement error, the degree to which target and comparison neighborhoods matched in a baseline period (1992-1994) would evidence dissimilar welfare receipt rate values in a follow-up period (1996-1998). While matchability speaks to the applicability of the matching scenario to many neighborhoods in many cities, measurement error speaks to the inferential integrity of the research design, the fundamental target and virtual comparison neighborhood comparability that supports causal attributions of impact in the former.

In this exercise, measurement error is defined as the absolute average percentage point difference in the pre-post change in welfare receipt rates in the target neighborhood and virtual comparison neighborhood. Simply, the appropriately weighted average pre-post change in welfare receipt among all successfully matched target tracts is compared to the average pre-post change in welfare receipt among all comparison tracts. The absolute value of this difference is an indication of the measurement error in a quasi-experimental design.

The frequency distribution of this error term provides us with a sense of the risk of making errors of given magnitudes for any one of the neighborhood matches yielded by a set of matching constraints. That is, the frequency distribution of this error term can be interpreted as the estimated probability distribution of a true error term. For example, the minimum and maximum errors provide a sense for the best and worse neighborhood matches, given the specified matching constraints. Table 2 below explains the interpretation of the frequency distribution of the pre-post measurement error.

¹¹Hollister and Hill, 1995.

The Neighborhood work Support Center Project

Table 2

Distribution of the Absolute Pre-Post Measurement Error

Diagnostic	Description
Absolute Pre-Post Measurement Error	This diagnostic provides the clearest assessment of the quality of the match. It shows the percentage point difference between the pre-post change in welfare receipt rates measured in target and comparison tracts matched at baseline. Since this difference was observed in the absence of an intervention, it provides a rough estimate of the outcome differences to be expected among otherwise matched tracts due to secular influences. In an evaluation, estimates would be inaccurate by this amount, expressed in probabilistic terms.
Minimum	The minimum measurement error provides an estimate of the best-case scenario, that is the best match obtained for any of the target neighborhoods.
25 th Percentile	One quarter of the matched target neighborhoods would be mismatched by this amount or less.
Median	Half of the matched target neighborhoods would be mismatched by this amount or less.
75 th Percentile	Three quarters of the matched target neighborhoods would be mismatched by this amount or less.
Maximum	The maximum measurement error provides an estimate of the worst-case scenario, that is the worst match obtained for any of the target neighborhoods.

Discussion of Measurement Error Requirements

Before reviewing the findings of this matching exercise, it may be useful to consider the range of matchability and measurement error demanded in an operational setting comparable to the NWASC intervention envisioned.

Improving matchability

Often target neighborhoods are selected *a priori* due to exemplary treatment programs already in operation. In such cases, evaluators are not afforded the luxury of choosing the best-matched target and virtual comparison neighborhoods. Instead, they are given the challenge to

find adequately matched virtual comparison neighborhoods for the specific target neighborhoods in specific cities.¹² Thus the *target neighborhood matchability* diagnostic responds to one of the first questions that will be asked of the evaluation team in a field application: what are the chances that the evaluation team can find an adequate virtual comparison neighborhood for a given target neighborhood in a given city? If target neighborhoods with promising programs were scarce, ideal target neighborhood matchability levels might range from 75 percent to 100 percent. If funders are committed to launching the initiative in particular cities, the proportion of all neighborhoods that qualify as target neighborhoods will be as important an indicator of feasibility as the target neighborhood eligibility. As programs are selected and implemented, target neighborhood coverage will prove an important indicator of the proportion of neighborhood residents included in the evaluation. Higher coverage rates will be demanded where community based organizations refuse to redefine their catchment areas or neighborhood residents demand to be served. Less obvious at first, the resiliency provided by comparison tract and neighborhood redundancies might safeguard demonstrations launched in urban areas where comparison tracts can be affected by unexpected, selective, localized events.

Reducing measurement error

While acceptable ranges of matchability will vary by situational concerns and subjective assessments requiring programmatic experience, existing empirical research clearly structures the expectations for impacts on welfare receipt and the consequences of measurement error. Welfare reform programs with mandatory participation requirements, for example, often lead to caseload declines. A recent MDRC cross-project synthesis of impacts on adult welfare receipt from experimental evaluations of five welfare-to-work programs in twenty sites, all of which implemented mandatory participation requirements, found impacts on welfare receipt rates ranged between minus twelve and minus two percentage points, and declined by five percentage points on average.¹³ Welfare reform programs with generous earnings supplements, on the other hand, tend to increase benefit receipt. Impacts on adult welfare receipt among six welfare reform programs with generous earnings supplements ranged from one to eleven percentage points, with an average increase of six percentage points.¹⁴ Welfare reform programs with time limits often generate both positive and negative impacts on welfare receipt. Impacts on adult welfare receipt among four programs with time limits ranged from minus four to four percentage points, and almost canceled out at an average of minus one half of a percentage point.¹⁵

¹²Hollister and Hill, 1995.

¹³Bloom and Michalopolous, 2001. Since none of the programs reviewed in this publication were targeted at the neighborhood level, some might expect neighborhood-level saturation interventions to yield greater impacts than those presented here.

¹⁴Bloom and Michalopolous, 2001.

¹⁵Bloom and Michalopolous, 2001.

Thus the range of impacts to be expected depends on the nature of the programs implemented. Declines of five percentage points are as common in successful welfare-to-work programs with mandatory participation requirements as increases of six percentage points or more in generous earnings supplement programs. For the purposes of this analysis, it would seem reasonable to expect successful welfare reform programs to generate impacts with absolute values ranging from four to five percentage points. What this means is that the evaluation team would try to identify (and then employ in a future intervention) a stringency of matching whereby there was an acceptably high probability that this measurement error would be considerably smaller than this range.

Findings

Matchability Diagnostics

As mentioned above, matchability refers to the applicability of the matching method to target neighborhoods and their constituent tracts within a specific city. Given that both target neighborhoods and virtual comparison neighborhoods consist of tracts, there are four dimensions of matchability.

- **Target Neighborhood Eligibility.** Target neighborhood eligibility refers to the proportion of target neighborhoods that were successfully matched.
- **Target Neighborhood Coverage.** Target neighborhood coverage refers to the average population residing in matched tracts within target neighborhoods. As mentioned above, the proposed NWASC initiative is designed to serve neighborhoods with total populations of 7,000-20,000 residents, so this analysis seeks coverage yielding that range.
- **Comparison Tract Redundancy for a Robust Design.** Comparison tract redundancy refers to the average number of unique comparison tracts matched to target neighborhoods. The greater this number, the greater the redundancy and thus the robustness of the research design. Redundancy describes any system or structure that, because of its internal duplication, is less vulnerable to failure when one or more components are compromised. One of the most common pitfalls of quasi-experimental designs is the attrition of comparison sites.¹⁶ For example, if a public housing development were razed or massive gentrification were to occur in a comparison tract, many low-

¹⁶Hollister and Hill, 1995.

The Neighborhood Work Support Center Project

Table 3

Matchability Diagnostics

Diagnostics		Description
Eligibility	Number of Target Neighborhoods with 7,000 – 20,000 Residents in 1990	The matching exercises are executed at the census tract level. Diagnostics for matched target tracts are aggregated to the neighborhood level. Neighborhoods with 7,000-20,000 residents in 1990 are eligible for selection as demonstration candidates.
	Number of Target Neighborhoods Matched	This diagnostic shows the number of neighborhoods containing at least one matched target tract.
Coverage	Average Population of Target Neighborhoods Matched	Even among the target neighborhoods selected, not all constituents tracts will be matched. In fact, some tracts will not be eligible for selection. This diagnostic shows the average total population of the matched tracts of the target neighborhoods.
Redundancy	Average Number of Unique Comparison Tracts Per Selected Target Neighborhood	The exercises employs a “matching with replacement” strategy, which means that comparison tracts can be matched to more than one target tract. Over the course of the evaluation, unusual events may occur in comparison tracts that would compromise their comparability with the target tracts, in which cases the compromised comparison tracts would be removed from the analysis. This diagnostic speaks to the neighborhood-level redundancy of the methods by providing an unduplicated count of the average number of unique tracts per selected target neighborhood.
	Average Number of Unique Virtual Comparison Neighborhoods	See description above. In addition to the tract level replacement discussed above, several unique comparison tracts may belong to the same neighborhood, which begs the question — how many unique neighborhoods would be selected as comparison sites. This diagnostic speaks to the neighborhood-level redundancy of the methods by providing an unduplicated count of the average number of unique actual comparison neighborhoods per selected target neighborhood.

income households would be displaced, and local outcome indicators might thereby appear to improve. Conversely, sewer or street repairs in a virtual comparison neighborhood tract may alter automobile and pedestrian traffic in a vital commercial district such that business is adversely affected, which might in turn adversely affect local employment. Matching scenarios that yield a greater number of comparison tracts increase the resilience of the research design to such exogenous events. A second dimension of comparison tract redundancy is the geographic spread of the comparison tracts across the cityscape. Comparison tracts could be clustered in a few neighborhoods or scattered throughout the city. The latter is preferable. As mentioned above, just as temperature readings from multiple areas of a swimming pool give a more accurate indication of water conditions, so comparison tracts drawn from more neighborhoods throughout the city will yield a more accurate read of comparison outcomes. Matching scenarios that draw comparison tracts from throughout the city likely will also be more resilient to local exogenous events.

How similar should target and comparison tracts be in their 1990 poverty rates, ethnic composition, and median home purchase loan amounts? Presumably, more stringent matching constraints would reduce measurement error, but at what cost in matchability? Similarly, how far apart is far enough? Farther distance requirements between target and comparison tracts might more effectively guard against inter-neighborhood contamination, but would probably impose costs, both in terms of matchability and measurement error. This exploration permits us to empirically gauge the tradeoffs involved in answering these questions.

Establishing a Prototype for Cleveland

The matching exercise was first developed in Cleveland, where multiple permutations of different matching constraints were evaluated. The results were encouraging. As shown in Table 4, the surface of matchability and measurement error spanned the following ranges of each matching constraint:

- a $\pm 2\%$ to $\pm 5\%$ match in 1990 poverty rate,
- a minimum separation distance of one mile,
- a $\pm 5\%$ to $\pm 10\%$ match in ethnic composition,
- a ± 0.5 to ± 1 standard deviation match in median home loan amount.

Matchability

Within this range of matching constraints as shown in Table 4 columns five through eight, between 36 percent and 77 percent of target neighborhoods were matched. Within those neighborhoods, the average total population of the matched neighborhoods ranged from 11,500 to 12,600. Eight to twenty unique comparison tracts were available, on average, for each matched target neighborhood. These comparison tracts were drawn from an average of six to ten unique neighborhoods, depending on the particular parameters used for matching variables.

In order to gain perspective on these results, two less stringent alternative matching scenarios are introduced in the last two rows of Table 4. The first alternative scenario matches target neighborhoods to the average of all other poor, urban, residential tracts at least one mile away. The second alternative matches target neighborhoods to ten randomly selected poor, urban, residential tracts at least one mile away. Note that these less restrictive alternatives match all twenty-two target neighborhoods and yield a much larger number of unique comparison tracts and neighborhoods. It stands to reason that the selectiveness imposed by the more stringent runs poses some costs in terms of matchability in order to achieve higher quality matches, as measured below.

Measurement Error

Columns five through nine of Table 5 show the distribution of the error among Cleveland target neighborhoods that were successfully matched. Half of the neighborhoods matched yield an error between 1.0 and 1.8 percentage points or less, depending on the stringency of the matching constraints. Seventy-five percent of target neighborhood matches would yield an error between 1.8 and 2.2 percentage points. Ninety percent would yield an error between 2.3 and 2.9 percentage points, and all matched neighborhoods would yield an error less than 4.2 percentage points. These findings suggest that the minimum detectable effects of the matches achieved in Cleveland were likely within the range of the impacts expected of successful welfare reform programs (roughly 5 percentage points).

As above, comparing these results against those afforded by less stringent alternative matching scenarios provides perspective on the risk of measurement error. The second to last row of Table 5 shows the distribution of error yielded when target neighborhoods are matched to all other poor, urban, residential tracts at least one mile away. Note that the risk of measurement error is attenuated in the more stringent runs employing this battery of matching criteria. For example, the measurement error at the 75th percentile of target neighborhoods matched to all poor tracts is 2.29, suggesting that 25 percent of the matches achieved by this method will

The Neighborhood Work Support Center Project

Table 4

Matching Constraints and Matchability: Cleveland

Matching Scenario Number	<u>Matching Constraints</u>			<u>Matchability</u>			
	Poverty Rate in 1990	Ethnic Composition 1990 ^a	Standard Deviation of Home Purchase Loan Amount 1993-1995 ^b	Number of Target Neighborhoods Matched ^c	Avg. Population of Target Neighborhoods Matched ^d	Avg. Number of Unique Tracts Per Target Neighborhood ^e	Avg. Number of Unique Comparison Neighborhoods Per Target Neighborhood ^f
1	±5%	±10%	±1	17	11,748	19.7	9.7
2	±2%	±10%	±1	14	11,802	11.9	7.9
3	±5%	±5%	±1	14	11,587	18.9	9.6
4	±5%	±10%	±0.5	14	11,905	14.0	7.9
5	±2%	±5%	±1	10	12,624	12.4	8.2
6	±2%	±5%	±0.5	8	11,479	7.6	6.4
All poor tracts	-	-	-	22	11,652	133.7	32.8
Ten randomly selected tracts	-	-	-	22	11,652	10.0	8.6

Table 4 (Continued)

SOURCE: MDRC calculations using data from the Urban Change Neighborhood Indicators Database.

NOTES: ^a To be eligible for a match, the proportion of each of three ethnic groups in the comparison tract- White, Black, and Hispanic- must match the respective proportion in the target tract within the number of percentage points specified in the column.

^b Based on data from Home Mortgage Disclosure Act of 1975 (HMDA) reports. Thresholds of \pm \$9,669 and \pm \$4,835 represent one standard deviation and 0.5 standard deviation of the distribution of home purchase loans among the 129 target tracts in Cleveland. Thresholds of \pm \$23,629 and \pm \$11,815 represent one standard deviation and 0.5 standard deviation of the distribution of home purchase loans among the 123 target tracts in Philadelphia.

^c The matching exercises are executed at the census tract level. Diagnostics for matched target tracts are aggregated to the neighborhood level. Target neighborhoods had 7,000-20,000 residents in 1990. There are a total of 22 target neighborhoods in Cleveland and 45 target neighborhoods in Philadelphia.

^d Even among the target neighborhoods selected, not all constituent tracts will be matched. In fact, some tracts will not be eligible for selection. This column shows the average total population of the matched tracts of the target neighborhoods.

^e The exercises employ a "matching with replacement" strategy, which means that comparison tracts can be matched to more than one target tract. Over the course of the evaluation, unusual events may occur in comparison tracts that would compromise their comparability with the target tracts, in which case the compromised comparison tracts would be removed from the analysis. This diagnostic speaks to the tract-level redundancy of the method, i.e. its resilience to comparison tract attrition, by providing an unduplicated count of the average number of unique comparison tracts per matched target neighborhood.

^f See note above. In addition to the tract level replacement discussed above, several unique comparison tracts may belong to the same neighborhood, which begs the question- how many unique neighborhoods would be selected as comparison sites. This diagnostic speaks to the neighborhood-level redundancy of the method by providing an unduplicated count of the average number of unique comparison neighborhoods per selected target neighborhood.

The Neighborhood Work Support Center Project

Table 5

Matching Constraints and Match Quality: Cleveland

Matching Scenario Number	Poverty Rate in 1990	Ethnic Composition 1990 ^a	Standard Deviation of Home Purchase Loan Amount 1993-1995 ^b	Distribution of the Absolute Pre-Post Measurement Error				
				Minimum	Median	75th Percentile	90th Percentile	Maximum
1	±5%	±10%	±1	0.05	1.75	2.11	2.52	2.65
2	±2%	±10%	±1	0.07	1.45	2.02	2.70	2.97
3	±5%	±5%	±1	0.04	1.20	2.14	2.32	2.54
4	±5%	±10%	±0.5	0.21	0.99	1.86	2.87	4.09
5	±2%	±5%	±1	0.31	1.16	1.76	2.43	2.66
6	±2%	±5%	±0.5	0.24	1.33	2.18	2.81	2.81
All poor tracts	-	-	-	0.24	0.83	2.29	3.40	4.87
Ten randomly selected tracts	-	-	-	0.10	1.28	2.47	3.62	5.18

Table 5 (Continued)

SOURCE: MDRC calculations using data from the Urban Change Neighborhood Indicators Database.

NOTES: ^a To be eligible for a match, the proportion of each of three ethnic groups in the comparison tract- White, Black, and Hispanic- must match the respective proportion in the target tract within the number of percentage points specified in the column.

^b Based on data from Home Mortgage Disclosure Act of 1975 (HMDA) reports. Thresholds of +\$9,669 and +\$4,835 represent one standard deviation and 0.5 standard deviation of the distribution of home purchase loans among the 129 target tracts in Cleveland. Thresholds of +\$23,629 and -\$11,815 represent one standard deviation and 0.5 standard deviation of the distribution of home purchase loans among the 123 target tracts in Philadelphia.

yield a measurement error of this magnitude or larger. This error is larger (by at least 0.11 percentage points) than any of those reported in the rows above, suggesting that even the worst 25 percent of the more stringent runs yield more precise matches. Similarly, the last row of Table 5 shows the distribution of error yielded by 100 iterations of a scenario that matches target neighborhoods to ten randomly selected tracts. The average measurement error at the 75th, 90th, and 100th percentiles is greater (by about 0.29 to 1.08 percentage points) than the measurement error yielded by the more stringent scenarios, again suggesting that the increased stringency of these runs affords some protection against risk of measurement error.

Effects of Changing the Stringency of Matching Constraints

Increasing the stringency of the matching constraints tends to reduce measurement error at the cost of matchability, as one would expect. Increasing the stringency of the poverty constraint from 5 percent to 2 percent decreased the median error by 0.3 percentage points, while reducing the number of target neighborhoods with at least one match from 17 to 14 (out of 22 total target neighborhoods in Cleveland). Increasing the stringency of the ethnicity constraint from 10 percent to 5 percent decreased the median error by 0.6 percentage points, while reducing the number of target neighborhoods with at least one match from 17 to 14. Increasing the stringency of the home loan constraint from one standard deviation to half of a standard deviation decreased the median error by 0.8 percentage points, while reducing the number of matched neighborhoods from 17 to 14. Simultaneously increasing the stringency of the poverty and ethnicity constraints decreased the median error by 0.6 percentage points, while reducing the number of target neighborhoods with at least one match from 17 to 10. Thus, while within the surface of matching constraints bulleted above, increasing the stringency of the matching constraint tends to improve measurement error by nearly a percentage point, it also costs 3-7 matches. Finally, within this range of constraints no ideal run, no “silver bullet” was identified that simultaneously maximized matchability and minimizes measurement error.

Testing the Prototype in Philadelphia

To assess the robustness of the Cleveland prototype the matching algorithm was replicated using data for Philadelphia. Based on the Cleveland results, two hypotheses were specified for Philadelphia. First, within the aforementioned ranges of each of the four matching criteria identified for Cleveland, a similar proportion of matched tracts and a similar range of the absolute error were expected. Second, it was hypothesized that within this range of constraints there would be no ideal set of constraints that would simultaneously improve matchability and measurement error. Both of the hypotheses specified for Philadelphia were supported and the results seemed acceptable in Philadelphia, if not quite as favorable as in Cleveland.

Matchability

As shown in columns five through eight in Table 6, within the same range of matching constraints specified for Cleveland, between 29 percent and 69 percent of Philadelphia target neighborhoods were matched. Within those neighborhoods, the average population of the matched tracts ranged from 11,500 to 12,500 residents. Six to fifteen unique comparison tracts were available, on average, for each matched target neighborhood. These comparison tracts were drawn from an average of five to nine unique neighborhoods.

As for Cleveland, two less stringent alternative matching scenarios are introduced in the last two rows of Table 6. The first alternative scenario matches Philadelphia target neighborhoods to the average of all other poor, urban, residential tracts at least one mile away. The second alternative matches target neighborhoods to ten randomly selected poor, urban, residential tracts at least one mile away. As was true for Cleveland, these less restrictive alternatives match all forty-five Philadelphia target neighborhoods and yield a much larger number of unique comparison tracts and neighborhoods. Obviously, the selectiveness imposed by the more stringent runs imposes some nontrivial costs in terms of matchability in order to achieve higher quality matches, as measured below.

Measurement Error

Columns five through nine of Table 7 show the distribution of the error among Philadelphia target neighborhoods that were successfully matched.¹⁷ Half of the target neighborhoods matched would yield an error no greater than between 1.1 and 1.8 percentage points. Seventy-five percent would yield an error between 1.9 and 4.2 percentage points. Ninety percent would yield an error between 2.3 and 2.9 percentage points, and all matches would yield an error less than 13.3 percentage points. Thus, most but not all of the matches obtained in Philadelphia would support the detection of welfare receipt impacts of at least ± 5 percentage points.

As for Cleveland, these results are compared against those afforded by two less stringent alternative matching scenarios. The second to last row of Table 7 shows the distribution of error yielded when Philadelphia target neighborhoods are matched to all other poor, urban, residential tracts at least one mile away. As in Cleveland, the risk of measurement error is attenuated in the more stringent Philadelphia runs. For example, the measurement error at the 75th

¹⁷Note that the distribution of the outcome variable, change in average monthly welfare receipt rate, is wider than the distribution for Cleveland. The standard deviation of the outcome variable is 4.5 in Philadelphia and 3.2 in Cleveland. Because the outcome measure is more variable in Philadelphia, both the risk and magnitude of measurement errors are greater.

The Neighborhood Work Support Center Project
Table 6
Matching Constraints and Matchability: Philadelphia

Matching Scenario Number	<u>Matching Constraints</u>				<u>Matchability</u>			
	Poverty Rate in 1990	Ethnic Composition 1990 ^a	Standard Deviation of Home Purchase Loan Amount 1993-1995 ^b	Number of Target Neighborhoods Matched ^c	Avg. Population of Target Neighborhoods Matched ^d	Avg. Number of Unique Comparison Tracts Per Target Neighborhood ^e	Avg. Number of Unique Comparison Neighborhoods Per Target Neighborhood ^f	
1	±5%	±10%	±1	31	12,386	14.5	9.2	
2	±2%	±10%	±1	21	12,162	8.8	6.7	
3	±5%	±5%	±1	22	11,487	11.5	8.0	
4	±5%	±10%	±0.5	27	12,440	12.8	8.5	
5	±2%	±5%	±1	14	11,943	6.6	5.4	
6	±2%	±5%	±0.5	13	11,985	6.2	4.8	
All poor tracts	-	-	-	45	12,470	130.0	53.4	
Ten randomly selected tracts	-	-	-	45	12,470	10.0	9.2	

Table 6 (Continued)

SOURCE: MDRC calculations using data from the Urban Change Neighborhood Indicators Database.

NOTES: ^a To be eligible for a match, the proportion of each of three ethnic groups in the comparison tract- White, Black, and Hispanic- must match the respective proportion in the target tract within the number of percentage points specified in the column.

^b Based on data from Home Mortgage Disclosure Act of 1975 (HMDA) reports. Thresholds of +\$9,669 and +\$4,835 represent one standard deviation and 0.5 standard deviation of the distribution of home purchase loans among the 129 target tracts in Cleveland. Thresholds of +\$23,629 and +\$11,815 represent one standard deviation and 0.5 standard deviation of the distribution of home purchase loans among the 123 target tracts in Philadelphia.

^c The matching exercises are executed at the census tract level. Diagnostics for matched target tracts are aggregated to the neighborhood level. Target neighborhoods had 7,000-20,000 residents in 1990. There are a total of 22 target neighborhoods in Cleveland and 45 target neighborhoods in Philadelphia.

^d Even among the target neighborhoods selected, not all constituent tracts will be matched. In fact, some tracts will not be eligible for selection. This column shows the average total population of the matched tracts of the target neighborhoods.

^e The exercises employ a "matching with replacement" strategy, which means that comparison tracts can be matched to more than one target tract. Over the course of the evaluation, unusual events may occur in comparison tracts that would compromise their comparability with the target tracts, in which case the compromised comparison tracts would be removed from the analysis. This diagnostic speaks to the tract-level redundancy of the method, i.e. its resilience to comparison tract attrition, by providing an unduplicated count of the average number of unique comparison tracts per matched target neighborhood.

^f See note above. In addition to the tract level replacement discussed above, several unique comparison tracts may belong to the same neighborhood, which begs the question- how many unique neighborhoods would be selected as comparison sites. This diagnostic speaks to the neighborhood-level redundancy of the method by providing an unduplicated count of the average number of unique comparison neighborhoods per selected target neighborhood.

The Neighborhood Work Support Center Project

Table 7

Matching Constraints and Match Quality: Philadelphia

Matching Scenario Number	Matching Constraints			Distribution of the Absolute Pre-Post Measurement Error				
	Poverty Rate in 1990	Ethnic Composition 1990 ^a	Standard Deviation of Home Purchase Loan Amount 1993-1995 b	Minimum	Median	75th Percentile	90th Percentile	Maximum
1	±5%	±10%	±1	0.08	1.57	4.22	7.79	13.27
2	±2%	±10%	±1	0.12	1.09	2.22	3.01	6.08
3	±5%	±5%	±1	0.11	1.09	1.91	2.95	11.95
4	±5%	±10%	±0.5	0.05	1.63	3.33	9.55	10.19
5	±2%	±5%	±1	0.21	1.75	2.38	2.82	2.93
6	±2%	±5%	±0.5	0.45	1.83	2.57	2.93	3.28
All poor tracts	-	-	-	0.13	1.73	4.19	8.31	15.84
Ten randomly selected tracts	-	-	-	0.09	2.42	4.40	7.94	16.19

Table 7 (Continued)

SOURCE: MDRC calculations using data from the Urban Change Neighborhood Indicators Database.

NOTES: ^a To be eligible for a match, the proportion of each of three ethnic groups in the comparison tract- White, Black, and Hispanic- must match the respective proportion in the target tract within the number of percentage points specified in the column.

^b Based on data from Home Mortgage Disclosure Act of 1975 (HMDA) reports. Thresholds of +\$9,669 and +\$4,835 represent one standard deviation and 0.5 standard deviation of the distribution of home purchase loans among the 129 target tracts in Cleveland. Thresholds of +\$23,629 and +\$11,815 represent one standard deviation and 0.5 standard deviation of the distribution of home purchase loans among the 123 target tracts in Philadelphia.

percentile of target neighborhoods matched to all poor tracts is 4.19, suggesting that 25 percent of the matches achieved by this method will yield a measurement error of this magnitude or larger. This error is larger than all but one of those reported in the rows above, suggesting that in all but one case, even the worst 25 percent of the more stringent runs are likely to yield more precise matches. Similarly, the last row of Table 5 shows the distribution of error yielded by 100 iterations of a scenario that matches target neighborhoods to ten randomly selected tracts. The measurement error at the 75th, 90th, and 100th percentiles is greater (by about 0.15 to 2.9 percentage points) than the measurement error yielded by the more selective scenarios, again suggesting that the increased stringency of these runs affords some protection against risk of measurement error.

Effects of Changing Stringency of Matching Constraints

Independently increasing the stringency of the poverty or ethnicity matching constraints in Philadelphia tends to improve measurement error at the cost of matchability, as it did in Cleveland. Increasing the stringency of the poverty constraint from 5 percent to 2 percent decreased the median error by 0.5, while reducing the number of matched target neighborhoods from 31 to 21 (out of a total of 45). Increasing the stringency of the ethnicity constraint from 10 percent to 5 percent decreased the median error by 0.5, while reducing the number of matched target neighborhoods from 31 to 22. Contrary to Cleveland, however, increasing the stringency of particular matching constraints in Philadelphia can lead to losses in both matchability and measurement precision. For example, increasing the stringency of the home loan constraint from one standard deviation to half of a standard deviation increased the error by 0.1 percentage points, while reducing the number of matched target neighborhoods from 31 to 27. Simultaneously increasing the stringency of the poverty and ethnicity constraints increased the median error by 0.2, while reducing the number of matched target neighborhoods from 31 to 14.

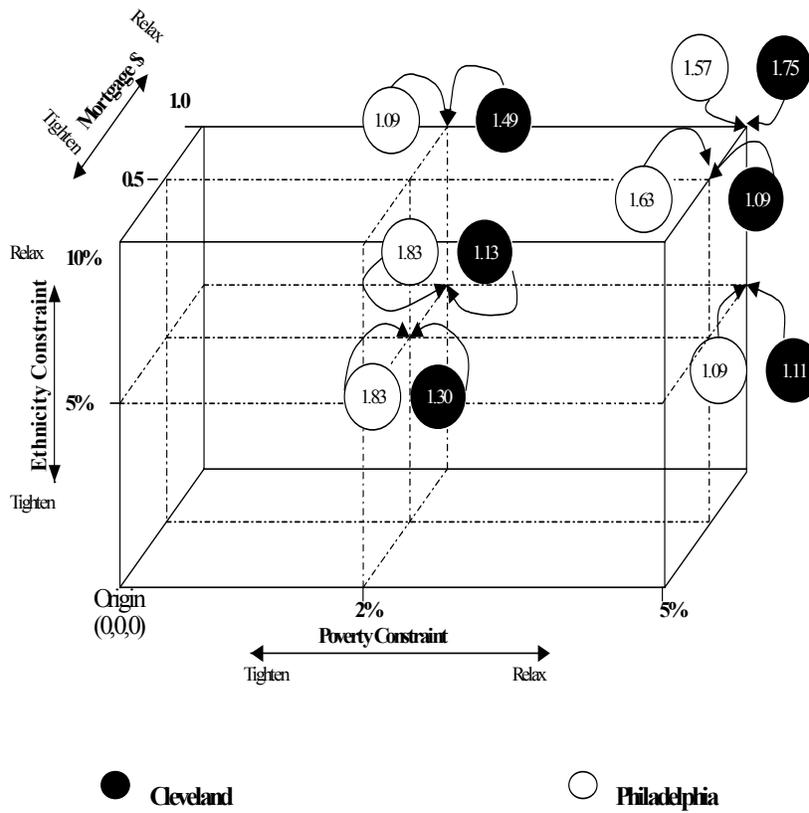
Discussion

Figure 2 illustrates two points regarding the tradeoffs of increasing and relaxing the stringency of each dimension of the matching constraints to reduce the measurement error in Cleveland and Philadelphia. First, within the range of constraints illustrated, the variations in measurement error are modest. Second, there is no global, cross-site, ideal run. That is, there is no common set of matching constraint parameters that minimizes measurement error for both cities. Less stringent matching constraints on poverty tend to yield better matches in Cleveland, but not necessarily for Philadelphia. Less stringent home purchase loan constraints yield better matches in Philadelphia, but not necessarily for Cleveland.

The Neighborhood Work Support Center Project

Figure 2

A Portrait of the Median Measurement Error Surface
Given Variations in Three Matching Constraints



Next Steps

The encouraging results of this preliminary exercise hold some promise for the field, but its full knowledge-building potential will only be realized as further analysis extends the current understanding of the possibilities and limitations of matching in other sites, with other outcomes, while applying other matching constraints.

Replication in other sites

As the Project on Devolution and Urban Change completes administrative records data collection for Miami-Dade and Los Angeles counties, new opportunities will emerge to validate the matching exercise in those counties. Levels of segregation and poverty concentration are much lower in the sunbelt cities of Miami-Dade and Los Angeles than in the rustbelt cities of Cleveland and Philadelphia, raising questions about the adequacy of eligible target tracts, the relevance of the welfare receipt rate outcome, the similarities of neighborhood change, and the relevance of place-based employment initiatives more generally in those counties.

Replication with other outcomes

In addition to public benefit information, the Urban Change Neighborhood Indicators database has a wealth of annual neighborhood indicators of crime, domestic violence, child maltreatment, and vital records, raising questions about whether the lessons learned from this matching exercise would apply to neighborhood-level initiatives focused on these other outcomes.

Replication for other time periods

The variability in neighborhood change trajectories generally, and the exceptionality of the 1990s in particular, would seem to warrant a degree of tentativeness concerning the application of these findings to other time periods.

Applying the Model to Other Cities

Exploratory application of the matching models described here in other cities need not wait for the further empirical validation exercises outlined above. Although the matching exercise described here can only be validated in the four Urban Change sites where annual welfare receipt rate outcome measures are available, the matching algorithm can be applied in other sites. Even this blind application of the matching algorithm in other sites could answer basic pragmatic questions concerning matchability, including:

- How many eligible target neighborhoods would be identified elsewhere?
- What proportion of eligible target neighborhoods would be matched?

- Would the comparison tract and neighborhood-level redundancies be comparable to those found in Cleveland and Philadelphia?
- How does the proportion of eligible target tracts matched vary by region, city size, the residential concentration of poverty and welfare receipt, etc.?

Conclusions

Although many evaluations of neighborhood-level interventions employ a quasi-experimental comparative design, to the authors' knowledge the validity of the target to comparison tract matching has never been rigorously evaluated.¹⁸ The Urban Change Neighborhood Indicators database presented a unique opportunity to contribute to such an evaluation. In both Cleveland and Philadelphia this report identifies four matching criteria and a multidimensional surface defined by narrow ranges of parameter values for each within which matchability and measurement error can be traded-off within relatively circumscribed bounds. The least stringent matching constraints on this surface yield the highest rates of target neighborhood matchability in Cleveland and Philadelphia (73 percent, and 69 percent respectively), but at some cost in measurement error (1.8, and 1.6 percentage points respectively). Conversely, the most stringent matching constraints yield proportionately fewer target neighborhood matches in Cleveland and Philadelphia (55 percent, and 47 percent respectively), but more accurate matches (1.1, and 1.1 percentage points respectively).

The results of these selective matching scenarios compare favorably with two less selective alternatives. Matching target neighborhoods to all other poor, urban, residential tracts would yield greater matchability, but at an increased risk of measurement error. Similarly, matching target neighborhoods to a random selection of ten comparison tracts would yield greater matchability at an increased risk of measurement error. Not only do the more stringent strategies yield consistently smaller measurement errors, they also offer greater face validity about the dimensions of similarity between target and comparison tracts.

Beyond these encouraging results, the exercise offered important practical lessons regarding matching constraints and the tradeoffs between matchability and measurement error engendered by more- and less-stringent matching constraints. Realizing the full potential of these important empirical lessons for the field will require replication in other sites, and with other outcomes, but the application of the matching methodology outlined here need not wait on further research. In fact, matching target and comparison sites in other cities would provide important practical information about their matchability and the applicability of these findings.

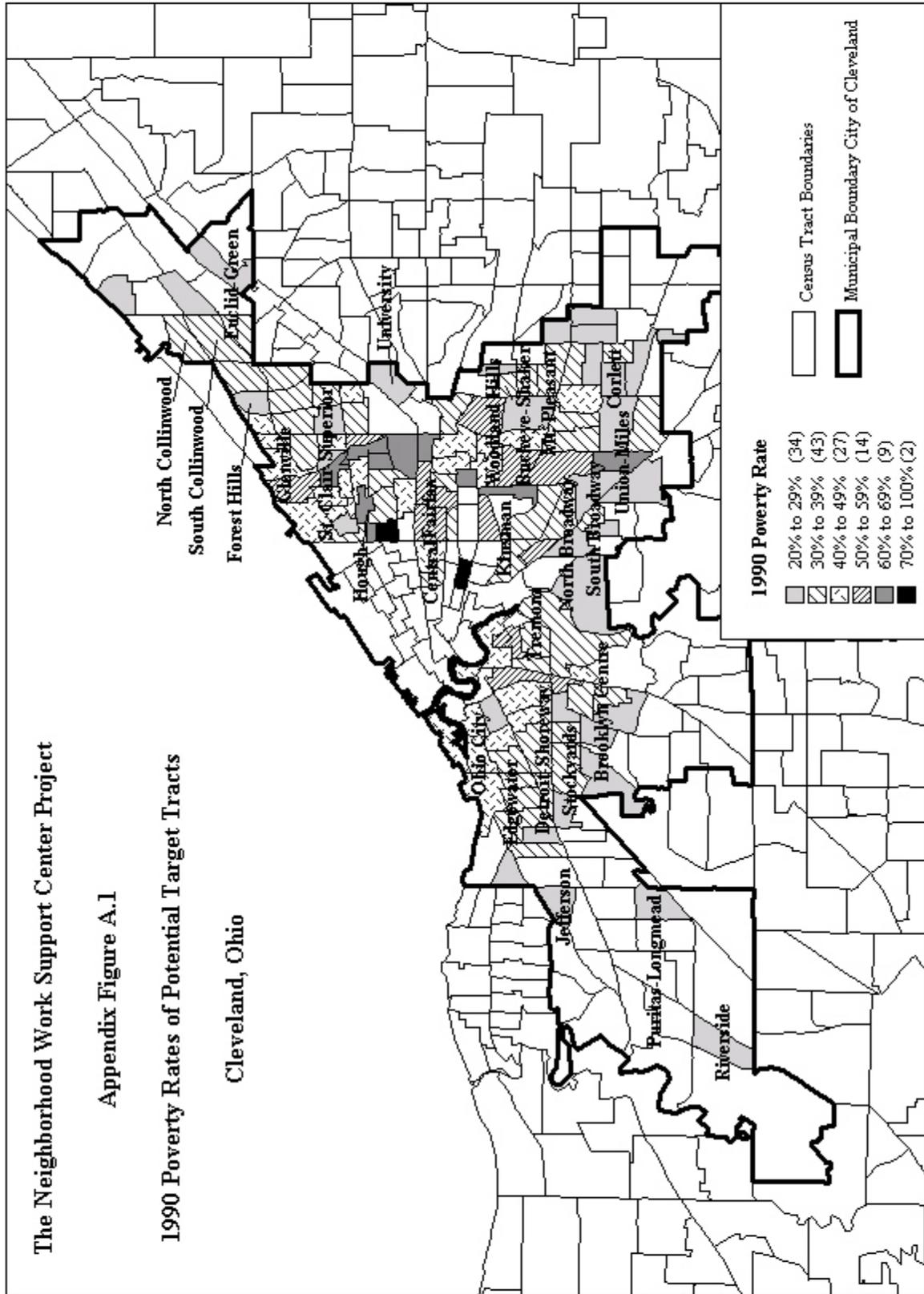
¹⁸Hollister and Hill, 1995.

The Neighborhood Work Support Center Project

Appendix Figure A.1

1990 Poverty Rates of Potential Target Tracts

Cleveland, Ohio

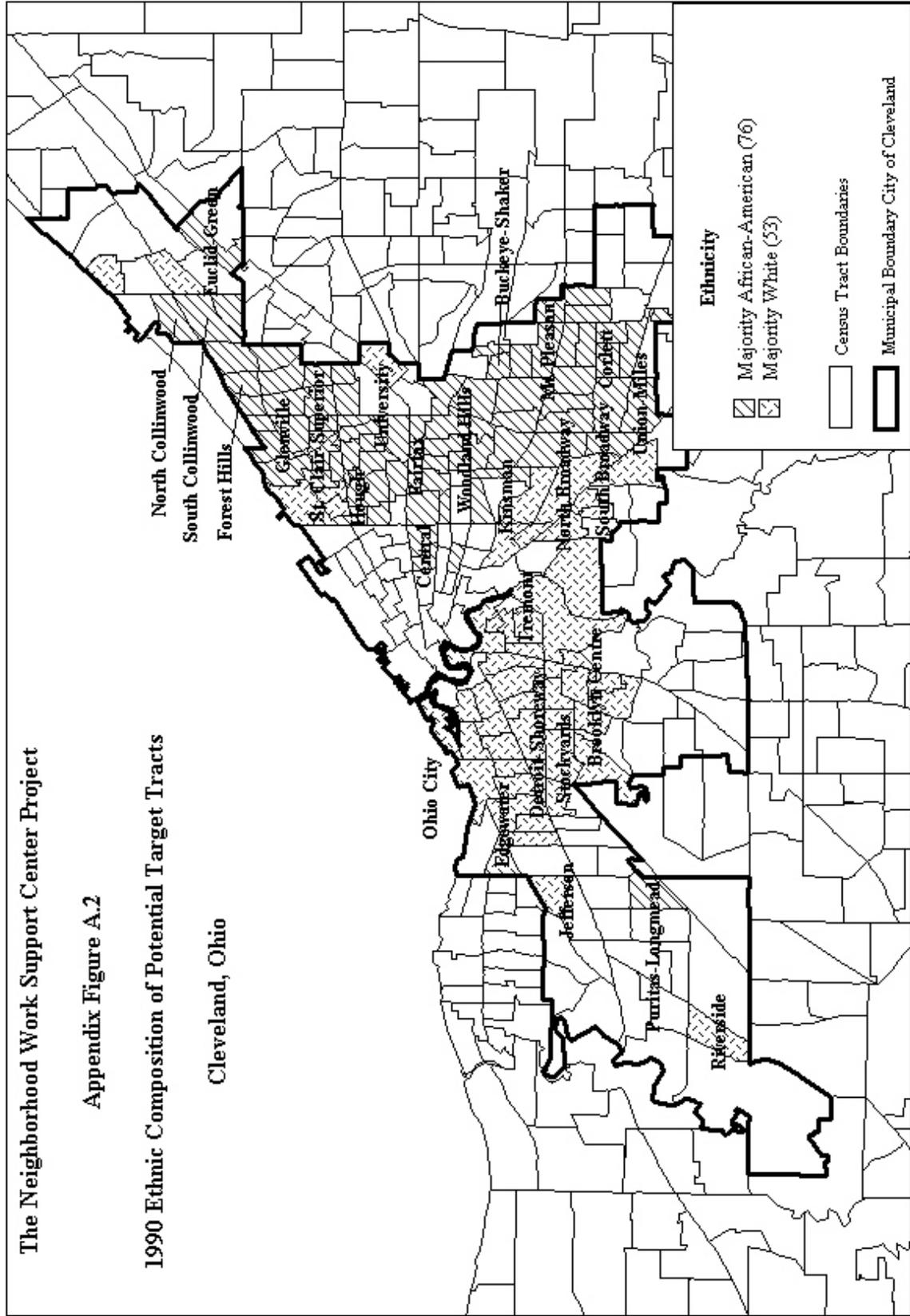


The Neighborhood Work Support Center Project

Appendix Figure A.2

1990 Ethnic Composition of Potential Target Tracts

Cleveland, Ohio

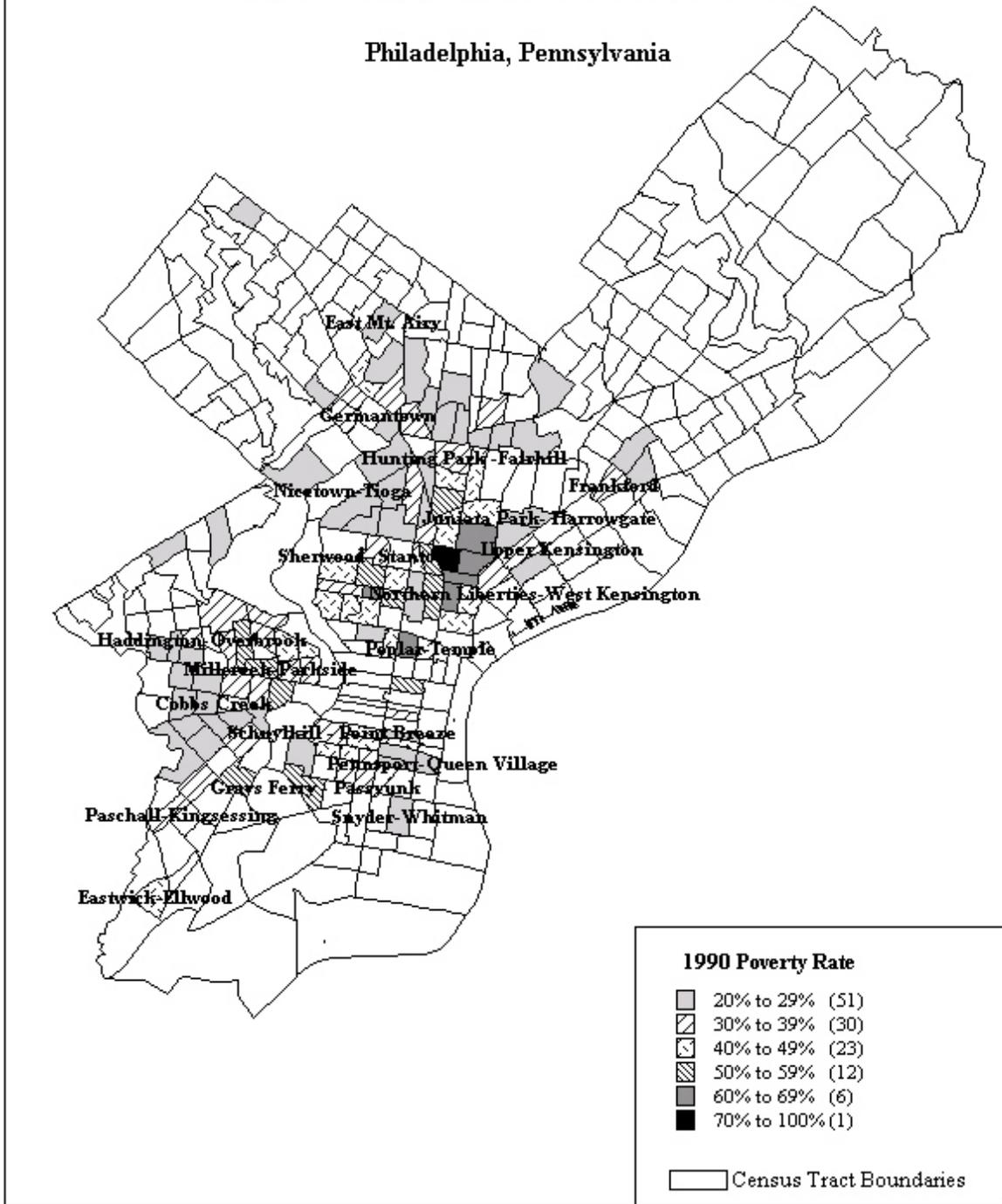


The Neighborhood Work Support Center Project

Appendix Figure A.4

1990 Poverty Rates of Potential Target Tracts

Philadelphia, Pennsylvania

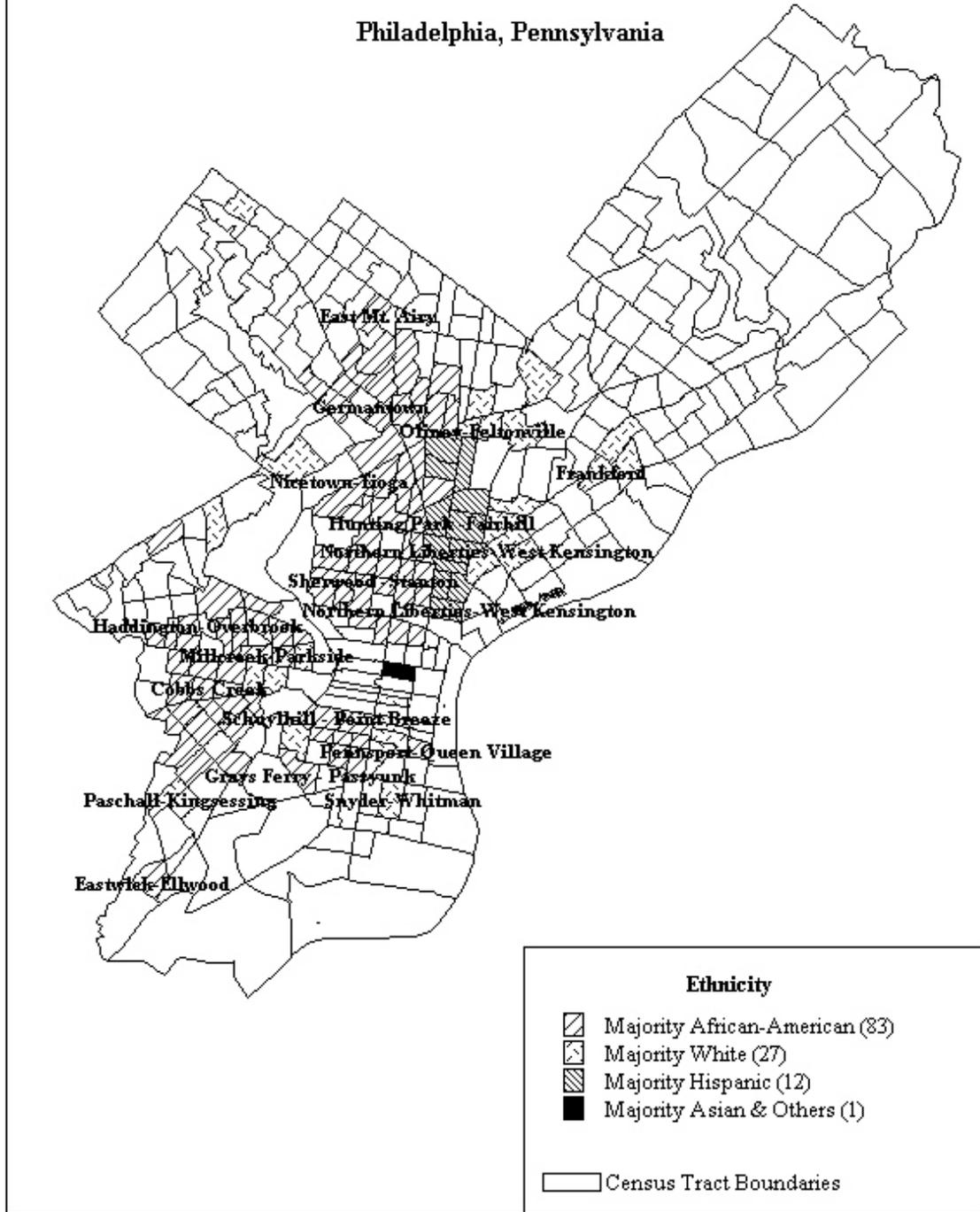


The Neighborhood Work Support Center Project

Appendix Figure A.5

1990 Ethnic Composition of Potential Target Tracts

Philadelphia, Pennsylvania

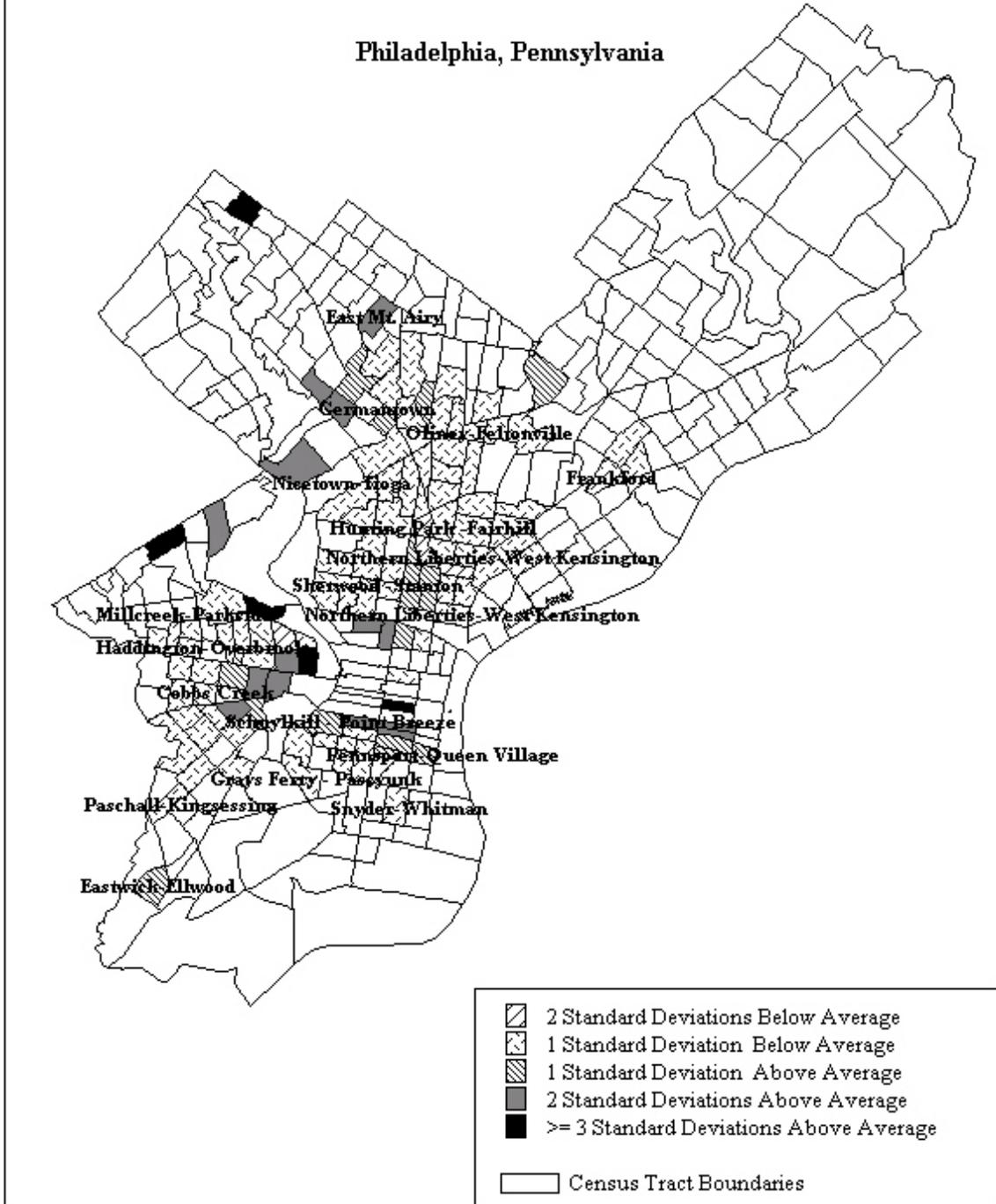


The Neighborhood Work Support Center Project

Appendix Figure A.6

1990 Median Amount of Home Purchase Loans Relative to Target Tracts

Philadelphia, Pennsylvania



References

- Bartik, Timothy. 1988. "Measuring the Benefits of Amenity of Improvements in Hedonic Price Models." *Land Economics*, vol. 64, May, pp. 172-183.
- Bloom, Dan and Charles Michalopoulos. 2001. "How Welfare and Work Policies Affect Employment and Income: A Synthesis of Research." New York: MDRC.
- Grieson, Ronald, and James White. 1989. "The Existence and Capitalization of Neighborhood Externalities." *Journal of Urban Economics*, vol. 25, no. 1, pp. 68-76.
- Hollister, Robinson G. and Jennifer Hill. 1995. "Problems in the Evaluation of Community-Wide Initiatives." *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts*, vol. 1, pp. 127-172.
- Palmquist, Raymond. 1992. "Valuing Localized Externalities." *Journal of Urban Economics*, vol. 31, no. 1, pp. 59-68.
- Polinsky, Mitchell, and Steven Shavell. 1976. "Amenities and Property Values in a Model of an Urban Area." *Journal of Public Economics*, vol. 5, January/February, pp. 119-129.
- Walker, Chris, Chris Hayes, George Galster, Patrick Boxall, and Jennifer Johnson. 2002. *The Impacts of CDBG Spending on Urban Neighborhoods*. Washington, DC: Urban Institute Report to the U.S. Department of Housing and Urban Development, Oct.; available at: <http://www.huduser.org/publications/pdf/CDBGSpending.pdf>

About MDRC

MDRC is a nonprofit, nonpartisan social policy research organization. We are dedicated to learning what works to improve the well-being of low-income people. Through our research and the active communication of our findings, we seek to enhance the effectiveness of social policies and programs. MDRC was founded in 1974 and is located in New York City and Oakland, California.

MDRC's current projects focus on welfare and economic security, education, and employment and community initiatives. Complementing our evaluations of a wide range of welfare reforms are new studies of supports for the working poor and emerging analyses of how programs affect children's development and their families' well-being. In the field of education, we are testing reforms aimed at improving the performance of public schools, especially in urban areas. Finally, our community projects are using innovative approaches to increase employment in low-income neighborhoods.

Our projects are a mix of demonstrations — field tests of promising program models — and evaluations of government and community initiatives, and we employ a wide range of methods to determine a program's effects, including large-scale studies, surveys, case studies, and ethnographies of individuals and families. We share the findings and lessons from our work — including best practices for program operators — with a broad audience within the policy and practitioner community, as well as the general public and the media.

Over the past quarter century, MDRC has worked in almost every state, all of the nation's largest cities, and Canada. We conduct our projects in partnership with state and local governments, the federal government, public school systems, community organizations, and numerous private philanthropies.