POVERTY DECONCENTRATION EFFORTS IN U.S. LOW-INCOME HOUSING POLICY: IMPACTS AND IMPLEMENTATION CHALLENGES

Joel Bolinger

Submitted to the faculty of the University Graduate School in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the School of Public and Environmental Affairs and the Department of Political Science Indiana University December, 2016
Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Doctoral Committee

____________________________________________________
Sean Nicholson-Crotty, Ph.D.

____________________________________________________
David Reingold, Ph.D.

____________________________________________________
Jill Nicholson-Crotty, Ph.D.

____________________________________________________
Matthew Hayes, Ph.D.

____________________________________________________
Maureen Pirog, Ph.D.

November 11, 2016
Acknowledgements

There are many people without whom this dissertation would never have been written, and thanking all of them would be impossible, but I will do my best to acknowledge some of the most important. First, I would like to thank my mother, Pat, and my late father, Bill, for instilling in me a love of learning. Tremendous thanks are due to Sean Nicholson-Crotty for shepherding and marshalling me through the process of writing the dissertation. Without his encouragement and occasional “tough love,” I never would have been able to make it through. Thanks are due also to my other committee member (David Reingold, Jill Nicholson-Crotty, Matthew Hayes, and Maureen Pirog) for serving on my committee and offering helpful encouragement and guidance. I would like to thank my fellow student and great friend, Gordon Abner, for his invaluable advice and insight as I crafted the idea for this project; he was truly instrumental, and I will forever be indebted to him. I would like to thank the late Dr. Evan Ringquist for teaching me the technique of meta-analysis and for modeling such dedication to scholarly professionalism. I will be forever grateful to my friends who put up with listening to my bellyaching about graduate school for countless hours, and my bandmates, who were gracious in encouraging me and accommodating my unpredictable graduate student schedule. Finally, I would like to thank my girlfriend, Audrey William, for her constant love and support.
The United States has implemented a variety of strategies to deconcentrate poverty through low-income rental subsidies. These policies are based largely on theoretical and empirical work from a variety of social science disciplines on how neighborhood characteristics condition life outcomes. This dissertation uses three different methods (meta-analysis, standard regression analysis, and an online experiment) to investigate these deconcentration policies. In chapter two of this dissertation, two quantitative meta-analyses of studies that examine the effects of these programs on individual economic and behavioral outcomes find little evidence that the programs have a meaningful impact. Chapters three and four explore areas that might offer explanations for poverty deconcentration’s disappointing results. Both of these areas are related to an area that few other housing researchers have explored: program implementation. The second empirical chapter uses regression analysis of administrative data to test whether the managerial quality among local housing authorities is associated with program performance. The third chapter uses an online experiment to assess how much potential middle-income residents of mixed-income housing developments might be influenced to choose those types of residences based on how their potential low-income neighbors are framed. The experiment tests the relative influences of frames for hypothetical neighbors based on sympathy (sympathetic or unsympathetic) compared to frames based on the implied race of potential neighbor. The dissertation concludes that management does matter for housing authority performance and that framing low-income neighbors
sympathetically or unsympathetically can indeed influence over the short run stated willingness to live in mixed-income developments, though racial frames have no effect.

Sean Nicholson-Crotty, Ph.D.

David Reingold, Ph.D.

Jill Nicholson-Crotty, Ph.D.

Matthew Hayes, Ph.D.

Maureen Pirog, Ph.D.
Table of Contents

Chapter 1: A Summary of Research Questions, Methodologies, and Findings........................................1

Chapter 2: The Economic and Behavioral Outcomes of Poverty
   Deconcentration Strategies in U.S. Low-Income Housing Policy:.........................................................8
   An Updated Meta-analysis

Chapter 3: Management Quality and Organizational Performance in .....................................................50
   U.S. Public Housing Authorities

Chapter 4: Recruiting Non-Poor Households for Mixed-Income Housing:
   A Survey Experiment on the Effect of Sympathy-Based........................................................................70
   and Affective Racial Framing

Chapter 5: Policy Implications and Directions for Future Research.........................................................88

References..........................................................................................................................................................94
Chapter 1: A Summary of Research Questions, Methodologies, and Findings
Overview of Research Questions and Policy Background

This dissertation investigates the failure of U.S. Federal poverty deconcentration policies to achieve improvements in the target population’s life outcomes. Specifically, the project investigates two explanations related to problems of policy implementation. The three empirical chapters in this dissertation (chapter 2-4) represent three different empirical approaches to investigating poverty deconcentration policies. Chapter two presents a meta-analysis of studies that estimate the impact of poverty deconcentration policies on economic and behavioral outcomes. Chapter three is a regression-based analysis of the relationship between housing authority management quality and program performance. Chapter four uses a survey experiment to assess the impact of implicit racial biases and attitudes toward housing subsidy recipients on the ability to recruit non-poor households to live in mixed-income housing developments. Chapter two presents evidence that poverty deconcentration policies have had disappointing results on two important life outcomes, and chapters three and four offer preliminary explorations of two different areas of policy implementation (management and marketing) that may have the potential for explaining the null findings from chapter two.

To contextualize the investigation into implementation failure in poverty deconcentration policy, a brief aside on policy history is necessary. In the United States, Federal rent subsidies for low-income rental housing have taken three general forms (Hayes 1995). The earliest form is what is commonly called public housing, which is housing that is owned and operated by the government. There have been no new public housing units constructed in the United States for decades, and the total stock continues to decline as older units are closed or demolished. The second and most common form of rent subsidy is federally funded voucher programs, most commonly known as Section 8, but which is now officially called the Housing Choice Voucher program. The HCVP subsidizes rent in the private market. Recipients of these vouchers are required to pay 30 percent of their monthly income in rent,
but the voucher covers the remainder. The third form of rental housing subsidy is construction of units which are at least partially reserved for low-income households; these are often referred to as “mixed-income developments.” The two major programs of this type are the HOPE VI program (Housing Opportunities for People Everywhere), which provides funding for local housing authorities to demolish dilapidated public housing and rebuild mixed-income developments, and the Low Income Housing Tax Credit program (LIHTC), which is often administered through state housing finance agencies and channels funding (as tax credits) to private developers to build units that are at least partially reserved for low-income tenants. The important factor to note in relation to this dissertation is that vouchers and mixed-income developments have as part of their stated goals the geographic deconcentration of poor households.

The scholarly case for the validity of the underlying causal theory in poverty deconcentration policies is quite striking. The “neighborhood effects” family of scholarly literature is fairly extensive, and it has been produced by a wide variety of academic disciplines. The idea that neighborhood segregation by race and class in the United States is harmful for residents of low-income, minority neighborhoods is supported by work from economics, sociology, geography, education, public health, and other fields. One of the key variables in the neighborhood effects literature is the degree of poverty concentration that residents face. A growing social science literature suggests that poverty concentration has a variety of detrimental effect on the residents of these areas, both in terms of their current well-being and their future opportunities (Wilson, 1987, 1996; Jencks and Mayer, 1990; Brooks-Gunn, et al., 1993). Scholars have posited a number of mechanisms through which neighborhood poverty might influence residents’ life outcomes, including spatial proximity to jobs, social capital, educational opportunities, neighborhood culture, public health, political influence, and others. Whatever the precise mechanisms through which neighborhood conditions effects residents’ outcomes, there seems to be wide endorsement of the idea that living and growing up in poor neighborhoods carries a variety of risks.
Meta-Analysis of Poverty Deconcentration Impacts on Economic and Behavioral Outcomes

Among the life outcomes that poverty deconcentration policies are intended to improve are economic well-being and problematic behaviors (criminal, delinquent, or risky activity). Empirical investigations into the effects of poverty deconcentration policies on participants’ outcomes in these areas have yielded conflicting results, with some studies finding strong positive effects, some studies finding negative effects, and some studies finding no effects. Of particular note are the conflicts between the Gautreaux program studies, which found strong positive effects, and the Moving to Opportunity program studies, which generally found no effects. When primary studies come to differing conclusions about the efficacy of an intervention, quantitative meta-analysis is an appropriate way to assess the weight of the empirical evidence.

Chapter two in this dissertation builds on a formal meta-analysis I completed in 2012 along with co-author Lanlan Xu on the effects of poverty deconcentration policies on low-income households’ economic and behavioral outcomes (Bolinger and Xu 2013). In the earlier meta-analysis, we identified 37 studies containing a total of 1755 effect sizes related to the effect of poverty deconcentration on individual economic well-being. We identified 13 studies containing a total of 1260 effect sizes related to the effect of poverty deconcentration on individuals’ tendency to engage in criminal, delinquent, or risky behaviors. Using the same search techniques and inclusion criteria, chapter two updates that meta-analysis with 434 new effect sizes from 11 new studies published since the previous meta-analysis and one study from the previous analysis which excluded some effect sizes that should have been included. As before, the updated meta-analysis uses descriptive meta-analysis techniques to examine overall average effects, and it uses meta-regression to look for mediating and moderating variables at the study level and model level that may vary systematically with effect size estimates.
The findings from the updated meta-analysis are consistent with the earlier findings from Bolinger and Xu 2013. The chapter will show that, while the programs have statistically significant impacts overall, they too miniscule to have any substantive significance. Substantively, the effects are average effects of these programs are nil, and while there are certain variables in the meta-regressions that will predict higher or lower effects, the changes in predicted effect sizes are also too small to have substantive significance. For all practical purposes, the effect of poverty deconcentration policies on economic and behavioral outcomes will be shown to be negligible.

Exploring Implementation Issues as Potential Explanations of Policy Failure

Scholars have been fairly puzzled by the lackluster performance of poverty deconcentration policies. A variety of explanations for this failure have been posited by scholars and practitioners, from the practical challenges of the housing search, to social pressures and desires to remain in familiar neighborhoods, to reluctance by landlords to accept housing vouchers. There is little consensus on the primary drivers of policy failure, and much of what has been written has been speculative.

Chapter three and four of this dissertation examine two implementation problems as potential explanations for the policies’ failure to produce desired outcomes. One common distinction in scholarly work on policy failure is between policies that are flawed in their underlying theories and ones that are flawed “downstream” somewhere in the implementation process (Hupe and Hill 2009; Mazmanian and Sabatier 1989). All policies are built upon a causal logic that links some government-based intervention with the achievement of some desired outcome or outcomes. If the underlying theory is faulty, the policy will likely fail. For instance, will banning smoking advertisements reduce teen smoking? The underlying theory is that teens smoke because they are influenced by advertisements, but if in reality they are influenced more by peers and role models, the policy will probably have no effect. The underlying causal theory in that case would be flawed. On the other hand, the policy might fail
“downstream” from the initial design of the policy in what would typically be called a problem of implementation. For instance, if tobacco companies could make more money by advertising than they would pay in fines for advertising, the policy might have no effect. Or if the agency responsible for fining violators did not have enough staff to effectively monitor the problem, the policy might fail to achieve its goal. The implementation literature has identified a wide variety of places in the causal chain where implementation might fail and many different mechanisms through which failure might occur. In poverty deconcentration policies, the analog distinction would be whether the underlying theory about the harmfulness of poverty concentration is flawed, or whether something in the implementation of the policies has been going awry. Because the causal theory that says that moving into neighborhoods with more affluent neighbors should improve low-income households’ lives is so well-supported by underlying theory, it is reasonable to look to “downstream” implementation problems to explain the failure of these deconcentration policies to improve life outcomes.

**Housing Authority Management Quality and Program Performance**

Chapter three examines whether poor management of public housing authorities might be one of the implementation problems that could help explain the null effects of deconcentration policies found in chapter two. One possible explanation for the failure to find positive behavioral or economic changes from poverty dispersal programs is that these programs were implemented poorly by the agencies responsible for their implementation. Most Federal rent subsidies for low-income households are administered by local government entities called housing authorities, of which there are approximately 4,000 in the United States. I use administrative data on the total population of public housing authorities in the U.S. (the local government entities responsible for implementing Federal housing subsidies) to examine whether management quality of housing authorities is correlated with program success. While this chapter is not able to assess directly the impact of housing authority management on the poverty deconcentration outcomes examined in chapter two, if management
matters for the outcomes examined in this chapter, it is suggestive of an area (management quality of housing authorities) for future researchers to examine in the attempt to explain that failure. Moreover, this area has been largely ignored by ex post analyses of poverty deconcentration policies, especially by quantitative research (Briggs et al 2010).

While there is a dearth of research on housing authority management per se, this chapter uses empirical and theoretical frameworks that are common in public management and implementation research. The question, “Does management matter?” is one of the central concerns of public management research. The relationship between executive compensation and organizational performance is also a focus of great attention in management literature (Meier and O’Toole 2002). Similarly, the idea that policy success or failure is conditioned by exogenous conditions is common in implementation research (Hill and Hupe 2009; Mazmanian and Sabatier 1989).

The data on management quality and program outcomes in the analysis comes from the Department of Housing and Urban Development (HUD). HUD utilizes two composite program quality measures corresponding to the two main housing authority functions, administering public housing and administering the Housing Choice Voucher program. For public housing, HUD uses the Public Housing Assessment System, or PHAS. For housing vouchers, HUD uses the Section Eight Management Assessment program, or SEMAP.

While it is impossible to draw firm causal claims from the data in this chapter, the regression analysis indicates that management quality in public housing administration is correlated with a substantively and statistically significant improvement in organizational performance, measured four different ways, and controlling for task complexity and state-level dummies. Management quality in public housing administration is also correlated with stronger performance in a separate program administered by housing authorities, the Housing Choice Voucher program (or Section 8). Though
study cannot directly link housing authority management to the failures of poverty deconcentration policies discussed in chapter two, it does provide support for investigating managerial quality’s impact in future research.

**Recruiting Middle-Income Households into Mixed-Income Housing: The Role of Perceptions of Low-Income Neighbors**

Chapter three utilizes an online survey experiment using a sample generated by the Amazon Mechanical Turk program to examine whether mixed-income projects might have trouble recruiting middle-class households due to race and class biases by unsubsidized households toward subsidized households. The experiment also tests whether stated willingness to live in mixed-income housing is susceptible to framing effects along dimensions of class race. The challenge of recruiting middle-income households for mixed-income housing has been under-recognized in studies of mixed-income projects. The policy objective of creating mixed-income developments—-to achieve improvements in low income households’ lives—requires the achievement of the “sub-objective” of recruiting non-poor households to live in the developments. If that sub-objective is not fulfilled, the program lacks an element that is crucial to its design. Very little scholarly work has been done on middle-income recruitment for mixed-income projects.

The experiment utilizes one control group and four treatment groups. The dependent variable in is the respondent’s stated willingness to live in mixed-income housing, and the treatment is a vignette about a hypothetical resident that might live in the development. For the dependent variable, respondents were told to imagine a scenario where they (and their family, if applicable) would be moving to an apartment in the near future. They were then asked to rate their willingness to live in this mixed-income project on a scale from 0 to 100. The treatment vignette offers either a sympathetic frame or unsympathetic frame, and both of these frames vary the name of the client to invoke different
racial attitudes. In all, there are five groups in the experiment: a control group with no frame and four treatment groups. The four treatment groups are a sympathetic story with a name designed to invoke the image of a white client, the same story with a name meant to invoke the image of a black client, an unsympathetic story with a white name, and an unsympathetic story with a black name.

The experiment shows that willingness to live in mixed-income housing is “frameable” based on how potential neighbors are pictured. Compared to the control group, respondents who were exposed to a sympathetic story about a potential low-income neighbor expressed much higher willingness to live in mixed-income housing on average. These treatment effects are both substantively and statistically significant. The sympathetic story featuring a white character raised expressed support by 18-22 points) compared to the control group. The sympathetic story featuring a black character raised expressed willingness to live in mixed-income housing by 15-19 points.

The unsympathetic story decreased expressed willingness to live in mixed-income housing compared to the control group, but not by nearly as much as the sympathetic story increased it. The unsympathetic white character decreased mean support by only 4-5 points compared to the control group, and this difference is not statistically significant. The unsympathetic story with a black character decreased mean support by 7-9 points.

The racial frame is shown to have no impact in either the sympathetic or the unsympathetic story groups. Varying the implied race of the character did not induce enough of a change in attitude to make the white and black character groups statistically distinguishable from each other. The chapter discusses possible explanations for this null finding, including whether this result stems from an underlying lack of racial animus in the population or whether the racial frame utilized in the experiment might be too weak to “tap” latent racial attitudes.
The findings from this survey experiment suggest several conclusions. One is that expressed willingness to live in mixed income housing is susceptible to framing, both in the positive and the negative direction. Managers of mixed-income projects should, when implementing marketing strategies, consider how non-poor households will view their potential low-income neighbors. While it is doubtful that a project would intentionally paint an unsympathetic picture of its low-income participants, the findings here suggest that the default attitude in the public is more congruent with an unsympathetic picture than a sympathetic picture. Marketers must contend with this default attitude, and sympathetic framing may be a useful strategy to do so. Much as with chapter three, this chapter does not establish a causal link between middle-income recruitment for mixed-income housing and the failure of poverty deconcentration policies, but it does suggest that this area offers potential to be more thoroughly explored in future research.

Summary

The three empirical chapters in this dissertation provide evidence for the following conclusions. One is that poverty deconcentration policies have been disappointing in their purported goal of improving economic and behavioral life outcomes. The second is that, as scholars continue to explore the reasons for these disappointing results, management quality among housing authorities is one area they should explore as a potential explanation. Management quality appears to be systematically related to program performance, so it follows that poor performance could possibly be explained by poor management. The third conclusion is that the ability of mixed-income housing projects to recruit middle-income households offers another area with potential for explaining poverty deconcentration results. The experimental analysis in this dissertation shows that the default attitude among the public toward the low-income residents of mixed-income projects is consistent with viewing them unsympathetically. However, it also shows that these attitudes are susceptible to framing effects, which indicates a possible policy solution to the implementation problem of recruitment.
Chapter 2:
The Economic and Behavioral Outcomes of Poverty Deconcentration Strategies in U.S. Low-Income Housing Policy: An Updated Meta-analysis
**Introduction**

This chapter presents an update to two meta-analyses completed in 2012 on the effects of participating in a U.S. Federal poverty deconcentration program on individual economic well-being and individual incidence of criminal, delinquent, and risky behaviors (Bolinger and Xu 2013). This update includes 434 new effect sizes from 11 new studies published since the previous meta-analysis and one study from the previous analysis which excluded some effect sizes that should have been included. Table 2-A presents a summary of the original two data sets, the two updated data sets, and the combined data sets, which were used to complete the two meta-analyses in this chapter. The conclusions from the chapter are much the same after updating the data set with newer studies. Participating in a poverty deconcentration policy appears to have on average almost no impact, positive or negative, on economic well-being or problematic behaviors. Meta-regressions are able to identify statistically significant relationships between the effect size estimates and study-level and model-level variables, but the associations are very small and are probably not of much substantive significance.
Brief History of U.S. Poverty Deconcentration Policies

For decades, U.S. housing policy has been moving away from government owned and operated public housing projects toward policy instruments that encourage recipients to live in communities with lower concentrations of neighborhood poverty (Hayes 1995, Orlebecke 2000). This approach has been motivated in part by a widely held belief among policy makers and the public that large-scale public housing developments, particularly high-rise projects, harm residents’ life chances by exposing them to high concentrations of poverty in their neighborhood environments. In the 1970s, Federal housing policy began to shift toward the use of vouchers and mixed-income developments to help ensure that recipients of housing subsidies could live in neighborhoods with lower concentrations of poverty. All of these programs have been motivated at least in part by the goal of spatially deconcentrating inner-city, assisted housing tenants, integrating those tenants into suburban, middle-

<table>
<thead>
<tr>
<th>Outcome Type</th>
<th>Number of Studies from Original Meta-Analysis</th>
<th>Number of Effect Sizes from Original Meta-Analysis</th>
<th>Number of New Studies from Updated Meta-Analysis</th>
<th>Number of New Effect Sizes from Updated Meta-Analysis</th>
<th>Total Number of Studies In this chapter</th>
<th>Total Number of Effect Sizes in this chapter</th>
<th>Median Number of Effect Sizes Per study</th>
<th>Mean Number of Effect sizes per study</th>
<th>Range of Effect Sizes Per Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economic Outcomes</td>
<td>37</td>
<td>1755</td>
<td>7</td>
<td>221</td>
<td>44</td>
<td>1976</td>
<td>6</td>
<td>44.9</td>
<td>1-776</td>
</tr>
<tr>
<td>Behavioral Outcomes</td>
<td>13</td>
<td>1260</td>
<td>5</td>
<td>213</td>
<td>18</td>
<td>1473</td>
<td>16</td>
<td>81.83</td>
<td>1-768</td>
</tr>
</tbody>
</table>

*Some studies contained both outcome types. The total number of studies coded is 55.*
class communities, and enabling them to exercise more choice in their housing decisions (HUD 1988, National Commission on Severely Distressed Public Housing 1992).

Since the Housing Act of 1937 until the mid-1970s, the Federal government’s main form of low-income rental assistance was the funding of government owned and operated housing units, usually referred to by the term public housing. Public housing projects are only available to low-income residents, so by definition, they lead to geographic concentrations of low-income persons. Good-faith efforts over the years to reserve public housing for only the most impoverished populations ended up exacerbating this concentration. The tendency of public housing to concentrate poverty has been especially true for very large public housing projects, which often constitute a large percentage of a neighborhood’s population (Hayes 1995; HUD 1988; National Commission 1992).

Concerns about the role of public housing in concentrating poverty have led to a variety of efforts by the US Department of Housing and Urban Development (HUD) to geographically disperse the recipients of housing subsidies. Section 8 of the United States Housing Act of 1937 (commonly referred as Section 8) was amended in 1974 to authorize the payment of rental housing assistance to private landlords on behalf of low-income households (Hayes 1995; Orlebecke 2000). Although not a major component of the initial motivation for the policy, over time, policy makers have come to see the Section 8 family of programs as a useful tool in the battle to deconcentrate poverty. By allowing recipients of housing subsidies to find private rental housing, Section 8 offers poor families the chance to move into less poor neighborhoods. It has now exceeded the public housing program in size and has become the largest public housing program in the U.S. Currently, the largest Section 8 program is the “tenant-based” voucher program (called the Housing Choice Voucher program), where recipients are free to live anywhere, as long as landlords are willing to accept the voucher and meet HUD’s basic qualifications. In the somewhat smaller “project-based” Section 8 program, the voucher is limited to a specific site, where the owner reserves some or all of the units for low-income tenants. In both
programs, the federal government pays the difference between the tenant’s contribution (30% of their monthly income) and the rent specified in the owner’s contract with the government.

In addition to these programs, there have been two large housing experiments conducted in U.S. cities to address poverty concentration. These two experiments, the Gautreaux program and The Moving to Opportunity for Fair Housing (MTO) program both placed former public housing residents into low-poverty neighborhoods. In 1976, a court order required the Chicago Housing Authority to desegregate its public housing. Their response, called the Gautreaux program, placed participants in low-poverty, racially integrated neighborhoods, accompanied by fairly intensive housing counseling services. Because the Gautreaux program appeared to show positive impacts (Rosenbaum and Popkin 1991; Popkin et al 1993; Keels 2005, 2008), in the early 1990s HUD authorized MTO, which was a much larger 5 city, 15 year longitudinal study with a random assignment, experimental design. Unlike Gautreaux, which placed residents in economically AND racially integrated neighborhoods, MTO’s main treatment group was only required to live in non-poor neighborhoods for a period of at least a year. The MTO experiment also had a second treatment group of households given a regular Section 8 voucher with no restrictions on neighborhood quality. The MTO experiment has now had three waves of follow up and has been subject to a wide variety of ex-post analyses by a number of research teams across the country, e.g. (Orr et al 2003; Briggs et al 2010; Goering and Feins 2003; Harcourt and Ludwig 2006; Kling et al 2005; Ludwig et al 2008. Hanratty et al 2003; Sanbonmatsu et al 2011).

One of HUD’s latest policy initiative to deconcentrate policy is a mixed-income housing development program called HOPE VI (Housing Opportunities for People Everywhere). HOPE VI is a major HUD plan to revitalize the worst public housing projects into mixed-income multi-family dwellings (Joseph and Feldman 2009; Cisneros 2009). The public housing units in the nation’s largest cities have long been seen as substandard and even dangerous living environments (HUD 1988, National Commission on Severely Distressed Public Housing 1992). This was particularly true of the high rise
public housing projects that emerged in large cities, such as Chicago and New York. Mixed-income housing is designed to improve subsidized renters’ immediate surroundings in hopes of combating the widely-believed pernicious effects of poverty concentration. The HOPE VI program began in 1992 and includes a variety of grant programs including Revitalization, Demolition, and the Main Street grant programs. Over the course of 15 years HOPE VI grants were used to demolish 96,200 public housing units and produce 107,800 new or renovated public housing units, of which 56,800 were to be affordable to the lowest-income households (Cisneros 2009; HUD 2010). Residents who were not re-housed in these mixed-income developments were granted Section 8 vouchers to use in the private rental market.

These policies in recent years have all attempted to deconcentrate poverty and promote mixed-income or more racially integrated neighborhoods. Whether or not they have succeeded in their goal of deconcentrating poverty, and whether, if so, that deconcentration has actually improved the life outcomes of those receiving public housing subsidies, is an empirical question on which there appears to be little consensus in the scholarly community. Although there is an extensive literature on different kinds of housing dispersal programs, the findings are often inconsistent and contradictory. By synthesizing the literature using the tool of meta-analysis, we might help answer these questions about neighborhood effects and provide guidance for future policy innovations.

**Theoretical and Empirical Foundations of Poverty Deconcentration Policies: Neighborhood Effects**

Poverty deconcentration policies are based in part on a body of scholarly research that links life outcomes to the characteristics of an individual’s neighborhood of residence. This work often goes by the shorthand term of “neighborhood effects.” The research on neighborhood effects has been a vital area of social science research across a variety of disciplines, because it touches on a question that is central to scholars of many stripes. Scholars have long sought to explain the mechanisms underlying the
reproduction of social class across generations. Though not mutually exclusive, explanations have divided into several camps, broadly speaking: structural explanations, cultural and familial background explanations, and human or social capital explanations (Corcoran 1995; Wilson 2010). In each of these families of theories (described below), a person’s neighborhood of residence may play an important role in transferring poverty and its consequences across generations.

One of the key variables in the neighborhood effects literature is the degree of poverty concentration that residents face. In 2000, 3.5 million poor people across the United States lived in neighborhoods with poverty rates in excess of 40 percent (Jargowsky, 2003). Wilson (1987), Jargowsky & Bane (1991), and Danziger & Gottschalk (1987) documented large and growing spatial concentrations of poverty in American cities.

In addition to segregation by income, racial segregation has been a delicate but important undercurrent in U.S. housing subsidies since its inception. Many scholars have argued that the negative effects of concentrated poverty interact particularly perniciously with high levels of racial segregation in U.S. cities (Wilson 1987; Massey and Denton 1993). Race may be contributing to these negative effects in a number of ways, such as increased likelihood of discrimination on the part of potential employers, lack of access to levers of political power, increased likelihood of bureaucratic neglect, and the development of “oppositional” cultural characteristics as a result of experiencing sustained, hostile discrimination by whites (Massey and Denton 1993; Wilson 1996; Anderson 2000). While the Gautreaux program is the only one of these poverty dispersal programs that explicitly focuses on racial integration, race is an important undercurrent in the others as well. Race has long been a major factor in the geographical dispersion of poverty in U.S. cities and in the development of Federal low-income housing policy, particularly in the site-selection of public housing projects (Hirsch 1983, 2000; Massey and Denton 1993; Dreier et al 2004; Turner et al 2009).
It is not a matter of dispute that extremely poor neighborhoods are often home to higher crime rates, higher rates of juvenile delinquency, underperforming public schools, poor housing and health conditions, limited access to public services, and limited access to job opportunities (Kneebone and Berube, 2008). The key question in the neighborhood effects literature is whether these neighborhood conditions have an independent impact on life outcomes over and above what would be predicted from individual and family characteristics alone. Much empirical work on this question suggests that poverty concentration does have a variety of detrimental effect on the residents of these areas, both in terms of their current well-being and their future opportunities (Wilson, 1987, 1996; Jencks and Mayer, 1990; Brooks-Gunn, et al., 1993). These conditions are thought to build on each other such that the neighborhood problems are greater than simply the sum total of the constituent residents’ problems. For instance, if large numbers of poor people cluster together, perhaps fewer businesses will locate in that area, leading to fewer jobs available to residents, whereas if the same number of poor people were dispersed across middle-class neighborhoods, they would enjoy closer physical proximity to available jobs.

The deleterious effects of high-poverty areas are thought to be especially severe for children, whose behavior and prospects may be particularly susceptible to neighborhood impacts. Children raised in poor, disadvantaged neighborhoods are at greater risk for developing anti-social behaviors (Catalano and Hawkins, 1996). Disorganized neighborhoods may have weak social control networks that allow criminal activities to go unmonitored and unnoticed (Sampson and Lauritsen 1994). Numerous studies have identified links between deviant peer group influences and adolescents’ test scores (Darling and Steinberg, 1997), mental health (Simon et al, 1996), as well as antisocial behavior and substance abuse (Case and Katz, 1991; Dubow, Edwards, and Ippolito, 1997; Gozales, 1996).

Neighborhood effects can be theoretically linked to many of the leading theoretical explanations for the intergenerational transfer of poverty. Structural theories of poverty transfer focus on exogenous
factors, like changes in the economy and the impact of racial discrimination, as being most explanatory of poverty transfer. Under this structural approach, a person’s neighborhood of residence is enormously important for understanding their life outcomes, because it determines what sorts of jobs, social networks, and services they can access physically. W.J. Wilson’s work (1987; 1996) has focused on structural changes in the economies of central cities, with high quality manufacturing jobs, stores and banks, and middle-class residents fleeing to suburbs, leaving behind a core of poor and minority individuals who lack physical access to high quality jobs for people with low skill levels and who lack the social connections with affluent people that may help them secure a job in the first place. This concept that the jobs and the people who need them are in different physical locations has been called “spatial mismatch.”

Cultural theories of poverty posit that poverty transfer is driven by the beliefs, values, behaviors, frames, social skills, and other components of culture that are transmitted across generations by families and communities. The connection to the neighborhood effects debate is the idea that communities, not simply families, are important for transferring culture across generations. Oscar Lewis’s (1966) concept of a “culture of poverty,” for instance, argued that poor communities across the world were characterized by similar sets of values, such as an emphasis on machismo, a “present time orientation,” and anti-intellectualism. While Lewis traced the development of these cultural characteristics to macro-economic conditions, his overall approach of examining the role of culture in transferring poverty across generations fell out of favor with the scholarly community. These types of theories were eschewed by many U.S. sociologists and anti-poverty scholars for some time as (perhaps unfairly) ignoring the role of social and economic structures in perpetuating class divisions, for painting an unrealistically monolithic picture of culture in poor communities, and for “blaming the victim,” i.e., laying the blame for poverty at the feet of the poor themselves. However, W.J. Wilson’s work helped revive cultural approaches to understanding poverty and poverty transfer, albeit in a much more
nuanced form, by incorporating them into the structural explanations mentioned above (Wilson 1987, 2010; Small 2001; Lamont et al 2010; Corcoran 1995). This approach posits that once poor, jobless, segregated neighborhoods (i.e., ghettos) are created by discriminatory housing policies, they tend to reinforce values, attitudes, beliefs, and behaviors that are inimical to successful “escape” from poverty.

For instance, Anderson’s (1999) ethnographic study paints a picture of young urban male culture where a violent “code of the street” has evolved as a response to sustained joblessness and material deprivation. Anderson theorizes that young men in these very poor communities must engage in violence, or at least portray themselves as being willing to resort to violence, as a way to earn respect from their peers, which then helps them avoid future victimization. This “cool pose” attitude helps protect these young men in their daily lives, but it is not valued at all in a customer-service oriented modern economy. If poor people do indeed learn behaviors, values, attitudes, and beliefs from the people and the conditions in their neighborhoods, then helping them move to a community with cultural norms that are more accepted by the mainstream economy might improve their life outcomes.

Human capital approaches to poverty transfer are also closely tied to the question of neighborhood effects. If labor market success is driven by a person’s stock of valuable skills, then it makes sense to look to schools as a potential player in why poverty is transferred across generations. Neighborhood of residence is important because it determines where a child will attend school and who his peers will be within that school. If a student lives in a neighborhood with a higher property tax base, their schools are more likely to receive higher per pupil expenditures. If a student lives in a neighborhood with high achieving peers, that student is more likely to do well in school (Coleman 1966; Case & Katz 1991). Regardless of which of these two theories of school failure one endorses, they both imply that a child’s neighborhood of residence should have important implications for that child’s development of marketable skills, which correlates with his or her future success in the labor market.
Thus, neighborhood of residence may perpetuate intergenerational poverty transfer through the mediating effect of neighborhood-based public school districts.

Social capital theory posits that a person’s social networks provide a range of direct and indirect supports that can enhance labor market outcomes. Granovetter (1973) analyzes the importance of “weak ties” that link individuals to a wider network of information and resources than is defined by that person’s network of family members, relatives, and close friends. Neighborhoods of concentrated poverty are seen as lacking bridging social capital, which can increase an individual’s access to resources that would allow him or her to “get ahead” economically. Moving people out of concentrated poverty into better neighborhoods might help poor people forge ties with people of higher social economic status and that these connections will help them find jobs.

Regardless of whether the main mechanisms of poverty transfer lie in structural constraints, cultural adaptations, educational disparities, or social capital gaps, if neighborhood effects are important in any of the hypothesized mechanisms of poverty transfer, it would follow that deconcentration policies are potential weapons against poverty. Moreover, if poverty deconcentration policies fail to have an impact on life outcomes, we would have evidence weighing against all of these theoretical models of poverty transfer. For instance, moving to a better neighborhood might improve people’s outcomes for a variety of reasons. It might increase their physical proximity to available jobs. It might improve their social networks with higher-income people, which in turn might help them find a job. It might increase a child’s exposure to positive peer influences and role models, promoting higher aspirations, better social skills, and more aversion to deviant behavior. And it might improve a child’s outcomes by allowing him or her to attend a higher quality school with a higher performing peer group. But if moving to a less poor neighborhood does not do anything at all (or if it makes things worse), one would have evidence that these theories of poverty transfer are deficient in some way.
Empirical Literature on the Effect of Poverty Deconcentration Policies on Economic and Behavioral Outcomes

The existing research on poverty dispersal policies offers conflicting stories about the effectiveness of these programs in improving individuals’ life outcomes. Studies of the Gautreaux program, which provided Section 8 vouchers to African-American families who were eligible for public housing in Chicago and required them to move to a low-poverty and racially integrated neighborhood, have revealed significant improvement in employment rates, academic attainment, and crime rates for families that used the program to move to suburbs rather than to neighborhoods in cities (Popkin et al. 1993; Rubinowitz and Rosenbaum 2000). Keels found that young boys who moved to suburbs seemed to reduce problem behaviors, while girls actually fared worse (Keels 2008). There were several weaknesses about the Gautreaux studies. The studies only followed families who actually moved, which turned out to be only about 20% of the sample. More importantly, families were not assigned randomly into different kinds of neighborhoods. Rather, the process involved a combination of self-selection and screening procedures, both of which create an endogeneity threat. Although the studies controlled for a number of important baseline characteristics, it was still possible that outcomes were driven by unobserved differences between suburban and urban movers, not by neighborhood type.

The Moving to Opportunity experiment (begun in 1994) used randomization to reduce the threat of endogeneity. Moving to Opportunity was much like the Gautreaux program, in that it granted Section 8 housing vouchers to public housing residents and conducted follow up studies to measure the program’s impact on a variety of life outcomes. The 5 city experiment had two treatment groups; one experimental group which was provided housing counseling and was required to move to a low-poverty neighborhood for at least one year, and one Section 8 group, which was offered vouchers according to the regular rules and services of the Section 8 program at that time, with no geographical restriction and no special assistance. Additionally, many of the analyses employ both Intent-to-Treat and Treatment-
on-the-Treated measures of program impact. Intent-to-treat measures compare the full sample of the experimental group to the full sample of the control group, including measures of the treatment group who never actually used their offered voucher. Treatment-on-the-Treated measures attempt to compare only those in the experimental group who actually used the treatment to the control group. In 2011, the third and final wave of post-randomization follow-up studies was released (Sanbonmatsu et al 2011). The findings from the three waves of MTO follow up are not entirely consistent across time, site, or measurement type (ITT or TOT). In general, using the full sample across all 5 sites, the MTO experiment has found no evidence of significant improvements for either treatment groups on measures of economic well-being, which is a stark contrast to the findings from the Gautreaux program (Goering and Feins 2003; Orr et al 2003; Briggs et al 2010; Sanbonmatsu et al 2011).

MTO’s evidence on problem behaviors among youth is mixed. The early MTO evaluations from Boston showed improvement on a behavioral problem index for boys only (Goering et al 2003; Katz et al 2001). Early findings from Baltimore showed reductions in violent crime rates for male youth, but increases in property crime (Goering et al 2003; Ludwig et al 2001). Early findings from New York showed behavioral improvements for boys, but no impacts for girls (Leventhal and Brooks-Gunn 2003). In general, these early findings seemed to show that housing mobility could improve youth behavioral outcomes for boys and would be harmless at worst for girls. However, for the interim evaluations, MTO data show that young men in the treatment group showed statistically significant increases in problem behavior, while young women showed statistically significant reductions in problem behavior (Kling et al 2005; Kling et al 2007; Briggs et al 2010). This finding is a direct contrast to Keel’s finding from the Gautreaux data that young female suburban movers fared worse than young male suburban movers. Another study of the entire MTO sample showed that when adults and youth are pooled together, there are no statistically significant impacts on individual crime rates, positively or negatively (Harcourt and Ludwig 2006). Clearly, MTO shows discrepancies in behavioral outcomes depending on site, time of
measurement, and subgroup (age and gender). Finally, the latest released MTO Final Impacts Evaluation (Sanbonmatsu et al 2011) shows virtually no impact on behaviors for either gender group.

Studies of the HOPE VI program thus far have reached conclusions that are similar to MTO. Briefly, HOPE VI funds the rehabilitation or demolition and reconstruction of highly distressed public housing projects, with the goal of creating mixed-income developments. By design, not all of the original tenants would be able to relocate into the new mixed-income developments, and those tenants were either allowed to transfer to another public housing site or were given a Section 8 voucher. The HOPE VI panel study (not included in this analysis because of insufficient reporting) has shown improvements on subjective sense of safety, but little improvement on individual economic outcomes (Buron 2002; Popkin et al 2009). Those who moved into other public housing units showed virtually no improvement, but those who moved via a Section 8 voucher did. To date, not enough of the mixed-income developments have come on-line to have studied their impact. Youth from families that used the voucher also displayed significant improvements in their level of problem behaviors compared to those who relocated to public housing (Gallagher & Bajaj, 2007). However, youth, particularly girls, who either used their vouchers to move to new public housing projects, or who remained in their current unit awaiting reassignment, showed significant increases in problem behavior. Unfortunately, most of the extant literature on the HOPE IV mixed-income developments is unusable in this meta-analysis because of a lack of reporting of statistics necessary to calculate effect sizes.

Method

A Brief Introduction to Quantitative Meta-analysis

Meta-analysis is a set of methods to quantitatively combine the results of different studies to assess the state of knowledge on a particular scientific question. Meta-analysis is most appropriate when different studies have provided different answers about a question (Ringquist 2013; Cooper and
Hedges 1994; Borenstein et al 2009). Studies on participation in poverty deconcentration policies have provided different answers about the size and direction of effects of these programs on individual economic and behavioral outcomes. By utilizing a rigorous method for assessing overall patterns of effects and even explaining variation in effects, quantitative meta-analysis is a useful contribution to research in this area.

The main dependent variable in meta-analysis is called an effect size. An effect size is an estimate of the magnitude and the precision of the association of one variable with another. Whether or not the association is causal in nature, the estimated relationship is still called an “effect size.” Meta-analysis was developed heavily in bio-medical sciences, in which random assignment designs and “treatment effects” are common, and it still bears the linguistic imprint of this history (Ringquist 2013).

While there are any number of measures of association utilized in science, meta-analysis offers techniques for converting them into a single comparable unit so that they can be quantitatively analyzed.

The unit of analysis in meta-analysis is a single quantitative model. Models are nested within studies, some of which are published in scholarly literature and some of which are found in the “gray literature.” The number of models published in a study varies greatly (see table 2-A). To generate the data for a meta-analysis, a researcher will attempt to identify all extant studies on the scientific question of interest and code the effect sizes generated by the model or models in those studies. Researchers may also code model-level and study-level variables that they believe may be associated with systematic variation in the effect sizes. These “mediating and moderating” variables as they are often called can then be used in techniques such as meta-regression to help explain systematic variation in effect sizes (Borenstein et al 2009; Ringquist 2013).
In meta-analysis, once effect sizes are converted to one comparable unit (for instance, this meta-analysis converts all effect sizes to Fisher’s Z), and an associated standard error is calculated for each, they are then weighted proportionally to the sample size in the underlying model that generated the effect size. A typical approach is to weight them with the inverse of the variance. The variance of Fisher’s Z 1/n-3, and so the inverse variance weight is n-3. The logic behind the weighting is that effect sizes that are estimated more precisely should be accorded more consideration when assessing the overall state of knowledge on a given question than effect sizes that are estimated imprecisely.

Meta-analysts much choose between a fixed effects framework and a random effect framework. A fixed-effects framework assumes that there is only one population parameter for the average effect size, and the estimates from each study vary from it only because of sampling error. The random effects framework assumes that different studies are actually sampling from different populations with different population parameters (Borenstein 2009; Ringquist 2013). The estimates of average effect sizes from random effects meta-analysis are for the mean of a distribution of population means, rather than the mean of a single population. There are techniques for assessing which framework is most appropriate, but Ringquist (2013) argues that there is almost always theoretical reasons for preferring the random effects framework for studies in the social science and policy disciplines. The two meta-analyses presented here utilize the random effects framework.

The meta-analysis techniques employed in this chapter can be loosely divided into two types: descriptive meta-analysis and meta-regression. Descriptive meta-analysis produces an estimate of the overall average effect from the effect sizes in the study, and it explores patterns in that overall average effect by utilizing visual tools called forest plots. Meta-regression is, as the name implies, the use of linear regression techniques to explore systematic associations between study and model-level variables with the size and valence of the estimated effect sizes. In the two meta-analyses in this chapter, I utilize
three different regression models (meta-regression, manually weighted OLS with clustered standard errors, and generalized estimating equations).

**Research Questions and Inclusion/Exclusion Criteria**

Every meta-analysis must clearly define the criteria by which studies (and the specific models in them) are included and excluded from the analysis. In doing so, a meta-analysis must strike a balance between being expansive enough in the inclusion criteria to offer a useful summation of current research and while avoiding being so overly expansive that the analysis runs afoul of the “apples-to-oranges” problem and dilutes the substantive conclusions to the point of being useless.

Table 2-B presents a summary of the inclusion criteria used for the two meta-analyses in this chapter, and it notes some of the major categories of research that are excluded by those criteria. It is important to note that there is much high-quality research that is relevant to the issue of poverty deconcentration that is not included in these studies. One should not take from the fact that it does not fit the inclusion criteria that a particular study is irrelevant or of low-quality. It simply may be asking a question that is different than the questions motivating this study, or it may not offer the quantitative information necessary for calculating an effect size.

<table>
<thead>
<tr>
<th>Inclusion Criterion</th>
<th>Explanation</th>
<th>Notable Exclusions from this criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S. Studies</td>
<td>Only studies on U.S. populations were included.</td>
<td>Studies on poverty deconcentration using non-U.S. data.</td>
</tr>
<tr>
<td>Non-Repetition</td>
<td>Only one version of each study was included.</td>
<td>Multiple published versions of the same studies or models.</td>
</tr>
<tr>
<td>Unit of Analysis</td>
<td>Only studies with individual-level data were included.</td>
<td>-Studies on neighborhood-level outcomes of poverty deconcentration policies.</td>
</tr>
<tr>
<td>Sufficient Information Provided</td>
<td>Only studies for which enough information to calculate an effect size.</td>
<td>-Qualitative studies of poverty deconcentration policies.</td>
</tr>
</tbody>
</table>
effect size was provided were included.

-Quantitative studies that did not provide sufficient information to calculate an effect size.

**Dependent Variable**

- Only outcome measures of economic well-being and outcome measures of problematic behaviors were included.
- Mental health outcomes.
- Physical health outcomes.
- Neighborhood quality outcomes.
- Educational outcomes.
- Social network outcomes.

**Independent Variable and Treatment Group**

- Participation in a government-run program designed to deconcentrate participants’ exposure to neighborhood poverty concentration.
- Studies that focus on unsubsidized moves from worse to better neighborhoods.
- Neighborhood effects studies where the independent variable is a measure of exposure to neighborhood quality (poverty, etc), even when this change is achieved via a government program.

**Control group**

- The “control population “must be either a population that remained in the higher poverty area, or it must be the treatment group at an earlier time point.
- Studies in which the control group are unsubsidized households (for example, a study on the impact of housing voucher receipt on previously unsubsidized households).

One important criterion to highlight is that this chapter is studying the impact of *participation in policy interventions* on individual outcomes, not the impact of *exposure to neighborhood characteristics* on individual outcomes. While the neighborhood effects literature is important for understanding the theoretical foundation and the motivation behind these programs and for their hypothesized results, this chapter should not be construed as a meta-analysis of neighborhood effects. Two major types of studies are excluded because of this criterion. One type examines the impact of private-market (i.e., unsubsidized) moves from worse to better neighborhoods on individual outcomes. Examples of this
type of study include Schwartz (2015), Gasper (2010), Metzger (2015), and Porter and Vogel 2014. Another type examines the relationship between neighborhood characteristics (like poverty rate) with individual outcomes. This would be the classic “neighborhood effects” analysis, where the independent variable is a measure of a particular neighborhood characteristic and the dependent variable is some outcome of interest. There are a large number of examples in the literature of this type of study, such as Brooks-Gunn et al (1993) and Darling and Steinberg (1997). These two types of studies are both clearly relevant to the topic, but nevertheless, the questions they attempt to answer are different than the questions under examination in this study. A null finding in these meta-analyses may indicate flaws in the underlying neighborhood effects theories, but it may just as well indicate a failure of the policies to effectively alter participants’ exposure to the relevant neighborhood characteristics. If the policies in question fail to induce enough of a “dosage,” then one would not expect them to alter life outcomes. Thus, the reader is cautioned to interpret this chapter as a meta-analysis of program effects, not a meta-analysis of neighborhood effects.

Another important criterion to highlight is the particular outcome types chosen. Economic outcomes (employment, income, welfare receipt) and problematic behaviors (crime and delinquency) were chosen because much of the motivation for housing dispersal projects was based initially on these goals (Briggs et al 2010). While clearly important, economic and behavioral outcomes are not the only important outcome measures that one might choose to study. Other outcomes that have been studied in the literature include mental health, physical health, neighborhood satisfaction, educational outcomes, and social network changes. Examples of studies of other types of outcomes include Deluca 2007; Deluca and Rosenblatt 2010; Acevedo-Garcia et al 2004; Damm and Dustman 2014; and Cook 2008. The reader is cautioned to remember that a null finding from this meta-analyses does not necessarily indicate that a different meta-analysis on poverty deconcentration would yield null findings on other outcome measures.
One final criterion to highlight is the constitution of the control group in the models. One important family of studies that is not examined in this meta-analysis are those that assess the impact of housing choice vouchers on life outcomes compared to a control group of individuals receiving no housing assistance. Examples of studies excluded by the criterion include Carlson et al (2011a; 2011b) and Lens (2014). While it is true that giving housing vouchers to previously unassisted households might allow them to relocate to lower poverty neighborhoods, we would argue that this is a fundamentally different policy intervention than deconcentrating households that are starting out as public housing residents. The initial degree of poverty concentration is likely to be much higher for public housing residents than for unassisted households, and the impact of the subsidy on work incentives is going to be completely different depending on whether the recipient is starting out as a public housing resident or as someone receiving no housing subsidies. Consequently, this meta-analysis only examines studies that offer a comparison group consisting of public housing residents (either the treatment group at a previous time point, or a cross-sectional study comparing voucher receivers or mixed income residents to public housing residents).

**Literature Search Process**

The literature search process begins with ex-ante defining key words and phrases to be used in data base searches and identifying data-bases and organizations that are likely to lead one to research on the relevant scientific question. The updated meta-analysis search was conducted using the following key word and phrases: HOPE VI, Section 8, Moving to Opportunity, Gautreaux, Poverty Dispersal, Housing Choice Voucher, Neighborhood Effects, Poverty Deconcentration, and Mixed-Income. Table 2-C presents the major databases and organizations that were searched.
Table 2-C: Data Bases and Sources Searched for Updated Meta-analysis

<table>
<thead>
<tr>
<th>Academic Search Engines</th>
<th>Think tanks,</th>
<th>Working</th>
<th>Government</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic Search Premier (EBSCO)</td>
<td>Urban Institute</td>
<td>SSRN</td>
<td>HUD</td>
</tr>
<tr>
<td>JSTOR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISI Web of Knowledge</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Google Scholar</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eric</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>British Library Document Supply Centre BLDSC</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The search process will inevitably identify hundreds if not thousands of “hits.” From these initial hits, the next level of sorting occurs by examining titles. Finally, these “relevant” studies winnowed into a collection of “acceptable” studies by examining, first, abstracts and ultimately by a close examination of the study’s content (Ringquist 2013). Figure 2.1 presents a “study flow diagram” of the search process for the earlier version of this study and for the updated version. In conjunction with the inclusion and exclusion criteria listed earlier, the results of this meta-analysis should be replicable (with allowances for intercoder disagreement and coder error). Ultimately, 44 studies from the first version of this meta-analysis were identified as containing at least one effect size that met the inclusion criteria. 11 studies were identified in the update as containing effect sizes that met the inclusion criteria, for a total of 55 studies. These sets are further explained in table 2-1. Appendix 2-A lists all of the studies that
were coded for this meta-analysis, divided into their outcome type (economic or behavioral). Again, some studies contained both type and are thus listed under both outcome categories.

**Analytical Strategy, Models, and Variables**

**Effect Sizes, Variance, and Weights**

The effect size chosen for the economic and behavioral meta-analyses was Fisher’s Z, a correlation, or r, based effect size that has been adjusted for small samples (Ringquist 2013; Borenstein et al 2009). All effect sizes in the studies examined were ultimately converted to Fisher’s Z. Effect sizes that were recorded as regression coefficients (least squares, probit, logit, tobit, etc) were converted to t-scores, then into r-based effect sizes, and finally into Fisher’s Z. Effect sizes that were recorded as differences of means were converted to a d-based effect size, then converted to r, and then into Fisher’s Z. Finally, effect sizes that were recorded as a difference of proportions were converted to log odds ratios, then converted to d-based effect sizes, and then converted to r and Fisher’s Z. Like the correlation coefficient, Fisher’s Z has a range of -1 to +1. Models that favored the control group (i.e., poverty deconcentration had a negative or null effect on the outcomes) would be converted ultimately to a negative Fisher’s Z, and studies that favored the treatment group were converted to a positive Fisher’s Z.
Each Fisher’s Z effect size is associated with a variance that is proportional to the inverse of the model’s sample size \((1/n-3)\). Because a random effects framework is appropriate for both statistical and theoretical reasons, the standard error is inflated slightly by an additional variance, \(\tau^2\), which captures the additional variability that stems from different underlying population parameters being estimated. The inverse of these random effects variance form the weights which are multiplied with the Fisher’s Z score. These weighted Fisher’s Z scores and their variances serve as the raw material for the calculation of overall average effects.

**Forrest Plots**

Forest plots offer a visual tool that sheds more light on the patterns underlying the overall average effect estimate. Ringquist (2013) argues that the overall average effect is not a very informative statistic in a meta-analysis that utilizes the sorts of studies found in social science and public policy. There is too much heterogeneity in the populations of interest (as indicated by the choice of the random effects framework) for the overall average effect to be useful. Forest plots help unpack the overall average effect. This analysis will use two types of forest plot (regular forest plots and cumulative forest plots) to help elucidate the estimate of overall average effect. These plots utilize the same basic data as the overall average effect size estimate (the weighted Fisher’s Z estimates and their standard errors), but they display visually how these estimates change over time or by study.

**Meta-Regression and Models**

Meta-regression is the use of regression techniques with a meta-analysis data set, where the effect size is the dependent variable and the independent variables are mediating and moderating variables from the models and studies that were coded for the meta-analysis. Like all regression techniques, the variation in the effect size estimates can be explained by variation in the study-level and model-level variables.
Meta-regression in the social sciences and public policy must account for complications with the data that are not found as often in the bio-medical sciences, where meta-analysis was most heavily developed. In the bio-medical sciences, each study typically produces only one effect size. In the social sciences and public policy, a given study may publish the results of many different models (with different parameterizations, dependent variables, estimation techniques, etc.) (Ringquist 2013, Borenstein et al 2009). Ringquist (2013) recommends that all of these effect sizes be included in the meta-analysis to preserve maximum information. However, doing so adds the complexity that each observation is no longer independent. Effect sizes are now clustered within studies, complicating estimations of standard errors in regression analysis. This analysis utilizes three meta-regression models to estimate the relationship between the independent variables and the effect size estimates. The first is a standard meta-regression that does not account for clustering. The second two techniques provide two different strategies for accounting for study-level clustering of observations (Ringquist 2013). The first is a weighted least squares (WLS) model in which the variables are weighted according to the inverse-variance weight strategy outlined above and in which the standard error calculations are clustered at the study-level. The second is a generalized estimating equations (GEE) approach which relaxes the requirement for non-independence of observations found in normal OLS.

**Independent Variables**

Tables 2-D and 2-E describe the independent variables in the meta-regressions for the economic and behavioral outcomes data sets, respectively. These variables were chosen because of strong evidence in the extant theoretical and empirical literature that they would be predictive of variation in
effect size estimates. The variables used in the two different outcomes type are very similar but not identical. These differences are explained in the discussion of each variable grouping below.

**Program Type**

This set of variables represents what types of housing programs are investigated in the original studies. As discussed earlier, four major housing programs are the focus of this meta-analysis: the Gautreaux program, the Moving to Opportunity Program, which includes the MTO voucher program and the MTO Section 8 program; the general Section 8 voucher program; the HOPE VI program, which includes the HOPE VI voucher program and the HOPE VI mixed income housing developments, and an other/multiple category for programs that were difficult to classify but still clearly met the inclusion criteria.

For the meta-regressions, these 7 program types were sorted into five groups based on the degree of programmatic similarity. These groups were the Gautreaux program, the Moving to Opportunity low poverty voucher program, all the other voucher programs (which all utilized the Housing Choice Voucher, or Section 8), mixed-income developments, and other. There was no need for an “other” category for the behavioral outcomes meta-analysis, because none were coded. This grouping recognizes that these programs differ significantly. The Gautreaux studies followed households that moved from inner-city, largely racially segregated public housing in Chicago to affluent, largely white suburbs. The Moving to Opportunity experiment’s primary treatment group was required to move to a less-poor neighborhood (<20% neighborhood poverty rate) for a period of one year as a condition of their voucher receipt, but there was no explicit effort to promote racial integration. The second MTO treatment group simply received regular Housing Choice Vouchers, which are really identical to all the other Housing Choice Voucher (Section 8) studies in the data set, including displaced HOPE VI households who were given a voucher. These programs were therefore combined into one
category. Finally, mixed-income housing developments represent a completely different approach to deconcentrating poverty, thus meriting its own category.

**Outcome Type:**

This set of variables describes the outcome measures in the original studies to test whether the direction and magnitude of the estimated program effects are dependent on the way the outcomes are measured. Employment, earnings, welfare receipt, and some general economic well-being indices are the most common methods of measuring the economic well-being of the tenants that are impacted by the housing programs. For negative behaviors, we categorize the outcomes into general crime/delinquency index, violent crimes, property crimes, and risky behaviors of the youth, which includes drug use, alcohol use, smoking, and risky sexual behaviors.

**Demographic Variables:**

Past studies have shown that the program effects might differ along gender lines, for different age groups, and by race (Kessler 2014; Chetty et al 2016; Clampett-Lundquist 2007; Clampett-lundquist et al 2011). For gender, we divide the effect sizes into three groups: males, females, and pooled. Pooled effect sizes are the reference category. For race, there is a dummy indicator for effect sizes that specifically examine a particular minority group. There is not enough variation in this variable in the behavioral data set to warrant including it. Finally, we create a single dummy variable for effect sizes that specifically examine participants who were youth at the time they began the program. The reference category is therefore studies that are on adults or which do not specify an age category.

**Study Quality**
There is evidence that more rigorous study designs tend to estimate effect sizes of a smaller magnitude. The effect sizes are sorted into three categories according to their level of rigor. One dummy variable codes whether an effect size was generated from a study that utilized a random assignment design. Another variable is a dummy codes whether a study utilized a different form of endogeneity control, such as matching or instrumental variables. Finally, studies with no control for endogeneity (e.g., single group pre-post designs) we coded as zero and form the reference category.

**Amount of Time Elapsed Since Start of Treatment:**

Evidence in the extant literature suggests that time may be an important variable to include. There may be exposure effects or delayed effects, so that effects would get stronger over time, or conversely, fade-out effects, where effects would decline over time. Unfortunately, fine grained assessments of these time variables were not impossible with the level of detail offered in the studies. This variable represents this author’s best estimation of the average amount of time elapsed for the study population since the initiation of treatment. However, whether the interim time consisted of treatment exposure, or whether the treatment occurred at a discreet moment and then was then ended, cannot be determined. One can regard this variable therefore as containing significant measurement error.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Number of Effect Sizes in Data Set (for Dichotomous Variables).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2-D: Economic Outcomes Data Set: Variable Descriptions</td>
<td></td>
</tr>
</tbody>
</table>

37
Average and Standard Deviation (for Continuous Variables).

<table>
<thead>
<tr>
<th>Programs</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gautreaux Program</td>
<td>34</td>
</tr>
<tr>
<td>Moving to Opportunity-Low Poverty Neighborhood Voucher</td>
<td>902</td>
</tr>
<tr>
<td>Moving to Opportunity—Section 8 Voucher</td>
<td>865</td>
</tr>
<tr>
<td>HOPE VI-Section 8 Voucher</td>
<td>33</td>
</tr>
<tr>
<td>Other Section 8 Voucher</td>
<td>46</td>
</tr>
<tr>
<td>Mixed-Income Development</td>
<td>52</td>
</tr>
<tr>
<td>Other/Multiple Programs</td>
<td>44</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>1675</td>
</tr>
<tr>
<td>Male</td>
<td>135</td>
</tr>
<tr>
<td>Female</td>
<td>166</td>
</tr>
<tr>
<td>Pooled</td>
<td>144</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age at Treatment</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult</td>
<td>1,450</td>
</tr>
<tr>
<td>Youth</td>
<td>382</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Race</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall/White</td>
<td>1926</td>
</tr>
<tr>
<td>Black</td>
<td>42</td>
</tr>
<tr>
<td>Latinos</td>
<td>8</td>
</tr>
</tbody>
</table>

| Outcome Type                                 |       |


<table>
<thead>
<tr>
<th>Variables</th>
<th>Effect Size Count (for Dichotomous Variables)</th>
<th>Average and Standard Deviation (for Continuous Variables)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Programs</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gautreaux Program</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Moving to Opportunity-Low Poverty Neighborhood Voucher</td>
<td>789</td>
<td></td>
</tr>
</tbody>
</table>

Table 2-E: Behavioral Outcomes Data Set: Variable Descriptions
<table>
<thead>
<tr>
<th>Program Type</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moving to Opportunity—Section 8 Voucher</td>
<td>648</td>
</tr>
<tr>
<td>HOPE VI-Section 8 Voucher</td>
<td>18</td>
</tr>
<tr>
<td>Other Section 8 Voucher</td>
<td>0</td>
</tr>
<tr>
<td>Mixed-Income Development</td>
<td>8</td>
</tr>
<tr>
<td>Other/Multiple Programs</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pooled</td>
<td>571</td>
</tr>
<tr>
<td>Male</td>
<td>453</td>
</tr>
<tr>
<td>Female</td>
<td>449</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age at Time of Treatment</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pooled</td>
<td>136</td>
</tr>
<tr>
<td>Adult</td>
<td>246</td>
</tr>
<tr>
<td>Youth</td>
<td>1091</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Race</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall/White</td>
<td>1463</td>
</tr>
<tr>
<td>Black</td>
<td>10</td>
</tr>
<tr>
<td>Latinos</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome Type</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Crime/Delinquency Index</td>
<td>849</td>
</tr>
<tr>
<td>Violent Crime</td>
<td>164</td>
</tr>
<tr>
<td>Property Crime</td>
<td>161</td>
</tr>
<tr>
<td>Risky Behaviors</td>
<td>299</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Years Since Treatment Began</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Years</td>
<td>8.9 (3.9) Avg (Std. Deviation)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study Rigor</th>
<th></th>
</tr>
</thead>
</table>
Random Assignment | 1407
---|---
Other Type of Control for Endogeneity | 55
No Control for Endogeneity | 11

**Sample Size in the Model Generating the Effect Size**

<table>
<thead>
<tr>
<th>N</th>
<th>2920 (3695.6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg (Std. Deviation)</td>
<td>1582 / 2364 / 4376</td>
</tr>
<tr>
<td>25% / 50% / 75% quartile</td>
<td></td>
</tr>
</tbody>
</table>

**Results: Economic Outcomes**

**Descriptive Meta-analysis**

**Overall Average Effect**

The descriptive meta-analysis identifies the overall average effects of the data set and helps identify patterns in the overall average effect using figures called forest plots. Before one conducts the descriptive meta-analysis, one must choose whether to adopt a fixed effects framework or a random effects framework. A fixed effect framework assumes that the effect of the underlying treatment in question differs from study to study for reasons of sampling error only. The random effects framework assumes that, due to idiosyncrasies of program implementation, historical conditions, target populations, and other variables, there is heterogeneity in the population parameter for mean effect size across studies. Table 2-F presents the Q and \( i^2 \) statistics calculated for the two economic outcome data sets, and based on these, a random effects framework was chosen (Ringquist 2013; Borenstein et al 2009). These two statistics both indicate that more variation exists in effect sizes than one would expect from sampling error alone, indicating that the studies in the meta-analysis are actually estimating different underlying population parameters.
The overall average effect from the random effects meta-analysis of all economic outcomes is negative and of a very small magnitude (-.00147), and it is highly statistically significant. In the context of random effects analysis, technically this means that the estimated mean of the distribution of population effect sizes in the universe of populations that could be studied is -.00147. This indicates that, overall, in the average population that could be sampled for a study, poverty deconcentration policies actually have an extremely small but negative impact on individual economic well-being.

To provide a sense of how the Moving to Opportunity studies are influencing these results, the overall average effect has also been calculated for a data set that excludes studies from MTO. For this much-reduced data set, the overall average effect on economic outcomes is 0.0132, and it is also highly statistically significant. Dropping the MTO studies changes the overall average effect size from very slightly negative to slightly positive. Although the magnitude is comparably 9x larger, it is still substantively quite a small effect. Fisher’s Z is comparable in magnitude to a correlation coefficient, ranging from -1 to 1, so the overall effects from both data sets is small. Nevertheless, the change in sign and magnitude illustrates that the MTO studies are quite influential on the data set.

<table>
<thead>
<tr>
<th>Data Set and Outcome</th>
<th>N</th>
<th>Average Effect Size Estimate</th>
<th>Std Deviation</th>
<th>Z score</th>
<th>P value (2 tailed)</th>
<th>95% Confidence Interval</th>
<th>Q statistic</th>
<th>I statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Economic Outcomes</td>
<td>1976</td>
<td>-0.00147</td>
<td>0.000416</td>
<td>-3.54</td>
<td>&lt;.001</td>
<td>[-0.00229 / -0.00066]</td>
<td>4519.08</td>
<td>56.3</td>
</tr>
<tr>
<td>Economic Outcomes, Non-MTO Studies Only</td>
<td>209</td>
<td>0.01318</td>
<td>0.002355</td>
<td>5.6</td>
<td>&lt;.001</td>
<td>[0.008564497 / 0.017794509]</td>
<td>72824</td>
<td>99.71</td>
</tr>
</tbody>
</table>
Forest Plots and Cumulative Forest Plots

Forest plots in meta-analysis display effect sizes, their estimated standard errors, and the degree to which the effect size is influencing the estimated overall average effect. Because there are 1976 effect sizes, a forest plot listing all of these would be too complex to be useful. However, figures 2.2 and 2.3 utilize data that has been averaged at the study level. All the effect sizes and their standard deviations in a given study are averaged, yielding one effect size estimate and one standard error estimate per study. Each row in the forest plots represents one study. Figure 2.2 lists study-averaged effect sizes in order of publication date, and figure 2.3 presents the same data in order of size. In both cases, the %weight column represents the amount of influence that the study-averaged effect size has on the overall average effect size. The weight is also displayed visually by the size of the shaded boxes around an effect size. Two patterns stand out from figures 2.2 and 2.3. One is that studies with average effect sizes of a larger magnitude, whether negative or positive, tend to be estimated less precisely (which in the case of random effects meta-analysis is equivalent to saying they have a smaller average sample size). The other is that there is very large heterogeneity in the weights of different studies, stemming from large heterogeneity in the underlying sample sizes in the models from each study, meaning about four studies with much larger sample sizes dominate the calculation of overall average effect.

To explore the degree to which the overall effect is being dominated by the MTO studies, the same study-averaged forest plots are calculated for a data set that excludes MTO studies. Those results are presented in figures 2.4 and 2.5. These figures display a similar pattern to the forest plots that
include MTO, though the pattern is not as stark. Still, studies with average effect sizes of a larger magnitude (negative or positive) tend to be estimated with less precision (i.e., smaller sample sizes), and the results still have a high degree of heterogeneity in the degree to which each study-averaged effect size contributes to the overall average effect size.

A cumulative forest plot in meta-analysis displays a running average of the overall average effect calculation over time. To facilitate this display, figures 2.6 and 2.7 present the cumulative forest plots presented collapsed into year-averaged effect sizes and standard deviations (rather than the study-averaged effects presented in the forest plots above). These cumulative forest plots show how, as time progresses, the average effect estimated by the literature changes as each year’s worth of studies is added to the data. Again, I present two figures, one that includes all studies on economic outcomes and one that excludes MTO. The patterns have similarities. Both begin with large positive effects that are estimated imprecisely and move toward effects of a much smaller magnitude estimated precisely. The difference between the data sets with and without MTO is that the one with MTO actually switches valence from a positive effect to a negative effect, while the data without MTO simply moves to a much smaller positive effect.

Several conclusions can be drawn from these six forest plots. One is that there is tremendous variation in sample sizes across studies in this area of research. As indicated in table 2-D, sample sizes range from an n as small as 15 to one outlier study with effect sizes that have an n of 17.9 million. In meta-analysis, both the weights of each effect size and the variance of each effect size are closely related to sample size. Effects from models with large sample sizes are accorded more precision (the variance of Fisher’s z is 1/n-3), and the weights are the inverse of the variance. Therefore when there is large variation in sample sizes as exists in this data set, studies with larger sample sizes dominate the overall average effect.
Another important conclusion is from these forest plots is that there is significant variation in both the size and the valence of effect size estimates, and studies with larger magnitude effect sizes tend to be estimated imprecisely. That is, there appears to be a relationship between sample size and the magnitude of the estimated effect size (though not the valence). Why should that be? One possibility is that small-n studies also tend to be less rigorous in design. Studies that tend to have small N’s may also tend to fail to account for threats to internal validity, such as self-selection, maturation, and excluded variables. Controlling for these threats will tend to result in lower effect size estimates.

A final conclusion is that there appears to be a time-trend in the effect size estimates. Early studies with small sample sizes found larger effects that were wiped away as later studies with larger sample sizes were added to the literature. This pattern too could be indicative of a relationship between study rigor and estimated effect size, as studies in a given area of science tend to become more rigorous over time.
Figure 2.2: Forest Plot of Study-Averaged Effect Size in Order of Year Published, All Economic Outcomes

Figure 2.3: Forest Plot of Study-Averaged Effect Sizes in Order of Size, All Economic Outcomes
Figure 2.4: Forest Plot of Study-Averaged Economic Effect Sizes in Order of Date Published, Excluding Moving to Opportunity Stories

Figure 2.5: Forest Plot of Study-Averaged Economic Effect Sizes in Order of Size, Excluding Moving to Opportunity Stories
Random Effects Meta-Regression: Economic Outcomes

Table 2-G presents the results of three meta-regression models examining the relationship of various mediating variables on estimated effect sizes. Very few of the variables in the three meta-regression models are statistically significant or of a very large magnitude. As a reminder, the three models represent increasing rigor with which standard errors are estimated, accounting for clustering of observations at the study level in the weighted least squares model (model 2) and relaxing the distributional assumptions in OLS using Generalized Estimating Equations (model 3). The GEE model also deflates the influence of studies that produce more effect sizes and accords comparatively more influence to studies that produce less.

In all three models, the constant is positive and statistically significant, indicating that effect sizes from studies where all the variables are zero (in the case of dummy variables, this represents the
reference category), the effect of poverty deconcentration on economic outcomes is positive but modest. Besides the constant, there are three variables that are statistically significant at $\alpha<.1$ in all three models and one variable that is statistically significant in two out of the three. In two out of the three models, effect sizes from models that examine male subjects will on average yield effect size estimates that are .008 lower than the reference category (pooled gender) at $\alpha<.01$. The effect is very small but is highly statistically significant. This variable is not significant, however, in the GEE model.

There are three variables that are significant in all three models. Effect size estimates from models that are based on a random assignment design, on average, produce an effect size that is .03 lower than models that use no endogeneity control. While the coefficient is small in absolute terms (the range of Fisher’s $Z$ is -1 to +1), relative to the constant and to the other coefficients in the model, it is proportionally substantial. This result lends support to speculation about the hypothesis about the forest plot results that speculated that models with larger effect estimates could result from less rigorous designs. Models that examine welfare receipt on average produce effect size estimates that are about .01 lower than models that examine the reference category, employment outcomes. Again, the effect is substantively very small, though in proportional terms is somewhat sizeable. The third variable that is significant in all three models is also a measure of outcome type. Models that examine a general economic well-being index of some sort will be on average about 0.01 larger than the reference category, employment outcomes.

While the valences of the coefficient estimates in the model align well with theoretical expectations, the relationships of most of these variables with predicted effect size cannot be distinguished from zero at conventional levels of alpha, especially once the standard errors are inflated to account for clustering of observations at the study-level. Moreover, even for the variables that are statistically significant, the effect is almost certainly too small to be of much substantive significance in terms of real-world outcomes. In other words, even under the best of circumstances, the impact of
participating in a poverty deconcentration program on individual economic well-being is likely to be positive but substantively small. Under the worst of circumstances, the impact is likely to be slightly negative but also substantively negligible.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Random Effects Meta-Regression (Model 1)</th>
<th>Weighted Least Squares with Clustered Standard Errors (Model 2)</th>
<th>Generalized Estimating Equations (Model 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.0457*** (0.01153)</td>
<td>0.0472*** (0.01186)</td>
<td>0.0453*** (0.0124)</td>
</tr>
<tr>
<td><strong>Program Type (Reference Category: Gautreaux Program)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moving to Opportunity-Low Poverty Voucher</td>
<td>-0.0194 (0.01354)</td>
<td>-0.0209 (0.02177)</td>
<td>-0.0183 (0.02099)</td>
</tr>
<tr>
<td>Housing Choice Voucher (Section 8)</td>
<td>-0.0188 (0.01352)</td>
<td>-0.0202 (0.02222)</td>
<td>-0.0174 (0.02141)</td>
</tr>
<tr>
<td>Mixed Income Housing</td>
<td>-0.0074 (0.01366)</td>
<td>-0.0114 (0.02222)</td>
<td>-0.0115 (0.0215)</td>
</tr>
<tr>
<td>Other/Multiple Programs</td>
<td>0.0092 (0.0153)</td>
<td>0.0072 (0.02537)</td>
<td>0.0049 (0.02319)</td>
</tr>
<tr>
<td><strong>Race of Treatment Group (Reference: Non-minority)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minority</td>
<td>.0058 (.00515)</td>
<td>0.004 (0.007)</td>
<td>0.0038 (0.00662)</td>
</tr>
<tr>
<td><strong>Age at Time of Treatment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Youth</td>
<td>0.0017 (0.00155)</td>
<td>0.0023 (0.00142)</td>
<td>0.0018 (0.00332)</td>
</tr>
<tr>
<td><strong>Gender (Reference Category: Both Genders)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>-0.0081***</td>
<td>-0.0077***</td>
<td>-0.0031</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td>0.0055**</td>
<td>0.0036</td>
<td>0.0011</td>
</tr>
<tr>
<td>------------</td>
<td>-----------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td></td>
<td>0.0028</td>
<td>0.0039</td>
<td>0.00344</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Time Elapsed Since Beginning of Treatment</strong></th>
<th>0.0</th>
<th>0.0</th>
<th>0.0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.00016</td>
<td>0.00033</td>
<td>0.00032</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Strength of Research Design (Reference: No Control for Endogeneity)</strong></th>
<th>-0.0314***</th>
<th>-0.0311*</th>
<th>-0.0297*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Assignment</td>
<td>0.0073</td>
<td>0.01897</td>
<td>0.01766</td>
</tr>
<tr>
<td>Other Form of Endogeneity Control</td>
<td>-0.0237***</td>
<td>-0.024</td>
<td>-0.0227</td>
</tr>
<tr>
<td></td>
<td>0.0073</td>
<td>0.01879</td>
<td>0.01767</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Outcome Type (Reference Category: Employment)</strong></th>
<th>-0.0086***</th>
<th>-0.0096**</th>
<th>-0.0079**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Welfare Receipt</td>
<td>0.00182</td>
<td>0.00402</td>
<td>0.00352</td>
</tr>
<tr>
<td>Income</td>
<td>0.0041**</td>
<td>0.0023</td>
<td>0.0004</td>
</tr>
<tr>
<td></td>
<td>0.00167</td>
<td>0.0016</td>
<td>0.0011</td>
</tr>
<tr>
<td>Economic Well-being Index</td>
<td>0.0114**</td>
<td>0.0102***</td>
<td>0.0105***</td>
</tr>
<tr>
<td></td>
<td>0.00526</td>
<td>0.00363</td>
<td>0.00289</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>N</strong></th>
<th>1941</th>
<th>1941</th>
<th>1941</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>R^2 / Adjusted R^2</strong></td>
<td>.46 (adjusted)</td>
<td>.1092</td>
<td></td>
</tr>
<tr>
<td><strong>Wald χ^2</strong></td>
<td>2168.89***</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Pvalue < 0.1  **Pvalue < 0.05  ***Pvalue < 0.01

**Results: Behavioral Outcomes**

**Descriptive Meta-Analysis**

**Overall Average Effect**
Once again using the random effects framework indicated by the Q and $I^2$ statistics, the overall average effect of participation in poverty deconcentration programs on individual behavioral outcomes is .0013 (table 2-F). The p-value of this estimate is .044, which makes it statistically significant at conventional levels of $\alpha$. This indicates that in the average population that could be sampled for a study, poverty deconcentration policies would typically have an extremely small but positive impact on individuals’ likelihood of engaging in problematic behaviors, but this impact is too small to be of much substantive significance.\footnote{Because there are so many fewer studies in the behavioral analysis, I do not conduct separate MTO and non-MTO analyses for behavioral outcomes.}

**Forest Plots**

Figures 2.8 and 2.9 present forest plots of study-averaged effects listed in order of date published and size, respectively. Once again, one can detect a pattern where larger magnitudes in effect size estimates are associated with larger standard errors (i.e., smaller sample sizes). And once again, this large heterogeneity in sample sizes is reflected in the large heterogeneity in the relative influence of each study on the calculation of overall average effect.
As with the economic outcomes meta-analysis, the cumulative year-averaged forest plot for behavioral outcomes (figure 2.10) shows how the overall average effect estimate changes over time as one incorporates the effect sizes from newer studies. The pattern here looks quite similar to the pattern from the economic outcomes. An initial large and imprecisely estimated positive effect quickly becomes subsumed by studies with much smaller, more precisely estimated (and more heavily weighted) positive effects.

Because the patterns are similar in the forest plots of economic and behavioral outcomes, it is not surprising that their interpretations are also similar. Once again, there is tremendous variation in sample sizes across studies. Sample sizes from the effect sizes coded range from an n as small as 61 to as large as 56,629. The 25th, 50th, and 75th percentile are 1582, 2364, and 4376 respectively. As with the economic outcomes data set, when there is large variation in sample sizes across the observations,
studies that use data sets with much larger sample sizes dominate the overall average effect. The pattern of variation in the size, valence, and precision of the effect size estimates is similar to that found in the economic outcomes data set, but the variation in size is not quite as large. Again, there appears to be a relationship between sample size and the magnitude of the estimated effect size, with effect size estimates of a larger magnitude (positive or negative) estimated less precisely. Finally, the cumulative year-averaged forest plot for behavioral effects shows that the time trends are also similar across the two outcome types. Initial large effects, imprecisely estimated, are subsumed by smaller, more precise, and more heavily weighted effects over time. Here there is a consistent pattern of the overall effect size estimate growing smaller as new studies are added to the literature.

Figure 2.10: Cumulative Forest Plot of Year-Averaged Effect Sizes, All Behavioral Outcomes

<table>
<thead>
<tr>
<th>datepublished</th>
<th>ES (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>0.03 (0.02, 0.04)</td>
</tr>
<tr>
<td>2003</td>
<td>0.00 (0.00, 0.01)</td>
</tr>
<tr>
<td>2005</td>
<td>0.00 (0.00, 0.01)</td>
</tr>
<tr>
<td>2006</td>
<td>0.00 (0.00, 0.00)</td>
</tr>
<tr>
<td>2007</td>
<td>0.00 (-0.00, 0.00)</td>
</tr>
<tr>
<td>2008</td>
<td>0.00 (-0.00, 0.00)</td>
</tr>
<tr>
<td>2010</td>
<td>0.00 (-0.00, 0.00)</td>
</tr>
<tr>
<td>2011</td>
<td>0.00 (0.00, 0.00)</td>
</tr>
<tr>
<td>2013</td>
<td>0.00 (0.00, 0.00)</td>
</tr>
<tr>
<td>2015</td>
<td>0.00 (0.00, 0.00)</td>
</tr>
<tr>
<td>2016</td>
<td>0.00 (0.00, 0.00)</td>
</tr>
</tbody>
</table>
Meta-Regression: Behavioral Outcomes

Table 2-H presents the results of three meta-regression models examining the correlates of effect size in the behavioral outcomes data set. Again, three models are presented. One is a standard meta-regression that does not account for study-level clustering; one is a weighted least squares model that accounts for clustering at the study-level; and one is a GEE model that also accounts for clustering. The magnitudes of the estimates are very similar across all three models, but the different ways that multiple effects within studies are treated across the three models results in substantial instability in standard errors and statistical significance. This can be understood as decreasing the influence of studies with more effect sizes and increasing the influence of studies with fewer effect sizes.

The constant in the models represents the estimated average effect size when continuous variables are set to zero and dummy variables are set to the reference category. The magnitude is similar across the three models and, while small in absolute terms, is fairly substantial relative to the unconditional overall average effect estimate presented in table 2-F. The constant is statistically indistinguishable from zero in the first model, but it is highly significant in the manually weighted clustered OLS and GEE models.

The program-type dummy variables once again are relatively stable in in terms of magnitude and valence, but the different estimation techniques result in substantial changes in which variables are and are not statistically significant. Here, the reference category is the Moving to Opportunity low poverty voucher program instead of the Gautreaux program, which was the reference category in the economic outcomes meta-regression. This change was made due to the small number of effect sizes from Gautreaux in this data set. Compared to the MTO low poverty program, effect sizes from Gautreaux are on average .05 lower. The reason that this difference becomes statistically significant in the second and third models is that these models inflate the relative influence of the Gautreaux studies
and decrease the influence of the large-N MTO studies. The effect is small in absolute terms, but
compared to the other coefficient estimates in the models and to the overall average effect, it is
substantial.

The change in statistical significance follows the opposite pattern for the Housing Choice
Voucher program variable and the mixed-income housing variable. Studies on these programs tended
to produce a larger number of effect sizes, as does the reference category, so clustering at the study
level makes it more difficult to distinguish effect size differences with the reference category from zero.

The association of age at treatment with effect size is indistinguishable from zero in the first two
models. In the GEE model, it is statistically significant at \( \alpha < .1 \), but even then, it is too miniscule in
magnitude to be of any substantive significance.

In the all but the first, unclustered regression model, effect sizes that focus on males cannot be
distinguished from those that examine both genders. However, effect sizes from samples of females are
slightly larger than effect sizes that pool genders. This result is consistent in all three regression models,
but the magnitude (.01) is very small.

In the second and third regressions but not the first, more rigorous studies tend to produce
lower effect size estimates than studies with no control for endogeneity. These differences are once
again small in absolute terms but substantial relative to the average effect estimate and the other
coefficients in the models. The change in statistical significance from the first model to the second and
third ones likely is a result of the clustering in these latter models according more influence to smaller
studies that also tend to be less rigorous.

Finally, in all three models, each outcome type is associated with a statistically distinguishable
lower effect size estimates than the reference category, violent crime outcomes. In other words,
poverty deconcentration is associated with more of a reduction in violent crime outcomes than in
general behavior index outcomes, property crime outcomes, or risky behavior outcomes. However, these differences are quite small for each outcome type (approximately .01).

The conclusions one can draw from these meta-regressions on behavioral outcome effect sizes is similar to the conclusions drawn from the economic outcomes data. While the valences of the coefficient estimates in the model align well with theoretical expectations, even for the variables that are statistically significant, the effect is almost certainly too small to be of much substantive significance in terms of real-world outcomes. In other words, even under the best of circumstances, the impact of participating in a poverty deconcentration program on the incidence of individual criminal or problematic behavior is likely to be substantively negligible, even under the “best” or “worst” conditions.

<table>
<thead>
<tr>
<th>Table 2-H: Meta-Regressions of the Impact of Poverty Deconcentration Programs on Behavioral Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
</tr>
<tr>
<td>Constant</td>
</tr>
<tr>
<td>Program Type (Reference Category: MTO Low Poverty Voucher Program)</td>
</tr>
<tr>
<td>Gautreaux</td>
</tr>
<tr>
<td>Housing Choice Voucher (Section 8)</td>
</tr>
<tr>
<td>Mixed Income Housing</td>
</tr>
<tr>
<td>Age at Time of Treatment</td>
</tr>
<tr>
<td>Youth</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

*N 1473 1473 1473

Pvalue < 0.1 **Pvalue < 0.05 ***Pvalue < 0.01
Based on the analysis of average effects, poverty deconcentration programs on average yield miniscule reductions in problematic behaviors of participants and miniscule decreases in their economic well-being. These average effects, though, are so too small to be of much substantive significance. Evidence from the meta-regressions suggests that effect size estimates DO vary with some study-level and model-level variables. In both behavioral and economic outcomes, the exact nature of the outcome variable matters. There is also strong evidence from the meta-regression and the visual analysis of the forest plots that studies with stricter controls for endogeneity tend to produce smaller effect estimates. The evidence on whether demographics matter is somewhat mixed, but the meta-regression results seem to suggest that, for behavioral outcomes, effect sizes are better for females and worse for males. However, it must be stressed that even for the variables that did vary systematically with effect size, the actual amount of predicted change in average effect size estimates they produce is very small and is probably of little substantive significance. The lack of relevance of the program type is particularly notable (other than the Gautreaux program in the behavioral meta-regression). As different as the details of these programs are from one another, one would expect to find systematic differences in their effect sizes, but none were detectable.

Overall, then, poverty deconcentration programs seem not to be very successful in achieving some of the central policy outcomes that many policy makers were hoping for: more work and earnings, greater independence from welfare, less crime and delinquency, and reductions in youth risky behaviors.

**Discussion: Poverty Deconcentration Skeptics and Apologists**

The disappointing results of poverty deconcentration programs have surprised many in the scholarly and policy communities, given the large body of theoretical and empirical work that informed them. Explaining a null finding is difficult, but scholars have offered some explanations. A small but
growing contingent of scholars could be labeled “poverty deconcentration skeptics.” Some argue against poverty deconcentration on normative grounds, likening it to the urban renewal program’s displacement of poor and minority city-dwellers in the mid-20th century (Lipman 2009; Darcy 2010). These critics argue that tearing down public housing and replacing it with mixed-income housing simply serves the interest of the middle-class while displacing the poor.

Others among these skeptics base their arguments on methodological and empirical grounds (Goetz and Chapple 2010; Manley et al 2012; Fraser et al 2012; Hedman and Van Hamm 2012; Van Hamm et al 2012. Galster 2012; Cheshire 2012). They frequently point out the ever-present self-selection problem in studies on neighborhood effects. Are there really separate impacts of neighborhood-level characteristics on people, or is it just that poor people and their associated problems tend to self-select into poor neighborhoods? Many of the scholars from these camps will point to the fact that studies that control for endogeneity such as MTO tend to find small-to-null effects from changes in neighborhoods. Perhaps the fact that certain neighborhood characteristics are associated with worse life outcomes is simply an artifact of the fact that poor households tend to be corralled by budget constraints and the housing market to closer proximity to one-another. If so, poverty deconcentration would not change individual outcomes, because those outcomes are driven by household and individual-level factors.

Oddly, the neighborhood effects literature seems to not have had much interaction with another well-grounded finding from social research: that moving is often financially and socially costly, especially for children (Metzger 2015). Moving can disrupt supportive social networks, erode the value of neighborhood-specific cultural knowledge (such as how to avoid getting in trouble), disrupt transportation strategies, etc. If neighborhood effects are simply spurious, or if they are real but small, then poverty deconcentration may bring the negative effects of disruption to households and neighborhoods without offering enough benefits to compensate. Clampet-lundquist and colleagues
(2007; 2011) for example argue that the negative behavioral consequences for boys in the MTO experiment can be traced to the fact that moving into a new school with new norms and expectations can create conflict with peers and teachers. The literature on mixed-income developments (see chapter 4 of this dissertation) indicates that simple physical proximity between more and less affluent households rarely results in their forming meaningful ties (Brophy and Smith 1997; Chaskin and Joseph 2009, 2011; Levy et al 2010, 2013).

Another scholarly camp might be called “poverty deconcentration apologists,” and their apologetics take several different forms. One form (and one obvious caveat about this meta-analysis noted in the description of inclusion criteria above) is that economic and behavioral outcomes are not the only possible outcome of interest. Some apologists point to a strong, consistent impact on poverty deconcentration movers’ mental health (Briggs 2010). Others argue that the “treatments” that been administered are deficient in some way and thus fail to adequately test either the theory of neighborhood effects or the ability of government to correct them. Casciano and Massey (2012), for example, argue that the null findings from poverty deconcentration programs can be explained by the programs’ lack of focus on racial segregation. One important way in which the Gautreaux program designs differ from MTO and other programs is that the Gautreaux program was part of a legal settlement involving racial discrimination and was designed specifically to promote racial integration. In contrast, Moving to Opportunity targeted only class. It provides families with opportunities to move to more affluent neighborhoods. In fact, most MTO families indeed remained in highly racially segregated, if more affluent, neighborhoods (Orr et al., 2003).

Other scholars have argued that MTO was a particularly weak treatment relative to the strength and duration of neighborhood effects. Aliprantis (2011, 2015), for example, reanalyzes the MTO data and looks for “dosage” effects, that is, the impact of neighborhood poverty on outcomes, and finds a significant effect. He argues that this indicates that too few of the households in the MTO experiments
actually utilized their voucher to move to a non-poor neighborhood for a significant enough amount of
time to make a difference. Much of the ex-post analysis of the MTO experiments has speculated
similarly.

Other poverty deconcentration apologists argue that overall average effects hide much sub-
group heterogeneity. For example, Chetty and Katz (2016) argue that the impact of deconcentration is
dependent on age at a finer grain than has typically been studied. Youth who were under the age of 13
at the time of relocation from public housing enjoyed positive effects on future economic outcomes,
while youth over 13 at the time of relocation suffered negative effects. Nguyen and colleagues (2016)
use model-based recursive partitioning to identify complex micro sub-groupings of the MTO data,
looking for heterogeneous effects on youth mental health.

The scholarly discussion of the strength of neighborhood effects and the efficacy of poverty
decentration policies continues apace despite strong evidence of null findings from studies of
decentration programs. The disagreement between the poverty deconcentration skeptics and
apologists will likely continue. The rest of this dissertation will attempt to contribute to that discussion
by exploring one area of explanation for the null effects found in this meta-analysis that has received
little attention in the scholarly literature: policy implementation. Chapter two will examine managerial
quality among housing authorities, the local government bodies responsible for implementing Federal
housing policies. Chapter three will examine middle-class attitudes toward mixed-income housing and
whether conceptualizations of potential neighbors’ class and race might be influencing the willingness of
middle-class residents to live in mixed-income developments.
Appendix 2-A: Studies Coded for Meta-analysis.²

Economic Outcomes

Original Meta-analysis


² Some studies are listed twice because they contain both types of outcomes.


Updated Meta-Analysis


Behavioral Outcomes

**Original Meta-analysis**


Updated Meta-Analysis


Chapter 3

Management Quality and Organizational Performance in U.S. Public Housing Authorities
Introduction

Given the theoretical and empirical weight behind the neighborhood effects family of theories, the disappointing outcomes of poverty deconcentration policies are surprising. Scholars have theorized and empirically explored a number of explanations for the disappointing results of poverty deconcentration policies on economic, educational, and behavioral outcomes. One explanation is that voucher recipient households that move out of public housing rarely relocate to neighborhoods whose economic and social conditions substantially improve upon that of their previous locations, and even when they do, these moves are often followed by further relocations that bring families back to poor, segregated neighborhoods. Explanations for this phenomenon range from social and psychological ties that anchor households to high-poverty areas, to the prohibitive financial and time demands of a wide-ranging housing search, to the unwillingness of landlords in affluent areas to accept Section 8 vouchers. Others have pointed to households’ labor market barriers (low levels of education and work experience, e.g.) that, at least in the short term, relocation cannot remedy. In conjunction with racial and class discrimination by employers, these barriers may make it unrealistic to expect relocating to a better neighborhood to improve household economic well-being. Regarding the lack of behavioral changes, some scholars have questioned whether closer physical proximity between social classes translates automatically into meaningful social interaction. Studies of mixed-income housing developments, for instance, have found that subsidized and market rate renters rarely interact. Without meaningful social interaction between classes, some of the hypothesized mechanisms of behavioral change, such as those posited by social capital theories, would not be able to occur.

Largely absent from the post-hoc analysis of poverty deconcentration policy failures is a focus on the front-line agencies responsible for administering the programs, public housing authorities. Most Federal rent subsidies for low-income households are administered by local government entities called
housing authorities, of which there are approximately 4,000 in the United States. One possible explanation for the failure to find positive behavioral or economic changes from poverty dispersal programs is that these programs were implemented poorly by the agencies responsible for their implementation. This chapter will use administrative data collected by the Department of Housing and Urban Development to assess the relationship between housing authorities’ managerial quality and their program performance. This analysis will investigate whether better management practices correspond to better performance along four dimensions: property maintenance, financial health, capital fund stewardship, and voucher program administration. If it can be established that “management matters” for the performance of housing authorities in these areas, it may offer both a potential diagnosis of the problems of poverty deconcentration policies and a prescription for their improvement. Admittedly, this chapter will not be able to assess directly the impact of housing authority management on the poverty deconcentration outcomes examined in chapter two, but if management matters for the outcomes examined in this chapter, it would be suggestive of an explanation for poverty deconcentration policy failures that implementation researchers might examine in the future.

Public Housing Authority Management: A Brief History and Review of Literature

Housing authorities are units of local government charged with administering Federal funds for public housing and for two private market subsidy programs, project-based and tenant-based Section 8. When the Federal government began the public housing program with the Housing Act of 1937, pressure from real estate interests and local governments induced legislators to provide local control over the administration of Federal public housing funds (Hays 1995; Bratt 1986, 1998a, 1998b; Hirsch 1983, 2000). To that end, the Act called for the creation of a new entity of local government called a housing authority. These entities would be responsible for the day-to-day management and
maintenance of public housing properties, with Federal funding and oversight (now conducted through the Department of Housing and Urban Development). In 1974, Congress created the Section 8 voucher program and also charged housing authorities with its local implementation.

The prominence of low-income housing policy in the national discourse has waxed and waned over the 75 years since the inception of the Federal public housing program. Perhaps the height of public scrutiny came during the years of Urban Renewal (1949-1974), when the Federal government channeled funds into cities for, among other things, the construction of new public housing projects, often with a high-rise design (Teaford 2000). These high-rise projects were built initially with much fanfare, touted as a cutting-edge solution to the problem of providing adequate and decent low-income housing to replace tenement-style slums in urban areas. The prominence of these projects in urban communities, both figuratively and literally, exacerbated the strength of the public’s disappointment as, over time, they became dilapidated and crime-ridden (Hays 1995; Hirsch 2000; Teaford 2000;). They were perceived as both enabling and creating gang violence, idleness, drug addiction, sexual promiscuity, welfare dependence, and others social ills. Ultimately, most of the high-rise projects were deemed unsalvageable and were demolished.

The high-profile failures of these large-scale urban housing projects, fairly or not, helped cement the public’s image of Federal housing subsidies as examples of misguided, counter-productive social engineering (Bratt 1986, 1991b; Teaford 2000). Since 1980, HUD has not added any new units of public housing to the existing stock (though subsidized units have been constructed through other Federally-funded programs), and as older units deteriorate and become unusable, the nation’s public housing stock continues to shrink (Hays 1995; Orlebecke 2000). While rental vouchers enjoy somewhat higher levels of support due to their more market-oriented mechanisms, they still suffer from a largely negative
reputation. Both public housing and voucher programs continue to struggle against a negative public image and to maintain a base of political support and funding (Katz and Turner 2001; Marcuse 1998).

One common target of criticism to explain perceived housing policy failures has been housing authority management (National Housing Task Force 1988; Hula 1991; Connerly 1986; Kingsley 1991; Peterman 1989; Bowie 2004). Much of this criticism has focused on the lack of performance incentives for housing authorities. A typical form of that argument would be that because housing authorities encounter such high demand for their programs, they can fail to maintain and improve buildings without sacrificing their revenue streams from rents. In addition to the lack of market-based incentives, arguments also point to housing authorities’ lack of accountability from the local electorate or from Federal oversight. Housing authority officials are usually unelected, and Federal oversight from HUD tends to be lax. Both factors insulate housing authority bureaucrats from the consequences of poor performance. Properties can become dilapidated and clients can be poorly served, the argument goes, without consequence for housing authority bureaucrats. Commonly-cited specific management problems include nepotism, indifference and even hostility toward clients, lack of effort to remove or screen out problem tenants, and financial mismanagement (Bratt 1991).

While these criticisms follow a thread of logic common to the wave of neo-liberal theory and policy reforms that arose during the 80’s and 90’s (Hood 1995; Kemensky 1996), housing authorities have also been criticized from the left. Liberal critics have documented housing authorities’ role in promoting racial segregation (Hirsch 1983; Massey and Denton 1993; Turner et al 2009; Bratt 1991). Arguably, maintaining racial homogeneity was one of the goals of mandating local control of Federal housing subsidy funds; local officials would be more likely to avoid introducing mixed-race developments when doing so would be locally unpopular (Hays 1995; Hirsch 1983, 2000). Some local public housing authorities in the early decades of the program intentionally created racially segregated
housing projects and sited them in places that maintained racial homogeneity, often in close cooperation with local governments and real estate interest. A number of housing authorities have even been on the losing end of Federal lawsuits over these racially biased practices. Perhaps the most notable of these is the 1976 *Hills v Gautreaux* case in Chicago, where the Chicago Housing Authority was found to have illegally promoted racial segregation in public housing and was ordered to provide vouchers to allow some public housing residents to move into white, suburban neighborhoods (Mendenhall 2006; Polikoff 2006; Keels 2008; Turner et al 2009).

Whether or not the blame for the failure of U.S. housing subsidies belongs to housing authority management, and even whether or not the perception that housing subsidies should be considered failures at all, are both matters of public and scholarly debate. Some housing experts argue for instance that the well-publicized failures of high-rise projects in large urban areas were outliers. They argue that the majority of the country’s public housing projects were always and continue to be of a low-rise design and rarely suffer the same degree of crime and dysfunction that arose in the urban high-rise projects. These critics contend that, contrary to popular opinion, most public housing fulfills its mission of providing safe and decent shelter for low-income households who otherwise would struggle to find affordable housing in the private market (Connerly 1986; Bratt 1986, 1998; Hayes 1995; Stockard 1998).

Other scholarly critics have accepted the failure hypothesis but have pointed to faulty policy design as the primary culprit. Housing authorities were compelled to follow two Federal dictates that were ultimately contradictory; public housing was to be reserved only for very low-income families, and the day-to-day operation and maintenance of the properties was to be funded by tenant rents (Hays 1995; National Commission on Severely Distressed Public Housing 1992; Turner et al 2001). The policy designers, though, underestimated the expense of building maintenance requirements and overestimated the strength and reliability of the revenue stream coming from clients. Many housing
projects, particularly the high rises, struggled to generate enough revenue through rents to cover maintenance costs. As the buildings deteriorated, they became less and less attractive to all but the most desperate families, further exacerbating the rent problem and inviting a concentration of only very poor people with their concomitant social problems. This vicious cycle of deterioration leading to low-rent households leading to further deterioration was compounded by the cheap building materials and techniques that were required by policy makers who were leery of making public housing seem too luxurious. Other housing experts have indicted Federal regulations from HUD that have limited housing authorities’ ability to adapt to local and even site-specific challenges.

Many of these policy design issues have been addressed in policy changes in recent decades that have, for instance, increased funding for operating expenses, increased the flexibility given to local housing authorities, provided more funding for building modernization, and encouraged a wider range of incomes in public housing projects (National Commission on Severely Distressed Public Housing 1992; Bratt 1998; Turner et al 2009). Still, there remains a perception among some scholars and segments of the population that housing authorities are classic examples of inefficient public bureaucracies. The possibility that poor management of housing authorities has contributed to failures in low-income housing policy deserves careful consideration. Despite the importance of this question in diagnosing the problems of low-income housing policy, little quantitative empirical work has been published in the scholarly literature on the management practices of U.S. housing authorities. What little quantitative empirical research exists on public housing management comes mostly from Great Britain and so is of limited applicability to the peculiar U.S. system. No large-n empirical studies of U.S. public housing authority management quality have been published in scholarly literature. Whether housing authorities are managed incompetently or are more the scapegoat for incompetent policy design is an important policy question. Those two diagnoses call for very different treatments. Consequently, an
analysis of the quality of housing authority management would be a useful contribution to scholarly and public discussions of low-income housing policy.

Existing Empirical Literature on Managerial Quality and Performance In Public Organizations

While there is little quantitative research on housing authority management per se, this chapter will use empirical and theoretical frameworks that are common in public management and implementation research. The question, “Does management matter?” is arguably the central concern of public management research (Meier and O’Toole 2002). Much empirical work has looked for the connection between management decisions and organizational performance. This work has spanned the public and private sectors, and has been conducted on local, state, and national levels of government.

As basic as the question is, quantitative empirical scholarship on whether management matters in public organizations is still an active area of public management research. While ample empirical research demonstrates the connection between management and performance for private firms, the extent to which these findings translate to public organizations is contested (Boyne 2004; Boyne et al 2006). Public management scholars who argue that they do not translate would cite the contrast between the relatively simple goal of the profit motive for private firms with the complex, ambiguous, and conflicting goals of public organizations, although some management scholars have argued that the extent of the difference has been overstated (Walker and Andrews 2013).

A very large number of quantitative and qualitative studies (often case-studies) have been published on the performance impact of specific management practices and qualities in public organizations. Out of this large body of work, some of the management variables that have emerged to show the most promise for improving performance include networking practices, organizational culture, goal-directed management, human resource management, and leadership (for reviews of empirical
findings on effective management practices, see Wolf 1997; Boyne 2003; or Rainey 2009). Despite these findings that specific managerial traits and practices tend to correlate with better performance, scholars still aspire and struggle to understand the performance impact of a more general underlying construct: “managerial quality”. This task is difficult for both construct validity and measurement validity reasons (Meier and O'Toole 2002, O'Toole and Meier 2011).

First, as a construct, it is difficult to define what “managerial quality” actually means. As noted, the potential avenues through which management might impact organizational performance are extremely numerous and often contingent on moderating variables (Walker and Andrews 2014; O'Toole and Meier 2011). What then exactly is managerial quality? Is it a set of decisions, systems, and “best practices” that research has found to be frequently effective? Given that the effectiveness of specific variables appears from the literature to be interactively contingent on moderating conditions, it would be difficult to specify a list of managerial traits or behaviors that universally promote performance in all programs in all organizations in all environments (Boyne et al 2006; O'Toole and Meier 2011). Is it better conceptualized as a personal trait that a particular manager can possess (like charisma, vision, or adaptability)? Is it something about the training and experience of managers? The construct seems simple on its face but upon further examination is difficult to define.

A distinct but related complexity is the measurement of managerial quality’s impact on performance. Measuring the impact of management quality is something of a question begging problem; how can one tell what decisions or practices constitute quality management other than through the use of performance indicators? But if performance indicators are used to define management quality, it becomes difficult to test the impact of management quality on performance, because their correlation would then be tautological. Some scholars have tried to circumvent this problem by using employee surveys to assess the quality of management, but this method too has
problems. It is unclear what management practices or characteristics would inform employees’ opinions of their supervisors (O’Toole and Meier 2011). If some part of that opinion is also based on organizational performance, then the tautology problem remains. Another technique is to measure managerial training and/or experience, but doing so requires one to demonstrate these variables do correlate with better management, which brings one right back to the problem of defining quality management. Moreover, the extent to which managerial experience and education actually correlate with performance is itself a matter of debate. Because of these difficulties, there is no commonly accepted, unproblematic way to measure the construct “managerial quality.”

Measuring the performance construct also presents major challenges. Public programs and public organizations have multiple goals, including the well-known “three e’s” of efficiency, equity, and effectiveness, and they typically have multiple constituencies. It is rarely self-evident which goals and which constituencies should be prioritized when measuring organizational performance (Mazmanian and Sabatier 1989; Rainey 1993; Stone 2001; Andrews et al 2006; Walker and Andrews 2013; Chun and Rainey 2005;). Similar to the problem of measuring management quality, debate exists over whether performance should be measured with subjective or objective measures and whether performance indicators should be more specific or aggregated (Andrews et al 2006). These ambiguities have led some scholars to take a constructivist view of the concept of performance, concluding that “all measures of performance are subjective” (Brewer 2006). Regardless, there is no consensus among policy analysts and public management scholars on how best to measure performance.

The difficulty in measuring the constructs of managerial quality and performance means that public management scholarship is still actively engaged in the task of understanding whether, to what degree, and under what conditions managerial quality promotes organizational performance (Boyne et al 2006). While scholars have been quite active on the problem of defining performance and in linking
specific management variables to it, work that addresses the latent managerial quality construct has received much less attention. Wolfe (1993) used employee perceptions of their managers’ abilities as a measure of managerial quality in Federal agencies and found a positive relationship between better ratings and better performance. Meier and O’Toole (2002) found that school superintendents with higher-than-predicted pay (a measure that they argue reflects their abilities) correspond to better measures of performance. Despite this good scholarly work, capturing the underlying construct of management quality and measuring its impact on organizational performance is still underexplored enough in the extant literature to merit further exploration. For instance, in their chapter on setting an agenda for future research in their 2006 edited volume on management and performance in public organizations, four pre-eminent scholars in the field, Boyne, Meier, O’Toole, and Walker, identified this methodological problem as one in need of more scholarly work. In another work, O’Toole and Meier (2011) call on scholars to investigate the managerial quality-organizational performance link in a variety of organizations, thereby extending the external validity of their own extensive work on Texas school districts.

How this Study Addresses Methodological Concerns

This study engages the existing literature by utilizing measures that mitigate many of the methodological problems that previous scholarship has identified. One strength of this study is its large-N design, which contributes to strong external validity. Much public management scholarship utilizes case-study designs, which by their nature are vulnerable to concerns that findings only apply to the particular case being studied. Leading scholars in the field have called for more large-N quantitative scholarship on the question of managerial quality and performance (Boyne et al 2006; Ingraham et al 2003; O’Toole and Meier 2011). By using a large-N design (and, moreover, a sample that is very nearly
exhaustive of the population of organizations being examined), this project has very strong external validity.

Another strength of this study is that the measures of managerial quality and performance it utilizes are all objective. While there is as yet no scholarly consensus on whether subjective or objective measures are superior, at least some major public management scholars (Ingraham et al 2003; O’Toole and Meier 2011) have expressed a preference for objective measures due to their being less distortion-prone and more conducive to quantitative analysis. Another advantage of the objective measurements is that the danger of common source bias is mitigated. While not exclusively a concern for subjective measures of managerial quality and performance, subjective measures are particularly vulnerable to common source bias (Andrews et al 2006; Brewer 2006). When the source of the managerial quality measurement and the source of the organizational performance measurement are the same, problems can arise. For instance, a survey that measures an employee’s perception of management and the employee’s perception of organizational performance can be simultaneously influenced by a confounding variables (e.g., their personal affection or dislike of their manager, their personal optimism or pessimism, etc). In this analysis, each measure of performance and managerial quality is assembled through a complicated system that is constructed from divergent areas of the organization’s functions and personnel. While not completely eliminating the danger of common source bias, this method does mitigate it.

The measures used in this study are also strong for, in a number of ways, mitigating concerns about construct validity. Contingency theory looms over this area of scholarship, presenting concerns about the comparability and the accuracy of measures of managerial quality and performance. According to contingency theory in public management, the performance impact of particular managerial behaviors or traits may depend on idiosyncratic variables such as the structure of the
organization, the task at hand, the specific goal being examined, aspects of the external environment, and other variables (Lawrence and Lorsch 1967). In other words, there is substantial evidence from public management scholarship that there is no “one size fits all” prescription for good management.

One way that this contingency issue complicates the study of managerial quality’s impact on performance is by presenting a challenge of comparability. Finding measures of performance that are commensurable across very different types of public organizations is difficult, given how different those organizations’ structures, activities and goals can be (O’Toole and Meier 2011). Finding an outcome measure that would allow one to compare, for example, the Army and the EPA, can be problematic; they do very different things and have very different substantive goals. Though many efforts at comparing disparate organizations’ performance have been attempted, such as the Program Assessment Rating Tool used by the U.S. government, or the Office of Management and Budget’s “agency scorecard” method, concerns about the possible over-aggregation of these methods are difficult to avoid (O’Toole and Meier 2011). Housing authorities, in contrast, while not perfectly identical to each other, are still relatively standardized in terms of programs and goals, thus mitigating concerns that the units of analysis might simply be fundamentally incomparable.

In addition to mitigating the problem of comparability, this study mitigates another construct validity concern raised by contingency theory: the accuracy of the managerial quality measure. If the impact of specific managerial traits and practices is contingent on the particulars of the organizations being studied, then possessing a firm grasp of the nuances of those organizations is necessary for designing accurate measures of managerial quality. What constitutes managerial quality in housing authorities may look very different than what constitutes managerial quality in a different type of organization. Any attempt to define managerial quality “a priori,” without specific knowledge of how these particular organizations functions would be speculative at best. The performance and managerial
quality measures in this study are defined by experts in the field (the Department of Housing and Urban Development). Rather than an *a priori*, speculative measure of managerial quality designed by those unfamiliar with the organization, this measure is designed by professionals who are steeped in the inner-workings, histories, and goals of housing agencies and are tailored to these organizations specifically. One should have confidence, therefore, that this measure of managerial quality accurately captures the construct.

Finally, this study promotes strong construct validity by utilizing multiple measures of performance. Much criticism has been levied at overly-naïve uses of single indicators as measures of performance, both for their use in academic literature and also their use by policymakers as tools for assessing, ranking, and comparing public organizations. For instance, many studies of school reform efforts use test scores on reading and math to assess performance, but that method has been criticized for ignoring educational gains in other areas, such as art and history (Smith 2003). Any single performance indicator is vulnerable to criticism that it does not reflect the totality of the potential goals that important constituencies may desire. In light of this concern, public management scholarship on performance has moved increasingly toward the use of multiple indicators of performance (Behn 2003; Kravchuck and Schack 1997; Ingraham et al 2003). Rather than simply utilizing test scores, Meier and O’Toole (2003), for instance, use 11 measures of performance in their analysis of superintendent quality on school performance. Other scholars use various taxonomic schemes to organize types or aspects of performance and utilize different measures for different types (e.g., inputs, outputs, outcomes; efficiency, equity, effectiveness) (Heinrich 2012). The analysis in this chapter uses 5 measures of performance across the two major program areas administered by housing authorities. While it does not capture all outcomes that could be of interest to any constituency, it does examine several, thus mitigating (though not completely eliminating) concerns about the robustness of the analysis to the specification of performance (Andrews et al 2006; Boyne et al 2006). Another advantage is to allow a
test of whether the influence of managerial quality on performance is contingent on the particular measure of performance chosen, or whether it is more broadly influential on organizational performance as a whole.

To summarize, this study utilizes measurements of managerial quality and performance that address many of the concerns raised by previous scholarship. These measures promote external validity (by encompassing almost the entire population of organizations being examined); reduce concerns about common source bias (by utilizing measures of that are objective and are drawn from different functional areas of the organization); promote comparability (by examining organizations with very similar structures, programs, and goals); promote accuracy (by using measures of quality and performance defined by experts in the field), and demonstrate robustness (by utilizing multiple measures of performance). These strengths position this study to substantially contribute to furthering the scholarship on the impact of managerial quality on organizational performance.

**Data and Variables**

The Public Housing Assessment System is HUD’s method of evaluating the performance of public housing authorities in their management of “public housing,” that is, government owned and operated housing units reserved for low-income households. The PHAS has four sub-components that score each housing authority on different indicators of quality: the physical condition of the properties, the operating financial health of the housing authority, the utilization of a capital fund, and the quality of the housing authority’s management practices. In turn, each of these four sub-components is itself determined by a variety of further sub-scores, but only the four main sub-component scores are publicly available.
The main independent variable in this analysis is the Management Assessment Sub Score (MASS), which captures the construct of management quality. HUD has determined the outcomes measured by this score (occupancy rate, rent and fee collection, and the ratio of monthly operating expenses to current and outstanding payments to vendors) to be indicative of quality management. HUD’s use of these particular outcomes as indicators of managerial quality follows the logic of much of the criticism in the past that has been levied at housing authority management from public housing’s critics and from several high-profile national reports (HUD 1998; National Commission on Severely Distressed Public Housing 1992). Failure to maintain high occupancy levels has been cited by experts as one of the central reasons housing projects suffered revenue problems, leading to maintenance and security problems. Lax rent collection from overdue tenants is also a commonly cited problem, with critics arguing that housing authority managers were “too soft” on tenants, or alternatively, simply negligent in enforcing rules. These are both variables over which managers have high levels of control. Quality managers should be able to quickly fill vacancies from the extensive waiting lists that all public housing authorities maintain, rather than letting vacant units sit idle. They also have control over how strict or lax are the collection of rent and fees. Finally, paying vendors on time every month is an indicator of management quality because it shows that organizations are living within their means, are not building up debt for operating expenses, and are acting responsibly toward third-party vendors. Responsible spending decisions are obviously an important managerial function. Certainly, the correspondence of this MASS variable to the construct of management quality is arguable. One could undoubtedly think of other indicators that would also reflect the quality of management in a housing authority, but using HUD’s score has the advantages of being commonly applicable across housing authorities, being constructed by subject-matter experts intimately involved in the day-to-day operations of the organizations being examined, and being directly tied to prominent critiques of how housing authorities have been managed in the past.
The maximum score on the management assessment measure is 25, and the mean score for housing authorities is 19.8 (79.2%), with a median of 21 (84%) . The standard deviation is 4.88. As can be seen in Figure 3.1, the number of housing authorities with high scores significantly outpaces the number of housing authorities with low scores. A sense of how HUD uses these scores may help provide an intuitive understanding of this distribution. HUD uses the overall PHAS to rank housing authorities as high performers, standard performers, substandard performers, or troubled. While this grading system is complex, one factor that automatically earns a classification of sub-standard performance is to fall below a 15 in management assessment score. To earn a high performer designation, the overall PHAS score must be 90%, with none of the four sub-component scores falling below 60%. 90% for the MASS score would be a score of 22.5. Thus, while it appears at first glance as if the distribution of scores in the sample is unusually high, the way HUD utilizes these scores is akin to a grading distribution. Much like with a grading distribution, there will be few zeros, and seemingly small absolute changes in scores can correspond to significant differences in assessed quality. This particular grading distribution appears to be centered around an average of 80%.
The six main dependent variables in this analysis are all measures of organizational performance. As explained in the preceding section above, these 6 outcome variables are designed to provide robustness to this analysis. Descriptive statistics for all of these outcome variables can be found in table 3-A. Much like the independent variable, all of the dependent variables are skewed to the high end of their range. Again, this makes sense if one conceives of them as a “grade,” where zero scores are very unlikely. One outcome variable is the “remainder” of the Public Housing Assessment Score, that is, the housing authority’s PHAS score minus its MASS score. In other words, this is the sum of each housing authority’s performance on the financial, physical, and capital fund scores. The maximum score is a 75. The next three dependent variables are the PHAS remainder variables disaggregated. Describing those in turn, the financial assessment sub score (FASS) is determined by three different indicators that measure organizational liquidity in relation to both operating expenses and outstanding debt service obligations. The maximum FASS score is 25 points. The physical inspection score (PASS) is calculated via a complex assessment process, but can be summarized as measuring the presence and severity of physical property problems in five categories: the surrounding property, building exteriors, building systems, common areas, and dwelling units. The maximum PASS score is 40. The capital fund program (CFP) score is determined by measuring how quickly a housing authority obligates capital fund monies and by the housing authority’s year-end occupancy rate, and its maximum is 10 points.

The next dependent variable measures housing authority performance on a separate program they administer, housing vouchers. One of these variables is the housing authority’s score on the Section Eight Management Assessment Program (SEMAP). SEMAP is a complex score composed of 14 sub-indicators that HUD uses to measure a housing authority’s performance in the administration of its Housing Choice Voucher program, commonly known as Section Eight. The maximum SEMAP score is 100 points. Table 3-B lists the sub indicators that constitute the SEMAP score, but only the final score is publicly available.
Table 3-A: Descriptive Statistics for Performance Indicator Variables

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public Housing Assessment System “Remainder”</td>
<td>0</td>
<td>75</td>
<td>66.66</td>
<td>7.92</td>
</tr>
<tr>
<td>Property Assessment Sub Score</td>
<td>15</td>
<td>40</td>
<td>34.47</td>
<td>4.25</td>
</tr>
<tr>
<td>Financial Assessment Sub Score</td>
<td>0</td>
<td>25</td>
<td>23.08</td>
<td>4.27</td>
</tr>
<tr>
<td>Capital Fund Program Score</td>
<td>5</td>
<td>10</td>
<td>9.35</td>
<td>1.04</td>
</tr>
<tr>
<td>Section Eight Management Assessment Program</td>
<td>25</td>
<td>100</td>
<td>89.2</td>
<td>10.4</td>
</tr>
</tbody>
</table>

Table 3-B: 14 Section Eight Management Assessment Program Sub Indicators

- Proper selection of applicants from the housing choice voucher waiting list
- Sound determination of reasonable rent for each unit leased
- Establishment of payment standards within the required range of the HUD fair market rent
- Accurate verification of family income
- Timely annual reexaminations of family income
- Correct calculation of the tenant share of the rent and the housing assistance payment
- Maintenance of a current schedule of allowances for tenant utility costs
- Ensure units comply with the housing quality standards before families enter into leases and PHAs enter into housing assistance contracts
- Timely annual housing quality inspections
- Performing of quality control inspections to ensure housing quality
- Ensure that landlords and tenants promptly correct housing quality deficiencies
- Ensure that all available housing choice vouchers are used
- Expand housing choice outside areas of poverty or minority concentration
- Enroll families in the family self-sufficiency (FSS) program as required and help FSS families achieve increases in employment income.

Source: HUD.gov

Control Variables and Constructs

Isolating the impact of management quality on performance requires one to control for variables that impact performance that are outside the control of management. Moreover, even where evidence exists on the effectiveness of certain management variables, the degree to which their effectiveness is conditioned on the external environment is largely unexplored. In other words, the external environment complicates investigation into managerial quality’s impact on performance by
being both a necessary control variable and also potentially an interaction variable. Broadly speaking, the central reason that the external environment complicates an analysis of managerial quality is by being a source of variation on two constructs, the difficulty of the task and the resources available to the managers (Mazmanian and Sabatier 1989; Hill and Hupe 2009; Walker and Andrews 2013). These variables are regularly identified as important in both public management and policy implementation scholarship.

Specifically in the analysis of management on the performance of public housing authorities, controlling for differences in the external environment in local jurisdictions and states is necessary. Factors that might vary from housing authority to housing authority that could impact performance would include the unique characteristics of the local client population, the local climate (and its impact on the physical plant of housing projects), the strength of the local housing market, the local economy, local political conditions, and local and state differences in budget allocations. To the extent possible in the available data, the proceeding analysis controls for these variations in the external environment. Still, compared to many local government entities, housing authorities have a relatively standardized funding process and standardized sets of programs and structures, which mitigates the danger of local idiosyncratic factors confounding the analysis of management quality.

This analysis uses three variables to capture task difficulty and complexity: the number of units managed by the housing authority, the poverty rate of the housing authority’s jurisdiction, and the percentage of a housing authority’s client households classified as “hard-to-house.” Households are classified as hard-to-house when they have any number of characteristics that make finding housing difficult. These include disability status, household size, advanced age, certain family structures (grandparents caring for young children or very large households), very low education, or members with a criminal record (Cunningham et al 2005). The hard-to-house rate is only available for housing
authorities that operate voucher programs, so it is only included in the models analyzing voucher program outcomes.

There may be state-level variation in variables that impact organizational performance. For instance, if there is state-level funding variation for housing authorities, one would expect greater resources to correspond to better performance. To capture these possible state-level effects, state dummy variables are included in the analysis.

**Data Sources**

The data for this analysis comes from the U.S. Census Bureau and from the U.S. Department of Housing and Urban Development. The PHAS scores in this analysis are an average of the most recent scores available from HUD, with up to 3 rounds of scores per housing authority, ranging from 2012-2014. The SEMAP score variable in this analysis is an average of each housing authority’s SEMAP score from 2005-2009. The unleashed voucher rate and the hard-to-house family rate are drawn from HUD’s Voucher Management System data. These variables represent the average of the two different point-in-time monthly measurements in 2010 and 2014. The total units under management for a housing authority is from HUD’s Executive Compensation Survey for 2013.

The poverty rate for each housing authority’s jurisdiction was taken from the Census’s American Community Survey and represents a 5 year average from 2009-2013. There is undoubtedly some measurement error in this variable, because housing authority jurisdiction is not one of the boundaries for which the Census calculates a poverty rate. The author attempted to match the housing authority jurisdiction to the appropriate Census boundary based on the jurisdictions cited in housing authorities’ names. When possible, housing authority websites were consulted for clarification on geographical boundaries. Where housing authorities appear to cover more than a two-county area, the poverty rate for the state was used.
Sample

This analysis has attempted to capture every housing authority in the United States. 3,069 of the nation’s 3,879 housing authorities are represented in four out of the six models in this analysis. All of the models that regress public housing program performance on public housing management utilize this sample. These 3,069 housing authorities represent every housing authority in the United States that manages any number of units of housing covered by an “annual contribution contract” with HUD, or about 1.08 million units. In other words, this analysis is exhaustive of the population of organizations being examined. The housing authorities in this sample manage 2.8 million out of the total 3.5 million (71.4%) subsidized living units currently under management by housing authorities in the public housing, project based Section 8, and Housing Choice Voucher Program. The housing authorities not included in this sample only operate a Housing Choice Voucher program (the tenant-based voucher) and do not manage any public housing. There are 768 such housing authorities, managing 700,000 living units. These housing authorities are excluded from the analysis because, without any “acc” units under management, they cannot and do not participate in the Public Housing Assessment System, from which the variables in the analysis are drawn. Additionally, 26 housing authorities are listed as having neither acc units nor Housing Choice Voucher program units. It is unclear what would cause that result. By capturing every housing authority in the U.S. that operates public housing, one can have maximum confidence in the external validity of the results of these four models.

Two of the models utilize a sample consisting of all housing authorities in the U.S. that operate both “acc” units and also Housing Choice Voucher program units. These models regress performance in the Housing Choice Voucher program on management quality as measured by the MASS score of the PHAS. In other words, they test whether the managerial quality measure from the Public Housing Assessment System correlates with better performance in a different program, the Housing Choice Voucher Program. That sample consists of 1,434 housing authorities, which, again, is exhaustive of the
population of housing authorities in the U.S. that manage both public housing and a tenant-based voucher program.

In both samples, a very small number of observations have been dropped due to missing data. Table 3-C provides a breakdown of the various samples being analyzed or excluded.

<table>
<thead>
<tr>
<th>Table 3-C: Sub-samples Used for Different Regression Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample used for Models 1-4</td>
</tr>
<tr>
<td>Housing Authorities that Manage Public Housing</td>
</tr>
<tr>
<td>N=3069</td>
</tr>
</tbody>
</table>

Housing authorities vary greatly in size. The smallest housing authority in the sample manages only 6 living units, and the largest (New York City) manages 274,606. The median number of units is 170, and even the 95th percentile in size is only 3,302. Most housing authorities therefore are of a moderate size, with a small number of very large organizations in the tails. Figure 3.2 shows the size distribution of housing authorities graphically.
Analytical Strategy

The basic model under investigation is a censored regression model (tobit) of housing authority performance on housing authority management quality, controlling for task complexity/difficulty and for any state-level variations that might impact performance. This basic model is depicted in equation 1, where $Y$ is the measure of organizational performance, $x_1$ is the measure of management quality, $X_2$ is a vector of indicators for task complexity, and $x_3$ is a state-level dummy.

\[
\text{Eq.1: } Y = \beta_0 + \beta_1 x_1 + \beta_2 X_2 + \beta_3 x_3 + \epsilon
\]

Tobit regression is utilized because the dependent variable for each specific model is a score, a data structure that has the characteristics of censored observations (the scores cannot fall below zero,
nor can they go above the maximum points). (Long 1997; Institute for Digital Research and Education ND). Robust standard errors are employed.

**Results**

In 5 out of 6 models, better management as measured by HUD’s own criteria for management quality (the MASS score within the Public Housing Assessment System) is correlated with better organizational performance, controlling for task complexity and state-level dummies. The PHAS “remainder” increases by 0.55 points as the MASS score increases by 1. A one standard deviation increase in MASS score would correspond to a predicted increase in PHAS remainder of 2.68 points (out of 75 possible points), or about 1/3\textsuperscript{rd} of a standard deviation. The result holds when examining the correlation of the MASS score with sub-components of the PHAS remainder. The predicted PASS score (physical assessment of public housing properties) increases by 0.23 for each 1-point increase in MASS score. A one standard deviation change in MASS score would correspond to a 1.12 increase in the predicted PASS score (out of a maximum score of 25, or a little more than one quarter of a standard deviation. The predicted FASS score (financial assessment score) increases by .47 with every 1 point increase in MASS score, with a one standard deviation increase in MASS corresponding to a 2.29 increase in predicted FASS, or about half of a standard deviation. The predicted CFP score (capital fund score) increases by .24 with each 1 point increase in MASS score. A one standard deviation increase in MASS corresponds to an increase in predicted CFP score of 1.17, which is a little more than one standard deviation. All of these coefficient estimates are highly statistically significant. While it is impossible to draw firm causal claims from this data, management quality in public housing administration is correlated with a substantively and statistically significant improvement in organizational performance, measured four different ways, and controlling for task complexity and state-level dummies.
Management quality in public housing administration is also correlated with stronger performance in a separate program administered by housing authorities, the Housing Choice Voucher program (or Section 8). An increase of 1 point in a housing authority’s MASS score is correlated with an
increase in predicted Section Eight Management Assessment Program (SEMAP) score of 0.53. This result is highly statistically significant. A one standard deviation change in MASS corresponds to a change in SEMAP of 2.45, which is just under ¼ of a standard deviation change. While clearly a less substantively significant correlation than that of MASS with the other PHAS scores, MASS’s correlation with Housing Choice Voucher performance is both statistically and substantively significant.

The correlation between managerial quality and performance in all of these models is clear and highly statistically significant. Substantively, the correlation is modest but larger than it may seem on first glance. Much as there is major substantive difference between a grade of 80% and a grade of 90% (a C+/B- vs a B+/A- in standard grade scales), a small magnitude change in performance score for these organizations probably has a greater substantive significance than it may seem. For a public housing authority with a management assessment score of 15, moving to a maximum score of 25 would correspond to a predicted change increase in PHAS remainder of 5.5 points, a “grade” increase of 7.3%. Remember that an overall PHAS score of 90% is required for a HUD designation of “high performer,” and a PHAS of 60% is required for a PHAS score of “standard performer.” Thus, while managerial quality is clearly not the only important predictor of performance, its correlation is both highly statistically significant and also substantively significant.

<table>
<thead>
<tr>
<th>Table 3-F</th>
<th>Tobit Regression of Section Eight Management Assessment Program on Management Assessment Sub Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
<td>Coefficient (standard error)</td>
</tr>
<tr>
<td>MASS</td>
<td>.53 (.07)**</td>
</tr>
<tr>
<td>Total Units Managed</td>
<td>0 (0)**</td>
</tr>
<tr>
<td>Jurisdiction Poverty Rate</td>
<td>-.07(.04)*</td>
</tr>
<tr>
<td>Hard-to-House Family Rate</td>
<td>-3.62 (2.386)</td>
</tr>
</tbody>
</table>

*P<.1 **P<.05 N=1397  (State dummies not displayed).
Limitations and Directions for Future Research

Caution should be used in interpreting the results of this analysis as causal. The data and analytical strategy here does not allow for strong causal claims. The correlation between managerial quality and performance is notable, but more research would be required to show that the correlation is causal in nature.

Construct validity is a concern for this analysis. HUD’s reasoning appears to be that a well-managed housing authority would exhibit the behaviors that are scored in their MASS indicator. They would be thorough in their collection of rent and fees; they would work to ensure that their units were occupied consistently; and they would pay their bills in a timely way and not rely on credit. This measure makes sense given the criticisms of housing authorities levied by the public and the government over the years. But certainly other ways of defining managerial quality would be feasible. The advantage of this measure is that it is applicable across all of housing authorities in the U.S. that operate a public housing program, and that it is the result of HUD’s collective expertise. The PHAS and its sub scores have undergone the full Federal rulemaking process and presumably represent the collective expertise of those in the public housing industry. One might argue that the substantive experts in the field are the ones best positioned to define what constitutes quality management in that field.

One limitation of this analysis is that for one of the outcome measures, the Capital Fund Program score, there is an endogeneity issue. One component of the CFP score is occupancy rate, which is also one of the components of the MASS score. This issue is also present then, though to a much lesser degree, in the regression of the “PHAS remainder” on the MASS score, because the CFP score is a component of the PHAS remainder. However, the correlation between MASS and both the physical and
financial health of the housing authorities does not suffer from that limitation. For these, the management quality indicator and the performance indicator are determined from distinct processes, reducing the danger of endogeneity and providing a safeguard against the tautology problem discussed above.

Another problem with this analysis is the time the data were gathered and the degree to which they represent averages across years differs from variable to variable. While this was a function of the data that were publicly available, the result is still that the timing of the variables does not match as precisely as one would like. A more precise match of scores by years would strengthen the analysis. Having enough data to conduct a panel analysis would be ideal, because organizations could serve as their own controls. However, the currently publicly available data does permit such an analysis, and it is unclear if data sufficient for that purpose will be available in the future.

A more direct measure of resources available to each housing authority would be ideal. HUD’s budget expenditures should be standardized, but there will still be some variation in the budgets available to different housing authorities stemming from state and even local budget expenditure decisions, as well as possibly private fundraising by the housing authority itself. Some of this variation will be captured with the state-level dummy variables included in this analysis. However, there is probably variation in available resources that are not captured in this analysis. I am unaware of any comprehensive source for data on housing authority budgets. If there is a correlation between higher size-adjusted budgets and the practices that constitute the MASS score, the coefficient for the MASS score in these regressions will be biased.

Conclusion

Limitations notwithstanding, this analysis has provided empirical support for the hypothesis that housing authority management quality can have a substantively significant impact on program
performance. While the data and analytical strategy do not allow for strong causal claims, there is a strong correlation between management quality and several different measures of program performance, controlling for several important confounding factors. For public management scholarship, this analysis contributes another piece of evidence that management does indeed matter. This conclusion is strengthened by a nearly exhaustive sample of the population of organizations under examination. This study encompasses housing authorities that are huge, tiny, and moderate. It encompasses housing authorities in liberal states and conservative states, in rural areas and urban, with minority and non-minority populations. By utilizing an exhaustive sample, external validity is maximized. One need not be concerned that this sample somehow misrepresents housing authorities.

This study takes up O’Toole and Meier’s call to extend managerial quality research beyond their extensive work on Texas School Districts to include other types of agencies (O’Toole and Meier 2011). Housing authorities, with their large population and relatively standardized programs and structures, are similar to schools, but they are also quite different in programs, goals, structures, and populations served. Insofar as they have similar properties to school districts, they possess many characteristics that make them valuable for public management scholarship. Most importantly, as described above, they are commensurable. But they are also different enough from schools that that this study is able to offer a valuable contribution into research on managerial quality’s impact on performance. Moreover, this analysis does so with a set of important organizations of tremendous substantive importance in U.S. social policy. Housing authorities manage some 3.5 million low income housing units. At an average household size of 2.1, that equates to approximately 7 million individuals served. Given the centrality of housing authorities to the lives of so many individuals and families, their scarcity as subjects of quantitative public management scholarship is notable. As this study demonstrates, whether housing authorities are managed well or poorly is predictive of whether they perform better or worse. In other words, in subsidized housing, management matters.
While it is impossible from this study to directly link housing authority management to the failures of poverty deconcentration policies discussed in earlier chapters, this study does provide support for investigating managerial quality’s impact in future research. If and when more and better data become available to allow for such an investigation, future scholarship should consider exploring it. Moreover, as policymakers and Federal bureaucrats consider reforms that may improve the outcomes of poverty deconcentration efforts, they should be careful to not ignore managerial improvements as a potential avenue of improvement.
Chapter Four:

Recruiting Non-Poor Households for Mixed-Income Housing: A Survey Experiment on the Effect of Sympathy-Based and Affective Racial Framing
Introduction

Despite decades of both scholarly attention and policy response to the problem of residential segregation, the United States remains highly segregated by race and class. Despite hopeful signs of improvement in neighborhood integration during the 1990s, improvements in neighborhood segregation slowed drastically since 2000 (El Nasser 2010). Of the many policy responses that have been lobbed at the problem of segregation, publicly subsidized mixed-income housing is among the newest, most promising, and yet least studied (Levy 2011; Levy et al 2013). While there is some promising research on the ability of mixed-income housing to improve low-income residents’ lives, the novelty of these programs and their heterogeneity in location, design, and implementation mean that they warrant much more research.

This paper focuses on the problem of recruiting nonpoor tenants for mixed-income housing, a step necessary for the causal logic of mixed-income housing to be successful. Very little research exists on the recruiting practices and recruiting success rates of mixed-income projects. This paper investigates whether individuals’ willingness to consider living in mixed-income housing is susceptible to framing effects on how the individuals view their potential neighbors. For managers and implementers of mixed-income housing, this paper provides insight into challenges they may face and opportunities they might exploit in designing marketing strategies for recruitment. Specifically, it provides insight into whether potential residents’ attitudes about the race and the life history of their potential low-income neighbors may present a challenge to recruitment, and if so, whether those attitudes might be altered through marketing strategies that attempt to “frame” attitudes toward low-income neighbors in a more favorable direction.

This is also a paper about communication frames and their use in policy design. By utilizing two different types of frames that nevertheless pull in the same direction simultaneously, this paper helps fill
a gap in the framing literature. Existing literature has focused on frames in isolation or in competition, but not on frames in interaction. It is not known how frames that pull in the same direction will interact. Might one frame “overpower” the other, such that only the more powerful frame will show an effect? Might they interact additively? Might they interact multiplicatively? These dynamics have not been studied. In this paper, the two frame types that are utilized (sympathy and race) invoke considerations that scholars have established as being important in determining individual attitudes toward the poor. These frames are interacted in ways that reinforce each other and in ways that do not, providing evidence on the dynamics of frame interaction.

A final contribution of this paper is to the use of framing theories in scholarship on policy design. While the impact of framing on policy participation and policy design has received some scholarly attention, particularly in the area of international relations and on some domestic issues like retirement savings decisions and participation in energy efficiency programs, these studies have more often drawn on psychological and economic framing literatures, not those from political science (Camerer and Kunreuther 1989; Mettler 2011). Framing in those contexts typically focuses on cognitive biases, such as anchoring effects and loss aversion, rather than the fundamental values and affective considerations that are more common in political science framing research. This paper investigates whether type of frames typically examined in political science can also be important in shaping the public’s decisions toward policy participation.

This paper utilizes a randomized survey-based experiment that presents one of four different stories (or a control group with no story) about a hypothetical low-income tenant of a mixed-income housing project and then asks respondents to rate their willingness to live in mixed-income housing on a scale from 0 to 100. Each story varies the implied race of the character in the narrative by using names
that are typically associated with a particular race. Each story also varies whether the character has a sympathetic or an unsympathetic story.

The findings suggest that frames that invoke sympathy can substantially increase expressed willingness to live in mixed-income housing. Unsympathetic frames appear to invoke responses that are closer to the control group attitude, indicating possibly that the default attitude toward low-income persons is more congruent with an unsympathetic view. On the question of whether the presence of a racial frame can condition the effect of a sympathy-based frame, the evidence indicates that affective racial frames had no detectable impact beyond the impact of the sympathy-based frame.

**Policy Background: Mixed-Income Housing as Poverty Deconcentration Policy**

Mixed-income housing projects are apartments or housing developments where some of the units are reserved for low-income households that pay reduced rents, while the rest are rented or sold at market rates. Often these projects are privately owned and operated but receive government subsidies. This relatively new approach to subsidized housing has been driven in part by a recognition that traditional public housing, where large tracts of housing are reserved exclusively for low-income households, results in high concentrations of poverty. Driven by a large body of social science that links neighborhood characteristics to life outcomes, Federal subsidized housing policy in the United States has explicitly adopted the goal of poverty deconcentration. The scholarship behind these policies is commonly known by the shorthand term of “neighborhood effects.”

Two of the key variables in the neighborhood effects literature are the degree of poverty concentration and racial segregation that residents face. In 2000, 3.5 million poor people across the United States lived in neighborhoods with poverty rates in excess of 40 percent, and many of these poor neighborhoods are also highly segregated by race (Hirsch 1983; Massey and Denton 1993; Jargowsky 1997, 2003; Dreier et al 2004). A growing social science literature suggests that the combination of racial
segregation and poverty concentration has a variety of detrimental effect on the residents of these areas, both in terms of their current well-being and their future opportunities (Wilson, 1987, 1996; Jencks and Mayer, 1990; Brooks-Gunn, et al. 1993, 1997). Extremely poor, segregated neighborhoods are often home to higher crime rates, higher rates of juvenile delinquency, underperforming public schools, poor housing and health conditions, limited access to public services, and limited access to job opportunities (Kneebone and Berube 2008).

While the Department of Housing and Urban Development has pursued poverty deconcentration through the Housing Choice Voucher Program (formerly known as Section 8) , the results have been largely disappointing (Bolinger and Xu 2013; Briggs et al 2010; Goering et al 2003). The large randomized longitudinal field experiment known as Moving to Opportunity (MTO), where some public housing residents were randomly assigned to receive a voucher, found almost no improvements on mover households’ economic or behavioral outcomes. Why the policy failed to produce the expected results is not entirely understood. One thing that is clear is that over time, voucher recipients, both within the MTO experiment and in general, tend to use their voucher in poor neighborhoods. They do this for a wide variety of reasons, including not least landlord discrimination against voucher holders (Beck 1996; Popkin 1999; Briggs et al 2010; Daniel 2009; Freeman 2013). Regardless of the cause, housing vouchers appear to be fairly ineffective at deconcentrating poverty, at least as they are currently implemented.

HUD has also pursued a different strategy for deconcentrating poverty: mixed-income housing. HUD’s major mixed-income housing program is called HOPE VI (Housing Opportunities for People Everywhere). HOPE VI is a major HUD plan to revitalize the worst public housing projects into mixed-income multi-family dwellings. Over the course of 15 years HOPE VI grants were used to demolish 96,200 public housing units and produce 107,800 new or renovated public housing units, of which 56,800 were to be affordable to the lowest-income households (HUD 2010). Residents who were not re-
housed in these mixed-income developments were granted Section 8 vouchers to use in the private rental market.

While there is yet no scholarly consensus of the efficacy of these projects, there are both reasons to be hopeful and reasons to be skeptical. On the hopeful side, some of the early longitudinal studies of residents show some improvements for low-income households (Popkin et al. 2009). Conversely, some evidence suggests that the programs are not working as intended. The benefits to low-income households of living in mixed-income developments are predicated on certain assumptions and theoretical foundations. One of these foundations is that idea that if low and moderate income households live in close physical proximity, they will forge social bonds, thus improving low-income households’ social capital. This might help low-income families in a variety of ways, including for instance leading to a job reference for the low-income household. Emerging research is finding that this vision of an egalitarian, tight-knit “urban village”, with a more equitable distribution of social capital, may have been unrealistic; evidence is showing that meaningful inter-class interactions in mixed-income developments have been more elusive than initially hoped, with some studies even finding antagonistic relationships developing between different class groups (Kleit 2005; Chaskin and Joseph 2009).

**Literature Review: Recruitment of Nonpoor and White Households for Mixed-Income Housing**

While the overall efficacy of mixed-income projects is still a major research area, a more proximate concern for the implementation of mixed-income projects is whether projects can successfully recruit nonpoor and white households. Very few studies have addressed this issue. What little evidence exists suggests that mixed-income developments tend to be attractive to middle-income families without children and that developments tend to be most successful in tight housing markets (Varady et al. 2005). However, those variables are largely beyond the control of policy implementers.
This study is the first to examine whether variables that are susceptible to manipulation by managers might improve successful recruitment for mixed income housing.

Varady et al in their four-site case study of mixed-income projects note that none of the sites explicitly marketed to families with children, resulting in three out of four sites being unsuccessful at recruiting those types of families. It is notable, though, that in all of the sites they studied, recruitment of middle-income households without children was successful. However, because the empirical research so far has been based on case-study methodology, the external validity of this finding is unverifiable. We simply do not know yet how difficult or easy the task of recruiting for mixed-income projects might be for the universe of all mixed-income projects in the U.S.

Regarding the specific problem of recruiting families with children, some studies argue that middle-class perception of poor school options in mixed-income development neighborhoods is driving away families with children (Varady et al 2005; Joseph and Feldman 2009; Levy et al 2013). An inability to attract non-poor households with children eliminates several of the hypothesized mechanisms through which poverty deconcentration might improve poor households’ lives. Lack of non-poor children, for instance, would eliminate the possibility of positive youth peer effects, both socially and in community schools.

Even in projects where all the market-rate units are rented or sold successfully and vacancy rates are low, the range of non-poor incomes present can range drastically. According to one of the earliest large studies on mixed-income developments and their financial viability and social policy goals, “we know virtually nothing about the optimal combination of incomes needed for successful mixed-income housing,” and little progress on that question has been achieved since (Brophy and Smith 1997; Levy et al 2013).
Finally, the racial makeup of non-poor households in mixed-income developments might either contribute to or reduce the level of racial segregation that residents face. There is no empirical scholarship on the issue of race in recruitment for mixed-income housing. However, there is ample scholarship on the role of race in determining the racial makeup of neighborhoods. Historically, racial prejudice has been documented as a key element in the post-war white flight from urban neighborhoods and in the creation of racially segregated urban ghettos whose borders have often persisted into the present (Hirsch 1983; Massey and Denton 1993; Dreier et al 2004). Racial neighborhood segregation, while intersecting with economic neighborhood segregation, has been shown to be harmful in its own right (Massey and Denton 1993; Dreier et al 2004), and racial segregation persists even when controlling for income differences. Moreover, while on average, people of all racial backgrounds express support for living in racially diverse neighborhoods, most still prefer a majority of the neighborhood be constituted by the in-group race (Charles 2000). This in-group preference is particularly strong for whites (Charles 2000). There is strong empirical evidence to suggest that mixed-income housing projects will face a challenge in promoting racial diversity if mixed-income projects fail to promote racial diversity and that this will weaken their ability to improve low-income minority residents’ lives.

To summarize, almost nothing is known about how potential nonpoor and white residents of mixed-income housing make decisions about participating in those projects or whether those decisions can be influenced. This study examines whether expressed attitudes toward living in mixed-income projects can be influenced by manipulating how potential residents’ think and feel about their potential low-income neighbors. These findings can inform how project managers design marketing strategies to recruit nonpoor households.
Existing Literature on Framing Effects in Political Science and Policy Design

Framing is the use of communication strategies to highlight mental considerations that favor one side of an argument or decision while downplaying elements that would favor the other side. This paper utilizes experimental design to detect framing effects on expressed attitudes toward participation in a voluntary policy, and in the process it contributes two theoretically novel elements. First, it examines the interactive effects of two different types of frames simultaneously. Frames can differ markedly in the type of considerations they are attempting to make accessible. Some frames appeal to core values, such as fairness, equality, or family. Some frames appeal to affective feelings toward individuals or groups. Some frames appeal to self-interest. The types of frames that can be utilized are as numerous and diverse as the universe of considerations available to individuals (Meffert et al 2004). Most existing research focuses on one type of frame at a time. While newer research has been turning to the question of competing frames (Chong and Druckman 2007), little has been published on the interactive dynamics of different frame types.

One of the frame types in this study is what I will call “sympathy-based,” which is a frame that invokes fundamental values like equality, deservingness, and compassion. Scholars of many stripes, from historians to sociologists to public opinion researchers to scholars of social welfare policy have long identified these elements as being powerful predictors of attitudes toward anti-poverty programs, finding that Americans’ support tends to rise and fall depending on whether the poor are seen as “deserving” or “undeserving” (e.g., Levine 1988; Iceland 2003; DeSante 2013). Typically, the poor are seen as more deserving if their poverty is presumed to be the result of outside circumstances and less deserving if their poverty is seen more as the result of their own poor choices. The intent of the sympathy-based frame in this study is to frame the experimental subjects’ perception of their potential low-income neighbors as being “the deserving poor.” The other type of frame utilized in this study is affective; it taps into respondents’ affective attitudes toward whites vs African Americans, which also
has been found to impact support for social programs (DeSante 2013). How these two frames will interact is unknown. Framing theory has yet to offer a theory for predicting the interaction of different frame types. In this paper, I hypothesize that the two frames will combine to amplify each other’s impacts.

The second theoretically novel contribution of this paper is to bring to bear framing concepts developed in political science to the topic of policy design. Framing as a subject of study in social science was popularized initially by the work of Daniel Kahneman and Amos Tversky. Their work sought to explain the existence of seemingly irrational behavior in which decisions with the exact same risk/reward profile would elicit different responses depending on how they were worded (Kahneman and Tversky 1979). This “prospect theory” as they called it would lead to investigation of other “cognitive biases” and would become a major theoretical contribution to many fields, such as marketing and behavioral economics. Policy scholars have utilized framing theory in both descriptive and prescriptive ways to understand and influence decision-making points in the policy process, including decisions about voluntary participation or nonparticipation in programs. Framing in those contexts typically focuses on cognitive biases, such as anchoring effects and loss aversion, rather than the fundamental values and affective considerations that are more common in political science framing research.

In political science, framing has taken on a somewhat broader meaning. It typically denotes the use of political communication strategies that target certain considerations in a person’s psyche to be made more accessible, thus influencing a person’s political attitudes and decisions (Sniderman and Theriault 2004; Chong and Druckman 2007). Many empirical studies confirm the existence of framing effects in political communication (e.g., Jacoby 2000; Nelson et al. 1997; Chong and Druckman 2007). Typically, these studies have been implemented as experimental designs that administer a framing
“treatment” and look for a resulting change in attitudes, either on pre-post measures, or cross-sectionally between experimental and control groups. These experiments consistently find that frames in isolation can impact expressed attitudes, though they are often criticized for poor external validity.

With notable exceptions (see Achen 1975), most framing scholars with a political science orientation argue that framing effects are possible because most Americans do not possess well-ordered ideological mental frameworks through which they have formed conscious ex ante attitudes on most specific political questions (Converse 1964; Zaller 1992; Zaller and Feldman 1992). Rather, most Americans possess a mental “soup” of considerations regarding political values, issues, affective attitudes toward groups, haphazardly gathered political information, and other elements that might inform a political attitude if marshaled to do so, but which usually lie relatively dormant and unorganized. Framing attempts to bring these favorable considerations to the forefront of a person’s attention, making them the “low hanging fruit” that a person grabs first when reaching for considerations to inform an opinion or political decision. Framing therefore can be used to influence the public to adopt a particular political opinion or to make some particular political decision.

This study fills a gap in the public policy literature by utilizing a conceptualization of framing that is more common in political science and applying it to the issue of decision making toward voluntary policy participation. Affective and sympathy-based frames have been shown to be powerful influences on expressed public opinion; it is reasonable to believe that they very well may be powerful influences on policy participation as well. If so, policy scholars will have reason to expand their conception of framing to include notions developed in political science.

**Hypotheses**

The theoretical background on framing and on recruitment for mixed-income projects and on framing effects suggests the following hypotheses:
Hypothesis 1: Treatment groups that are exposed to sympathetic and unsympathetic frames about potential low-income neighbors in a mixed-income project will respectively display increased and decreased willingness to live in mixed-income housing compared to respondents who receive no frame.

Hypothesis 2: Treatment groups that are exposed to frames that are intended to invoke the image of an African-American neighbor will display less willingness to live in mixed-income housing than treatment groups that are exposed to frames that are intended to invoke images of a white neighbor, and also to the control group which is not exposed to any frame.

Hypothesis 3: Treatment groups that are exposed to two frames simultaneously that pull respondents in the same attitudinal direction will display more support or nonsupport (depending on the direction of the pull) than groups exposed to two frames simultaneously that pull in opposite directions.

Method
Research Design

This study utilizes a survey experiment that randomly assigns respondents to one of four treatment groups or to a control group. The treatments in this experiment are fictional anecdotes about a low-income woman who decides to move into a mixed-income housing project (see appendix 4-A for the wording of each treatment). There is also a control group in the experiment that was not exposed to a story. Each respondent in each group was first presented with a brief definition of mixed-income housing before being presented with the treatment. Each respondent in a treatment group is then randomly assigned to receive one story that varies in two different ways. One way is to vary whether the narrative is sympathetic or unsympathetic, and the other is to vary the name of the character to imply whether the person is African American or white.

Varying the names of unseen individuals to invoke and measure implicit racial bias is an increasingly common technique in social science research (Bertrand and Mullainathan 2004; Widner and
Chicoine 2011) and has been used specifically to examine rental discrimination by landlords (Carpusor and Loges 2006; Ahmed and Hammerstedt 2008). This technique invokes racial attitudes that may be “unconscious” and dampens any social desirability bias that might arise from explicitly mentioning a subject’s race. I draw my two names, Lakisha and Jill, from the pioneering work in this area by Bertrand and Mullainathan (2003). They generated a list of common black and white names based on national data on baby name frequencies, and they verified the reliability of the racial connotations of the names with pre-testing. The names in this study were borrowed from their list of black and white names (i.e., names that in the absence of other information typically invoke black or white images in survey respondents’ minds).

For the outcome variable, respondents were told to imagine a scenario where they (and their family, if applicable) would be moving to an apartment in the near future. They were told to imagine that they found a desirable apartment in a decent neighborhood and that the apartment was part of a mixed-income building. They were then asked to rate their willingness to live in this mixed-income project on a scale from 0 to 100. The exact text of the outcome variable questions can be found in Appendix 4-A.

Each respondent was also asked a battery of questions concerning their demographic background and political leanings.

**Data, Sampling, and Randomization**

Amazon Mechanical Turk was used to construct the sample. MTurk (as it is commonly known) is a service wherein individuals sign up to participate in online tasks in exchange for monetary compensation. It is becoming increasingly popular in social science as a way to gather data quickly and inexpensively. Research into the representativeness of MTurk samples has shown that it is more representative of the general U.S. population than the usual population used for experiments, college
students, and it compares reasonably well with nationally representative samples, such as the GSS and ANES (Burhmester et al 2011; Paolacci et al 2010; Berinsky et al 2012). Compared to national probability samples, Mechanical Turk’s subjects are more likely to be female and have higher education levels, neither of which is particularly problematic for the external validity of my research question.

The survey was limited to U.S. residents. Efforts were made to avoid retaking of the survey, as were attention checks to ensure thoughtful answers. The final sample for this analysis after exclusions for possible repeats and attention check failure was N=1,262 from an initial sample of 1,544 respondents. As such, this can be conceived as a treatment-on-the-treated analysis. Table 4-A displays the sample sizes of the various treatment groups and the control group.

Randomization was conducted via the survey software (Qualtrics), with individuals randomly assigned to one of four treatment groups or the control group. A brief description of the survey and the compensation amount were used to recruit existing Mturk “workers.” The compensation was $0.50, and survey took approximately 10 minutes to complete. The survey was administered June 5-6, 2015.

Table 4-A: Sample Sizes of Experimental Groups

<table>
<thead>
<tr>
<th>Experimental Group</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group</td>
<td>253</td>
</tr>
<tr>
<td>Sympathetic Story/White Character</td>
<td>254</td>
</tr>
<tr>
<td>Sympathetic Story/Black Character</td>
<td>252</td>
</tr>
<tr>
<td>Unsympathetic Story/White Character</td>
<td>247</td>
</tr>
<tr>
<td>Unsympathetic Story/Black Character</td>
<td>256</td>
</tr>
</tbody>
</table>

Analytical Strategy

The analytical strategies employed are regression-adjusted treatment effects and pairwise difference of means. Due to randomization, any statistically significant differences between the experimental groups can be interpreted as the causal impact of the treatment on the outcome variable, i.e., as a treatment effect. While it is appropriate to use a simple difference of means test, such as a T
test, a better strategy is to use a regression model that includes other relevant covariates. This improves the precision of the treatment effect estimates. However, regression only allows for each experimental group to be compared to one reference group at a time. Typically this would be the control group. In this experiment, because there are multiple treatment groups, there are other interesting comparisons to examine between different treatment groups. To test the differences between different treatment groups, this analysis uses two techniques. One technique is a pair-wise difference of means test, which is a t-test adjusted for multiple hypothesis being tested simultaneously. The other is to use regression-adjusted treatment effects, using a group other than the control group as the reference category.

For each model tested, two types of regression are used. One is OLS, and the other is a censored regression model (tobit). The data structure for the outcome variable is analogous to a percentage, though it is not exactly the same (see the wording of the question in Appendix 4-A) Like a percentage, the range of the variable is from 0 to 100. For that type of outcome, OLS can be conceived as the linear probability model (LPM). The LPM can be appropriate if most of the data fall in the center of the distribution. This data is not a bad candidate for the LPM; the mean of the outcome variable is almost exactly in the center of the range, which would favor the LPM, but the large standard deviation means that a significant number of responses will be at in the tails of the distribution, where the LPM can be particularly misleading. In another sense, the outcome variable can be better conceived as a continuous variable that is truncated at both the lower and upper ends, or at 0 and 100 respectively. Truncated data is best analyzed with a tobit model. The two methods produce similar results. Only the tobit results will be displayed, but one can view the OLS results in Appendix 4-B.

The other covariates in the model both improve the precision of the treatment effect estimates and also provide insight on factors that correlate with greater or less support for living in mixed income
housing. Because randomization ensures that the treatment variables are statistically independent from
the other covariates in the model, their inclusion or exclusion does not impact the unbiasedness of the
estimates of treatment effects, but it does add precision to those estimates. These variables also offer
some interesting insights of their own accord on the factors that correlate with more or less support for
living in mixed-income housing, holding the treatment effects constant. The covariates in the model are
age, sex, race, education, political ideology, partisanship, parental status, housing status (rent vs own),
and household income (adjusted for household size).

As another robustness check, the analysis presented here was also conducted with a sample
that only included white respondents, due to the possibility that the treatment effects would vary
depending on race, and due to the theoretical importance of racial integration in poverty
deconcentration efforts. The results from the whites-only sample are extremely similar to the results
from the full sample in terms of the magnitudes, directions, and statistical significance of the estimated
coefficients. Those results are available upon request.

Results
Variables that Correlate with Willingness to Live in Mixed-Income Housing

The covariates in the regression model offer some interesting insights on the correlates of
willingness to live in mixed-income housing, controlling for exposure to the treatment. These results are
displayed in Table 4-A. The treatment effects are excluded from the table for ease of presentation, but
were included in the model. Age, sex, race, education, partisanship, and current housing status all had
small coefficient estimates that were also not statistically significant. Household income per capita (that
is, adjusted for household size) had a statistically significant correlation with willingness to live in mixed-
income housing, but this coefficient was relatively small. A $10,000.00 increase in household income
per capita (which would be a $20,000 increase for a household of two, a $30,000 increase for a
household of three, a $40,000 increase for a household of four, etc) would correspond to a reduction of
3 points in a respondent’s willingness to live in mixed income housing. Wealthier households are less likely to live in mixed-income housing, but the effect of income is not as drastic as one might expect.

Political ideology and parental status are the two variables whose coefficients are both statistically significant and of a notable magnitude. Having children correlates with nearly a 10 point reduction in willingness to live in mixed income housing, and being one point more conservative on the ideology scale correlates with nearly an 8 point reduction. Further analysis of the political ideology variable shows that much of the effect is driven by the high willingness of liberals to live in mixed-income housing compared to moderates, whereas conservatism correlates with a reduction in willingness, but of a smaller magnitude. Table 4-B displays these results, which were reached using a regression model where ideology is treated as a categorical variable, rather than ordinal, which more clearly reveals the distribution of support for living in mixed income housing across each ideological category.

These correlations suggest that, while mixed-income projects will face something of a challenge in recruiting wealthier households, the bigger challenge will be in recruiting families with children. They will also face a challenge in recruiting moderates and especially conservatives, though it is doubtful that diversity in political ideology would be an explicit goal of most projects. Conversely, the easiest route to creating the conditions for mixed-income housing to work would be to recruit liberal, educated households without children. On the other hand, the presence of children is a necessary precondition for some of the theorized benefits of mixed-income housing to accrue to low-income residents.

Table 4-A: Blocking Variables from Tobit Regression

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient Estimate and Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>-.03 (0.1)</td>
</tr>
<tr>
<td>Sex</td>
<td>2.8 (2.3)</td>
</tr>
<tr>
<td>Race (white vs nonwhite)</td>
<td>.37 (2.57)</td>
</tr>
<tr>
<td>Education</td>
<td>-1.1 (0.9)</td>
</tr>
<tr>
<td>Partisanship (7 point scale)</td>
<td>-1.3 (1.0)</td>
</tr>
<tr>
<td>Political Ideology (5 point scale)</td>
<td>-7.9 (1.7)**</td>
</tr>
<tr>
<td>Has Children</td>
<td>-9.8 (2.6)**</td>
</tr>
<tr>
<td>Household Income per Capita (in thousands)</td>
<td>-0.3 (.07)**</td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Current housing status (rent vs own)</td>
<td>-3.6 (2.2)</td>
</tr>
<tr>
<td>*α&lt;.1  **α&lt;.05</td>
<td></td>
</tr>
</tbody>
</table>

(Treatment effects not displayed).

**Table 4-A: Willingness to Live in Mixed Income Housing by Ideological Category**

<table>
<thead>
<tr>
<th>Ideological Category</th>
<th>Mean Value of Dep. Variable</th>
<th>Regression estimate and SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Liberal</td>
<td>64.6</td>
<td>23.5 (3.92)**</td>
</tr>
<tr>
<td>Liberal</td>
<td>51</td>
<td>10.5 (3.1)**</td>
</tr>
<tr>
<td>Moderate</td>
<td>41.3</td>
<td>Reference Category</td>
</tr>
<tr>
<td>Conservative</td>
<td>35.8</td>
<td>2.0 (4.2)</td>
</tr>
<tr>
<td>Very Conservative</td>
<td>31</td>
<td>-3.8 (6.9)</td>
</tr>
</tbody>
</table>

**Treatment Effects**

The mean response for the outcome variable for each treatment group is displayed in Figure 4.1. The mean from the control group was 43. Again, that is the mean self-reported willingness to live in mixed-income housing on a scale of 0 to 100. The control group was given the definition of mixed-income housing but was not exposed to a narrative. Compared to the control group, respondents who were exposed to a sympathetic story about a potential low-income neighbor expressed much higher willingness to live in mixed-income housing on average. These treatment effects are both substantively and statistically significant. The sympathetic story featuring a white character raised expressed support by 18-22 points (OLS and Tobit estimates, respectively) compared to the control group. The sympathetic story featuring a black character raised expressed willingness to live in mixed-income housing by 15-19 points (OLS and Tobit). These estimates are displayed in Table 4-C. The other covariates in the model are excluded for ease of presentation but can be seen in table 4-A.
Though the sympathetic frame clearly influenced willingness to live in mixed-income housing, the racial frame did not condition this impact. Table 4-A displays the pairwise comparison of means between different experimental groups and the standard error of these estimates. The p-values are adjusted using Tukey’s method to account for multiple T tests being conducted simultaneously. While support increased more when the story was about a white character than a black character, the difference between mean support in the sympathetic/white treatment group and mean support in the sympathetic/black treatment group is small and not statistically significant. At least in the presence of a sympathetic frame, the frame invoking affective attitude toward white vs African American neighbors had no statistically significant impact.

Now we will turn to the two treatment groups that were exposed to an unsympathetic story. The unsympathetic story did decrease expressed willingness to live in mixed-income housing compared to the control group, but not by nearly as much as the sympathetic story increased it. This result is
particularly surprising given the *strength* of the unsympathetic frame. The character described in the narrative was very unsympathetic. The unsympathetic white character decreased mean support by only 4-5 points (OLS and Tobit) compared to the control group, and in neither model was this difference statistically significant. The unsympathetic story with a black character decreased mean support by 7-9 points (OLS-Tobit) and was statistically significant at alpha=.05 in both the OLS and Tobit models. In either case, it is clear that the overall impact of the unsympathetic story was weaker in the negative direction than the sympathetic story was in the positive direction, though it still had an impact in one of the two unsympathetic story treatment groups. In both of them, though, the unsympathetic story elicited support levels that were much closer to the default control group attitude.

Turning to the interaction of the racial frame with the unsympathetic frame, the difference between the unsympathetic black character group and the control group is statistically significant, whereas the difference the unsympathetic white character group and the control group is not. However, the difference between the white character group and the black character group is not statistically significant.

As a robustness check on the effect of race as a frame, I conducted OLS and Tobit regressions with different reference categories. This method will add precision to the estimates of treatment group mean differences. These results are shown in Table 4-E. The results confirm that the difference between stories that have a black character and stories that have a white character is not statistically significant.

<table>
<thead>
<tr>
<th>Treatment Group</th>
<th>Tobit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sympathetic Story/White Character</td>
<td>22.1 (3.3)**</td>
</tr>
<tr>
<td>Sympathetic Story/Black Character</td>
<td>19.0 (3.3)**</td>
</tr>
</tbody>
</table>

Table 4-C: Regression Adjusted Treatment Effects (compared to control group)³

³ Models were also tested to check for interaction effects between treatment group and ideology and treatment group and parental status. None of the interaction terms were statistically significant.
Unsympathetic Story/White Character   -5.2 (3.3)
Unsympathetic story/Black Character   -9.4(3.3)**
*α<0.1  **α<.05

Table 4-D:  Pairwise Comparison of Means between Experimental Groups

<table>
<thead>
<tr>
<th>Experimental Group Comparison</th>
<th>Difference of means and Standard Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>White Sympathetic vs Black Sympathetic</td>
<td>.87 (2.8)</td>
</tr>
<tr>
<td>White Unsympathetic vs Black Unsympathetic</td>
<td>2.46 (2.8)</td>
</tr>
</tbody>
</table>
*α<0.1  **α<.05

Table 4-E:  Regression Adjusted Treatment Effects of Black Character Vs White Character

<table>
<thead>
<tr>
<th>Experimental Group Comparison</th>
<th>Tobit</th>
</tr>
</thead>
<tbody>
<tr>
<td>White Sympathetic vs Black Sympathetic</td>
<td>3.1 (3.4)</td>
</tr>
<tr>
<td>White Unsympathetic vs Black Unsympathetic</td>
<td>4.3 (3.4)</td>
</tr>
</tbody>
</table>
*α<0.1  **α<.05

Conclusion

The findings from this survey experiment suggest several conclusions. One is that expressed willingness to live in mixed income housing is susceptible to framing, both in the positive and the negative direction. Mixed-income housing is a policy that requires voluntary participation by non-poor households as a precondition for achieving many of its purported goals. The findings here suggest that managers of mixed-income projects should, when implementing marketing strategies, consider how non-poor households will view their potential low-income neighbors. While it is doubtful that a project would intentionally paint an unsympathetic picture of its low-income participants, the findings here suggest that the default attitude in the public is more congruent with an unsympathetic picture than a sympathetic picture. Marketers must contend with this default attitude, and sympathetic framing may be a useful strategy to do so.
The findings also suggest that recruiting in more conservative areas and recruiting families with children and that are wealthier will be more difficult. However, no interaction effects were found between the treatment categories and conservatism, parental status, or income. The lack of a significant interaction effects suggests that households that are conservative, have children, and have higher incomes are just as susceptible to framing as households without children. This finding is notable for potential marketers, who may want to consider utilizing sympathetic framing of potential neighbors in marketing strategies for mixed-income projects.

Another major conclusion is that affective racial frames had no impact on stated willingness to live in mixed-income housing when combined with sympathy-based frames. It is impossible to distinguish the effects of the racial frame from chance at conventional levels in either the sympathetic case or the unsympathetic case. While it is possible that using a black-sounding name was too weak of a frame to “tap” latent racial prejudices in the respondents, it is a method that has been used successfully to detect racial prejudice in other contexts. Two other explanations are possible. One is that the sympathy-based frames simply overwhelmed the affective racial frame. That is, it is possible that a racial frame would be more impactful in isolation. Further research on interactive dynamics of multiple frame types would be appropriate to help clarify the causal mechanisms at play.

It is also possible that respondents simply did not harbor much racial prejudice, at least not for the issue of considering who their future neighbors might be. If so, the result is surprising, given existing evidence of the impact of racial prejudice on housing decisions. Designers and managers of mixed-income projects can interpret this as “good news,” if their goal is to recruit a racially and economically diverse population. While they must contend with pre-existing attitudes toward the poor in general, they do not have to contend with the complicating factor of racial bias. One would presume that one type of bias is easier to overcome than two. Further research should also investigate whether the
apparent lack of racial bias found in this experiment is limited to support for living in mixed-income housing, or whether it applies more broadly to other sorts of policy preferences and attitudes about policy participation.

Finally, this research suggests that policy analysts should take account of the full spectrum of frame “types” when studying policy participation. In this analysis, frames invoking core values were shown to be potential ways to induce policy participation. While research on policy participation has drawn usefully from framing concepts developed in psychology and economics, scholars should consider augmenting these “cognitive bias” types of frames with frame types typically examined in political science.
Appendix 4-A

Survey Experiment Questions

Informational Paragraph Presented to All Respondents

Before you answer the following question, I need to make sure you know what mixed-income housing means. In mixed-income housing, some of the apartments in a building are reserved for low-income households only. The rent on these units will be lower than others in the building, even though they may be the same in size and quality. For the rest of the units in the building, the landlord can choose who to rent to and how much to charge, just as with any other apartment building.

Control Group Question

Presume that you and (if applicable) your family needs to move in the coming year, and presume that you choose to rent an apartment rather than buy a house. Presume that you find an apartment you like and can afford in a decent neighborhood. Next you find out that the building is a mixed-income development and that 30 percent of the units in the building will be reserved for low-income households who will pay 50 percent less for rent than the market-rate households like yours will pay.

Please rate how willing you would be to rent this apartment, with 0 representing complete reluctance and 100 representing complete enthusiasm.

(Respondents used a horizontal sliding scale to answer the question).

Treatment Group 1: Sympathetic Story/Black Character Name

Please read the following carefully.

Lakisha is a low-income widowed mother with two children. She married her husband just out of high school, but when the children were young, he was killed in a car accident by a drunk driver. Lakisha did well in high school and has pursued education from the local community college to improve her skills, but family and work obligations have kept her from graduating. Despite the fact that she works two jobs, she has trouble making rent some months. She and her children currently reside in an apartment in a neighborhood where the crime rates are high and a lot of the students end up dropping out of high school. She would like to move to a better neighborhood if she could find one where she could afford a decent apartment for her family.

Lakisha recently learned about an apartment available in a mixed-income housing development. Remember, this means that some of the apartments in the building will be reserved for low-income families who will pay less in rent, while the rest of the apartments will be rented at the normal market rate. If Lakisha moves, her children will live in the same building as middle-income families and will attend the same public schools.

(Respondents asked to click a box to indicate they have read the story).

Having heard Lakisha’s story, now presume that you and (if applicable) your family needs to move in the coming year, and presume that you choose to rent an apartment rather than buy a house. Presume that you find an apartment you like and can afford in a decent neighborhood. Next you find out that the building is a mixed-income development and that 30 percent of the units in the building will be reserved for low-income households like Lakisha’s who will pay 50 percent less for rent than the market-rate households like yours will have to pay.

Please rate how willing you would be to rent this apartment, with 0 representing complete reluctance and 100 representing complete enthusiasm.

(Respondents used a horizontal sliding scale to answer the question).
Treatment Group 2: Sympathetic Story/White Character Name.

Same as treatment group 1, but with “Lakisha” replaced by “Jill.”

Treatment Group 3: Unsympathetic story/Black Character Name

Please read the following carefully.

Lakisha is a never-married low-income single mother with two children from two different fathers. Lakisha never put much effort into school and finally dropped out as a teenager, and she has struggled with maintaining a job, partially due to a drug habit that is both expensive and frequently results in legal trouble. She is currently unemployed and surviving on government benefits. She and her children currently reside in an apartment in a neighborhood where the crime rates are high and a lot of the students end up dropping out of high school.

Lakisha recently learned about an apartment available in a mixed-income housing development. Remember, this means that some of the apartments in the building will be reserved for low-income families who will pay less in rent, while the rest of the apartments will be rented at the normal market rate. If Lakisha moves, her children will live in the same building as middle-income families and will attend the same public schools.

(Respondents asked to click a box to indicate they have read the story).

Having heard Lakisha’s story, now presume that you and (if applicable) your family needs to move in the coming year, and presume that you choose to rent an apartment rather than buy a house. Presume that you find an apartment you like and can afford in a decent neighborhood. Next you find out that the building is a mixed-income development and that 30 percent of the units in the building will be reserved for low-income households like Lakisha’s who will pay 50 percent less for rent than the market-rate households like yours will have to pay.

Please rate how willing you would be to rent this apartment, with 0 representing complete reluctance and 100 representing complete enthusiasm.

(Respondents used a horizontal sliding scale to answer the question).

Treatment Group 4: Unsympathetic Story/White Character Name

Same as treatment group 3, but with “Lakisha” replaced by “Jill.”
Appendix 4-B

OLS (Linear Probability Model) Results for Regression Models

Table 4-F: Covariates from Regression Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient Estimate and Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.0 (0.1)</td>
</tr>
<tr>
<td>Sex</td>
<td>1.2 (1.8)</td>
</tr>
<tr>
<td>Race (white vs nonwhite)</td>
<td>0.4 (2.1)</td>
</tr>
<tr>
<td>Education</td>
<td>-1.2 (0.7)</td>
</tr>
<tr>
<td>Partisanship (7 point scale)</td>
<td>-1.0 (0.8)</td>
</tr>
<tr>
<td>Political Ideology (5 point scale)</td>
<td>-6.3 (1.3)**</td>
</tr>
<tr>
<td>Has Children</td>
<td>-7.2 (2.0)**</td>
</tr>
<tr>
<td>Household Income per Capita (in thousands)</td>
<td>-0.2 (0.1)**</td>
</tr>
<tr>
<td>Current housing status (rent vs own)</td>
<td>-2.9 (1.7)</td>
</tr>
</tbody>
</table>

*α<.1 **α<.05
(treatment effects not displayed)

Table 4-G: Regression Adjusted Treatment Effects (compared to control group)

<table>
<thead>
<tr>
<th>Treatment Group</th>
<th>OLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sympathetic Story/White Character</td>
<td>18.3 (2.7)**</td>
</tr>
<tr>
<td>Sympathetic Story/Black Character</td>
<td>15.6 (2.7)**</td>
</tr>
<tr>
<td>Unsympathetic Story/White Character</td>
<td>-3.8 (2.7)</td>
</tr>
<tr>
<td>Unsympathetic Story/Black Character</td>
<td>-6.9 (2.7)**</td>
</tr>
</tbody>
</table>

*α<0.1 **α<.05

Table 4-H: Regression Adjusted Treatment Effects of Black Character Vs White Character

<table>
<thead>
<tr>
<th>Experimental Group Comparison</th>
<th>OLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>White Sympathetic vs Black Sympathetic</td>
<td>2.7 (2.7)</td>
</tr>
<tr>
<td>White Unsympathetic vs Black Unsympathetic</td>
<td>3.0 (2.7)</td>
</tr>
</tbody>
</table>

*α<0.1 **α<.05
Chapter Five:

Policy Implications and Directions for Future Research
Summary of Dissertation Findings

This dissertation has investigated the economic and behavioral outcomes of poverty deconcentration policies on participant households, and it has explored two areas of low-income housing policy implementation. There are several major findings from this research. The major findings from the meta-analysis chapter are that poverty deconcentration policies on average have very little effect (positive or negative) on economic and behavioral outcomes of participants and that studies that are more rigorous in their internal validity tend to find effects of a smaller magnitude. The major finding from the chapter three examining managerial quality of housing authorities is that better management of public housing authorities is correlated with better program performance, controlling for task complexity and state level dummies. From the chapter four, which used experimental design to assess the influences and the malleability of people’s willingness to consider living in mixed income house, there are several major findings. One is that the default attitude in the public about living in mixed-income housing is more consistent with unsympathetic views of future neighbors than with sympathetic views. Another is that stated willingness to live in mixed-income housing is susceptible to change based on whether future low-income neighbors are described sympathetically or unsympathetically. And finally, that stated willingness to live in mixed-income housing is not susceptible to change based on the implied race of future neighbors.

Each of these findings yields potential policy ramifications and has implications for future research, which are explored in this chapter.

Null Effects of Poverty Deconcentration Policies: Research Directions and Policy Implications

The policy implications of the null finding for the effects of poverty deconcentration policies are that housing vouchers alone are not likely to achieve improvements in recipients’ economic well-being or improve their chances of avoiding problematic behavior. As discussed in chapter two, there is still
vigorou academic debate about what exactly went awry with these policies, with some believing that
they were too weak of a treatment, some believing they were fundamentally flawed in their underlying
theories, and some looking to implementation problems (as in this dissertation). One need not be able
to precisely diagnose what went wrong to conclude that continuing to run the programs in the same
way they have been run so far is unlikely to achieve different results. Simply offering a family a housing
voucher to move out of public housing is unlikely to improve their economic situation or behavioral
outcomes. The policies may still be justified based on other criteria, but improving these life outcomes
should not be among them.

Mixed-income projects are still in need of more studies before a definitive position should be
taken on their efficacy. While the results of the meta-regression in chapter two indicate that skepticism
toward their effects is warranted, more studies would make those conclusions more secure. As more
and more HOPE VI projects are created, there will be more and more opportunities to study the impacts
of these programs. Doing so with rigorous research design will be of utmost priority, as the meta-
analysis in chapter two indicated.

Many of the types of studies that did not meet the inclusion criteria of this meta-analysis would
themselves be strong candidates for meta-analysis. One avenue for researchers to explore would be to
complete a meta-analysis on neighborhood effects. Poverty deconcentration may have failed for many
reasons, but one possibility is that the underlying neighborhood effects theories are simply flawed. A
meta-analysis that examined the impact of neighborhood conditions (independent variable) on life
outcomes (dependent variable) would be very valuable, given the large number of such studies that
exist. Recall that these studies were excluded from this meta-analysis because, while similar, they ask a
fundamentally different question. Similarly, a meta-analysis examining the impact of program
participation (as with the meta-analysis included here) on different outcome measures, like mental
health, physical health, safety, or education would be valuable. A meta-analysis of voucher receipt on life outcomes compared to unassisted households would also be valuable, though it is unclear whether voucher receipt improves neighborhood conditions or not for such households.

More research is necessary before it will be possible to know decisively whether changes in geography can be an effective social policy intervention. Nevertheless, there is ample evidence that as they have been practiced heretofore, the answer is that on average they cannot, at least not for improving economic and behavioral outcomes. However, as discussed in chapter two, there are still “apologists” who believe the extant empirical literature is based on flawed methodologies. Those who believe the treatments so far have been too weak might advocate for several different changes to poverty deconcentration policies, both as a matter of research and a matter of policy. One would be to institute thorough “housing counseling” to help low-income households identify eligible housing in the best possible neighborhoods. Many of the post-hoc analyses of the Moving to Opportunity Studies, and many studies of the Section 8 program, have indicated that identifying voucher-eligible housing in good neighborhoods is a major challenge for many low-income households. Searching for housing is time-consuming and expensive, and intensive housing counseling has shown promise as a way to ensure that households receiving vouchers actually end up in better neighborhoods than the ones they left behind.

Another might be to add racial integration as a component of the program. Casciano and Massey (2012) argue that the weak outcomes of the MTO experiment can be traced to its failure to address racial segregation. If racial segregation has an independent (or interactive) effect on life outcomes compared to segregation by income, and if poverty deconcentration policies fail to achieve increases in racial integration, it would not be surprising if program outcomes were weak.

Yet another approach to improving research and policy in this area might be to institute and study more wide-reaching requirements for landlords to accept housing vouchers (a decision which in
many localities is totally voluntary). There is ample evidence from studies of the Section 8/Housing Choice Voucher program that landlords are often resistant to accepting vouchers. While in most areas of the country, participation by landlords is entirely voluntary, in some areas, source-of-income laws exist that require landlords to participate. These laws need to be studied more intensively than they have been to-date. If effective, they could increase the potency of the “dose” that poverty deconcentration policies are able to provide.

More fine-grained sub-group analysis is one promising direction for research in this area. One example is the recent piece by Chetty and Katz (2016) which shows positive effects of relocation on children who were under age 13 at the time of their move, but negative effects for children over 13. Other researchers are focusing re-analyzing the moving to opportunity data to see if households that were the most compliant with the research protocols might have seen more improvements. These researchers give up some of the internal validity that comes from random assignment in exchange for insight into how a heavier “dose” of poverty deconcentration might induce better outcomes.

While there are still ample areas for research and policy experimentation in the area of poverty deconcentration policies, the evidence of these programs’ ineffectiveness at achieving improvements in economic or behavioral outcomes is strong and must be reckoned with. If the Federal government is going to spend more money on efforts to deconcentrate poor households, the burden of proof that the programs can be effective now rests with the apologists for these programs. Millions of dollars and dozens of studies with thousands of models indicate that inducing a household to move to a new neighborhood will not improve their economic well-being or reduce their likelihood of engaging in problematic and risky behaviors.

Management Matters in Public Housing Authority Performance: Research Directions and Policy Implications
As researchers continue to investigate why poverty deconcentration policies have failed, they should consider investigating the degree to which problems with management quality among housing authorities might be contributing to the problem. Chapter three showed that managerial quality as defined by HUD varies systematically with housing authority performance in public housing and in the Housing Choice Voucher Program. While the data are not ideal, the sample is nearly exhaustive. Future research into the link between management and performance in public organizations should continue to leverage these HUD data sets on management quality and program performance to monitor this correlation over time. Tying these managerial variables directly to poverty deconcentration outcomes would further strengthen this research direction and would provide more direct evidence that failures in poverty deconcentration policies are related to program implementation.

The policy implications of this finding are that strengthening managerial quality at the local housing authority level could provide substantial improvements to low-income housing program performance throughout the country. Exactly how that strengthening would or should occur is another question altogether, but the evidence presented in chapter three indicates that one would predict an improvement in program performance from an improvement in management. However, because the analysis in chapter three cannot distinguish between a causal relationship between managerial quality and performance and simple association, any direct policy implications of this chapter should be regarded with caution. If management quality and program performance are merely associated, investing in major improvements in management might not yield a corresponding improvement in performance. Adding more control variables and taking and adding time points so that a panel model can be implemented would strengthen the internal validity of the study.

Recruiting for Mixed-Income Housing Projects: Research Directions and Policy Implications
Chapter four showed that the public’s default attitude toward living in mixed-income housing is lukewarm at best (43 out of 100). Moreover, this level of willingness is more consistent with an unsympathetic view of hypothetical low-income neighbors than with a sympathetic view of them. The policy implications are that mixed-income housing projects need to be aware of how potential middle-income residents may view their future neighbors and market their project accordingly. As discussed in chapter four, this experiment does not establish that real-world mixed-income projects struggle to recruit non-poor households. However, if projects do encounter that problem, they should consider whether their marketing strategies address this concern among non-poor households. The research in chapter four indicates that painting potential neighbors sympathetically can induce substantively significant increases in willingness to live in mixed-income housing, at least in the short run. Marketing managers should adopt a sympathetic frame for describing low-income residents in their marketing strategies.

One of the encouraging findings from chapter four is that the public’s willingness to live in mixed-income developments does not appear to depend on the race of their potential neighbors. While the backstory of their future neighbors mattered quite a bit, their implied race did not. While one could argue that the racial frame used in the experiment was weak, it has been used in a number of other published experiments. Another explanation is that the public does not harbor significant racial biases once one controls for class-biases. The policy implication of this finding is that racial integration in mixed-income projects is feasible. If marketing managers can frame potential neighbors sympathetically, the perceived racial makeup of the development should not be a problem.

This experiment detected short-term effects, and it is uncertain whether these framing effects would endure over the long run. This caveat is a common one in framing experiments. Future research should detect how the frames utilized in this experiment last (or do not last) over time by doing long-
term follow ups with experimental participants. Another common issue with framing research is that “lab” experiments cannot replicate real-world conditions. Research into how real-world mixed income housing projects are recruiting non-poor residents, and even field experiments on different the efficacy of different frames and marketing strategies, would strengthen this area of research. This sort of research would also shed light onto whether, to what degree, and under what conditions non-poor recruitment into mixed-income developments is actually a problem that real-world projects are encountering.

**Concluding Remarks**

Are poverty deconcentration policies forever doomed to fail, or might they be made more effective with improvements to their implementation? This dissertation suggests that, although skepticism toward them is warranted, it is too early to completely write them off as failures. More and more sophisticated methods will continue to try to disentangle neighborhood effects from selection effects in the primary social science literature, helping policymakers predict under what conditions and for which demographics of people poverty deconcentration can be helpful, and when it is likely to fail. On the policy side, as researchers continue to puzzle over the null results presented in chapter two of this dissertation, they should explore challenges to policy implementation as one area with potential for explaining the apparent failure of poverty deconcentration policies. This dissertation has shown that managerial quality and recruitment of non-poor households into mixed-income developments are two areas worth exploring, but there are still many other areas in the implementation chain that could be explored. At the end of the day, researchers may indeed come to the conclusion either that the underlying neighborhood effects theories are flawed and that individual and household effects explain so much of individual outcomes that neighborhoods are unimportant, or that neighborhood effects are simply too deeply entrenched to be mutable by conventional policy mechanisms. If so, U.S. low-income
housing policy should reconsider its emphasis on poverty deconcentration. On the other hand, if more research shows that, under the right conditions, poverty deconcentration policies can be effective and efficient, policymakers should turn to focusing on how to make those conditions prevail for the majority of America’s housing subsidy recipients. More research is warranted before the book is entirely closed on poverty deconcentration policies.
References


Converse, Philip E. (1962). *The nature of belief systems in mass publics*. Survey Research Center,


Joe A. Bolinger
Curriculum Vitae

Education

Indiana University                        Bloomingt on, IN
**Ph.D. in Public Policy**                           2008-December 2016
School of Public and Environmental Affairs/Department of Political Science
Major Fields:  Public Policy, American Politics, and Policy Analysis

Dissertation Title:  *A Multi-Method Investigation of Rental Housing Subsidies and Poverty Deconcentration: Impacts and Implementation Challenges*

Dissertation Advisors:  David Reingold (co-chair)
Sean Nicholson-Crotty (co-chair)
Jill Nicholson-Crotty
Matthew Hayes

**Master of Public Affairs**                          2005-2007
School of Public and Environmental Affairs
Concentrations:  Policy Analysis and Nonprofit Management

Earlham College                        Richmond, IN
**Bachelor of Arts**                           1999-2003
Religion and Philosophy (Double Major)
Graduated with honors from the college and in both majors.

Publications and Papers Under Review


Conference Presentations


Joe Bolinger. “Effectiveness of Community Colleges in the United States.” Presentation at the SPEA Graduate Student Conference Bloomington, IN March 2013


Grants and Awards

SPEA/Political Science graduate student fellowship award. Tuition, fees, and stipend. 2008-2016.

Center on American Politics. $500.00 research grant. 2015.

SPEA Conference Travel Award. $600.00 2015

SPEA Conference Travel Award $500.00 2012

SPEA Service Corp Fellowship 2006-2007

Lilly Endowment Scholar (4 year full college scholarship) 1999-2003

Teaching Experience

As Instructor of Record (26 total sections)

V161: Urban Problems and Solution 18 sections Fall 2010-Fall 2015

V160: National and International Public Policy 7 sections Summer 2010-Summer 2014

K300: Statistical Techniques 1 section Fall 2011

As Teaching Assistant (3 total sections)

Y100: Post Elections 2008 1 section. Spring 2009

Y103: Intro. to American Politics 1 section Fall 2008

Independently lead 3 discussion sections.

H320: Health Systems Administration 1 section Fall 2007
Pedagogical Training

V621: Seminar in Teaching Public Affairs
Frequent trainings from Indiana University’s Center for Innovative Teaching and Learning.

Disciplinary and Public Service

Political Science Graduate Student Association (PSGSA)       June 2010-May 2011
Offices Held: Co-President

Association of SPEA Ph.D. Students (ASPS)       Jan. 2010- April 2012
Offices Held: Social Chair and Graduate and Professional Student Organization Representative

SPEA Graduate Student Conference Planning Committee  2012-2013

Selected Professional Experience

School of Public and Environmental Affairs       Bloomington, Indiana
Research Assistantship for Dr. David Reingold                2009-2010
Research Assistantship for Dr. Matthew Auer                2005-2006

Area 10 Agency on Aging       Ellettsville, Indiana
SPEA Service Corp Fellow       2006-2007

Locality/Sangre de Cristo Youth Ranch       Taos, New Mexico
Assistant Director                  2007-2008

South Central Community Action Program       Bloomington, IN

Musician       Bloomington, IN
Banjoist and songwriter for two active bands.                  2009-Present