PREVENTING HOMELESSNESS IN ALAMEDA COUNTY, CA AND NEW YORK

CITY, NY:  INVESTIGATING EFFECTIVENESS AND EFFICIENCY

By

ANDREW LOUIS GREER

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirement

for the degree of

DOCTOR OF PHILOSOPHY

in

Community Research and Action

August, 2014

Nashville, Tennessee

Approved:

Marybeth Shinn, Ph.D.
Sandra Barnes, Ph.D.
Paul Speer, Ph.D.
Sonya Sterba, Ph.D.

**ABSTRACT**


Due to lack of rigorous evaluations, there is limited evidence that homelessness

prevention programs effectively reduce rates of homelessness and efficiently direct

services where they can make the most difference. Effectiveness is the ability to reduce

rates of homelessness among people who would otherwise experience it. Efficiency is the

ability to direct services to those who would benefit most. Evidence of effectiveness

requires a counterfactual – typically a comparison between a treatment group and a

similar group that does not receive treatment. Evidence of efficiency necessitates

development of a risk model and investigation of the levels of risk where services make

the most difference. Investigations sometimes confound effectiveness and efficiency:

evaluators might believe that services are effective when those services are imprecisely

targeted.

The current study examines effectiveness and efficiency for prevention programs

in two sites. It develops risk models for homelessness using Cox proportional hazard

models for 2,761 applicants for Homelessness Prevention and Rapid Re-Housing in

Alameda County and for 10,220 individual applicants for HomeBase prevention services

in New York City. Further, it uses a regression discontinuity design for the sample in

Alameda County to examine the effectiveness of services. The findings provide limited

evidence that prevention programs can reduce entries into homelessness and stronger

evidence that programs can work better by focusing on individuals and families at highest

risk.  Triage models that exclude some applicants as too risky to serve are not supported

by the data. The studies also contribute to the understanding of the causes of

homelessness, via the examination of risk factors in the two sites. The results suggest that structural issues are the driving forces of homelessness in two housing markets with limited access to affordable housing. Future research is necessary to obtain more precise estimates of prevention effects and to examine similarities and differences in findings across housing markets. Prevention programs might do better not only to provide immediate prevention services for individuals and families but also to combat the structural forces that lead to high rates of homelessness.

# ACKNOWLEDGEMENTS

conversations; Lauren Brinkley-Rubenstein for her open ear and unforgettable Seder celebrations; and Holly Wegman Karakos for her discussions and "Nash-" time.

I am grateful for the love and support of my parents and relatives. I am so lucky to have them in my life. In light of my recent employment, they will be happy to know that when I am 65, I will retire from a career rather than from graduate school.

Finally, I acknowledge the people represented in each row of my data. One of the strongest weaknesses of my work was the lack of direct interaction with homeless people. I am truly sorry that we live within a system where homelessness is a reality for anyone.

**TABLE OF CONTENTS**

**LIST OF TABLES**

# LIST OF FIGURES

**PREFACE**

This dissertation investigates the effectiveness and efficiency of homelessness prevention programs for at-risk populations. However, before I introduce the studies in the dissertation, I acknowledge an important contradiction between how homelessness prevention is done (practice) and how homelessness is understood (theory). Homelessness prevention is most often carried out at the individual level, and the investigations in this dissertation are no exception. The causes of homelessness, however, are not solely or even primarily individual-level phenomena (Burt, 1991; Shinn, 1992). Instead, much of the homelessness literature reports structural-level causes of homelessness (Apicello, 2010; Byrne, Munley, Fargo, Montgomery & Culhane, 2012, Shinn, 2007, 2010).

Such a contradiction may seem like an insurmountable obstacle to prevention efforts, but structural factors that push people into homelessness often manifest at the individual level (c.f., Apicello, 2010). For example, income inequality and housing costs - which render housing unaffordable to many people at the bottom of the income distribution - may manifest at the individual level as the inability to pay rent. Further, effective prevention programs may help people avoid homelessness, even if programs offer services based on individual characteristics without addressing structural-level causes of homelessness. For example, a housing subsidy that makes rent affordable can counteract the structural causes of homelessness for an individual household. Accordingly, targeting risk factors at the individual level can effectively prevent

homelessness for certain people at risk, even if structural causes of homelessness remain in place.

Targeting risk factors at the individual level creates limited change in the structural causes of homelessness, and these structural causes directly influence rates of homelessness (Burt, 1991; Koegel, Burnam & Baumohl, 1996; Rossi, 1994; Shinn, 2007; 2010; Toro, 2007). In environments with minimal income inequality and strong safety nets for poor people, homelessness rates are much lower compared to environments with heightened income inequality and weaker safety nets (Toro, 2007). Such safety nets include policies that provide financial assistance to poor households or that aim to increase the availability of affordable housing. For example, in many European countries, family homelessness occurs at much lower rates than in the U.S., which has fewer protections for at-risk families (Toro, 2007).

Although the targeting of structural factors that manifest at the individual level may seem like a temporary solution, it is a necessary step to keep people housed while prevention advocates fight for changes to structural-level causes of homelessness. Recent discussions explore homelessness prevention strategies that address both individual and structural factors associated with homelessness, but such strategies remain absent in practice (c.f., Apicello, 2010). The root causes of homelessness are likely to remain unaltered if prevention providers address only individual-level consequences of structural causes. However, homelessness prevention efforts can nonetheless alleviate much suffering, because many families and individuals experience homelessness while structural change occurs slowly.

Although structural-level causes of homelessness mostly determine rates of homelessness, individual-level risk factors reflect the types of people who become homeless. Thus, in the absence of structural-level change, one danger of targeting individual-level characteristics of homelessness is that prevention programs might target at-risk groups at the expense of exposing other groups to homelessness risk (e.g., giving scarce housing subsidies to people in shelter may lengthen the waiting list for other poor people) (McChesney, 1990; Sclar, 1990; Shinn, Baumohl, & Hopper, 2001). The most direct way to combat this threat would be to change the structural forces that direct people into homelessness. However, with limited structural-level changes, prevention programs should be routinely evaluated to accurately direct services to those most at risk for homelessness.

To best combat homelessness, both structural and individual-level considerations should be included in homelessness prevention research (Apicello, 2010; Lee, Tyler, & Wright, 2010; O'Flaherty, 2004). To evaluate homelessness prevention that is carried out at the individual level in the absence of structural-level change, researchers should frame their studies in ecological context. Accordingly, investigations should include considerations of the characteristics of homeless people (individual level) and the root causes of high homelessness rates (structural level). For this reason, this dissertation includes a discussion of the structural forces contributing to homelessness before presenting individual level and ecological investigations of homelessness in subsequent chapters.

# CHAPTER 1


# AN INTRODUCTION TO THE STRUCTURAL CAUSES OF HOMELESSNESS

# AND INDIVIDUAL-LEVEL HOMELESSNESS PREVENTION


## Who is Considered Homeless?

The current sub-section explores definitions of homelessness, because

homelessness lacks a uniform definition. In the United States (U.S.), definitions tend to

focus on the lack of a suitable residence. HUD's definition includes four categories of

homelessness:

- People who are living in a place not meant for human habitation, in emergency shelter, in transitional housing, or are exiting an institution where they temporarily resided [or]… if they are exiting an institution where they resided for up to 90 days … and were in shelter or a place not meant for human habitation immediately prior to entering that institution.
- People who are losing their primary nighttime residence, which may include a motel or hotel or a doubled up situation, within 14 days and lack resources or support networks to remain in housing...
- Families with children or unaccompanied youth who are unstably housed and likely to continue in that state. This … category of homelessness … applies to families with children or unaccompanied youth who have not had a lease or ownership interest in a housing unit in the last 60 or more days, have had two or more moves in the last 60 days, and who are likely to continue to be unstably housed because of disability or multiple barriers to employment.
- People who are fleeing or attempting to flee domestic violence, have no other residence, and lack the resources or support networks to obtain other permanent housing… (NAEH, 2012, p. 1).

HUD does not provide as comprehensive a typology as that found in other developed

countries.

The European Union (E.U.) offers a more inclusive definition of homelessness

that the authors describe as "home"-based. For example, the European Typology of

4

Homelessness and Housing Exclusion (ETHOS) includes multiple domains (i.e.,

physical, social, and legal) to generate a broad classification of homelessness into four

broad categories:

- *rooflessness* (without a shelter of any kind, sleeping rough)
- *houselessness* (with a place to sleep but temporary in institutions or shelter)
- living in *insecure housing* (threatened with severe exclusion due to insecure tenancies, eviction, domestic violence)
- living in *inadequate housing* (in caravans on illegal campsites, in unfit housing, in extreme overcrowding) (FEANTSA, 2005).

Although the HUD definition restricts the category of persistent housing instability to

families and youth, the ETHOS typology does not distinguish homeless families from homeless

individuals. Further, the much broader ETHOS definition of homelessness includes not only

those lacking a suitable residence (i.e., by the HUD definition), but also those living in

overcrowded conditions. As the differences above demonstrate, definitions of sociopolitical

concepts, such as homelessness, are not always straightforward and may depend on

political agendas and research goals (Hopper, 1991). The E.U. plans to survey households

about a range of homelessness indicators in future censuses. Researchers in the U.S.

could expand databases, and thus research findings, by following a similar approach,

albeit with a typology tailored to their goals.

Some definitions of homelessness are restricted because of challenges

operationalizing the concept. For example, point-in-time (PIT) estimates offer a useful

snapshot of homelessness, but they are problematic because the most vulnerable groups

tend to be overrepresented; they are more likely to be homeless on any given night than

those that experience brief or one-time encounters with homelessness (Shinn & Greer,

2011). The 2013 Annual Housing Assessment Report to Congress (AHAR) reports a

nationwide PIT estimate of 610,042 people on a single night in January 2013,

representing a decrease from 671,888 on a night in January 2007 (HUD, 2013a). In 2013,

slightly less than two-thirds of homeless people were individuals (i.e., single adults or

unaccompanied youth) and the remaining households were families (i.e., at least one

adult and one child). The count of chronically homeless individuals (i.e., individuals with

extended bouts of homelessness and disabilities) decreased by 25.2% from a single night

in January 2007 to a single night in January 2013 (HUD, 2013a). Although such data

provide a useful snapshot of homelessness, the numbers of people staying in shelters over

the course of a year offer additional insight into the problem of homelessness. Annual

shelter stays (i.e., stays in emergency shelter or transitional housing) are a second

commonly reported measure of homelessness. For example, in 2012, 1.49 million people

stayed in shelter for at least one night, representing a 0.9% decrease from 2011. In other

words, from October 1, 2011 to September 30, 2012, one in 209 people in the U.S. stayed

in a homeless shelter (HUD 2013b). Measuring shelter stays provides useful information

about shelter entry rates, but fails to include homeless people who stay on the streets.

Some people are more likely to become homeless than others, and characteristics

of homeless groups differ. In 2011, African Americans were overrepresented among both

families and individuals who experienced homelessness, but especially among families.

Moreover, people experiencing homelessness had higher rates of disabilities than the U.S.

population overall (HUD, 2013b). Other studies found that baby boomers have

maintained highest risk status for single adult homelessness over the last three decades,

meaning that the average age of single adults experiencing homelessness has risen over

that period, whereas a similar aging trend is not apparent for families (Culhane et al.,

2013). Further, among childless adults, longer durations of homelessness were associated

with older age and arrest history (Caton et al., 2005). In summary, empirical evidence indicates differences across homeless groups, and researchers should continuously strive to uncover the reasons for these differences.

**What Causes Homelessness?**

Historically, two broad classes of theories frame the causes of homelessness. The first class posits that structural-level factors cause homelessness. Such factors include policy, social exclusion, and other macro- and community-level domains. The second class posits that individual factors cause homelessness. Some individual factors, such as mental illness, might originate at this level, but other individual-level risk factors for homelessness, such as higher risk for racial minorities, are manifestations of structural factors at the individual level (c.f., Koegel, Burnam & Baumohl, 1996). However, even supposed individual factors like mental illness - that hinge on society's definitions of the factor - are heavily influenced by policy decisions. Such decisions include the level of disability benefits and the requirements for obtaining and maintaining them.

Until the 1990s, structural and individual-level theories were viewed as competing, however, they address two different research questions (Koegel, Burnam & Baumohl, 1996). Structural-level theories address the question of homelessness rates: why do so many people become homeless? Alternatively, individual-level theories address the question of who becomes homeless: why do certain people become homeless? Currently, the most accepted theory of homelessness causes is a combination of structural and individual factors (Apicello, 2010; Lee, Tyler, & Wright, 2010). Sociologists discuss the causes of homelessness with a similar macro/micro framework (for review see Lee et al., 2010), with O'Flaherty (2004) suggesting homelessness is "a

conjunction of unfortunate circumstances (p.1)." Some studies take an explicitly ecological view, suggesting that the structural and individual factors are interrelated and that individual factors tend to be manifestations of structural-level forces (Nooe & Patterson, 2010).

The current section discusses the structural factors associated with homelessness: policy (Shinn, 2007), social exclusion (Shinn, 2010), and community-level domains (Byrne, Munley, Fargo, Montgomery & Culhane, 2012). Structural factors originate from sources beyond households. Examples of the origins of structural factors include economic systems, governments, institutions, schools, and neighborhoods. Below I present each of the structural factors in detail.

**Policy.** In the current subsection, I discuss universal and targeted policies, sometimes across countries, to illustrate how policy type relates to homelessness. Universal policy applies to everyone – it defines services that are available to the whole population and are a responsibility of the public sector. Targeted policy, in contrast, assumes that the market addresses the needs of most of the population and focuses on particular at-risk groups. With targeted approaches, the adequacy of funds allocated to services differs across locales (Czischke & Gruis, 2007). The links between economic inequality and universal social policy are well documented (McFate, Lawson, & Wilson, 1995; Smeeding, 2005), and some homelessness researchers argue that developed countries with less generous universal policies and greater inequality have higher homelessness rates than developed countries with more generous universal social policies (Shinn, 2007, 2010; Toro, 2007). Although social expenditures may reduce rates of homelessness by keeping people out of poverty, cross-country comparisons are difficult

due to the diversity of social and financial structures. An exhaustive analysis of social policy across developed countries is beyond the scope of this chapter. Nevertheless, I highlight policies most relevant to homelessness.

Universal social policy can adopt a variety of forms, such as housing or welfare-related policy. Housing policy can be further divided into tax benefits and direct assistance. United States housing policy favors homeownership via tax benefits and offers many fewer resources to low-income households. For example, in 2008, homeowner tax benefits (e.g., mortgage interest deductions) totaled more than $171 billion, whereas benefits for direct housing assistance totaled just over $40 billion. Further, households earning more than $100,000 annually receive the vast majority of the tax breaks (Schwartz, 2010). Policy that is intended to reduce poverty rates (e.g., social welfare policy) also impacts homelessness rates. Shinn (2007, 2010) argues that developed countries with progressive tax policy and transfer programs that reduce inequality experience much lower rates of homelessness than countries without such policy. Further, the types of policy are directly linked to rates of specific types of homelessness. For example in countries, such as the United States, and secondarily the U.K. that have high financial inequality and limited income support for families, rates of homelessness for families exceed rates in other developed countries.

Rossi (1994) offers additional insight into the causes of homelessness, especially focusing on family homelessness. First, Rossi argues that family homelessness originated in the 1970s and 1980s from high unemployment, an economic recession, and lower levels of financial assistance than had previously been offered. He further posits that homeless people failed to realize the so-called "trickle-down" effects of Reagan era

economic policy. Further, welfare assistance is conditioned on employment, which is often difficult for young single mothers to maintain (Blank, 2010; Shinn, 2010). Some sociologists in the U.K. argue that welfare-to-work policies do not align with single parents' desires or capabilities (Rafferty & Wiggan, 2011). Based on the results of two quarters of the UK Quarterly Labour Force Survey in 2006 and 2007, Rafferty and Wiggan (2011) suggest that the primary reported reason that single parents do not want employment is that they are taking care of their children. Further barriers to employment include poor health, disability, and care for older dependent children. Rafferty and Wiggan (2011) argue that welfare-to-work regulations are government impositions of idealized parenthood and acceptable citizen behavior.

On the other hand, some sociologists argue that prevention programs should assist homeless people to find a place within the labor market. Fallis (2010) argues that progressive housing policy should focus on targeted homelessness prevention rather than on universal social housing policies, especially for countries with limited funding for housing programs. Supported by a thorough history of social housing in Canada, Fallis (2010) suggests that the economic conditions of today are quite different from the times of Keynesian economics, when social housing programs fit well with traditional families. Those at risk for homelessness were not required to have the same amount of mobility necessary to participate in today's economy. Instead of universal housing policy, the author recommends renovating social housing and targeting services to homeless people to give them access to the labor market – although he offers no consideration for the diminished economic benefits that homeless people might realize within this market.

To avoid homelessness, people require stable housing. To respond to the need for stable housing, the United States opts for targeted approaches instead of universal homelessness policy. However, targeting specific groups for an intervention is complicated. A central danger with targeted rather than universal prevention strategies includes shifting resources to at-risk groups at the expense of removing resources from a group that avoided homelessness because of them.  Thus a game of "musical chairs" (McChesney, 1990; Sclar, 1990) or "queue jumping" (Shinn, Baumohl, & Hopper, 2001) threatens effective homelessness prevention. In terms of homelessness prevention, researchers should search for at-risk groups continually, because the face of homelessness (i.e., groups most at risk of experiencing homelessness) changes over time (Hopper, 1991; Israel, Toro, & Oullette, 2010; Shinn, Baumohl, & Hopper, 2001).

**Social Exclusion.** If policy makers truly represent the citizens who voted for them, then policy should reflect underlying social and cultural beliefs. Based on psychological theory (Levine & Levine, 1970), Shinn (2007) argues that a structural framing of homelessness is found in societies that favor government intervention and more generous welfare programs. Alternatively, the dominant culture in the United States reflects a stronger emphasis on individual responsibility than in continental Europe and exhibits a preference for individual explanations of homelessness.

Some sociologists in the U.K. take similar stances by arguing that New Labour legislators attempted to link homelessness with employment strategies (Dobson & Mcneill, 2011). The authors suggest that personal responsibility was a key component of New Labour legislation and conclude that work programs were intended to create "empowered and responsible citizens" (p. 586), when, in fact, the programs perpetuate

the social exclusion of marginalized groups. Further sentiments of neoliberal reproduction arise in discussions of homelessness prevention programs that are framed with economic rather than social understandings (c.f., Willse, 2010). Dobson and McNeill (2011) argue that researchers should critically examine prevention programs to consider how much they benefit targeted recipients compared to how much programs reinforce existing power structures. Other studies concur (Willse, 2010).

Several researchers in North America advance the discussion of the relationship between homelessness and the labor market by considering employed homeless people (Shier, Jones, & Graham, 2012). The authors interviewed 61 employed homeless people from 2008 to 2009 in Calgary, Canada, to examine the government preference for "work-first" models (Iverson & Armstrong, 2006, in Shier et al., 2012). Such models are intended to increase individual responsibility to participate in the labor market. Shier et al. (2012) conclude that multiple factors associated with the labor market fail to prevent homelessness and may contribute to increasing rates. Many respondents reported insufficient work, inconsistent pay, weak relationships with employers, job loss resulting in loss of employment-based housing, and undesirable employment.

A similar study in the U.K. includes interviews with 30 mostly unemployed homeless individuals to understand their perceptions of employment seeking (McNeill, 2010). The authors find four pathways to employment: work focus (i.e., felt they were job ready), deferred focus (i.e., desired training before seeking employment), uncertain focus (i.e., felt unsure or uninterested in employment), and resettlement focus (i.e., focused on reestablishing stable housing rather than employment). The authors conclude that a one-

size-fits-all program tied to employment requirements fails to respond to the various pathways people experiencing homelessness desire.

Additional studies find more compassion in the public opinion of homelessness in countries with more progressive welfare policy (i.e., continental Europe) than in countries with less generous welfare policy (i.e., English-speaking countries) (Toro, 2007). Further, in societies with heterogeneous populations, a status sometimes called "ethnic fractionalization" tends to favor policy that benefits the status quo and does not transfer wealth to minorities (Alesina & Glaser, 2004; Shinn, 2007). Alternatively, more homogenous societies often have generous welfare policies, including policies for those who experience homelessness. It seems that the powerful are less willing to assist homeless people if they consider homeless persons to be different from themselves.

**Discrimination: Racism [1] and Stigma[2].** Racism is one of the most conspicuous forms of discrimination that affects who becomes homeless. The U.S. consistently reports higher rates of homelessness for people belonging to minority groups than to the majority group (HUD, 2013a, 2013b). For example, in the U.S., 36% of the overall population identified as non-white compared to 60% of sheltered homeless people (HUD, 2012b).

---

[1] Here, the term racism focuses on differences in treatment of racial minorities. I intend the term "racism" to acknowledge differences in power and opportunity for groups with diverse levels of social capital. I borrow Blumer's (1958) and Bobo's (1999) framework of racial prejudice as group position to define racism including: "1) …a feeling of superiority on the part of dominant group members, 2) …a belief that the subordinate group is intrinsically different and alien, 3) …a sense of proprietary claim over certain rights, statuses, and resources, and 4) …a perception of threat from members of a subordinate group who harbor a for a greater share of dominant group members' prerogatives" p. 449

[2] Here, I use Link and Phelan's (2001) definition that stigma is "…the convergence of interrelated components…stigma exists when elements of labeling, stereotyping, separation, status loss, and discrimination occur together in a power situation that allows them" p. 377.

However, overrepresentation of minorities among the homeless is not a unique

phenomenon to the United States. In the U.K., 14% of the overall population belongs to

a non-white ethnic minority (U.K. Census, 2013) compared to 35% of homeless people

(i.e., owed a homelessness duty[3]) (Office of Deputy Prime Minister, 2013).

Discrimination, in the form of stigma, also affects people with disabilities, such as

mental illness and substance abuse. For example according to the Substance Abuse and

Mental Health Services Administration, in 2011, national rates of mental illness (13.3%),

substance abuse (6.1%), or both (2.2%) (SAMSHA, 2013, US Census, 2011) were much

lower than rates of sheltered persons who had severe mental illness (26.2%) or of

sheltered adults who experienced chronic substance abuse (34.7%) (Paquette, 2011). For

chronically homeless individuals, the rates are even higher, with 30% experiencing

mental health problems and approximately 50% with co-occurring substance abuse

(Paquette, 2011). For homeless individuals with co-occurring substance abuse and mental

illness, providing housing without additional requirements has been shown to keep

people housed longer and at higher rates compared to housing with additional

requirements (Tsemberis et al., 2004).

The evidence for direct connections between some disabilities and homelessness

appears to be weak. For example, Jencks (1995) argues that the crack epidemic was

responsible for increases in homelessness. If crack use were a direct cause of

homelessness, investigators should have seen an explosion of homelessness followed by a

---

[3] A homeless duty is the determination that a local housing authority has a "duty to
house" a homeless person. Eligibility requires evidence that the person is 1)
unintentionally homeless, 2) connected to the local area, 3) eligible for public funds
(immigration status), and 4) considered to be in priority need of services (definition
differs by nation).

plateau and recession in rates as crack use rose and fell. However, cocaine use, including crack use, has declined dramatically by almost 42% from 2006 to 2011 (NIDA, 2012), while rates of homelessness declined by only 6.3% from 2007 to 2012 (HUD, 2012b). Although the time periods are slightly different comparing cocaine use and homelessness, there is little evidence for a strong relationship.

**Family composition.** Single parenthood has been blamed for increasing homelessness rates. Jencks (1995) argues that homelessness rates increased from 1969 to 1989 when "unskilled women not only married less but continued to have children [which] pushed more of them into the streets." (p. 58). Other studies support the idea of relationships between homelessness and a cultural shift towards declining marriage rates (Burt, 1991). Although family structure is related to poverty, among poor households, no studies suggest that two-parent households offer additional protection from homelessness (c.f., Shinn et al., 2013). Actual increases in homelessness rates for single mothers are more likely caused by the drastic restructuring of social policy, leading to decreased assistance for single mothers (Rossi, 1994). As described at the beginning of this chapter, rates of family homelessness are much lower in countries that have strong social welfare nets and generous family policies compared to higher rates in countries that lack such policies. Further, families that receive housing subsidies tend to stay housed (Khadduri, 2008; Wood et al., 2008).

**Structural Causes at the Community-Level.** Structural determinants of homelessness exist at multiple levels. Investigations of homelessness at the community level tend to include at least some of the following community-level correlates of homelessness: housing markets, economic conditions, demographic composition, safety

net, climate, and transience (Byrne et al., 2012; Lee et al., 2003). When affordable

housing is out of reach, the poorest households can be at heightened risk for homeless.

For example, multiple studies find positive associations between the cost of rent and rates

of homelessness at the community level (Early & Olsen, 2002; Lee et al., 2003). It is

important to note that estimating homelessness rates is challenging. However, rental cost

was still positively associated with homelessness under different model assumptions, for

example that shelter counts were accurate and street counts undercounted homelessness

by one-fifth and one-tenth. Further, Early & Olsen (2002) were unable to find many other

significant predictors of homelessness, most likely due to inaccurate counts of street

homelessness.

Local revitalization efforts can lead to displacement of already homeless and

extremely poor people. For example, Jencks (1995) argues that the destruction of skid

rows led to higher homelessness rates. Jencks' argument is especially relevant for people

with mental illnesses and suggests that displacement from deinstitutionalization did not

lead to homelessness until cheap housing and skid rows largely disappeared.

Local economic conditions have been shown to influence homelessness rates in

previous studies, with higher unemployment and poverty rates associated with higher

rates of homelessness (Burt, 1991; Quigley et al., 2002). Demographic characteristics of

communities provide, perhaps, the least consistent associations with rates of

homelessness. In terms of individual traits, African Americans and female-headed

households are overrepresented groups that experience homelessness. However, at the

community-level findings are inconsistent in terms of demographic characteristics (Byrne

et al., 2012). As with all investigations that span multiple levels of analysis, researchers

should be aware of the temptation of the ecological fallacy when associations from one level are assumed to be true at other levels.

Local social safety nets can have negative associations with homelessness rates. Further, as the safety net gets wider and more substantial, protection becomes stronger (Byrne et al., 2012). For example, Honig and Filer (1993) used HUD estimates of homelessness rates across 50 metropolitan areas in 1984. They found that welfare programs, such as higher rates of Aid to Families with Dependent Children, were associated with lower rates of homelessness, although the rate of SSI receipt was associated with higher homelessness rates (Honig & Filer, 1993).

Climate is another community-level factor that some investigations hypothesize is associated with homelessness. Studies find that lower rates of precipitation and higher temperatures contribute to higher rates of homelessness, and higher rates of people staying in unsheltered locations (Byrne et al., 2012). Finally, some studies have investigated associations of rates of transience with rates of homelessness. For example, Lee et al., (2003) found increased homelessness rates in areas that experienced population growth. The authors argue that this finding may be due to higher competition in the housing market when increasing demand (i.e., higher in-migration) raised prices and left those unable to compete at risk for homelessness.

Despite the evidence for structural causes of homelessness, most homeless prevention efforts are framed at the individual level. Current frameworks include interventions for individual-level manifestations of structural-risk factor domains including demographics, human capital, housing conditions, disability (including criminal justice involvement), interpersonal discord, childhood experiences, and shelter

history (Apicello, 2010; Shinn et al., 2013). Instead of addressing policy that permits greater housing affordability, homelessness prevention normally targets the individual characteristics of those who experience homelessness. Even though such prevention practices undeniably fail to address the structural causes of homelessness, such efforts are vital to assist people who are becoming homeless currently. I explore this assertion further in the next section.

**Homelessness Prevention**

Homelessness prevention research builds on broader literatures on prevention in public health and epidemiology. Prevention reduces problematic outcomes by minimizing risk factors that lead to higher rates of unwanted outcomes and by maximizing protective factors that lead to lower rates (Coie et al., 1993; Mrazek & Haggerty, 1994). Exposure to risk factors can have additive effects or worse, multiplicative effects, on dysfunctional outcomes. Alternatively, protective factors can be combined to mitigate the harmful effects of risk factors (Coie et al., 1993).

Typologies of prevention that originated in public health have been extended to homelessness prevention (Culhane et al., 2011; Shinn, Baumohl, & Hopper, 2001). At least two overlapping prevention typologies have been used to frame homelessness research. The first describes the precision of strategies for targeting prevention programs and includes three categories: universal, selected, and indicated prevention (Mrazek & Haggerty, 1994; Shinn et al., 2001). Universal prevention gives an entire population access to a prevention strategy. For example, a society can implement a right to housing that would guarantee affordable housing to the lowest-income population. Universal prevention can be cost effective if the prevention strategy is cheap. However, for

expensive strategies, universal prevention of relatively uncommon conditions would be more costly than targeting at-risk groups or individuals.

The remaining two categories of prevention, selected and indicated, are types of targeted prevention (Mrazek & Haggerty, 1994; Shinn et al., 2001). In selected prevention, membership in a high-risk group makes one eligible for prevention. Targeting money to an at-risk community would qualify as selected prevention. For example, neighborhood revitalization programs target particular neighborhoods even though a range of income levels exists across households in any particular neighborhood. Indicated prevention is targeted prevention for households that are screened for high-risk characteristics. Instead of targeting larger groups that *may* contain households at risk of a particular unwanted outcome, indicated prevention targets specific at-risk households. Because of the specificity of indicated prevention, this category would be the most cost-effective choice for expensive interventions to alleviate relatively rare conditions.

Other researchers frame their discussion of prevention with a second widely used public-health based typology:  primary prevention, where individuals are prevented from an unwanted outcome (i.e., reducing incidence); secondary prevention, where individuals are assisted quickly to reduce the duration of an unwanted outcome (i.e., reducing prevalence); and tertiary prevention, where prevention efforts attempt to reduce secondary side-effects of an established problem (Culhane et al., 2011). Although the categories of primary and secondary prevention are straightforward descriptors of homelessness prevention strategies, tertiary prevention is not as clearly applicable to such strategies. In public health, tertiary prevention can refer to preventing the debilitating effects of a condition. In homelessness studies, multiple researchers equate tertiary

prevention with chronic homelessness prevention (Burt et al., 2007; Culhane et al., 2011). However, prevention strategies that reduce rates of chronic homelessness often address the unwanted outcome (i.e., homelessness), rather than addressing secondary unwanted side effects of homelessness. Tertiary prevention may be a better descriptor for interventions such as soup kitchens or health clinics, where housing is not provided, but the side effects of homelessness (e.g., hunger, health problems) are addressed.

Culhane et al. (2011) argue that the timing of prevention strategies should be as close as possible to the shelter door, or the point between primary and secondary prevention when individuals are about to lose or have just lost their homes. The authors claim that such a system should avoid the creation of specialized services in favor of maximizing support from mainstream agencies. They further argue that such strategies would maximize cost-effectiveness and prevention effectiveness. The authors posit that those threatened with losing their homes would avoid the high costs of shelter entry, and those at risk for long-term homelessness would avoid expensive long-term shelter costs. Culhane et al. (2011) also suggest that a gradient of prevention services, which offers people the least expensive services for their level of homelessness risk (and reserves more expensive services for those who are not helped by less expensive options), would be superior to existing systems of care. However, the authors do not discuss specific approaches to targeting homelessness-prevention strategies to those in greatest need of services.

As suggested earlier, to combat homelessness, prevention efforts should address structural *and* individual-level risk factors for homelessness. Apicello (2010) proposed the *population and high-risk framework* that offers such a multi-level approach. This

framework suggests that population-level interventions, such as increased affordable housing, can be combined with strategies targeted at high-risk households. The combined approach would effectively reduce homelessness without the "queue-jumping" problems associated with traditional targeted strategies. Unfortunately, the population and high-risk approach is limited by the lack of population-based interventions. In the absence of such programs, homelessness prevention and homelessness causes will remain somewhat mismatched. Current homelessness prevention is carried out with a focus on the individual level, and investigations of current prevention programs are vital to understand how well such programs work to reduce homelessness rates for people at highest risk of losing their homes. With the majority of homeless people experiencing brief but expensive shelter stays (Culhane, Metraux, & Byrne, 2011), a focus on community-based homelessness prevention may be a cost-effective strategy to combat homelessness.

**Effectiveness and Efficiency.** Successful homelessness prevention must be both effective and efficient (Burt et al., 2007). Effectiveness is the ability of a program to reduce rates of homelessness among people who would otherwise experience it. Efficiency is the ability of a targeting model to direct services to those who would benefit most from such services. Sometimes investigations confound these ideas: evaluators might believe that services are effective when those services are going to individuals who were not at risk to begin with. Evidence of effectiveness requires some reasonable counterfactual – typically a comparison between a treatment group and a similar group that does not receive treatment. Limited evaluations and their lack of sophistication lead to sparse evidence for effectiveness (Apicello, 2010; Greer & Shinn, in progress; Shinn et al., 2001). Nevertheless, some evaluations show evidence of effectiveness in community-

based homelessness prevention (Cragg & O'Flaherty, 1999; Culhane et al., 2002; Khadduri, 2008; Messeri, O'Flaherty, & Goodman, 2011; Padgett et al., 2011; Pearson et al., 2009; Perlman & Parvensky, 2006; Rolston, Geyer, & Locke, 2013; Sadowski et al., 2009; Stretch & Kreuger, 1993; Tsai et al., 2011; Tsemberis et al., 2004; Wong et al., 1997; Wood et al., 2008).

Few studies investigate efficiency explicitly. Although some scholars investigated the risk factors associated with homelessness (Bassuk et al., 2010; Crane & Warnes, 2000; Early, 1998; 2004; Edgar, 2006; Gubits et al., 2009; Nolan et al., 2005; O'Connell et al., 2008; Wong et al., 1997; Zlotnick et al., 1999), most risk factor investigations fail to combine risk factors into models and to assess how accurately the models target those most at risk. Exceptions were a New York City investigation of family homelessness (Shinn et al., 1998), a New York City investigation of HomeBase prevention services (Shinn et al., 2013) and a Hennepin County, MN investigation (Barnett et al., 2011). Efficiency investigations are needed to show who benefits from particular prevention strategies. Such models are just as vital in understanding homelessness prevention as models that examine intervention effectiveness.

**Dissertation Research Questions**

This dissertation addresses the dearth of evidence for effective and efficient homelessness prevention programs by empirically examining two community-based prevention programs - one in New York City (NYC), NY, and one in Alameda County, CA. I bring to these studies the perspective that research on effectiveness and efficiency can assist prevention programs to improve services and decrease rates of homelessness. Although structural causes of homelessness must be addressed to successfully prevent

homelessness overall (Apicello, 2010; Shinn, 2010), prevention programs that assist at-risk households individually can reduce rates of homelessness and mitigate its harmful effects on physical and mental well-being of people who experience it. With the perspective of community-based homelessness prevention in mind, I ask the following questions in three papers:

1. *Efficiency:* Which individual applicants for homelessness prevention services in New York City are at highest risk of entering shelter?  How does an empirical model compare with a similar model derived for families?

2. *Efficiency:*  What applicants for homelessness prevention services in Alameda County are at highest risk for different manifestations of homelessness? How do models compare with those in New York City?

3. *Effectiveness:*  Does a community-based homelessness prevention program (Homelessness Prevention and Rapid Re-housing Program [HPRP]) in Alameda County effectively reduce rates of homelessness for applicants who received services?

The current chapter now introduces the context of these homelessness prevention studies: NYC's HomeBase Program and Alameda County's EveryOne Home Program.

**Research Context**. NYC's HomeBase and Alameda County's EveryOne Home programs are the major providers of community-based homelessness prevention services in their respective geographies. The HomeBase program serves New York City with homelessness prevention and has participated in empirical research to examine its

effectiveness (Messeri, O'Flaherty, & Goodman, 2011; Rolston, Geyer, & Locke, 2013) and to improve the efficiency of its targeting strategies for families (Shinn et al., 2013). HomeBase administrators have requested a more efficient targeting model than the one in use for childless adults. The EveryOne Home program serves Alameda County, CA, with homelessness prevention and seeks to improve the effectiveness and efficiency of its targeting models.

Homelessness prevention in Alameda County and prevention services in NYC differ in at least two ways. The first difference is variation in homelessness service systems. The service system in Alameda County is less integrated than that of NYC. Further, New York City has a legal right to shelter; Alameda County does not. The second difference is climactic. The annual range of temperatures in Alameda County is much smaller than in New York. Chapter 4 includes an in-depth discussion of additional differences between Alameda County and NYC.

Across sites, people applied for homelessness prevention services, and their subsequent shelter entry was tracked in administrative records by the Homelessness Management Information System (HMIS) in Alameda County and records of the Department of Homeless Services (DHS) in New York City. Additionally, in Alameda County, if applicants reapplied for prevention services, staff recorded where individuals stayed since their last application. Alameda intake records indicated whether or not a household was unstable (e.g., overcrowded, with arrears), in imminent danger of losing housing (e.g., threatened with eviction), or literally homeless. Thus shelter entry is not the only potential indicator of homelessness in Alameda County.

Additional measures of homelessness are recorded only for a biased sample, namely those initial applicants for services who return to providers for additional services. Homelessness may also be underestimated to a greater extent than in New York City because not all shelters participate in HMIS and people might be more likely to stay on the street in a warm climate. From a cost standpoint, such an analysis is useful because we can predict homelessness as it impacts the Alameda County homeless service system (i.e., service providers and participating shelters). However, from a social justice standpoint, a more comprehensive assessment is necessary to more accurately investigate factors that lead to overall homelessness. Such data are not collected currently.

**Organization of the Dissertation**

The remainder of the dissertation investigates homelessness prevention programs and ends with policy recommendations based on the outcomes. Chapter 2 uses survival analysis to model factors that contribute to hazard of shelter entry after adult applicants without children applied for homelessness prevention services in NYC. Chapter 2 contributes to the literature by developing a screening model for individuals and then comparing the efficiency to a model developed for families in the same city. Chapter 3 uses survival analysis to model factors that contribute to hazard of shelter entry after applicants applied for homelessness prevention services in Alameda County, CA. Further, this paper expands outcomes to include additional definitions of homelessness: shelter entry, imminently losing housing, and unstable housing. Additionally, I compare both the substantive model and its efficiency to results in NYC. Chapter 3 contributes to the literature by expanding the definition of homelessness to incorporate an ecological framework. Further, the paper compares risk models for Alameda County, CA, with NYC

to determine the extent to which rick factors are consistent across these disparate

contexts. Chapter 4 uses a regression discontinuity approach to investigate the

effectiveness of the Homelessness Prevention and Rapid Re-housing Program (HPRP) in

reducing rates of homelessness in Alameda County, CA. Chapter 4 contributes to the

literature because it is the first evaluation of HPRP that includes a counterfactual, and the

chapter uses one of the first regression discontinuity (RD) designs to investigate

homelessness. Further, the design includes two discontinuities - one for people at high

and low risk - allowing researchers to examine differential effects of prevention at

different risk levels. Finally, Chapter 5 summarizes the findings across studies and

describes implications for community-based homelessness prevention.

**CHAPTER 2**

**PREVENTING HOMELESSNESS FOR INDIVIDUALS IN NEW YORK CITY:
TARGETING SERVICES TO THOSE MOST LIKELY TO BENEFIT**

**Introduction**

Nationwide in 2012, 1.48 million people stayed in shelter at least one night, and almost two-thirds were individuals not part of family units (HUD, 2013b). Shelter stays are expensive (Culhane et al., 2011a; Spellman et al., 2010) and associated with a variety of adverse outcomes (Grunberg & Eagle, 1990). Compared to shelter stays, accurately targeted and effective community-based prevention programs might be cheaper and less emotionally taxing (Culhane et al., 2011b). The current study develops a model to predict shelter entry for adult applicants for the HomeBase homelessness prevention program in New York City (NYC). The goal of this study is to assist service providers to target prevention services to applicants who can benefit most. Although the approach is similar to one used previously for family applicants, literature suggests that individuals and families have different risk profiles associated with shelter entry.

Here, families are defined as at least one adult and one child, and individuals are defined as adults without children. Individuals form the largest group of people experiencing homelessness nationwide but not in New York City (NYC) (HUD, 2012b). Of all homeless people staying in shelter or transitional housing in NYC on a single night in January 2013 (HUD, 2013b), less than 35% were individuals.  This is lower than the national rate of individuals in shelter both because of the high cost of housing in New

York City (which is particularly problematic for families) and because of the city's legal right to shelter (which means that families who might put up with extremely poor housing conditions in other jurisdictions overwhelmingly enter shelter as an alternative in New York). Individuals, on the other hand, more often become homeless on the streets. In the city, couples without children can be sheltered together but are considered part of the individual adult system. Such couples are included here, and I use the term *individuals* to describe the population.

In response to the vital need to improve targeting of prevention services, this paper answers the following five questions: What is the pattern of subsequent shelter entry for individual applicants for prevention services? Which risk factors contribute to shelter entry for this group? How do risk factors for shelter use vary between families and individual applicants? Are some applicants at such high risk that prevention services make little difference? How does the efficiency for an empirical model for individuals compare with decisions made by service providers in the absence of such a model?

**How Do Families and Individuals Differ?**

Homelessness plagues both families and individuals, but characteristics differ across these groups. When comparing NYC to the United States (US) as a whole, the trends in homelessness rates for individuals in shelter appear to be headed in opposite directions. From 2007 to 2012, homelessness rates for individuals in shelter fell by 6.5% nationally (HUD, 2013b), compared to a 15.2% increase in NYC (HUD, 2013a). Although the rates of national homelessness are based on Homelessness Management Information Systems (HMIS) data covering an entire year and the NYC rates are based on Point-in-time (PIT) estimates for a single night, the opposite trends are noteworthy.

Risk profiles from the national data illustrate *how* families and individuals

experiencing homelessness differ. For example, in 2011, African Americans were

overrepresented in both groups, but especially among families. Additionally, single

homeless people had higher rates of disabilities (HUD, 2012a). Other studies found that

baby boomers have maintained highest risk status for single adult homelessness, leading

to an increase in average age for this group over the last three decades, whereas a similar

aging trend is not evident for families (Culhane et al., 2013). Further, among individuals,

longer durations of homelessness were associated with older age and arrest history (Caton

et al., 2005).

**Why Do Families and Individuals Differ?**

A description of the structural forces that shape homelessness suggests *why* rates

of homelessness for families and individuals differ. Such descriptions illuminate how

rates of homelessness may shift with changes in economic and social policy. For

example, during the second half of the twentieth century, contemporary homelessness

emerged first for individuals and then for families in step with shifts in political, social,

and economic forces. Rossi (1994) attributed longitudinal shifts in the characteristics of

homeless people to macro-level changes. For example, in the 1950s and 1960s, older

single men living in urban "skid rows" exemplified homelessness. Contrastingly, family

homelessness appeared in the 1970s and early 1980s with greater unemployment, an

economic recession, and less financial assistance.

In spite of an economic boom in the late 1980s, the wealthiest Americans'

financial gains failed to "trickle down" to the poorest households (Rossi, 1994). In the

1990s, many states restructured welfare programs to require employment, a challenge for

single mothers (Shinn, 2010). Discrimination across multiple domains  (e.g., housing, employment, imprisonment) is another structural force that likely leads to greater rates of homelessness for minorities (Shinn, 2010). These macro-level factors relate strongly to the growing proportions of people experiencing homelessness throughout the second half of the twentieth century.  Additionally, the financial costs of housing further differentiate families and individuals. Families require larger units than individuals; young families are at particular risk because of the high costs of childcare. For individuals, financial burdens may be lower, but disabilities and more restrictions on public assistance than those for families impede housing affordability. Across at-risk groups, but especially for families with children, a lack of affordable housing pushes financially constrained households into homelessness (Shinn & Weitzman, 1994).

Disabilities, such as mental illness and substance abuse are factors that are often included in discussions of homelessness. In 2011, national rates of mental illness (13.3%), substance abuse (6.1%), or both (2.2%) (NSDUH, 2013, US Census, 2011) were much lower than rates of sheltered persons who had a severe mental illness (26.2%) or of sheltered adults who experienced chronic substance abuse (34.7%) (Paquette, 2011). For chronically homeless individuals, the rates are even higher - 30% experienced mental health problems and approximately 50% exhibited co-occurring substance abuse (Paquette, 2011). Further, the exclusion of substance abuse as a disability to qualify for Supplemental Security Income (SSI) likely exacerbated already constrained financial situations, especially for single adults (Burt, 2001). Alternatively, individuals with mental illness might alternate repeatedly between shelters, jails, and substance or mental health facilities to compensate for a lack of stable housing with supportive services, a

30

phenomenon that is sometimes called the "institutional circuit" (Hopper et al., 1997, p. 659).

Finally, single parenthood has been blamed for increasing homelessness rates. Jencks (1995) argues that homelessness rates increased from 1969 to 1989 when "unskilled women not only married less but continued to have children [which] pushed more of them into the streets." (p. 58). Although family structure is related to poverty, among poor households, to my knowledge studies do not suggest that two-parent households offer additional protection from homelessness (c.f., Shinn et al., 2013). Actual increases in homelessness rates for single mothers are more likely due to the drastic restructuring of social policy, leading to decreased assistance for single mothers (Rossi, 1994).

In sum, the characteristics of families and individuals who experience homelessness tend to differ in response to economic, social, and political structures. Structural barriers, however, cannot explain all causes of homelessness. Rather, the complex causes of homelessness likely originate from both structural and individual levels (Lee et al., 2010). For example, potential individual-level risk factors for homelessness include the death of a spouse, a mental disorder, and experiences of domestic violence (Bassuk et al., 2001). Ending homelessness requires making housing more affordable. In the absence of such structural changes, people continue to lose homes. Provision of rapid, effective targeted homelessness prevention to those at risk can reduce the immediate financial and emotional costs of shelter entry.

**Risk Factors and Model Efficiency**

Many studies assess risk factors for homelessness, but for purposes of prevention, investigations should also assess how efficiently a collection of risks organized into a targeting model can select people at risk for homelessness (Burt et al., 2007; Shinn & Greer, 2011). To test the efficiency of a model, evaluations should examine hit rates and false-alarm rates at various levels of assessed risk (Shinn et al., 1998; Shinn et al., 2013; Swets, 1996). The hit rate is defined as the proportion correctly predicted to enter shelter among all shelter entrants. The false-alarm rate is defined as the proportion of households incorrectly predicted to enter shelter among all people who avoid shelter entry.[4] Notably, the denominators of these rates differ. For the hit rate, the denominator includes all cases where the outcome occurs. For the false-alarm rate, the denominator includes all cases where the outcome is absent. In the case of a continuous risk model, service providers can provide services to all who exceed some cutoff of risk, with that cutoff suggesting a particular trade-off between hit rates and false-alarm rates (Shinn et al., 2013).

Models predicting homelessness tend to have low hit rates, unless researchers and policy makers are willing to tolerate high false-alarm rates. For example, in a nationally representative sample, Hudson and Vissing correctly predicted 2.6% of the people who self reported an experience of homelessness at a false-alarm rate of 0.1%. The authors chose such a low cutoff for risk because the false alarm rate applied to the entire population of the nation. This study used demographic, socio-economic, and mental illness predictors, but the authors ignored differences between families and individuals.

---

[4] In the language of epidemiology, the hit rate is synonymous with sensitivity. The false-alarm rate is one minus specificity, where specificity is the proportion correctly predicted to avoid the unwanted outcome among all who avoid it.

Other investigations modeled homelessness risk for families (Barnett et al., 2011; Shinn et al., 1998; Shinn et al., 2013) and chose higher false alarm rates because they applied to more select populations. With a targeting model, Shinn et al. (1998) correctly identified 66% of shelter entrants (i.e., the hit rate) with a false-alarm rate of only 10% of families receiving public assistance in NYC. Families receiving public assistance are a much more select group than the national population, nonetheless, offering services to 10% of the public assistance caseload at the time of the study would have meant that over 80% of services would have gone to people who would avoid shelter without them (Shinn et al., 2001). In an even more select sample of 2,602 homeless families, half of whom participated in the rapid exit program in Hennepin County, Barnett et al. (2011) attempted to predict shelter re-entry and found a hit rate of 48% of re-entrants with a false-alarm rate of 23%, or those who were predicted to re-enter shelter but who did not.

A recent investigation examined the efficiency of targeting models for families who applied to the HomeBase prevention program in New York City. Shinn et al. (2013) used Cox proportional hazards modeling to identify risk factors for shelter entry over three years among 11,105 families who applied for HomeBase services. The authors calculated that if HomeBase continued to serve the same percentage of applicants (66.5%) but selected them according to the targeting model rather than worker judgments, they would improve the hit rate to 90.4% from 71.6% among applicants who entered shelter, at the expense of a false-alarm rate of 65.7% among applicants who remained housed. However, targeting remains difficult: even in the highest decile of risk, only 44% of families who failed to receive services entered shelter.

To my knowledge, only one study considers the efficiency of predictive models for any population subgroups other than families. Greenberg et al., 2006 created a model that predicted rates of subsequent homelessness for previously homeless veterans. They found that for the lowest-risk group, 2.9% of veterans experienced subsequent homelessness, whereas 27.6% of the highest-risk group experienced homelessness again. The authors found that better housing outcomes originated from 1) entering the program without a status of homeless, 2) receiving treatment in a substance abuse or psychiatric program rather than a medical program, and 3) having greater income or access to financial assistance. The authors do not report false-alarm rates.

The current study adds to the literature by developing a risk model for subsequent shelter entry for individuals who applied for prevention services in NYC. This investigation then compares the risk factors for individuals and families. Additionally, the current study examines the efficiency of the targeting models for the two groups and whether some adults are at too high risk to benefit from prevention efforts. The investigation permits targeting of prevention services to individuals who will benefit most.

**Methods**

Participants were 10,220 individuals who applied for NYC's HomeBase prevention services from September 28, 2004, to December 29, 2010. Overall, the sample contained mostly females (61%), African Americans (56%), and high school graduates (59%). Further, the majority was middle-aged (median age = 46), employed (66%), without a veteran status (97%), unmarried (88%), without a history of a mental health

diagnosis (79%), and without a history of substance abuse (82%). Service providers determined the eligibility of applicants for services.

**Variables**

At application, intake workers surveyed participants about the following domains: demographic variables, human capital, housing conditions, disability, interpersonal discord, childhood experiences, and shelter history. (Variables used in analysis are shown in Table 1.) The New York City Department of Homeless Services (DHS) merged these survey results with administrative records of applicants' previous interactions with the DHS shelter system, and the date of any subsequent shelter entry.

**Analyses**

The current study develops a risk model predicting subsequent shelter entry. It uses survival analysis (Cox proportional hazards) to model time to any subsequent shelter entry in days, from risk factors among individual applicants for prevention services. The model includes a dichotomous variable indicating whether the individual avoided shelter throughout the follow-up period. Use of survival analysis, rather than logistic regression was important because individuals had different follow-up periods; additionally survival analysis models time to shelter entry and not simply whether shelter entry occurred. I impute 50 data sets for missing data with STATA, including auxiliary variables according to the literature (Graham, et al., 2007; Sinharay et al., 2001)[5].

I compare results to those found in a previous investigation of families that applied for HomeBase services (Shinn et al., 2013) with respect to both risk factors and rates of shelter entry among applicants who were judged eligible for services (and

---

[5] See appendix for additional notes about multiple imputation.

presumably received them) and those judged ineligible by level of risk. Next, I develop a short screening model to streamline service delivery. Finally, as described in more detail below, I examine the efficiency of the model.

Following methods developed in Shinn et al. (2013, the current chapter employs three survival analysis techniques: 1) Kaplan-Meier survival curves, 2) Hazard Functions, and 3) Cox proportional hazards models. First, Kaplan-Meier survival curves show the momentary probability that an individual has not experienced homelessness as a function of time, beginning at the point of application for prevention services. Kaplan-Meier estimates are computed as the product of the proportion of individuals avoiding an outcome at a particular interval and the proportion of individuals avoiding the outcome at all previous intervals (Kleinbaum & Klein, 2005). By definition, the cumulative proportion can only stay the same or decrease over time. Kaplan-Meier curves are useful because they provide the length of time avoiding outcomes. Steep slopes represent rapid occurrences of the outcome, whereas gradual slopes signal longer times avoiding the outcome (Luke, 1993).

While survival curves provide the cumulative proportion avoiding an outcome, hazard curves show the instantaneous rate of experiencing an outcome at each time, given that the outcome has not already occurred. The hazard function is slightly more complex than the survival function, and can be represented with the following formula:

$$h_i = 2q_i/w_i(1 + p_i), \qquad (1)$$

where $h_i$ is the instantaneous hazard rate, $q_i$ is the proportion experiencing the outcome at some time point, $w_i$ is the width of the time interval, and $p_i$ is the proportion avoiding the outcome, with the denominator for both proportions being the people who are still

eligible to experience the outcome for the first time (Luke, 1993). The hazard function is useful, because it shows the times for which risk for the outcome is highest and lowest, and how that risk changes over time.

Cox proportional hazards models are a popular semi-parametric technique to model the effects of multiple covariates that may shift the hazard function up or down. The model assumes that the ratio of hazards across levels of the covariate is constant (i.e., proportional) across time (Cox, 1972). The Cox proportional hazards model thus shows which risk factors increase or decrease the risk (i.e., hazard rate) for individuals over time. The current chapter uses Cox proportional hazards to model risk for experiencing shelter entry after households in applied for HomeBase services. Following the full model, I create a short screening model to address two issues with the full model. First, the full model is likely overfitted, and second, the challenges of collecting such a large number of variables led to large amounts of missing data. I create a parsimonious screening model by eliminating non-significant variables via backwards regression and then verifying that each remained non-significant when added back to the final model. I also check that each predictor remained significant in the robustness tests in two random subsamples.

Next, the current chapter examines the efficiency of these models – their hit rates relative to their false alarm rates. Hit rates and false alarm rates are related to type I and type II errors. Figure 2.1 shows the relationship of model predictions with actual experiences of the outcome of interest. Box A refers to hits, or correct model predictions of outcomes that actually occur. Type I errors (box B) refer to false alarms, or model predictions of outcomes that do not actually occur. Type II errors (box C) refer to misses

or failures of the model to predict outcomes that occur. Box D refers to correct model predictions that an outcome will not actually occur. The hit rate from Figure 2.1 would be: $\frac{A}{A+C}$. The false alarm rate from Figure 2.1 would be: $\frac{B}{B+D}$. It is important to note that the denominators of hit rates and false alarm rates are different; for the hit rate, it is all cases where the outcome actually occurs, and for the false-alarm rate, it is all the cases where the outcome is absent.

**Figure 2.1**. Contingency Table for Hit Rates and False alarm Rates

|  | **Actual Outcome Present** | **Actual Outcome Absent** |
|---|---|---|
| **Model Prediction Present** | (A) Correctly Predicted Hits | (B) False Alarms Type I error |
| **Model Prediction Absent** | (C) Misses Type II error | (D) Correctly Predicted Negative |

For the current chapter, the hit rate is the proportion correctly predicted to enter shelter among all who actually experience the outcome. The false-alarm rate is the proportion of households incorrectly predicted to enter shelter among all people who avoid it. Any model will generate multiple hit rates and false alarm rates depending on what cutoff is used for risk. When individuals with any risk factors are predicted to experience the outcome the hit rate will be high, but so will the false alarm rate. When only individuals with many risk factors are predicted to experience the outcome, both hit rates and false alarm rates will be lower.

A receiver-operating characteristic (ROC) is a graph of hit rates against corresponding false alarm rates for all possible cutoffs of risk scores. Computing an ROC curve requires a dichotomous outcome (e.g., shelter entry vs. not). In the present study

ROC curves were generated using logistic regression, where predicted scores from a logistic regression were averaged across the fifty imputed data sets to create an average risk score for shelter entry.

ROC curves can be used to compare competing models with the goal of selecting the model with the highest hit rates as compared to the lowest false-alarm rates (Swets, 1996) across levels of risk (see Figure 2.2). Where the model does no better than randomly placing households into positive or negative classifications, the ROC curve would fall directly on the 45-degree reference line in Figure 2.2. To the extent that the model predicts an outcome perfectly, the hit rate would be 1 and the false alarm rate would be 0. Thus the curve would be pulled to the upper left corner (see ideal curve below in Figure 2.2). Most ROC curves fall between these extremes, and ROC curves from different models can be compared to select the model that most closely approaches the upper left corner. The ROC curve can also be used to select cutoffs in risk scores to make people eligible for services. Because hit rates increase with increasing false alarm rates along the ROC curve, policy makers can decide how many false alarms they can tolerate in order to obtain as many correct hits as possible.

**Figure 2.2.** Example of ROC Curves



Example of ROC Curves

Ideal Curve

Example of
Model Curve

Hit
Rate

Reference Line of Zero Predictability

False-Alarm Rate (1 – Specificity)
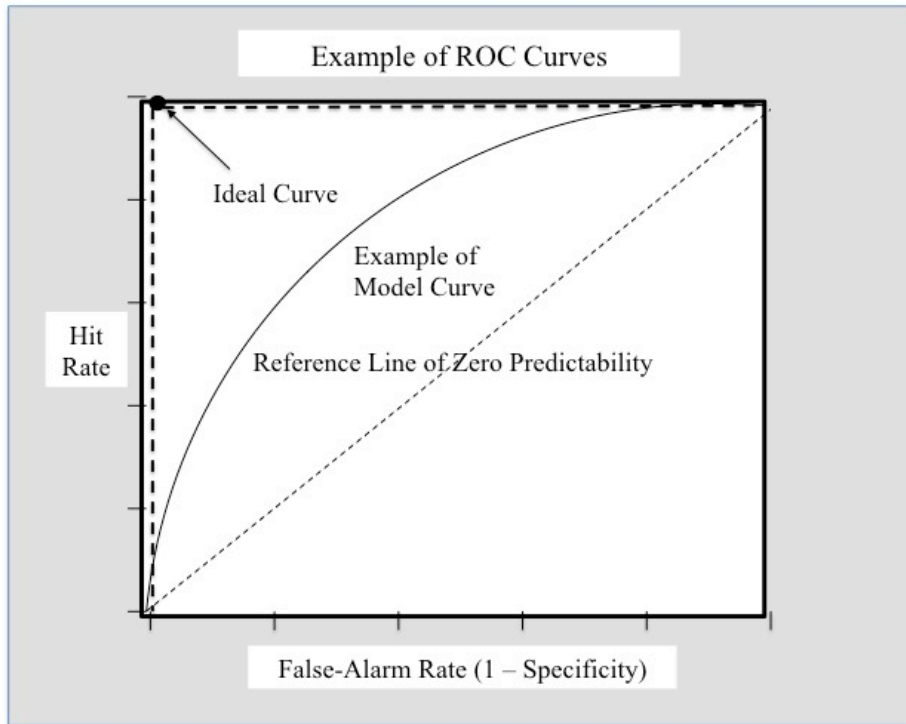
**Results**

Table 2.1 contains descriptive statistics, hazard ratios, and confidence intervals for the model predicting shelter entry for individuals. Only 5.4% of those who applied for services entered shelter subsequently (over the next 2 to 8 years), and the majority of people who entered shelter did so within one year of applying for services. The hazard ratio and 95% confidence interval for each predictor is adjusted for all other variables in the model. Hazard ratios represent the amount by which the rate of shelter entry is multiplied for people who exhibit the characteristic (or for continuous variables, the multiple for each additional increment such as year of age), adjusted for other variables.

Among demographic variables, only age made a reliable contribution to the model. Controlling for the other variables in the model, younger applicants were more likely to enter shelter. None of the human-capital variables contributed reliably to the full

model. For housing conditions, rent arrears and threats of eviction contributed to the model. Increasing arrears were associated with significantly higher rates of shelter entry. Findings for threats of eviction were mixed. Individuals who were threatened with verbal eviction entered shelter at a rate that was more than two times the rate of those who were not threatened verbally. On the other hand, those who faced a legal eviction action entered shelter at slightly less than two-thirds the rate of those who did not indicate a legal eviction threat.[6]

For disability/criminal justice variables, none of the variables in the categories of interpersonal discord or childhood experiences contributed reliably to the full model. For shelter-history variables, a self-reported shelter application in the last three months increased hazard for shelter entry by over 2.5 times. Individuals who were reintegrating into the community from an institution entered shelter at more than 1.3 times the rate of those who were not reintegrating. Individuals who had a previous shelter stay were over 18.5 times more likely to enter shelter than those without a previous shelter stay.

Some of these results were unexpected, so I explore them further. I start with the seemingly protective effect of legal eviction threats. If service providers are likely to target a factor, such as legal eviction, that increases risk in the absence of services, and they are able to counteract such risk, the net effect of the factor might be zero or even protective. Under these circumstances the factor might appear to confer risk for those who did not receive services and protection for those who received services. In statistical terms one would say that services interacted with the factor in predicting shelter entry. To determine whether this was the case for eviction or any other predictor I looked

---

[6] I reexamined the different types of eviction for families and did not find the same apparent protective effect of legal eviction as I found for individuals.

systematically for such statistical interactions showing differential associations of variables with shelter entry for individuals who were and were not eligible for services. After finding no significant interactions, I investigated the relationship between legal eviction and subsequent shelter entry for all ineligible individuals. The association, while still protective, approached zero (HR = .92).

The strongest support for the idea that legal eviction might appear protective only because of an association with services comes from the subset of ineligible individuals who lived outside of the community district (n = 907). For these applicants, who could not receive services, legal eviction was a risk factor for subsequent shelter entry (HR = 1.34). For this reason, I exclude legal eviction as a protective factor for homelessness in the screening model, even though it serves to reduce the predictive power of the model in a combined sample of those who did and did not receive services. Although it is not surprising that previous shelter stays are associated with subsequent shelter entry, the magnitude of the effect (HR= 18.6) is impressive. I explore this relationship further in the next section.

**Table 2.1.** Descriptive Data, Adjusted Hazard Ratios, & Confidence Intervals for Predictors of Shelter Entry in Cox Regression (n = 10,220) *Continuous Variables in Italics*

| Predictor[a] | No Shelter % or mean n = 9,663 | Shelter % or mean n = 557 | Hazard Ratio | 95% Confidence Interval |
|---|---|---|---|---|
| **Demographics** | | | | |
| Male | 38.1 | 45.5 | 0.979 | 0.735-1.303 |
| African American[a] | 55.1 | 69.8 | 0.859 | 0.544-1.356 |
| Hispanic | 37.5 | 22.5 | 0.684 | 0.418-1.117 |
| English Speaker | 71.6 | 95.0 | 1.506 | 0.865-2.623 |
| *Age* | *44.6* | *41.6* | *0.977*[***] | *0.967-0.987* |
| Married/partner | 11.5 | 21.1 | 1.013 | 0.726-1.413 |
| Veteran | 2.8 | 4.3 | 1.077 | 0.529-2.192 |
| **Human Capital** | | | | |
| High school /GED | 59.0 | 57.7 | 0.889 | 0.690-1.145 |
| Currently employed | 55.1 | 71.7 | 1.075 | 0.715-1.618 |
| Currently receiving public assistance | 56.8 | 62.0 | 1.630 | 0.969-2.742 |
| Lost benefits in past year | 10.4 | 17.5 | 1.075 | 0.636-1.814 |
| **Housing Conditions** | | | | |
| Name on lease | 45.7 | 65.3 | 0.627 | 0.343-1.146 |
| *Arrears[c]* | *$1600* | *$3429* | *1.018*[***] | *1.008-1.027* |
| Overcrowding or Discord [b] | 19.1 | 14.3 | 0.866 | 0.586-1.280 |
| Doubled up | 26.8 | 19.8 | 1.459 | 0.944-2.255 |
| Verbal Eviction threat | 13.2 | 29.6 | 2.085[***] | 1.353-3.212 |
| Legal Eviction action | 32.5 | 28.4 | 0.648[*] | 0.456-0.921 |
| Rent > 50% income | 38.0 | 47.6 | 1.211 | 0.809-1.811 |
| Unsafe conditions | 6.4 | 10.1 | 1.072 | 0.721-1.593 |
| *Level of disrepair* | *4.2* | *3.2* | *0.613* | *0.260-1.442* |
| *Moves in past year* | *0.7* | *0.6* | *0.797* | *0.629-1.010* |
| Currently receiving subsidy | 4.9 | 5.5 | 1.194 | 0.629-2.264 |
| **Disability/Criminal Justice** | | | | |
| Chronic health/hospitalization | 53.6 | 44.7 | 0.969 | 0.605-1.553 |
| Mental illness/ hospitalization | 21.6 | 18.4 | 0.615[7] | 0.386-0.981 |
| Substance problem/treatment | 17.2 | 25.0 | 1.195 | 0.464-3.077 |
| Criminal justice involvement[d] | 21.1 | 31.0 | 0.885 | 0.581-1.348 |
| **Interpersonal Discord** | | | | |
| Domestic violence[e] | 15.9 | 13.8 | 1.003 | 0.602-1.672 |

---

[7] To create a robust model, I estimated it initially in two independent random subsamples of 50% of the data. (For each, I imputed fifty datasets based only on the information in the subsample.) The resulting models were substantially similar to the complete model with the exception that mental illness failed to be a reliable predictor in either subsample. Accordingly, mental illness is omitted from the screening survey.

| | | | | |
|---|---|---|---|---|
| Protective services involvement[f] | 4.3 | 6.7 | 0.890 | 0.386-2.052 |
| *Discord rating[g]* | *2.1* | *1.6* | *0.962* | *0.819-1.130* |
| **Childhood Experiences** | | | | |
| *Adversity index[h]* | *0.5* | *0.6* | *1.058* | *0.870-1.286* |
| **Shelter History (self report)** | | | | |
| Shelter history as adult | 23.4 | 75.0 | 1.400 | 0.872-2.250 |
| Shelter app. last 3 mos. | 3.3 | 17.1 | 2.517*** | 1.758-3.604 |
| Reintegrating into community | 10.4 | 27.5 | 1.372* | 1.058-1.780 |
| **By Administrative Data** | | | | |
| Previous shelter stay | 7.2 | 70.6 | 18.561*** | 12.620-27.297 |

*Notes.*
[a] Omitted Race/ethnicity category is All Other
[b] Overcrowding and discord were combined in the original data set
[c] Truncated at $15,000, HR and CI in units of $100
[d] Any family member ever incarcerated or respondent on probation or parole
[e] Experienced domestic violence or violence in past year
[f] ACS investigation in past year, open case, child ever in foster care, currently in protective care
[g] Discord rating (9-point scale) with landlord, leaseholder, or household members
[h] Count of 5 experiences in childhood: family receipt of public assistance, abuse, shelter, foster care, 4 or more residential moves
 $* = p \le .05$ $** = p \le .01$ $*** = p \le .001$

**Previous Shelter Stays**

As shown in Table 1, 70.6% of HomeBase applicants who later entered shelter

had been in New York City shelters previously, compared to just 7.2% of those who

avoided shelter.  To explore this relationship further, I graphed the proportions of

individuals with and without prior shelter histories who stayed out of shelter over time

(see Figure 2.3). Among HomeBase applicants without previous shelter stays, the

probability of avoiding shelter entry remained above .95 throughout the study period. For

those with previous shelter stays, however, the probability of avoiding shelter entry was

just over .5 by about four years after the HomeBase application.[8] No other variable was

---

[8] These survival probabilities are lower than for the full sample, because the average person in the full sample was not observed as long. The survival graph for individuals with previous shelter stays levels off after just over four years because no further entries were observed among the small group of such individuals who applied for HomeBase early enough to have such a long follow-up observation period.

nearly so powerful a predictor of shelter entry, either taken alone or controlling for other

variables.[9]

**Figure 2.3.** Cumulative Survival Estimates for Shelter Entry
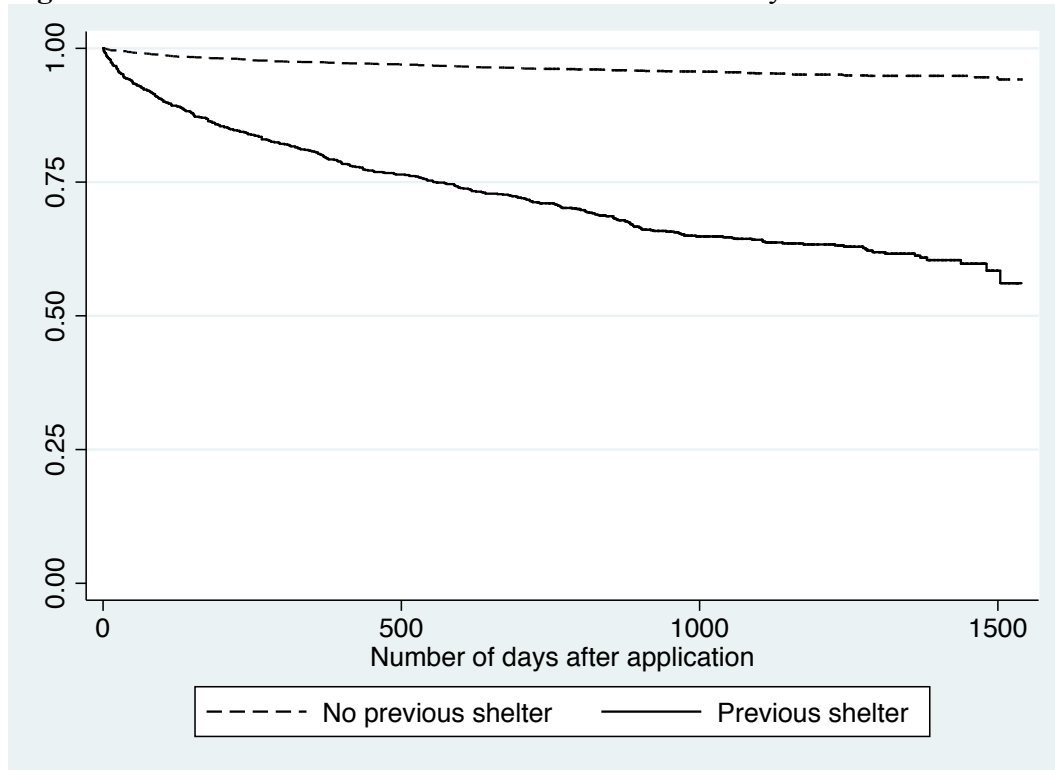


### Risk Model for Individuals

After looking at risk factors for homelessness, I investigated whether households

could be at such high risk that prevention would make little difference in subsequent rates

of shelter entry. This does not appear to be the case. Figure 2.4 shows the proportion of

individuals who entered shelter by decile of risk (calculated by averaging predicted

---

[9] I further investigated separate models for individuals with and without a previous
shelter stay. Individuals with a previous shelter stay (n = 1084) had similar predictors of
subsequent shelter entry to the overall model, with all significant variables remaining in
the model (in the same direction of risk) except currently receiving public assistance and
mental illness/hospitalization. Individuals without a previous shelter stay (n = 9125),
however, had only three significant predictors of subsequent shelter entry: age, arrears,
and verbal eviction threat. A combined model, stratified by shelter entry, produced
similar substantive results to the full model with the exception that mental
illness/hospitalization fell out of the stratified model.

scores across fifty imputed data sets) separately for adults judged eligible or ineligible for

HomeBase services, and Figure 2.5 shows the parallel model for families, from a

previous investigation (Shinn et al., 2013). Ignoring eligibility, the probability of shelter

entry was similar for families (Fs) and individuals (Is) at the lowest decile of risk (Fs =

1% and Is = 0%) and at the highest (Fs = 37% and Is = 38%).

The risk models for individuals and families differed at intermediate risk deciles.

The proportion of individuals who entered shelter remained close to 0 until the ninth

decile, when it rose rapidly. By way of contrast, risk rose more smoothly for families,

with a tenth or more of families entering shelter at each decile in the top half of the risk

distribution. Services seemed most helpful for individuals in the tenth decile of risk,

although they also appeared to make some difference for those in the ninth decile, as

indexed by the difference in shelter entry rates between those judged eligible and

ineligible for services. Services did not appear to matter for individuals or families in risk

deciles who rarely entered shelter, most likely because there was little risk to avert. On

the other hand, the fact that legal eviction was associated with staying out of shelter

might be taken as effectiveness of eviction prevention services. Similar to families, the

majority of individuals avoid shelter entry, even in the highest risk decile.

**Figure 2.4.** Rate of Shelter Entry for Deciles of Risk by Eligibility Status for Individuals
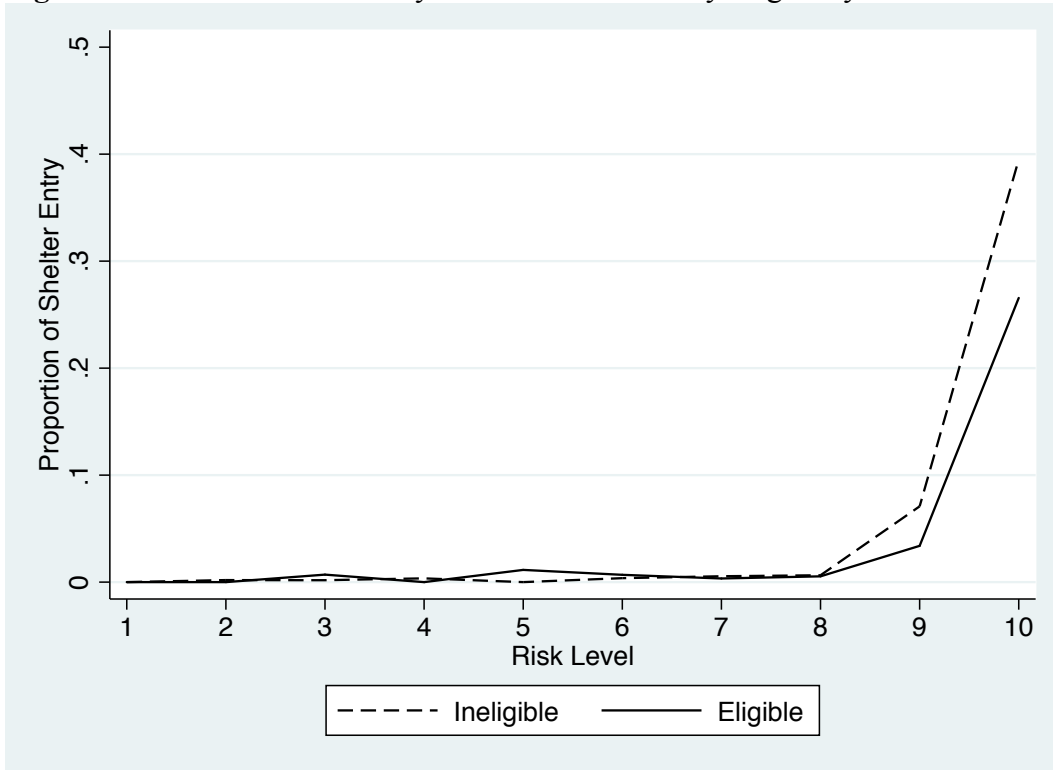


**Figure 2.5.** Rate of Shelter Entry for Deciles of Risk by Eligibility Status for Families
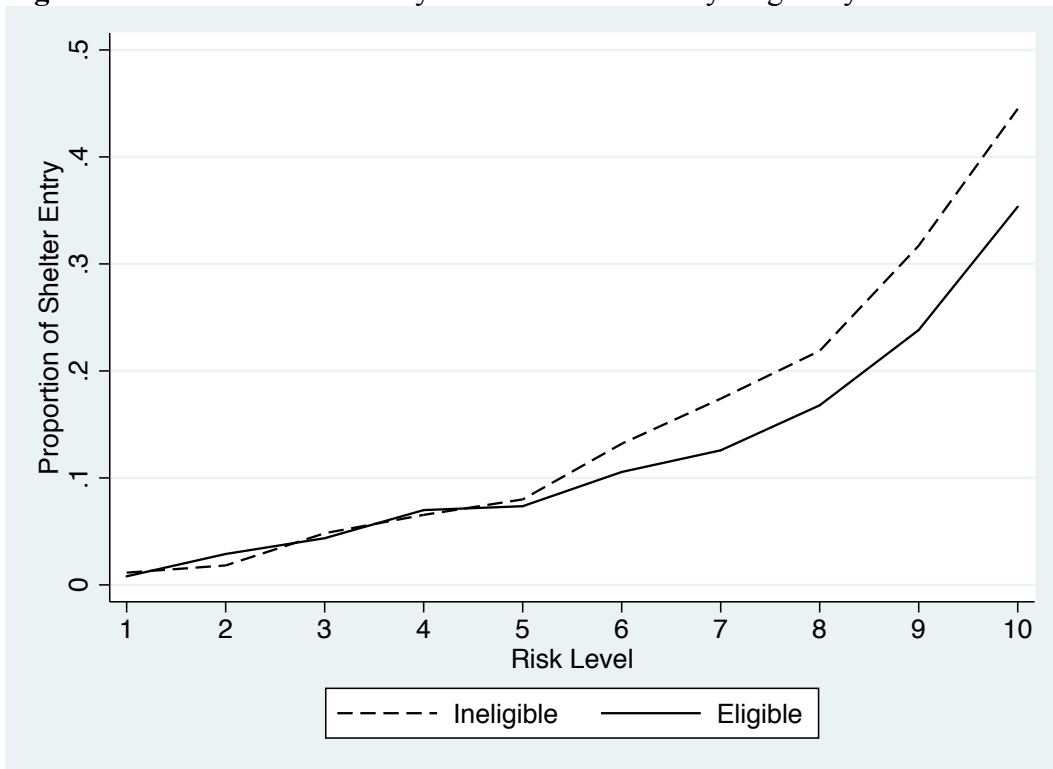
**Screening Model**

As described in the methods section, I created a short screening model by eliminating non-significant variables via backwards regression and then verifying that each remained non-significant when added back to the final model. I added public assistance to the model at this stage, because it became significant after other correlated variables were removed. In line with the body of forecasting literature (Dawes, 1979; Dawes & Corrigan, 1974; Dana & Dawes, 2004) and the previous study of homeless families (Shinn et al., 2013), I assigned from one to six points based on the comparative magnitudes of regression coefficients for dichotomous predictors and shelter entry rates at specific intervals for continuous predictors.

Table 2.2 introduces the screening model. Individuals could score from 0 to 16 points across seven variables, with increasing scores associated with increased risk of shelter entry. Previous shelter stays are worth the most points. Actual scores (averaged across 50 imputed data sets) ranged from 0 to 14.1 points with the median score of 1.4. Receiving a score of three or more placed individuals in the ninth decile of risk, and a score of seven or more placed individuals in the highest decile of risk.  Thus almost all applicants with a previous shelter stay (six points) were in the top decile of risk, and few applicants without such a stay reached the top decile.

**Table 2.2.** Screening Model Predicting Individuals Who Should Receive HomeBase Services

1 point
- Reintegrating into the community
- Currently receiving public assistance

2 points
- Verbal eviction threat
- Reports applying for shelter in last 3 months

6 points
- Administrative record of previous shelter stay

Age
- 29-32 years: 1 point
- 28 or under: 2 points

Arrears
- $5000-$8000: 1 point
- $8000 or greater: 2 points

**Model efficiency.** Figure 2.6 shows the efficiency of the resulting models. I plotted the hit rates compared to false-alarm rates for the full model and the screening model. In addition to the tradeoff in efficiency for the full and screening models, Figure 2.6 shows point estimates for one-variable models based on whether administrative records showed that the respondent had been in shelter previously (*previous shelter)* and whether the intake worker deemed the respondent eligible for services (*worker)*.

Several conclusions about the efficiency of various models are evident in figure 2.6. First, the full model is only slightly more efficient than the screening model at high levels of risk, but departs at lower levels. Second, the Previous Shelter Model exhibited a high hit rate (70.6%) compared to a low false-alarm rate (7.2%), and was far more efficient than worker decisions during the period under study (hit rate of 50.7% vs. false alarm rate of 43.4%) among those deemed eligible for services. This comparison

excluded those who were ineligible because they were outside of the service area or

refused services (n = 1,137). Additionally, service providers would serve fewer applicants

with the Previous Shelter Model (10.7% of applicants for Previous Shelter model; 39.1%

currently). Holding the proportion of applicants served constant at 39.1%, the Screening

Model would increase the current hit rate to over 90% and misses would fall by over

85%.

**Figure 2.6**. ROC Curves for Model Efficiency
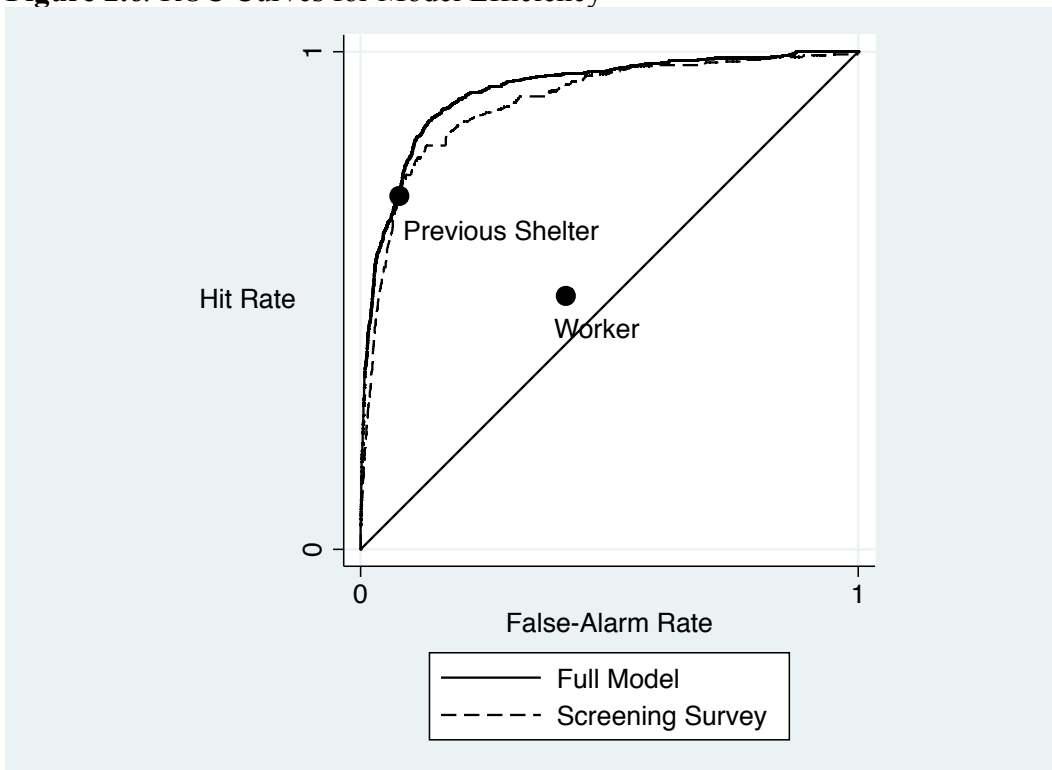


As a global comparison of efficiency, I investigated the Area Under the Curve

(AUC) values for full and Screening Models across families and individuals. If the

estimated curve fell along the 50% reference line (i.e., the diagonal line in figure 2.6), the

resulting AUC would be .50, whereas a perfect model would capture all of the area (i.e.,

the AUC value would be 1.00). I compared the resulting AUC estimates, after using Stata

software to estimate nonparametric ROC curves with bootstrapping for inference. The

full model for individuals had an AUC of .92 (CI .91 - .94), and the Screening Model had an AUC of .90 (CI .88 - .91). The full model for families had an AUC of .76 (CI .74 - .77), and the Screening Model had an AUC of .74 (CI .73 - .75). Accordingly, the Screening Model for individuals is substantially more efficient than the model for families.

As a further test of the robustness of the Screening Model, I examined how well it predicted shelter entry for people deemed ineligible for services for different reasons. By targeting 39.1% of applicants (the same proportion offered services currently) with the screening model, I identified 87% of the 295 applicants who were deemed ineligible for services but who entered shelter subsequently. This includes 89% of 46 individuals thought to have insufficient housing risk, 88% of 161 deemed eligible for a more appropriate program, 97% of 29 who did not comply with the intake process, 100% of 19 who refused services, and 89% of 27 who lived outside of the community district.

**Comparisons to Families**

I compared the model developed here with a model previously developed for families applying to HomeBase. Although the dates of applications were different for individuals and families, comparisons between the groups are useful to assess how the targeting of services may differ across groups. Individuals entered shelter at half the rate of families over a three-year period (6.4% vs. 12.8%).[10] Overall individuals received services at lower rates (39.1%) than did families (66.5%). Further, risk factors for individuals differed from those for families.

---

[10] Compared to previous analyses, the rate of shelter entry for individuals is slightly higher when I include only cases with three years of data.

Table 2.3 provides descriptive statistics for applicants for prevention services who avoided or entered shelter by household status. Some differences in characteristics of the two groups of shelter entrants are especially noteworthy – those reported are significant by t-test or chi-square analysis at the .01 level. For demographic variables individual shelter entrants were more likely to be male and African American and less likely to be Hispanic than heads of families who entered shelter. The individuals were also older, and more likely to be married, to speak English, and to be veterans.

For human capital variables, individual shelter entrants were more likely to have a high-school diploma, be employed, and be a recipient of public assistance.  For housing variables, individual entrants were twice as likely to be current leaseholders and a third as likely to be doubled up as family shelter entrants. Perhaps as a result, they had much lower rates of overcrowding or discord. They were more likely to pay over half of their incomes for housing and had higher arrears, on average.  Individuals reported higher levels of disrepair but had moved less frequently than families.  Additionally, for each of the housing variables, the trend in percentages across shelter status was in the opposite direction for individuals and families.

Rates of chronic health and of mental illness did not differ significantly across groups of shelter entrants. However, individual shelter entrants were more likely to have histories of substance abuse and criminal justice involvement than family entrants. Individual shelter entrants reported lower levels of all forms of interpersonal discord, including rated discord, domestic violence, and involvement with protective services

Individual shelter entrants had more prior involvement with the shelter system than their family counterparts.  They were more likely to report having stayed in shelter

previously, having applied for shelter in the last three months, and to be reintegrating into

the community from an institution (27.5% vs. 11.5%) than families. The rate of previous

shelter stays based on administrative records was much higher for individual shelter than

for family shelter entrants.

**Table 2.3.** Descriptive Comparisons of Individuals and Families, Continuous Variables in Italics

| | Individuals | | Families | |
|---|---|---|---|---|
| Predictor[a] | No Shelter | Shelter | No Shelter | Shelter |
| | % or *mean* | % or *mean* | % or *mean* | % or *mean* |
| | *n = 9,663* | *n = 557* | *n = 9,686* | *n = 1,149* |
| Demographics | | | | |
| **Male** | 38.1 | 45.5 | 9.8 | 6.7 |
| **African American** | 55.1 | 69.8 | 51.9 | 56.3 |
| **Hispanic** | 37.5 | 22.5 | 45.5 | 41.3 |
| **English Speaker** | 71.6 | 95.0 | 77.7 | 86.3 |
| ***Age*** | *44.6* | *41.6* | *33.7* | *30.1* |
| **Married/partner** | 11.5 | 21.1 | 13.7 | 13.5 |
| **Veteran** | 2.8 | 4.3 | .7 | .6 |
| Human Capital | | | | |
| **High school /GED** | 59.0 | 57.7 | 55.7 | 44.7 |
| **Currently employed** | 55.1 | 71.7 | 51.6 | 43.6 |
| **Currently receiving public assistance** | 56.8 | 62.0 | 26.9 | 37.5 |
| **Lost benefits in past year** | 10.4 | 17.5 | 14.3 | 19.9[ns] |
| Housing Conditions | | | | |
| **Name on lease** | 45.7 | 65.3 | 38.3 | 30.0 |
| ***Arrears*** | *$1600* | *$3429* | *$1507* | *$1163* |
| **Overcrowding or Discord** | 19.1 | 14.3 | 39.2 | 54.0 |
| **Doubled up** | 26.8 | 19.8 | 47.2 | 63.6 |
| **Verbal Eviction threat**[c] | 13.2 | 29.6 | | |
| **Legal Eviction action**[c] | 32.5 | 28.4 | | |
| **Eviction**[f] | | | 55.3 | 66.1 |
| **Rent > 50% income** | 38.0 | 47.6 | 33.5 | 28.6 |
| **Unsafe conditions** | 6.4 | 10.1 | 9.1 | 11.3[ns] |
| ***Level of disrepair*** | *4.2* | *3.2* | *2.2* | *2.4* |
| ***Moves in past year*** | *0.7* | *0.6* | *1.0* | *1.3* |
| **Currently receiving subsidy** | 4.9 | 5.5 | 10.3 | 8.5[ns] |
| Disability/Criminal | | | | |

| | | | | |
|---|---|---|---|---|
| Justice | | | | |
| **Chronic health/hospitalization** | 53.6 | 44.7 | 42.2 | 44.9$^{ns}$ |
| **Mental illness/ hospitalization** | 21.6 | 18.4 | 12.4 | 13.9$^{ns}$ |
| **Substance problem/treatment** | 17.2 | 25.0 | 7.0 | 11.2 |
| **Criminal justice involvement** | 21.1 | 31.0 | 11.9 | 17.4 |
| Interpersonal Discord | | | | |
| **Domestic violence** | 15.9 | 13.8 | 24.7 | 29.3 |
| **Protective services involvement**[f] | 4.3 | 6.7 | 8.7 | 16.1 |
| *Discord rating* | *2.1* | *1.6* | *2.3* | *2.7* |
| Childhood Experiences | | | | |
| **Young Mother**[f] | | | 22.3 | 33.6 |
| *Adversity index* | *0.5* | *0.6* | *0.6* | *1.0* |
| Shelter History (self report) | | | | |
| **Shelter history as adult** | 23.4 | 75.0 | 24.4 | 47.0 |
| **Shelter app. last 3 mos.** | 3.3 | 17.1 | 3.1 | 11.0 |
| **Reintegrating into community** | 10.4 | 27.5 | 6.2 | 11.5 |
| By Administrative Data | | | | |
| **Previous shelter stay** | 7.2 | 70.6 | 10.8 | 25.0 |
| *Number of Previous Shelter Applications*[f] | | | *0.3* | *0.7* |
| **Found Eligible Previously**[f] | | | 2.4 | 8.0 |
| **Exit to Subsidy**[f] | | | 2.9 | 5.7 |

f- these variables were used in the families study only

c- these variables were used in the individuals study only

ns- chi-squared analyses or two-tailed t-tests between individual and family shelter entrants were not significant at p = .01

    Table 2.4 compares risk factors that significantly predicted shelter entry across the two groups. Families had many more predictors than did individuals, and for the most part, risk factors for individuals were a subset of those for families. The only new risk factor for single individuals was the amount of rent arrears. With the exception of legal eviction, discussed earlier, variables that contributed to shelter entry for both groups did

so in the same direction.  Note that for families, self-reports of previous shelter were more predictive than administrative records, perhaps because families were more likely to include domestic violence shelters that would not be part of Department of Homeless Services records.  For individuals, the administrative records were stronger predictors.  In each case, these variables were correlated, so that only one entered the final model.

**Table 2.4.** Comparing Screening Models of Risk Factors across Families and Individuals

| Screening Model Risk Factors | Significant risk factors for individuals[*] | Significant risk factors for families[*] |
|---|---|---|
| **Demographics** | | |
| Age | Protective | Protective |
| Child Under 2 yrs old | NA | Risk |
| Pregnant | NA[11] | Risk |
| **Human Capital** | | |
| High school /GED | | Protective |
| Currently employed | | Protective |
| Currently receiving public assistance | Risk | Risk |
| **Housing Conditions** | | |
| Name on lease | | Protective |
| Overall Eviction threat | | Risk |
|     Verbal Eviction threat | Risk | |
|     Legal Eviction action | Protective[**] | |
| Arrears | Risk | |
| Moves in past year | | Risk |
| **Disability/Criminal Justice** | | |
| Mental illness/ hospitalization | Protective[**] | |
| **Interpersonal Discord** | | |
| Protective services involvement | | Risk |
| Discord rating | | Risk |
| **Childhood Experiences** | | |
| Adversity index | | Risk |
| **Shelter History** | | |
|   **By Self Report** | | |
| Shelter history as adult | | Risk |
| Shelter application last 3 mos. | Risk | Risk |
| Reintegrating into community | Risk | Risk |
|   **By Administrative Data** | | |
| Previous shelter | Risk | |

*- significant variables included in a parsimonious model arrived at by eliminating non-significant predictors one at a time, and then checking whether any variables that had been eliminated regained predictive power after other variables had been eliminated. Public assistance re-entered the model after other correlated variables were eliminated.

** - variables that were eliminated for the screening model based on robustness analyses

---

[11] By definition, individuals could not be pregnant or have a child under 2 years old.

**Discussion**

The study developed a model for shelter entry among individuals who applied for HomeBase prevention services that was more efficient than the decisions of workers and also more efficient than a comparable model for families. One predictor stood out: The rate of shelter entry was much higher for individuals with a previous stay in homeless shelters.

Few other predictors contributed reliably to the full model. Additional significant risk indicators included lower age, higher arrears, verbal eviction threat, absence of a legal eviction threat, an application for shelter within three months, reintegrating from an institutional setting, and receipt of public assistance. Subsequent analyses cast doubt on the robustness of legal eviction and mental health predictors. Accordingly, these variables were eliminated from the Screening Model.

In deploying the Screening Model, providers can choose cut-off scores on the model that correspond to tradeoffs of hit rates and false-alarm rates. As noted our earlier study of families (Shinn et al., 2013), a choice about cutoffs is not simply a technical decision. Moral and ethical considerations, and costs to homeless people as well as to the City, should be considered.

Individuals differed from families in several ways. Only 5.4% of those adults subsequently entered shelter, a lower rate than for families. The lower rate of shelter entry for individuals among HomeBase applicants is consistent with the lower overall rate of shelter use by individuals than by families in NYC (HUD, 2012c). This pattern might not generalize beyond NYC because nationally, more shelter users are single adults (HUD, 2013a). Further, predictors of shelter entry were fewer for individuals than for

families. Finally, characteristics of individuals and families differed descriptively, in ways consistent with the literature (for a full description, see Table 2.3). However, in the context of other variables, such as previous shelter entry, many of the variables that distinguished individuals and families failed to predict shelter entry. For example, criminal justice involvement and substance abuse were higher for individuals, but neither variable predicted higher levels of shelter entry.

The comparison of individuals and families suggests that HomeBase is especially beneficial for applicants at the highest level of risk in both groups. For individuals, services did not seem to matter for applicants below the eighth decile of risk, most likely because there was little risk to avert. On the other hand the fact that legal eviction was associated with staying out of shelter might be taken as evidence for the effectiveness of eviction prevention services. Services seemed more helpful for individuals in the ninth decile of risk and above, as evidenced by the difference in rates of shelter entry for eligible and ineligible applicants. Risk rose faster for families, and services began to make a difference at about the sixth decile of risk.

Some limitations of this investigation were similar to limitations of our study of family homelessness. For example, omitted variable bias likely exists for the current investigation. Further, inevitable data inaccuracies and effective services could weaken prediction. Additionally, the face of homelessness changes over time for both families and individuals. One primary challenge with targeting research includes a tradeoff between timely models with current risk factors and the allowance of sufficient time for at-risk applicants to enter shelter so that models can be created and evaluated.  Both

studies of individuals and of families suggest that following applicants for at least a year is useful – the majority of shelter entries happen within the first year.

Some limitations are more applicable to the current study than for the investigation of homeless families. For example, HomeBase workers may successfully target services to individuals for particular risk factors. To the extent that workers successfully neutralize these risks, the risk factors would not contribute to the overall risk model, so that the effectiveness of services may be undetectable. Finally, I make a similar caution about model uptake in other locales as we did for the investigation of homeless families: the model may be a good starting point in the absence of local data, but the approach to better efficiency rather than the specific model is the transferrable tool from the current investigation.

Efficiency, a vital component of successful prevention, can be increased dramatically for both individuals and families by means of an empirical model to target services to those who can benefit most. Serving the same proportion of individuals with a Screening Model instead of current decisions would have increased the hit rate to over 90% and reduced misses by over 85%. However, targeting remains imperfect as evidenced by the fact that most individuals, like families, avoid shelter entry, even in the highest risk decile. Although targeting is imperfect, even with the empirical models, a large body of research suggests that empirically based models tend to be superior to worker judgments (Ægisdóttir et al., 2006; Dawes, Faust, & Meehl, 1989; Grove et al., 2000), as was the case here.

In sum, as for families, the Department of Homeless Services could improve its targeting of HomeBase applicants likely to enter shelter by adopting an empirical risk

model. Directing services only to adults who have been in shelter previously according to Department of Homeless Services records would increase correct predictions while serving many fewer applicants. However, if the City can serve the same proportion of individuals as currently, the screening survey would correctly target an even higher proportion of subsequent shelter entrants (over 90%). Even a one-variable model based on administrative records of prior shelter experiences is far more efficient than current decisions.

# CHAPTER 3

## DOES CONTEXT MATTER? COMPARING RISK FACTORS AND TARGETING EFFICIENCY FOR DIFFERING MANIFESTATIONS OF HOMELESSNESS IN ALAMEDA COUNTY, CA, AND NEW YORK CITY, NY

### Introduction

Community-based programs that seek to prevent homelessness efficiently require identifying people who are at high risk for becoming homeless in the absence of services. Relatively few models evaluate risk for homelessness explicitly, and these are confined to particular locales (Barnett et al., 2011; Shinn et al., 1998, 2013) or populations (Greenberg et al., 2006; Hudson & Vissing, 2008). Little is known about how well models generalize across social contexts. Differences in the geography, demographics, housing and labor markets, as well as social policies could render models developed in one locale useless for application in another. Risk models may also depend on definitions of homelessness. The factors that put one at risk for entering shelter (the most common operational definition of homelessness) may not be the same as factors that put one at risk for housing instability. The present paper replicates a process used to generate a risk model for shelter entry in one location (New York City, NY) in another site (Alameda County, CA) and compares the resulting models both across geography and, in the case of Alameda County, across definitions of homelessness. It provides one window into the extent to which context matters in efforts to prevent homelessness.

The prevention programs in New York City and Alameda County targeted similar populations (homeless families and individuals), but used different risk factor assessments and different outcome variables. The Homebase program serves New York City with homelessness prevention and has participated in empirical research to examine its effectiveness (Messeri, O'Flaherty, & Goodman, 2011; Rolston, Geyer, & Locke, 2013) and improve the efficiency of its targeting strategies (Shinn et al., 2013). The EveryOne Home program serves Alameda County, CA, with homelessness prevention and seeks to examine its effectiveness and improve the efficiency of its targeting models. Because the programs are the major providers of community-based homelessness prevention services in their respective geographies, the programs will be referred to as: *Alameda County* and *NYC* prevention services.

Homelessness prevention in Alameda County and prevention services in NYC differ in at least two ways. The first difference is cultural variation in the systems set up to combat homelessness. The prevention system in Alameda County is less integrated than that of NYC. Further, NYC has an explicit right to shelter; Alameda does not. The second difference is variation in risk factors and outcomes, attributable to structural factors including housing, economics, demographics, and climate. These structural factors are sometimes modeled as community-level domains in homelessness investigations (Byrne et al., 2012; Lee et al., 2003). Community-level domains for each program are described further in the methods section.

For the purposes of comparison, part of this study is an explicit replication of the methods of previous studies in NYC that investigated risk factors for homelessness for families and childless adults. The current study adds to the existing efficiency

investigations by answering the following questions: What is the pattern of subsequent homelessness for prevention applicants in Alameda County? Which risk factors contribute to homelessness? How do risk factors differ for models predicting different manifestations of homelessness? How does model efficiency for prevention applicants in Alameda County compare with efficiency for models developed in NYC?

**Targeting Efficiency and Ecological Contexts of Homelessness**

Successful targeting requires the identification of risk factors that lead to homelessness *and* the evaluation of the efficiency of targeting models to correctly direct services to those who can benefit most (Shinn et al., 2013). Although services may be useful for all people in extreme poverty, targeting is meant to differentiate service recipients who need services to avoid homelessness from those who would avoid homelessness otherwise. Efficiency is defined as the extent to which a targeting model directs services to those who benefit most from services. Unless prevention services are affordable enough to be offered universally, both efficient targeting and effective services are vital to the prevention program's success (Burt et al., 2007).

Despite the expressed need of prevention programs to improve targeting services and some promising research (Barnett et al., 2011; Shinn et al, 1998, 2013), targeting efficiency is often disappointing and investigations of efficiency are mostly absent in the literature (Greer & Shinn, in progress). Often the efficiency of targeting models is underwhelming (Greenburg et al., 2006; Hudson & Vissing, 2010). The most common criterion for access to homelessness prevention services is eviction (US Conference of Mayors, 2008), but most people who are evicted do not become homeless. For example, a decision model based only on eviction proved to be only slightly better than chance in

predicting homelessness among families in NYC (Shinn et al., 2013). For community-based prevention programs across the nation, differentiating the characteristics of poor people who need services to avoid homelessness from those who will avoid homelessness without services remains a highly desired but difficult undertaking.

An ecological framework from community psychology (Bronfenbrenner 1979; Nelson and Prilleltensky 2005; Toro, Trickett, Wall, & Salem, 1991) offers a way to understand differences in risk factors and outcomes across settings. Nooe and Patterson (2010) put forward an *ecology of homelessness* with individual- and structural-level risk factors that lead to multiple types of homelessness. Such a framework calls for an understanding of risk that is rooted in environment. Further, previous research points out that interventions are always "interventions-in-context" (p. 124). The specificities of implementation are dependent on the nurturing or obstructing social context of their environments (Cornish and Campbell, 2009).

The ecological perspective suggests that inconsistencies in risk factors for homelessness across studies are likely attributable to environmental differences. Homelessness studies are often framed with at least some of the following community-level domains that have been associated with homelessness: housing markets, economic conditions, demographic composition, safety net, climate, and transience (Lee et al., 2003; Byrne et al., 2012). For example, Burt (1991) found the housing market domain, and specifically, a lack of affordable housing, to be the main contributor to homelessness in some cities. Other cities with high homelessness rates had ample affordable housing, but lacked job training, supported housing, or drug-treatment facilities (Burt, 1991). Further economic conditions, such as poverty rates, have been positively associated with

increased risk for homelessness (Early & Olson, 2002). Additionally, some structural

models include temperature as a predictor (i.e., climate domain) and find significant

relationships between warmer temperatures and increased homelessness (Raphael, 2010).

Finally, increased mobility into an area has been positively associated with increased

homelessness (Lee et al., 2003). Community-level domains are useful for framing

ecological determinants of homelessness. However, inconsistencies in defining risk and

homelessness make it difficult to link specific risk factors with homelessness rates for

individuals across diverse geographies.

In addition to the different characteristics associated with risk, the ecological

framework suggests that different definitions of homelessness might capture the spectrum

of housing instability in different settings. Definitions of homelessness can vary due to

research interests or political goals (Hopper, 1991) as well as the availability of data on

different indicators and interpretations of what it means to be homeless. HUD provides a

lengthy definition of homelessness that focuses on the absence of a suitable residence

(NAEH, 2012). The definition of homelessness includes a set of risk factors that permits

some additional families and youth to be classified as homeless and thus expands the pool

of eligibility for federal assistance from previous definitions. However, the European

Union (EU) definition has a greater variety of measures of housing instability

(FEANTSA, 2005), including the following typology:

- *rooflessness* (without a shelter of any kind, sleeping rough)

- *houselessness* (with a place to sleep but only temporarily in institutions or

  shelter)

- living in *insecure housing* (threatened with severe exclusion due to insecure

tenancies, eviction, domestic violence)

- living in *inadequate housing* (in caravans on illegal campsites, in unfit housing, in extreme overcrowding) (FEANTSA, 2005).

Predictors of homelessness might differ for varying manifestations of homelessness. The current chapter seeks to investigate risk factor differences across NYC and Alameda County, CA. Although risk factors that are assessed and available for modeling are not identical in Alameda and NYC, risk-factor domains will be compared across studies and similarities and differences will be discussed. Further, the current chapter explores differing manifestations of homelessness within Alameda County. The remainder of this paper compares the locales and the samples, develops the models for Alameda County, and then compares them to the models previously developed for New York City (Chapter 2; Shinn et al., 2013).

## Methods

### Ecological Characteristics

Table 3.1 explores the ecological context of Alameda County and NYC in comparison with each other and with the US as a whole. Starting with housing markets, NYC has lower rates of homeownership, a higher median rent, and a higher rental vacancy rate than Alameda County, although both communities have far lower rental vacancy rates than the national average. For economic conditions, NYC has a higher unemployment rate and a higher rate of people earning less than 50% of the poverty level than Alameda County. In NYC, there are higher proportions of African Americans, Hispanics and single-person households as well as lower proportions of baby-boomers than in Alameda County. In terms of the social safety net, NYC has a higher proportion

of households receiving SSI or public assistance, a higher maximum monthly TANF

payment, and a higher per capita mental health expenditure than in Alameda County. The

temperature differential from the average high in July to the average low in January and

the amount of annual precipitation are much higher in NYC than in Alameda County.

Finally, there are lower rates of people moving in NYC than in Alameda County.

**Table 3.1.** Community-level domains in Alameda County and New York City

| Domain | Alameda County | New York City | United States |
|---|---|---|---|
| *Housing Market* | | | |
| Owner occupancy rate[1] | 54.5% | 32.6% | 66.1% |
| Rent (median) | $892[2] | $1,191[2] | N/A |
| Rental vacancy rate (MSA) | 4.1%[3] | 4.9%[3] | 12.4%[1] |
| *Economic Conditions* | | | |
| Civil labor force unemployed[1] | 9.2% | 9.5% | 8.7% |
| Persons with incomes < 50% poverty level[1] | 5.4% | 7.8% | 6.6% |
| *Demographic Composition* | | | |
| Black[1] | 14.0% | 26.6% | 13.5% |
| Hispanic/Latina(o)[1] | 22.2% | 28.4% | 16.1% |
| Baby Boomers (45-64 y.o.)[1] | 25.9% | 24.2% | 26.1% |
| 1-person household[1] | 27.7% | 32.5% | 27.3% |
| *Safety Net* | | | |
| Households in poverty receiving Public Assistance or SSI[1] | 31.0% | 31.6% | 26.0% |
| Maximum monthly TANF payment for family of 3 (state)[4] | $638 | $753 | N/A |
| Per Capita expenditure to mental health agency (state)[5] | $152.60 | $256.31 | $120.50 |
| *Climate* | | | |
| July avg high-Jan avg low[6] | 25.6°F | 57.5°F | N/A |
| Annual Precipitation[6] | 20.8in | 46.6in | N/A |
| *Transience* | | | |
| Household moved before staying in home for 1+ years[1] | 16.0% | 11.4% | 15.4% |

[1]American Community Survey 2007-2011
[2]HUD FMR estimates for efficiency apartment 2013
[3]Flanagan & Schwartz (2013) Rental Housing Market Condition Measures: A Comparison of U.S. Metropolitan Areas From 2009 to 2011. *American Community Survey Briefs*
[4]Center on Budget and Policy Priorities 2011
[5]Kaiser-Family Foundation 2010
[6]NOAA National Climatic Data Center 1981-2010

**Samples**

Participants in Alameda County were 2,761 people who applied for the Homelessness Prevention and Rapid Re-housing Program (HPRP) from the county's eight housing resource centers between October 21, 2009, and April 28, 2012. Applicants provided details for the following domains: demographic variables, human capital, housing conditions, disability, interpersonal discord, and shelter history. Based on their cumulative score on an index developed by the County, 83% of participants were offered services using a triage approach.[12] For the analyses of the Alameda County data, the current study replicates the methods used in NYC (Chapter 2; Shinn et al., 2013). I impute missing data for predictors with Stata, including auxiliary variables according to the literature (Graham, et al., 2007; Sinharay et al., 2001).

Participants in NYC were 21,325 applicants for the Homebase prevention program, of which 11,105 were families and 10,220 were individuals without children. Participants applied for services from September 28, 2004, to December 29, 2010. At application, intake workers surveyed participants about the following domains: demographic variables, human capital, housing conditions, disability, interpersonal discord, childhood experiences, and shelter history. Workers then decided who should receive services, without following any particular model.

**Predicting Time to Homelessness in Alameda County**

The current study predicts time between the initial application for services and subsequent homelessness in Alameda County for three increasingly inclusive definitions

---

[12] For a detailed description of participant characteristics and the eligibility process, see chapter 4 of this dissertation. The risk-scoring variable in Alameda County was not highly correlated with the risk scores for the three models in the current study.

of homelessness: 1) literal homelessness, 2) literal homelessness or imminent housing

loss (termed threatened homelessness), and 3) literal homelessness, imminent housing

loss, or unstable housing (termed any housing instability). The broadest two definitions of

homelessness were created from the following variables that share some similarities with

the EU definition of homelessness.

- Literal homelessness (according to HUD's definition); similar to *rooflessness* and

  *houselessness*

    o In a shelter, transitional housing, or in a place not meant for habitation

    o Escaping a domestic violence situation (similar to *insecure housing*)

    o Just exited jail, substance abuse treatment, hospital, psychiatric facility, or

      foster care setting from shelter or the streets

- Imminent housing loss; similar to *insecure housing*

    o Being evicted, discharged, or otherwise notified of imminent housing loss

- Unstable housing and at-risk for losing housing; similar to *inadequate housing*

    o Doubled up, being evicted from public or assisted housing

The measure of literal homelessness derives from Alameda's Homelessness Management

Information System (HMIS) and is primarily shelter entry.  Only episodes of

homelessness that began after the date the participant applied for services are counted.

The other two outcomes include literal homelessness based on HMIS data and additional

information available only for those applicants for Alameda County's Homelessness

Prevention and Rapid Re-Housing Program who applied for services a second time

(whether or not they received them on the initial application that marked their entry into

the study).  Threatened homelessness includes self-reported literal homelessness or

imminent housing loss upon reapplication. Any housing instability includes the first two and additionally self-reported unstable housing or being at-risk for losing housing upon reapplication. Because the self-reported outcomes are available only for the subset that re-applied for services, results will be interpreted with the potential for bias in mind.

## Results:  Alameda County

### Pattern of Subsequent Homelessness

The patterns of subsequent homelessness are similar across outcomes in Alameda County. For that reason, only one Kaplan-Meier survival curve and one Hazard function are presented here. Time to subsequent homelessness was calculated by subtracting the date of subsequent homelessness from the date of application for HPRP. The main differences across outcomes are the final rates of those experiencing the outcome: HMIS homelessness (2.2%), threatened homelessness (9.0%), and any housing instability (14.2%). Figure 3.1 shows the Kaplan-Meier survival curve for the outcome of any housing instability. This curve suggests that most applicants for prevention services avoided subsequent instability. Figure 3.2 shows the hazard function for any housing instability. This function suggests that the greatest risk for housing instability occurred just after participants apply for services. These findings are similar to investigations of homelessness prevention in NYC (Greer et al., in progress, Shinn et al., 2013).

**Figure 3.1**. Kaplan-Meier Survival Curve for Any Housing Instability in Alameda County, from 2009 to 2012
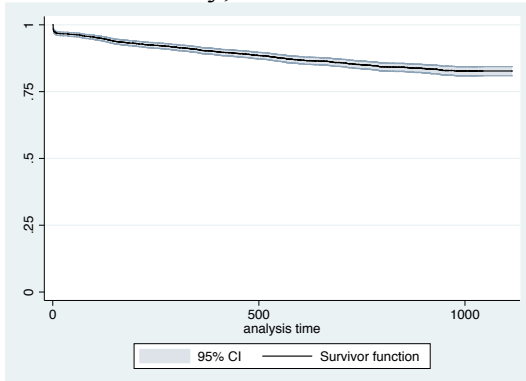
**Figure 3.2.** Hazard Function for Any Housing Instability in Alameda County, from 2009 to 2012





## Risk Factors for Alameda County

Table 3.2 shows risk and protective factors in Alameda County for each level of homelessness. Model 1 investigates predictors of literal homelessness according to HMIS records. Model 2 expands the outcome to include homelessness or imminent loss of housing. Model 3 expands the outcome further than the previous two models by including any report of housing instability as defined in the methods section. The Table includes all variables that reliably predicted any of the three outcomes, arrived at by the following trimming procedure.

For each model, predictors that failed to reliably contribute to the model were eliminated one at a time starting with the least reliable predictor. When only reliable predictors remained in the model, I added the eliminated variables back into the model one at a time to determine whether their contributions were reliable in the context of a parsimonious model. I then added back predictors that were reliable in either of the other models. Findings across all models control for all other variables in the model, and including all variables that mattered for any of the models allows a consistent set of controls. Including these additional variables from other models did not affect the

71

reliability of other variables already in the models. The results across models are reported in Table 3.2. Variables that failed to contribute to any model (p < .05) are listed in the table footnotes. Across models, the domains of demographic variables, disability/criminal justice, and interpersonal discord failed to reliably contribute to predicting the outcomes.

The human capital variables that were most important varied somewhat by outcome. Receiving public benefits[13] was associated with more than double the risk for literal homelessness, compared to not receiving such benefits. Part time (rather than full time or no employment) was associated with 1.5 times the risk for threatened homelessness and 1.3 times the risk for any housing instability. Earning less than 30% of the Area Median Income (AMI) was associated with almost double the risk for the two broader definitions of homelessness compared to earning more than 30% AMI.

The housing conditions domain included five risk factors that were reliable in one or more of the models, which is the highest number of significant covariates in any of the domains. Rates of literal homelessness were more than double for applicants who were living doubled-up with another household in the same dwelling unit compared to those without such living situations. Being doubled-up was not a reliable predictor of the broader homelessness outcomes. Having one or more evictions in the past five years was associated with almost double the risk for literal homelessness and over 1.4 times the risk for threatened homelessness compared to those without such histories, although a history of eviction was not significantly associated with the broader outcome of any housing instability. On the other hand, being evicted from public housing at the time of

---

[13] Defined as receiving any of the following: SNAP, MEDICAID, MEDICARE, SCHIP, WIC, VA services, TANF childcare, TANF transportation, other TANF (CalWORKS), or Section 8.

application was associated with 2.2 times the risk for threatened homelessness and over

1.7 times the risk for any housing instability compared to those without such an eviction,

but being evicted from public housing was not associated reliably with literal

homelessness. Similarly, paying more than 40% of one's income for rent was associated

with almost double the risk for the two broadest homelessness outcomes, but it did not

reliably predict literal homelessness.

Having outstanding debt in excess of $1000 was associated with about 1.5 times

the risk for all three homelessness outcomes compared to having no such debt. Finally,

for the shelter variables domain, those who reported spending the night before applying

for services as homeless[14] experienced rates of subsequent homelessness (that is an

episode of literal homelessness that began after the date of the application for services)

almost ten-fold higher than those who did not so report. The variable was not reliably

associated with the two broadest homelessness models.[15]

---

[14] Defined as 1) Emergency shelter, including hotel or motel paid for with emergency
shelter voucher, 2) Place not meant for habitation inclusive of non-housing service site,
or 3) Transitional housing for homeless persons
[15] I further investigated interactions of treatment with each risk factor to determine
whether risk profiles differed according to whether participants received treatment. I
found no significant interactions. Thus I omit interactions from the models above.

**Table 3.2.** Risk Factors for Homelessness: Alameda County

| Predictor | Model 1: Literal homelessness (n = 60) | | Model 2: Threatened Homelessness (n = 248) | | Model 3: Any Housing Instability (n = 393) | |
|---|---|---|---|---|---|---|
| | HR | CI | HR | CI | HR | CI |
| **Demographics[a]** | | | | | | |
| N/A | | | | | | |
| **Human Capital[b]** | | | | | | |
| Part-time employment | 1.72 | 0.93-3.18 | 1.49** | 1.11-2.00 | 1.34* | 1.06-1.70 |
| Receiving public benefits | 2.08* | 1.08-4.01 | 1.26 | 0.95-1.66 | 0.98 | 0.79-1.22 |
| < 30% AMI | 2.05 | 0.73-5.78 | 1.85** | 1.20-2.85 | 1.92*** | 1.37-2.70 |
| **Housing Conditions[c]** | | | | | | |
| Doubled up | 2.72** | 1.45-5.09 | 1.01 | 0.73-1.41 | 0.98 | 0.75-1.27 |
| Eviction history | 1.94* | 1.08-3.51 | 1.43* | 1.02-2.00 | 1.26 | 0.95-1.66 |
| Being evicted from public housing | 1.06 | 0.24-4.60 | 2.20** | 1.43-3.38 | 1.73** | 1.21-2.48 |
| Rent > 40% income | 1.53 | 0.44-5.31 | 2.18** | 1.36-3.50 | 2.26*** | 1.57-3.26 |
| Outstanding debt > $1000 | 1.69* | 1.00-2.86 | 1.37* | 1.06-1.76 | 1.40** | 1.14-1.71 |
| **Disability/Criminal Justice[d]** | | | | | | |
| N/A | | | | | | |
| **Interpersonal Discord[e]** | | | | | | |
| N/A | | | | | | |
| **Shelter History[f]** | | | | | | |
| Previous night homeless | 9.80*** | 5.40-17.81 | 1.21 | 0.89-1.66 | 0.83 | 0.63-1.10 |

[a] Unreliable variables included Female, African American, Hispanic, Married, Veteran, Pregnant, Age, Family vs. Individual status
[b] Unreliable variables included High School diploma/GED, Vocational training; AMI = Area Median Income
[c] Unreliable variables included Leaseholder history
[d] Unreliable variables included Disabled, Unhealthy compared to peers, Criminal justice history
[e] Unreliable variables included Domestic violence history, Escaping domestic violence currently
[f] Unreliable variables included Homeless more than one time in the last year, Previously homeless self reports, Previously homeless HMIS data
*** $p < .001$, ** $p < .01$, * $p < .05$

The models suggest that the populations subject to the various forms of risk are different, and Table 3.3 shows the percentages of people in alternate divisions of the groups: people who experienced literal homelessness, people who experienced imminent loss of housing, but not literal homelessness, and people who experienced unstable housing, but not imminent loss of housing or literal homelessness, and people who did not experience any of these forms of subsequent housing instability.

The literal homelessness group appears to be different than the groups that reported imminent housing loss or unstable housing, and all other groups appear to be different than the group that did not experience any subsequent instability. All three of the unstable groups had higher rates of part-time employment than the group with no housing instability. Those who experienced literal homelessness reported the highest rates of receiving public benefits, and both of the remaining groups that experienced instability reported lower rates of receiving public benefits than the group reporting no instability.

The proportion of those who were extremely poor was highest for the literal homelessness group, and the remaining instability groups had higher rates of extremely poor people than those with no housing instability. Being doubled up and having a history of eviction was highest for the group experiencing literal homelessness. Paying more than 40% of one's income for rent and having more than $1000 in debt were higher across the groups with housing instability than for the group with no housing instability. Finally, spending the previous night homeless before applying for services was much higher for the group that experienced subsequent homelessness, and those in the imminent housing loss or unstable housing groups experienced rates lower than the group with no housing instability.

**Table 3.3.** Proportions with Risk Factors for Alternate Groups: Alameda County

| Predictor[a] | Literal Homelessness (n = 60) | Imminent Housing Loss (n = 188) | Unstable Housing (n = 145) | No Housing Instability (n = 2368) |
|---|---|---|---|---|
| **Human Capital** | | | | |
| Part-time employment | 23.73 | 24.47 | 22.76 | 16.97 |
| Receiving public benefits | 79.66 | 64.36 | 64.14 | 67.24 |
| < 30% AMI | 93.22 | 89.36 | 88.97 | 81.72 |
| **Housing Conditions** | | | | |
| Doubled up | 26.32 | 16.67 | 19.44 | 17.53 |
| Eviction history | 29.31 | 14.12 | 13.33 | 12.55 |
| Being evicted from public housing | 3.51 | 13.11 | 6.99 | 7.24 |
| Rent > 40% income | 94.00 | 89.89 | 90.07 | 81.57 |
| Outstanding debt > $1000 | 47.46 | 44.68 | 48.97 | 38.54 |
| **Shelter History** | | | | |
| Previous night homeless | 64.41 | 9.57 | 6.90 | 21.70 |

[a] Only reliable predictors from Table 2 are included here.

## Risk Profiles

Next, I investigated whether some applicants were at such high risk that services ceased to be beneficial, as would be implied by a triage model, in which services are given to those at intermediate levels of risk. I calculated risk scores for each applicant for each model based on average predicted scores across fifty imputed data sets[16] and plotted the proportion of applicants who had the adverse outcome at each quintile of risk, separately for those who did and did not receive HPRP services. Results are shown in Figures 3.3, 3.4, and 3.5. Ignoring the receipt of HPRP services, rates of housing instability tend to increase as risk increases. However, the shapes of the risk profiles differ across outcomes.

In the risk profiles for Figure 3.3, rates of literal homelessness remain close to 0 for the first two risk quintiles. However, for risk scores of three to five, rates of homelessness gradually increase. Rates of literal homelessness are quite similar for those

---

[16] Although previous studies divide risk into deciles, the smaller sample size of the current study was better suited to quintiles of risk.

who did and did not receive HPRP services in the first four risk quintiles.  However, the profiles diverge for the highest level of risk. Here, the difference in rates of homelessness for recipients and non-recipients of HPRP is at its maximum. Accordingly, services seemed most helpful for those in the highest risk quintile. The risk profiles for Figure 3.4 show rates of threatened homelessness across quintiles of risk. Here services made little difference in the first two quintiles.  The risk profiles for recipients and non-recipients of services diverged in the third quintile of risk and were largest for the top three quintiles. Again, there was no evidence that any applicants were too risky to benefit from preventive services – applicants in the top three quintiles appeared to benefit from services, which seemed to make the largest difference for intermediate categories of risk.

Figure 3.5 contains the risk profile for rates of any housing instability across quintiles of risk. For each quintile, applicants who received services experienced housing instability at lower rates than those who did not receive services. Further, the differences between groups are most apparent for the lower and upper quintiles. None of the graphs suggest that applicants were at such high risk that services failed to be beneficial, and the graph of any housing instability suggests that services might reduce unwanted outcomes regardless of risk level.

**Figure 3.3.** Risk Profile for Literal Homelessness
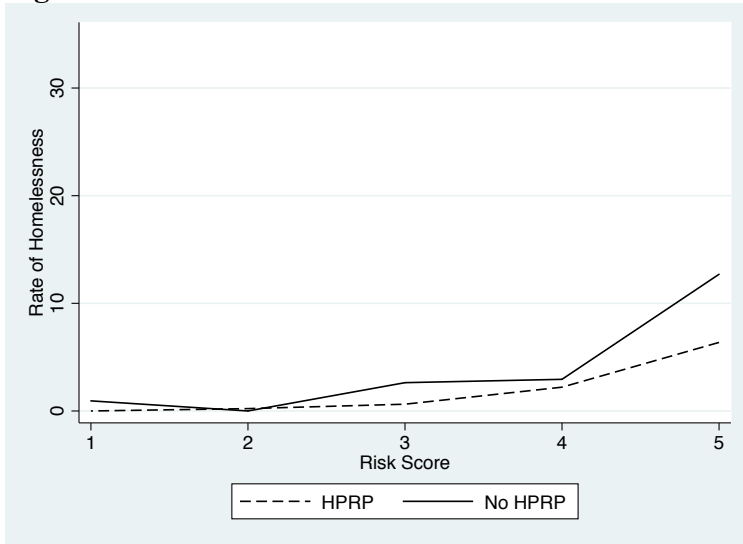


**Figure 3.4**. Risk Profile for Threatened Homelessness
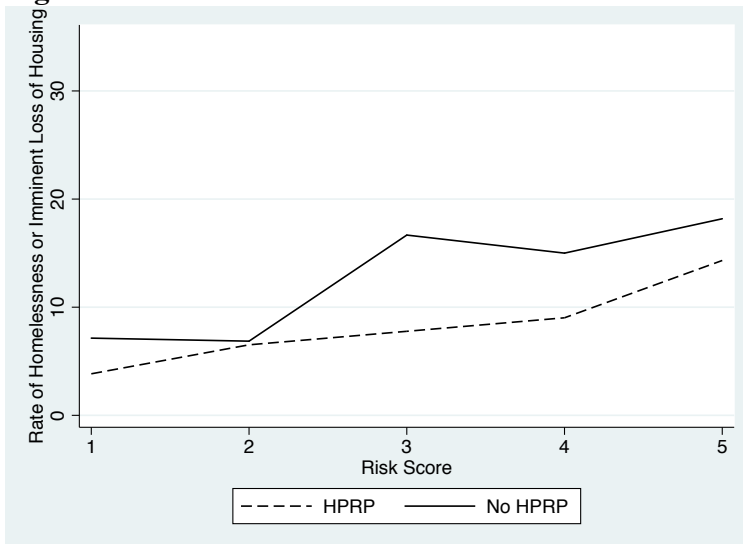
**Figure 3.5.** Risk Profile for Any Housing Instability
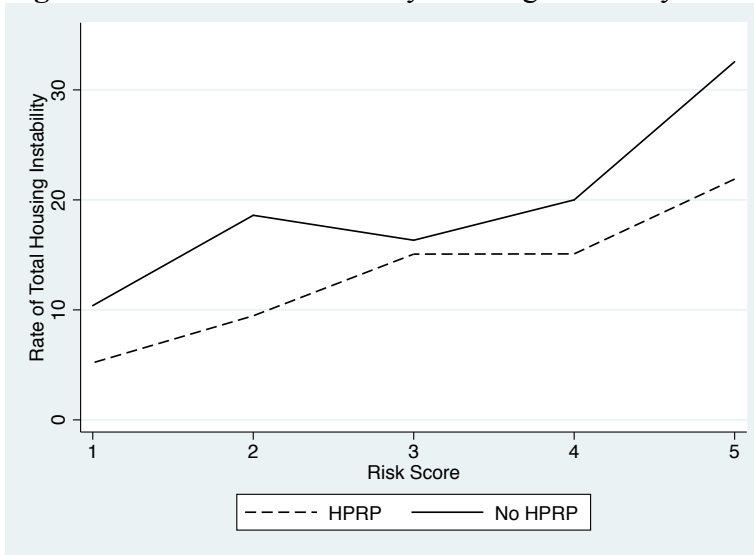


## Comparing Risk Factors for Alameda County to NYC

Table 3.4 compares risk factors for literal homelessness in Alameda County with

parallel models derived previously (Chapter 2; Shinn et al., 2013) to predict shelter entry

for individuals and families in New York City.[17]  Literal homelessness was chosen in

Alameda because it is most similar to shelter entry in NYC. Across locations, intake

questions were not identical. Accordingly, Table 3.4 includes categories of risk along

each row and specific variables under the column headings of Alameda County, NYC

Families, and NYC Individuals. Variables that mattered in one location but were not

available in the other are labeled *NA*. Those that were tested and did not predict are

shown with two dashes. The literal homeless model in Alameda County shares many risk

factors with the shelter entry model for individuals in NYC. The lack of age as a

---

[17] Separate analyses for families and individuals in Alameda yielded substantively similar results for the broader risk models that predicted threatened homelessness as well as any housing instability. Separate models predicting literal homelessness for the two groups exhibited a few differences (e.g., doubled up was a risk factor for individuals, but not families). However, fewer than 30 family applicants experienced literal homelessness and there were many estimation problems. Appendix Chapter 3 shows descriptive differences between individuals and families in Alameda.

protective factor for Alameda County constitutes the primary difference between the models. Having spent the previous night homeless at the time of an application for services is the strongest predictor of subsequent literal homelessness for Alameda County, and this variable is correlated with previous shelter records and self reports (i.e., the strongest predictors of shelter entry for NYC models), but the particular variable was not asked in New York City.

Some differences across models are due to the absence of variables in at least one location. For example, applicants in Alameda County were not asked about mental illness, a factor that approached significance for NYC individuals. Further, Alameda County did not collect information about adverse childhood experiences, which were combined to form a risk factor index for families in NYC. NYC applicants were asked many more interpersonal discord variables than Alameda County applicants, and these mattered for families. There was also some consistency in variables that were not associated with homelessness either location. These included self-reports of domestic violence and criminal justice histories.

**Table 3.4.** Comparing Literal Homelessness and Shelter-Entry Risk in Alameda to NYC

| Risk Categories[*] | Alameda | NYC families | NYC individuals |
|---|---|---|---|
| **Demographics** | | | |
| Family vs. Individual | -- | NA | NA[18] |
| Age | | Younger | Younger |
| Child age | NA | Child under 2 | NA |
| Pregnant | -- | Pregnant | NA |
| **Human Capital** | | | |
| Education | -- | No High school/ GED | -- |
| Employment | --[19] | Not Employed | -- |
| Currently receiving public assistance | Receiving public benefits | Receiving public assistance | Receiving public assistance |
| **Housing Conditions** | | | |
| Name on lease | Doubled up | Not a leaseholder | -- |
| Overall Eviction threat | Eviction in the last five years; Being evicted from public housing | Any eviction threat NA | Verbal eviction threat; No legal eviction threat[**] NA |
| Outstanding debt | Outstanding debt $1000 or more NA | -- | NA High Housing Arrears |
| Mobility | NA | Number of moves in last year | -- |
| **Disability/Criminal Justice** | | | |
| Disability | -- | -- | No history of mental illness or hospitalization[**] |
| **Interpersonal Discord** | | | |
| Protective services involvement | NA | History of protective services | -- |
| Discord rating | NA | High discord with landlord, leaseholder, or | -- |

---

[18] Separate models were estimated for individuals and families in New York City, and the data covered different time periods, so that family status cannot be formally tested. However, it is clear that among applicants for prevention services, families were at higher risk of shelter entry than single individuals (12.8% of families entered shelter vs. 6.4% of individuals).

[19] The hazard ratios for part-time employment and low income were even larger in the model predicting literal homelessness than in the other two models, but the confidence intervals included a ratio of 1, meaning no effect.

| | | | |
|---|---|---|---|
| | | within household | |
| **Childhood Experiences** | NA | | |
| Adversity index | | Multiple childhood adversity experiences | -- |
| **Shelter History** | | | |
| Shelter history as adult | Spent last night homeless | NA | NA |
| | | Shelter history as adult (self report); Shelter application in last three months; Reintegrating into community from an institution | Shelter history as adult (administrative record); Shelter application in last three months; Reintegrating into community from an institution |
| | NA | | |

*- significant variables included in a parsimonious model arrived at by eliminating non-significant predictors one at a time, and then checking whether any variables that had been eliminated regained predictive power after other variables had been eliminated. For individuals, public assistance re-entered the model after other correlated variables were eliminated.

** - variables that were eliminated for NYC individuals based on robustness analyses

**Model Efficiency**

Figures 3.6 through 3.10 show the ROC curves across models, computed with the parametric probit ROC command in Stata. Table 3.5 compares model efficiency in two ways: by reporting the estimated hit rates and confidence intervals that correspond to the tradeoff in false-alarm rates of .10, .25, and .50, and by reporting the Area Under the Curve (AUC) for each model. Better models have higher hit rates for a given false alarm rate and larger AUC values. If the AUC for a model were .50, the estimated curve would fall along the 50% reference line (i.e., each of the diagonals in Figures 3.6 through 3.10). However, a model that perfectly predicts an outcome would capture all of the area (i.e., the AUC value would be 1.00).

Several conclusions are apparent across models in Alameda County (Figures 3.6-3.8). First, the ROC curve for literal homeless (Figure 3.6) is far more efficient than the ROC curves for the two broader definitions of homelessness in Figures 3.7 and 3.8. The confidence intervals for the hit rates for the literal homeless model in Table 3.5 fall above values from the confidence intervals of the remaining two models at each level of false alarms examined. However, there is substantial overlap in the confidence intervals for the models with the two broader aspects of homelessness. Further, the AUC value for the literal homeless model is significantly larger than the AUC values for the remaining two Alameda models. Accordingly, it was possible to target literal homelessness more efficiently than broader forms of homelessness in Alameda County; the two broader models do not differ significantly in terms of efficiency.

**Comparisons with NYC.** Figures 3.9 and 3.10 present ROC curves for the two investigations of shelter entry in NYC (Chapter 2; Shinn et al., 2013)[20], and Table 3.5 includes parallel measures of model efficiency. In terms of efficiency, the model for Alameda County falls between the efficiency of the two models in NYC – it is more efficient than the model for families (especially at higher false alarm rates) and less efficient than the model for individuals.

---

[20] ROC curves in the current study were calculated slightly differently than the nonparametric ROC curves in the previous studies. This was done for two reasons. First, the parametric ROC estimate is smooth, and the overall shape of the ROC is more clearly seen compared to non-parametric estimations. Second, the parametric models permit estimates of hit rates with confidence intervals for given false alarm rates. Although the curves were estimated differently, the AUC values only differed slightly. For families, the non-parametric AUC value was .76 (compared to .75 in the parametric model). For individuals the non-parametric AUC value was .92 (compared to .93 for the parametric model).

**Figure 3.6.** ROC: Alameda Literal Homelessness
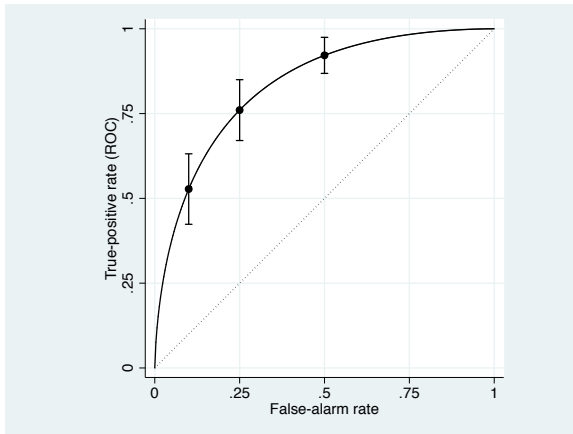


**Figure 3.7.** ROC: Alameda Threatened Homelessness
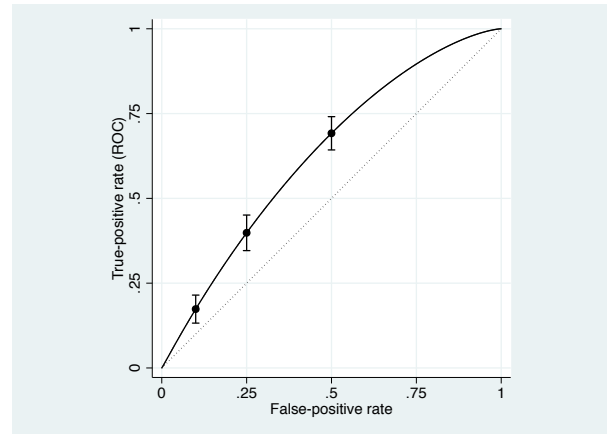


**Figure 3.8**. ROC: Alameda Any Housing Instability
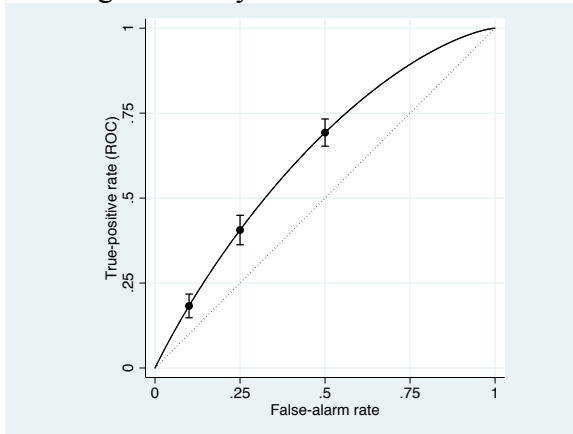


**Figure 3.9.** ROC: NYC Families Shelter Entry
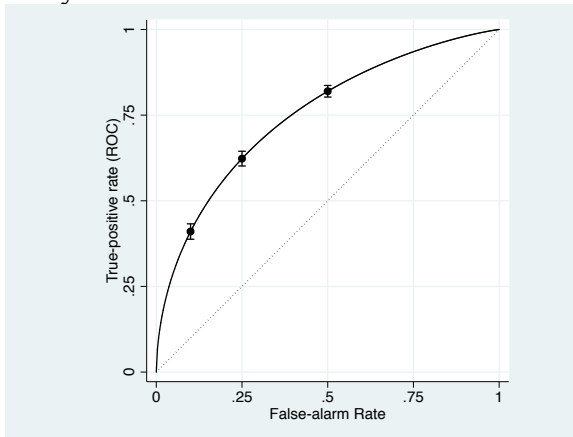


**Figure 3.10**. ROC: NYC Individuals Shelter Entry
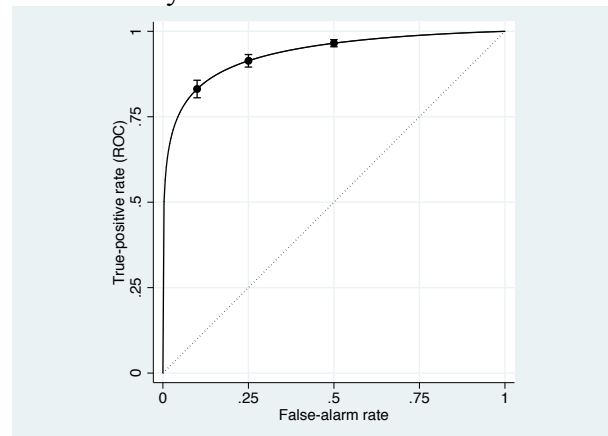
**Table 3.5.** Efficiency Comparisons for Alameda County and NYC

| | *Alameda County* | | | | | | *New York City* | | | |
| | Model 1: Literal Homelessness | | Model 2: Threatened Homelessness | | Model 3: Any Housing Instability | | Shelter Entry NYC Families | | Shelter Entry NYC Individuals | |
| | Est. | CI | Est. | CI | Est. | CI | Est. | CI | Est. | CI |
|---|---|---|---|---|---|---|---|---|---|---|
| **Hit Rate (FA = .10)** | .527 | .424-.631 | .174 | .132-.215 | .183 | .148-.217 | .410 | .388-.433 | .831 | .805-.857 |
| **Hit Rate (FA = .25)** | .760 | .671-.850 | .398 | .346-.451 | .406 | .363-.450 | .623 | .601-.645 | .914 | .895-.932 |
| **Hit Rate (FA = .50)** | .922 | .868-.975 | .692 | .643-.741 | .693 | .653-.733 | .820 | .803-.837 | .965 | .955-.976 |
| **AUC** | .835 | .787-.884 | .630 | .600-.664 | .633 | .605-.661 | .753 | .739-.766 | .934 | .922-.946 |

Notes:
Est. = Estimate
CI = Confidence Interval
FA = False Alarm Rate
AUC = Area Under the Curve

## Discussion

The current study investigated risk factors for homelessness in Alameda County and compared results to findings from previous studies in NYC. Receiving public benefits was positively associated with literal homelessness in Alameda County and across shelter-entry models in NYC, suggesting that benefit programs may not be enough for at-risk applicants to avoid homelessness. Previous homelessness stood out as the primary predictor across sites, however, some variables differed. NYC might ask applicants whether they spent the previous night homeless before applying for services, because this predictor was more reliable than HMIS records or self-reports of previous homelessness for Alameda County applicants. Further, eviction was associated with literal homelessness and shelter entry across locations, although the type of eviction differed across sites. NYC might ask applicants if they are being evicted from public housing, because such an eviction mattered for Alameda applicants above and beyond having an eviction history. Younger age predicted shelter entry in NYC, but not in

Alameda County. Additionally, having outstanding debt predicted literal homelessness in Alameda County, and high arrears predicted shelter entry for NYC individuals. The number of moves in the last year was a risk factor for families in NYC but not for individuals in NYC, and Alameda County applicants might be asked about mobility in future studies. Further, Alameda County could consider asking applicants about child protective services, discord, and adverse childhood experiences, because these were reliable predictors for NYC families.

Across sites, previous homelessness and a lack of affordable housing seem to be the strongest factors that lead to literal homelessness or shelter entry. Conversely, individual disabilities, criminal justice involvement, and domestic violence failed to predict homelessness across sites (except to the extent that they may have predicted housing affordability). Although these factors may exacerbate risk for homelessness among those who lack access to housing, the dearth of alternative, affordable housing seems critical to avoid homelessness. Additional reliable predictors appear to be linked to financial costs (e.g., having a young child in NYC) or to access to informal housing assistance (e.g., discord, childhood experiences in NYC). These findings support an understanding that the origins of homelessness are primarily structural (Byrne, Munley, Fargo, Montgomery & Culhane, 2012; Israel, Toro, & Oullette, 2010; Shinn, 2007; 2010).

This study asks whether context matters in regards to homelessness risk and found that, across two sites, risk models were remarkably similar, differing more sharply in the questions that were asked than the answers found. It is somewhat surprising that context did not matter more. Few minor contextual differences emerged: the lower proportions of

one-person households and the lower rates of rental vacancies in Alameda County compared to NYC may account for doubling up as a risk factor for homelessness (see Table 1). Being a leaseholder, which was negatively correlated with doubling up with another household, was protective for families in NYC. The biggest contextual difference is the rate at which families enter shelter in NYC (12.8%) compared to the rate of literal homelessness in Alameda County (2.2%), although the time periods differed. This difference may have to do with the right to shelter in NYC and the fact that most shelters are apartment style. Both Alameda County and NYC are located in more difficult housing markets than the U.S. at large. Possibly in more benign housing markets, individual factors might matter more (Burt, 1991).

Risk factors for the two broadest outcomes of homelessness differed from the models that predicted literal homelessness or shelter entry. This similarity is likely driven by the fact that 63% of the group that had any housing instability overlapped with the threatened homelessness group. Both risk models included multiple indicators of poverty that were associated with unstable housing. These variables included part-time employment, earning less than 30% AMI, rents higher than 40% of income, and outstanding debt in excess of $1000. Additionally, being evicted from public housing was a reliable predictor of threatened homelessness. Notably, previous shelter did not reliably predict broader definitions of homelessness. For the Alameda County models, no significant risk factors for homelessness originated from the following domains: demographic variables, disability/criminal justice, and interpersonal discord. Perhaps the European style typology of homelessness would be useful to distinguish outcomes of housing instability (FEANTSA, 2005).

The most efficient models were those that investigated predictions of literal homelessness or shelter entry, regardless of location in Alameda County or NYC. Many communities that took part in HPRP, including Alameda County, adopted a triage approach - providing services to households whose risk was considered intermediate and were encouraged by the Department of Housing and Urban Development to do so. The current study suggests that triage is not justified; rather, for literal homelessness, services make the most difference if given to households at highest risk. For broader outcomes, prevention seemed to help regardless of risk level. Literal homelessness is the most costly outcome both for people entering shelter as well as the public. Accordingly if resources are limited, it might make sense to focus on people at highest risk of homelessness, but services remain useful to diminish housing instability for a broader range of people. One important caveat is that this conclusion and all others apply only to people who apply for prevention services. Many people who experience homelessness do not apply for services, and strategies to expand applications might change conclusions.

This study was limited in several ways. First, the broader measures of homelessness in Alameda County required applicants who had previously applied for prevention services to return to service providers or to call an emergency assistance line to report such outcomes. This limitation would result in the underreporting of homelessness, since some people would fail to return to service providers. Accordingly, individuals who fail to report outcomes are counted as not experiencing it in targeting models, and the extent of housing instability might remain undetected. The current study suggests that HMIS systems should be as integrated as possible and should collect information systematically on broader definitions of homelessness, perhaps following the

European typology (FEANTSA, 2005). A second limitation is the short length of follow

up for the Alameda County models. This limitation would also lead to underreported

outcomes. In NYC, all participants were followed in administrative records for at least

two years (three years for families) after applying for services. On the other hand, some

participants in Alameda County were followed for only three months after applying for

services. However, the limitation seems slightly less problematic given the fact that the

highest risk for experiencing instability outcomes was shortly after applying for services.

Finally, the sample in Alameda was not large enough to distinguish reliably between

models for individuals and families given the small number of people who became

homeless in the short follow-up period, but differences by household type should be

explored where power is adequate to do so.

Empirical targeting models can direct services where they are most useful,

particularly in the case of literal homelessness. Targeting models are not a panacea – a

majority of those at highest risk avoid the unwanted outcome. Nevertheless, empirical

models are useful in identifying relative risk. This is consistent with longstanding

literature that supports empirical targeting models in multiple domains (Ægisdóttir et al.,

2006; Dawes, Faust, & Meehl, 1989; Grove et al., 2000).

Context mattered less than was expected. The similarities across sites and

populations suggest that it may be possible to develop a generic targeting model that is

broadly useful. More research is needed in more sites – particularly those with more

benign housing markets – to test the generality of the model. The risk factors that were

important in any of the three analyses would be an excellent starting point, along with

any additional variables that seem appropriate locally. By testing and employing

empirical targeting models, localities will almost assuredly improve the efficiency of their prevention programs and provide important information for other sites as well. Accordingly, the limited resources that support homelessness prevention could be better targeted where they are most needed – to individuals and families who would otherwise become homeless.

# CHAPTER 4

## PREVENTING HOMELESSNESS IN ALAMEDA COUNTY, CA: A REGRESSION DISCONTINUITY DESIGN

### Introduction

As the US began to experience the negative effects of the recent recession,
worries mounted that homelessness rates would increase, especially for families (Sard,
2009). To prevent homelessness rates from rising, the Obama administration initiated the
$1.5 billion Homelessness Prevention and Rapid Re-Housing Program (HPRP) (Witte,
2012). HPRP was intended to prevent homelessness by stabilizing households. HPRP
provided short-term housing subsidies and modest social services to people at risk for
homelessness or those who had recently become homeless. Subsidies lasted up to a
maximum of eighteen months, with quarterly determinations of eligibility for continued
assistance. These subsidies were typically not as "deep" as more permanent Housing
Choice Vouchers, which pay the difference between 30% of income and the Fair Market
Rent. Instead, HPRP permitted subsidies that were "shallow" (of lessor value, temporary,
or declining in value over time). The current study examines how well HPRP worked to
reduce homelessness rates for applicants for services in Alameda County, California.

Most people who experience homelessness do so once for a brief time period,
with the average length of stay in shelters varying across cities (Culhane, Metraux, &
Byrne, 2011). Thus, it is plausible that a short-term subsidy to stabilize people would be
sufficient to end homelessness for many. Rigorous evaluations of HPRP programs could

show whether this is the case. However, findings that most people who received HPRP stayed out of shelter fail to provide sufficient evidence that the program is successful. The fact that most people also avoid shelter re-entry in the absence of specialized services underlines the importance of counterfactuals showing what would have happened in the absence of the program. The current study attempts to uncover the effectiveness of HPRP with a regression discontinuity design (RDD).

**Policy Context of HPRP**

As a part of the American Recovery and Reinvestment Act of 2009, HUD rolled out the Homelessness Prevention and Rapid Re-Housing Program (HPRP) with allocations based on the formula for the Emergency Shelter Grants Program. HPRP had four requirements. First, an introductory consultation with applicants was required to determine eligibility for services. Second, households were required to have an income level at or below 50% area median income (AMI). Third, households were required to be either homeless or at risk for losing their housing. Finally, applicants were required to be recertified at least once every three months to reassess eligibility (U.S. Department of Housing and Urban Development, 2011).

HPRP regulations explicitly stated that the program was intended for people who would become homeless "but for" prevention services. Such "but for" criteria included a lack of housing options, insufficient money to obtain housing or remain stably housed, and inadequate support networks to sustain stable housing. The United States Department of Housing and Urban Development (HUD) stated that the assessment of these three criteria is "relatively easy" (2011). Further, HUD provided some suggested risk factors for homelessness (e.g., mental health and substance abuse histories, significant medical

debt), but did not provide methods to uncover locally relevant risk factors so that service providers might know which households were at greatest risk for homelessness.

In addition to the somewhat vague instructions for risk determination, HPRP had potentially contradictory goals of service provision. The "but for" criterion encouraged service providers to target people at high risk for homelessness. However, HPRP's second goal, that services should be capable of preventing homelessness, encouraged service providers to target people with fewer risk factors. Accordingly, many communities, including Alameda County, adopted a triage approach, providing services to households whose risk was deemed neither too high nor too low (see Figure 4.1). In cases like this, in which assignment to treatment is based on an individual's score, a regression discontinuity design can be an appropriate method to determine a treatment's effectiveness.

**Figure 4.1.** Eligibility for Services Based on Continuum of Housing Stability (Triage)
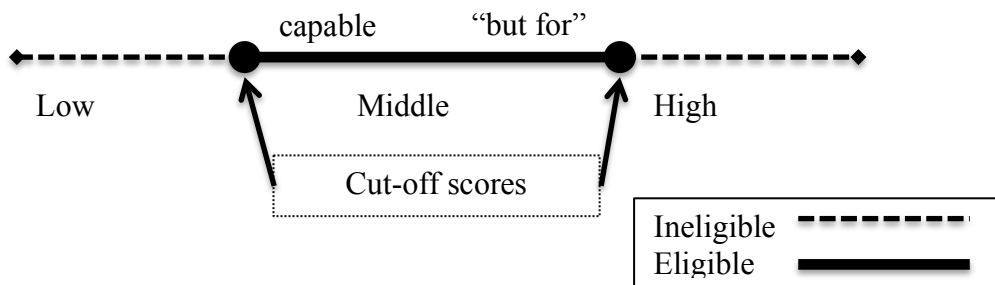


Existing HPRP investigations are limited, because no investigations of HPRP included a counterfactual. The current study addresses the limitation. It uses a regression discontinuity design to compare applicants who received homelessness prevention in Alameda County with those who did not. Although rigorous investigations of HPRP were not found in the literature, the current study builds on previous evidence of effectiveness from deep and shallow housing subsidies.

**Deep and Shallow Subsidies**

The current section reviews literature on evaluations of housing subsidies as a base from which to examine whether the shallow subsidies of HPRP helped people to sustain housing. Housing subsidies are one of the most commonly evaluated homelessness prevention strategies. Accordingly, it is not surprising that 93% of mayors in many of the nation's largest cities identified mainstream housing subsidies (e.g., Housing Choice Vouchers) as their primary homelessness prevention strategy moving forward (US Conference of Mayors, 2011).

Studies suggest that deep housing subsidies, typically Housing Choice Vouchers, reduce rates of homelessness. The most compelling evidence that deep subsidies reduce housing instability came from an experiment on the effects of housing vouchers for families (n = 8,731) who qualified for Temporary Assistance for Needy Families (TANF) in six sites: Atlanta, GA, Augusta, ME, Fresno, CA Houston, TX Los Angeles, CA, and Spokane, WA (Wood et al., 2008). Results from a treatment on the treated analysis indicated that 9% of the treatment group lacked their own housing (i.e., lived on the streets or in a shelter, or doubled up with friends or families) in the fourth year of the five-year study, as compared with 45% of the control group. The Housing Voucher Evaluation provided strong evidence that vouchers offered multiple benefits to families, including improved housing stability. Khadduri (2008) reviewed studies of family homelessness and concluded that across studies, vouchers reduced rates of family homelessness effectively. She argued that policy makers should increase the number of vouchers offered, as this would be the most effective means of reducing rates of family homelessness.

An important policy question is whether HPRP works close to as well as deeper subsidies. HPRP, with shorter terms and shallower subsidies, is much less expensive than most deep subsidies. Although not associated with HPRP, one study compared recipients of shallow subsidies to those who did not receive such assistance. Dasinger and Speiglman (2007) used a survival analysis to model subsequent homelessness for participants in Project Independence (PI), a program for low-income individuals with HIV in Alameda County, CA. The impressive findings reflected a 99% rate of stable housing for PI participants in contrast to a 32% rate for the comparison group after one year. Furthermore, 96% of PI participants were stably housed after two years compared with 10% of the comparison group. Similar studies should be conducted in other cities with different populations.

In previous homelessness studies that investigated risk for shelter entry (Chapter 2; Shinn et al., 2013), family and individual applicants entered shelter at higher rates if they were at heightened risk for homelessness and ineligible for prevention services compared to those who received services. Applicants who were at lowest risk for homelessness did not enter shelter at differential rates across eligibility status. These findings from previous studies inform my hypotheses that 1) homelessness prevention services will be associated with decreased homelessness rates for applicants with low housing stability at the time of application, and 2) rates of housing outcomes will not differ across treatment or comparison status for applicants with high stability. To examine the effectiveness of HPRP, the current study will compare results from a treatment group to a counterfactual.

**Regression Discontinuity**

The goal of the current study is to estimate homelessness outcomes for people who participated in HPRP relative to outcomes without the program, holding all else equal. For quantitative evaluations, a program's effect is evidenced by the difference in the outcome for these two groups, where outcomes in the group that does not receive the treatment form a counterfactual. Random assignment would estimate this difference directly. Randomized-Controlled Trials (RCTs) assign individuals randomly to treatment or control groups with the goal of producing unbiased estimates of program effectiveness. Random assignment minimizes the selection bias that might otherwise arise from unaccounted for differences when alternative placement strategies are used. For RCTs with sufficient sample sizes, the differences in the groups post treatment are likely attributable to the effectiveness of the intervention because random assignment equates groups in expectation, on all characteristics. RCTs are widely accepted as the "gold standard" to detect program effects between treatment and control groups (Rossi, Lipsey & Freeman, 2004).

Due to the challenges and ethical considerations inherent in RCTs, only a few homelessness prevention programs have evidence of effectiveness resulting from random assignment. For those studies that have used random assignment, *treatment* is often the random assignment to a housing intervention such as the receipt of housing vouchers (Tsai, Kasprow, and Rosenheck, 2011; Wood, Turnham & Mills, 2008), homelessness prevention (Rolston, Geyer, & Locke, 2013), or supportive housing for homeless individuals with a history of mental illness (Tsemberis, Gulcur, & Nakae, 2004). *Control*

*groups* typically receive treatment as usual, and the differences in outcomes (e.g., avoiding homelessness) are attributable to the effect of the treatment.

Multiple quasi-experimental strategies (i.e., where similar treatment and comparison groups are formed by some process other than random assignment) have been suggested to uncover prevention effectiveness. Such strategies do not have the same ethical dilemmas as random assignment, but the lack of random assignment can introduce selection bias, (i.e., biased results from selection strategies that do not eliminate differences between the treatment and non-treatment groups before the start of the experiment) (Bloom, Michalopoulos & Hill, 2005; Rossi, Lipsey & Freeman, 2004). Although the results from quasi-experimental designs may be biased, such designs often permit researchers to avoid unethical research practices and to investigate program effectiveness in useful ways (Cook & Campbell, 1979).

In RDD, assignment to treatment is based on the value of an observed covariate (e.g., stability score) on either side of a fixed cut-off score (Cook & Campbell, 1979; Rossi et al., 2004). Due to the fact that individuals immediately on either side of a cut-off score typically share many of the same attributes (so much so that they are nearly identical), differences in outcomes can be attributed to program effectiveness. More than one cutoff score is possible, if some are deemed at such low risk that they do not need treatment and others are deemed at such high risk that treatment will make little difference – this is the classic triage model, where resources are devoted to those in the middle for whom they are expected to make the most difference.

The regression discontinuity approach can produce the least biased estimates of the quasi-experimental designs, and it is sometimes considered the best alternative to the

"gold-standard" approach of random assignment (Rossi, Lipsey & Freeman, 2004). Imprecise control over assignment resembles randomization close to the cut-off score (i.e., local randomization) (Imbens & Lemieux, 2008). RDD is most similar to randomization close to the cut-off score (or scores) and less similar at extreme scores, where chance plays a lower role in treatment assignment (Cook, 2008). If individuals on two sides of a cut-off score are different because of precise self-selection into treatment, the assumptions of RDD do not hold (Lee & Lemieux, 2010). Accordingly, individuals should not have the ability to *precisely* manipulate their scores on a measure so that they do or do not receive treatment.

Historically, RDD has taken two basic forms: the parametric and the nonparametric models. The assumption of the parametric model is that one regression model accounts for the pretreatment association between the score variable and the outcome throughout the range of the data (Schochet, 2009). Alternatively, the non-parametric assumption is more flexible (Lee & Munk, 2008) and involves local linear regressions on either side of the cutoff(s) (Imbens & Lemieux, 2008). In RDD, there is a tension between obtaining a high enough sample size for adequate power and misspecifying the model by including individuals who scored far away from the cut-off score. The current paper uses both parametric and non-parametric RDD as recommended by Imbens and Lemieux (2008).

There are two primary RDD methods, based on how strictly treatment assignment depends on the covariate (c.f., Imbens & Lemieux, 2008). The sharp regression discontinuity (SRD) corresponds to the absolute dependence of treatment on a sharp cutoff on the covariate that determines eligibility (e.g., stability score). In a fuzzy

regression discontinuity (FRD), there is a shift in mean probability of treatment assignment at the cutoff(s) as opposed to the strict shift (from 0 to 1 or vice versa) of probability of assignment to treatment in the SRD. The current research paper uses a FRD design to examine the effectiveness of HPRP to increase housing stability in Alameda County, CA. Not only is the current study one of the first investigations of HPRP, but also it is the first known investigation of a triage approach to homelessness prevention. Specifically, the current study contributes to previous knowledge by answering the question: Did HPRP in Alameda County, CA, decrease rates of housing instability for applicants who received services?

**Data**

This section describes the data for the current study. The sample included 2,761 applicants for the Homelessness Prevention and Rapid Re-Housing Program (HPRP) at participating community-development agencies in Alameda County, CA. Applicants who received an intake survey between October 21, 2009 and April 28, 2012 are included in the analyses. One limitation of the intake survey is that the countywide implementation of the intake survey was inconsistent in some places. Some providers likely withheld the survey from applicants whom providers considered poor candidates for HPRP. Because HPRP was one of the most substantial services available, individuals who failed to receive services because of a lack of income were unlikely to receive deeper subsidies elsewhere.

For consistency with Alameda County's implementation of the HPRP program, the current paper discusses housing stability, as opposed to homelessness risk. Here, *housing stability* is defined as the inverse of homelessness risk, according to scores on the

County's intake survey. Those with low housing stability scores did not receive services because they were presumed to have so few resources that services would be ineffective to help them avoid homelessness. Instead, most of those with low housing stability scores were referred to other programs. Those with high housing stability scores did not receive services, because they were presumed able to avoid homelessness on their own. Those with intermediate scores received services.

**Outcome Variables**

Unlike investigations of homelessness that consider shelter entry as the sole homelessness outcome (Barnett et al., 2011; Chapter 2; Shinn et al., 1998; 2013), the current study includes multiple housing outcomes: 1) literal homelessness, 2) literal homelessness or imminent housing loss (hereafter referred to as *threatened homelessness*), and 3) literal homelessness, imminent housing loss, or unstable housing (hereafter referred to as *any housing instability*), as recorded in Alameda County's Homelessness Management Information System (HMIS). In the years leading up to HPRP's nationwide implementation, HMIS databases expanded to include most metropolitan areas, and the Department of Housing and Urban Development incentivized comprehensive data collection (Gutierrez & Friedman, 2005). Alameda County's HMIS system included three outcomes of homelessness:

Literal homelessness (according to HUD's definition):

- In a shelter, transitional housing, or in a place not meant for habitation
- Escaping a domestic violence situation
- Just exited jail, substance abuse treatment, hospital, psychiatric facility, or foster care setting from shelter or the streets

Imminent housing loss:

- Being evicted, discharged, or otherwise notified of imminent housing loss

Unstable housing:

- Doubled up, being evicted from public or assisted housing

The degree to which measures captured instances of each outcome is considered in the discussion.

**Covariates**

The current subsection describes covariates that were collected at the time of application. Intake workers computed a stability score – the primary covariate – from a set of housing stability subscales. Table 1 gives descriptive statistics for the subscales and variables that intake workers used to create the overall stability score. The subscales were: *employment and income potential, financial status,* as well as *housing and legal*. Further, table 4.1 includes additional baseline covariates and outcomes participants experienced after applying for services. If the assumptions of RD hold, people immediately on either side of a cut point should have nearly identical characteristics. Thus control variables should not differ significantly across cut points. I estimate models with and without the set of covariates.

**Table 4.1.** Descriptive Statistics for Outcomes, Baseline Covariates Included in the Stability Score, and Additional Covariates (n = 2761)

| Variables | Percent or *Mean* |
|---|---|
| **Subsequent Outcome Variables** | |
| Literal Homelessness | 2.1 |
| Homelessness or Imminent Housing Loss | 8.9 |
| Any Housing Instability | 14.1 |
| **Baseline Covariates** | |
| *Employment and Income Potential* | |
| Employed Part-Time | 17.9 |
| Unemployed | 53.1 |
| High school Diploma | 78.3 |
| Receiving Public Benefits | 67.1 |
| *Financial Status* | |
| <30% AMI | 82.8 |
| Rent > 40% Income | 82.9 |
| Outstanding Debt > $1000 | 39.7 |
| *Housing and Legal* | |
| Eviction History | 13.1 |
| Being Evicted from Public Housing | 7.5 |
| Leaseholder History | 79.1 |
| Homeless >once last year | 20.3 |
| Previously Homeless (Self Report) | 31.6 |
| Previously Homeless (HMIS Record) | 15.0 |
| Previous Night Homeless (Self Report) | 21.0 |
| Criminal Justice History | 16.6 |
| *Additional Covariates* | |
| Female | 73.2 |
| African American | 68.2 |
| Hispanic | 15.5 |
| *Age (mean)* | *39.3* |
| Veteran | 6.1 |
| Married | 12.8 |
| Pregnant | 3.8 |
| Escaping Domestic Violence Currently | 6.4 |
| Domestic Violence History | 30.6 |
| Living Doubled up | 17.8 |
| Disabled | 38.4 |
| Unhealthy Compared to Peers | 26.8 |

**Regression Discontinuity Design (RDD) - Graphical Representations**

Following recommendations from the literature (Imbens & Lemieux, 2008; Jacob, Zhu, Somers, & Bloom, 2012), the current study examines plots of 1) the probability of receiving treatment as a function of the scoring variable to investigate shifts in probability around the cutoffs, 2) the covariates as a function of the scoring variable to illustrate whether there are unexpected jumps in covariates around the cut-off scores, 3) the density of the rating variable to show the proportion of individuals who received particular scores, and 4) the outcomes as a function of the scoring variable to investigate preliminary evidence of a treatment effect at the cut-off score.

**1. Graphs of probability of receiving treatment as a function of the rating variable (to assess sharp vs. fuzzy discontinuities).** The graph of the probability of receiving treatment shows how rates of treatment receipt differed on either side of the cutoffs in the stability scores at time of application. As described in the introduction, absolute shifts in the probability of treatment receipt on either side of a cutoff (i.e., probability shifts from 0 to 1) correspond to a sharp regression discontinuity (SRD). Alternatively, if there is a significant shift in probability on either side of a cutoff, but the shift fails to be absolute, a fuzzy regression discontinuity (FRD) is appropriate. A FRD permits some control group members to receive treatment and some treatment group members to fail to take up treatment.

For the current study, two factors suggest that the discontinuities in the probability of treatment are not sharp. First, intake workers could override assignment to treatment. That is to say, intake workers could offer applicants treatment if they scored outside of the eligibility cutoffs. Second, applicants who received scores within the range

of eligibility sometimes failed to take up treatment. To the extent that intake workers altered scores to precisely control assignment to treatment, RDD will be invalid (Lee & Lemieux, 2010). However, scores on the scoring variable did not appear to be systematically manipulated to shift individuals into or out of eligibility, especially around the lower stability cutoff (described further in Figure 4.3).

Figure 4.2 shows the rate of treatment for individuals for each discrete stability score. The dashed lines signify the cut-off scores, and the size of each point is weighted by the sample size for each score. The triage approach is clearly apparent in this graph. Rates of treatment were highest for those between the cutoffs and much lower immediately on the other sides of the cut-off scores. This graph suggests that an FRD approach is the appropriate choice for analysis: although there was a shift in the rate of treatment receipt across the cut-off scores, the shift was not an absolute change. Specifically, the current study describes a type II FRD (Jacob, Zhu, Somers, & Bloom, 2012) where some members who are assigned to the treatment group do not actually receive treatment (i.e., no-shows) and some members who are assigned to the comparison group receive treatment (i.e., crossovers) (Battistin & Retorre, 2008).

**Figure 4.2.** Rate of Treatment Receipt by Stability Score (with Frequency Weights)



Note: Dots are sized by frequency weights
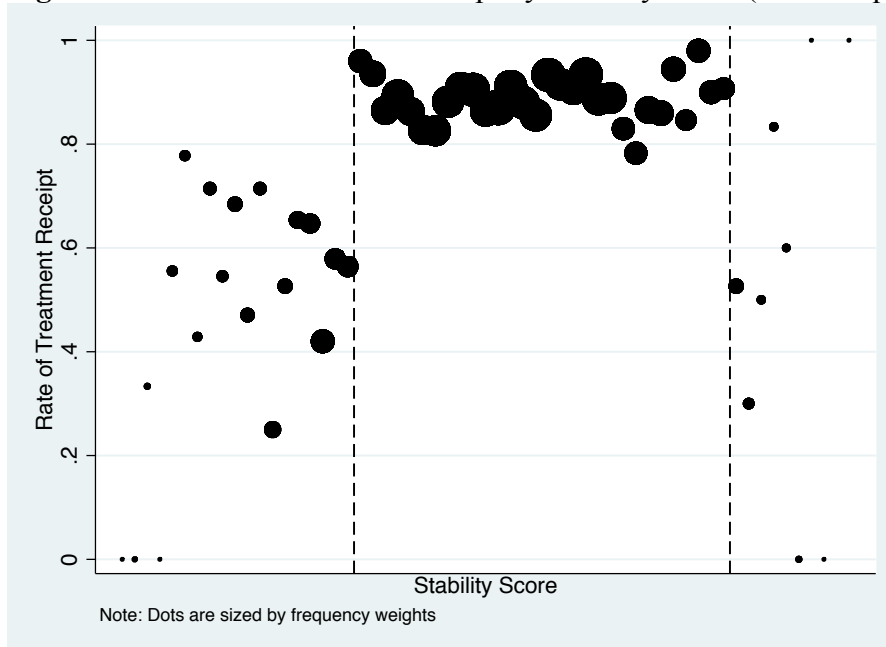
**2. Relationships of covariates & rating variable (to assess internal validity).**

Each covariate was inspected for evidence of a jump around the cut-off scores. If the

visual representation suggests a discontinuity at the cutoff for a covariate, then

differences between individuals on either side might not resemble random assignment,

and the assumptions of RDD would fail to be met. If a discontinuity occurs at the cutoff

for a particular covariate, then applicants on the two sides of the cutoff would differ on

that factor. Accordingly, differences in the outcome would no longer be solely

attributable to treatment. In addition to graphical analyses, covariates were tested

statistically for discontinuities similar to tests for the outcomes. In the current study, most

graphs and all analyses lacked evidence of a discontinuity for covariates.[21]

**3. Density of the rating variable (to assess manipulation of treatment).** The

systematic manipulation of scores to include or exclude individuals from treatment can

---

[21] Due to the large number of potential covariates, a table of estimates that shows no
evidence for discontinuities is presented in the appendix.

undermine the validity of RDD (Lee & Lemieux, 2010). Accordingly, a distribution of the rating variable's density should not contain drastic shifts on either side of the cutoffs. In the current study, the density of the rating variable (Figure 4.3) shows how smoothly the rates of scores transition across cutoff scores. Here, the higher cutoff score – the one corresponding to higher housing stability – appears to be somewhat problematic (i.e., scores seem to be gathered on the side of the cutoff that favors treatment). Consequently, evidence of manipulation seems more likely for the group that would have fallen to the right of the upper stability cutoff score (but who might have received lower stability scores to enable treatment) than for the group that fell to the left of the lower stability cutoff score. Further, the sample sizes are small for the group that fell above the higher stability cutoff score. Thus I am less confident in results from the discontinuity around the higher cutoff score than those around the lower cutoff score.

**Figure 4.3**. Density of Stability Score Overlaid with Normal Distribution (continuous curve) and Kernel Density Plot[22](dashed curve)
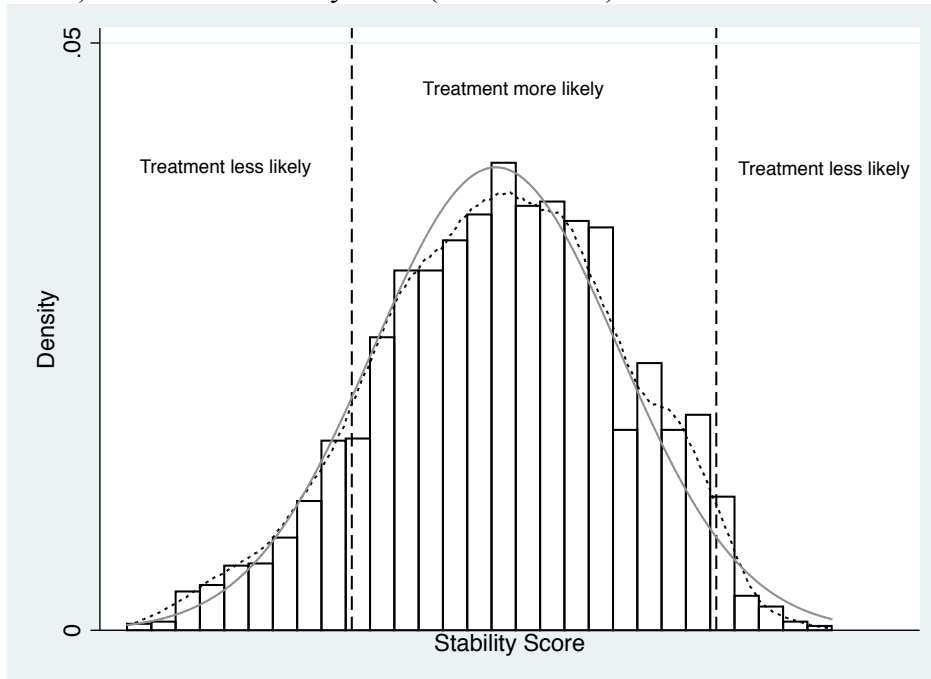


**4. Graph plotting the relationship between the outcomes and rating variable (to visualize the magnitude of the impact).** The graphs of outcomes as a function of stability score provide initial evidence of a discontinuity of the outcomes around the cutoff score. If the lines appear to be discontinuous at the cutoff, more sophisticated RDD investigations are then used to estimate treatment effects. Two graphs are presented for discontinuities around the lower stability cutoff score for each outcome to show the clearest visual representations of the data, while still allowing readers to examine possible differences at the cutoff scores. The first graph for each outcome takes into account 80% of the observations around the lower cutoff, with the most extreme 20% excluded. The second graph focuses on the closest 50% of observations around the cutoff

---

[22] A Kernel Density Plot is a non-parametric tool often used to show how well the data fit a normal distribution. For continuous variables, Kernel density estimates converge faster to the true underlying density than bins in a histogram (for more information, see: Scott, 1979).

score, with the most extreme 50% excluded. Following the graphs that correspond to low

housing stability scores, two graphs around the cutoff score that corresponded to high

housing stability are presented for each outcome. I use Lowess smoothers[23] to fit a curve

through the data points. As a reminder, the outcomes were: 1) literal homelessness, 2)

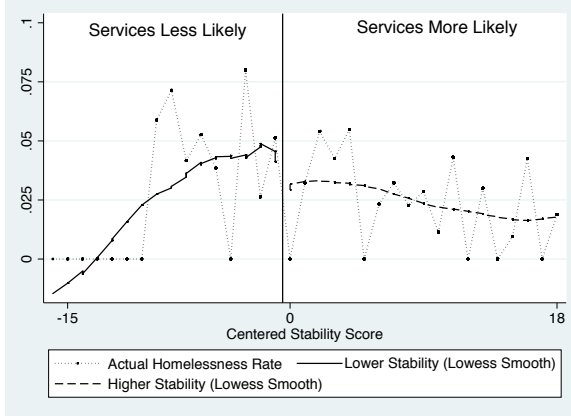threatened homelessness, and 3) any housing instability.

Figures 4.4 through 4.9 provide initial visual evidence of discontinuities in the

rates of unstable housing outcomes at the cutoff scores of applicants with low housing

stability scores. Each dot represents the average rate of an outcome for applicants for

each stability score. The solid lines are Lowess curves for applicants that were less likely

to receive treatment (i.e., their housing stability scores at baseline were too low and fell

below the cutoff). Dashed lines represent Lowess curves for applicants that were more

likely to receive services (i.e., they fell above the cutoff score). The vertical line

represents the cutoff score. The graphs indicate the possibility of discontinuities, with

curves that mostly suggest higher rates of subsequent unstable housing for those that

were less likely to receive treatment compared to those that were more likely to receive

treatment. This difference is especially apparent near the cutoffs, which is the area of

interest for regression discontinuity designs.[24] Note that the scale of the ordinate varies

across graphs, depending on the range of the data.

---

[23] Lowess (Locally weighted scatter plot smoothing) is a non-parametric regression
method that fits a smooth curve to the data via simple – yet computationally intensive –
local models (for more information, see: Cleveland, 1979).
[24] Lowess curves are most influenced by nearby scores, but more distant scores still affect
the overall shape of the curves. Accordingly the shape of the curve closest to the cutoff
differs when I consider 80% of the data compared to 50% of the data.

**Figure 4.4.** Outcomes Around Lower Stability Score Cutoff: Rates of Literal Homelessness (80% of data)



**Figure 4.5**. Outcomes Around Lower Stability Score Cutoff: Rates of Literal Homelessness (50% of data)
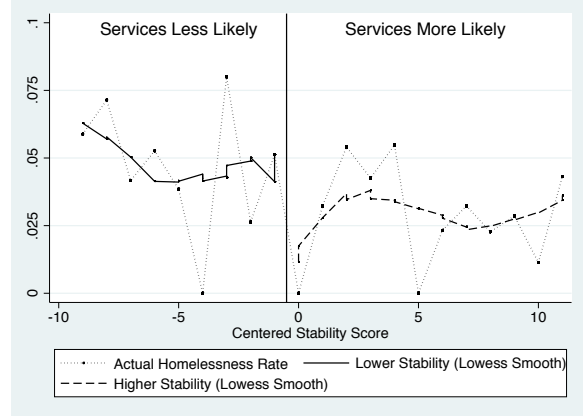


**Figure 4.6.** Outcomes Around Lower Stability Score Cutoff: Rates of Threatened Homelessness (80% of data)



**Figure 4.7.** Outcomes Around Lower Stability Score Cutoff: Rates of Threatened Homelessness (50% of data)



**Figure 4.8.** Outcomes Around Lower Stability Score Cutoff: Rates of Any Housing Instability (80% of data)



**Figure 4.9.** Outcomes Around Lower Stability Score Cutoff: Rates of Any Housing Instability (50% of data)



109
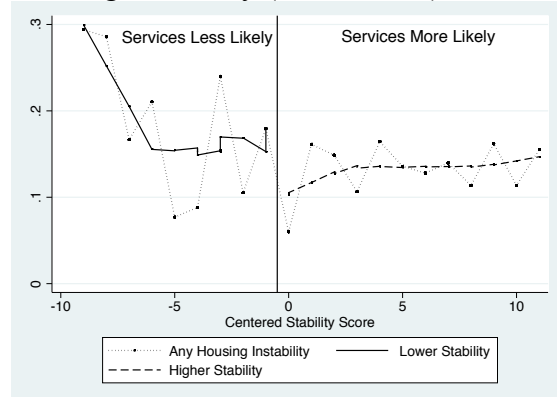
On the other hand, Figures 4.10 through 4.15 appear less suggestive of discontinuities in the rates of unstable housing outcomes at the cutoff scores of applicants with higher housing stability scores. The solid lines in Figures 4.10 through 4.15 represent applicants who were less likely to receive treatment (in this case, they were deemed too stable and fell above the cutoff score). The dashed lines represent applicants who were more likely to receive services (they fell below the cutoff score).

The findings for applicants with higher housing stability scores are less supportive of an effect of treatment than for those with lower scores. Unexpectedly, rates of unstable housing were lower for applicants immediately to the right of the cutoff score (who were less likely to receive treatment) compared to those just on the left. As scores extend further away from the cutoff score, rates of unstable housing for those to the right of the cutoff score rose briefly, before declining sharply. Further, sample sizes above the upper cutoff were small (n = 98), leading to sharp changes in average outcome when, in reality, few individuals experienced such outcomes.
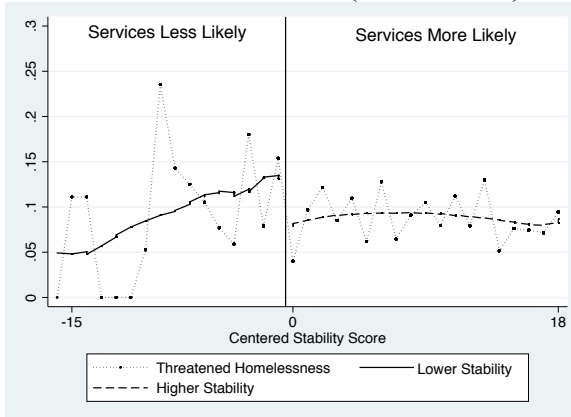
**Figure 4.10.** Outcomes Around Higher Stability Score Cutoff: Rates of Literal Homelessness (80% of data)
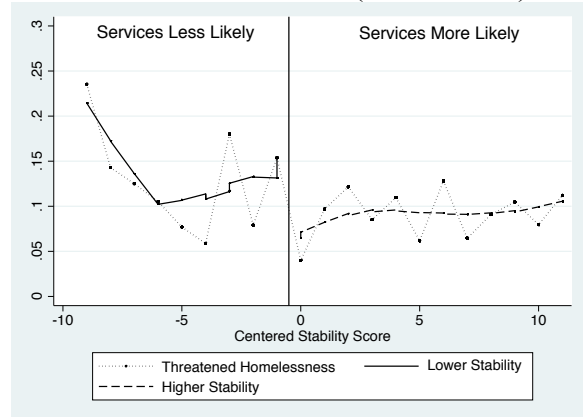


**Figure 4.11.** Outcomes Around Higher Stability Score Cutoff: Rates of Literal Homelessness (50% of data)
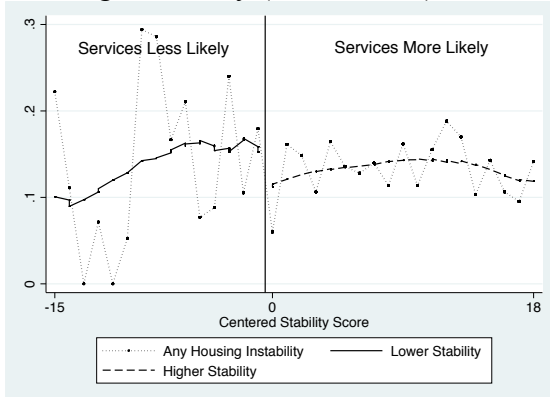
**Figure 4.12.** Outcomes Around Higher Stability Score Cutoff: Rates of Threatened Homelessness (80% of data)
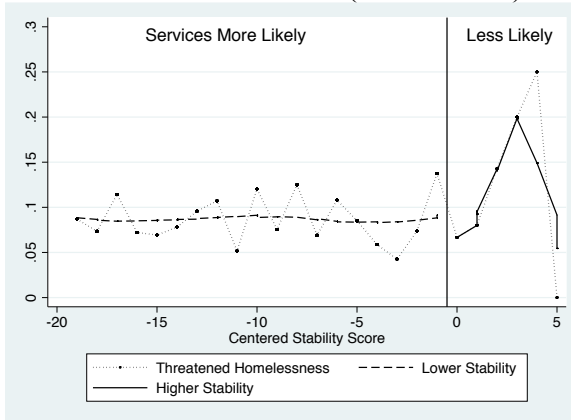


**Figure 4.13.** Outcomes Around Higher Stability Score Cutoff: Rates of Threatened Homelessness (50% of data)
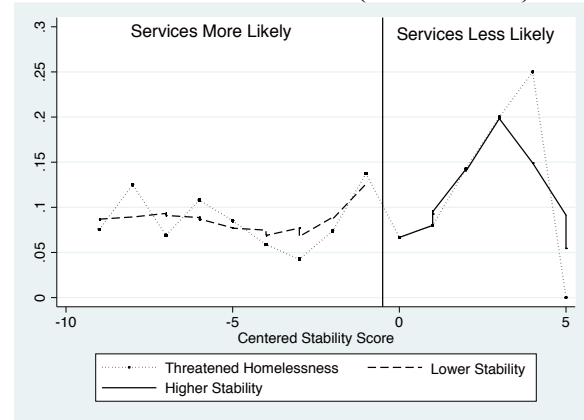


**Figure 4.14.** Outcomes Around Higher Stability Score Cutoff: Rates of Any Housing Instability (80% of data)
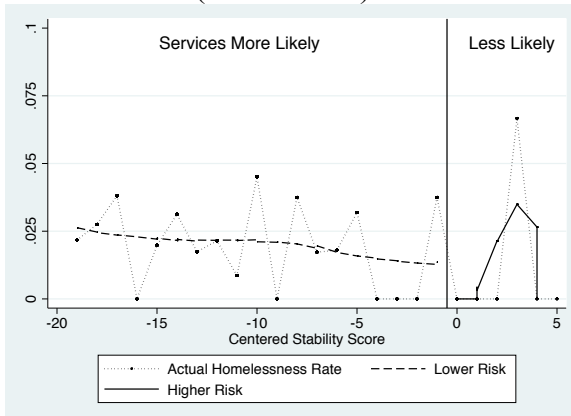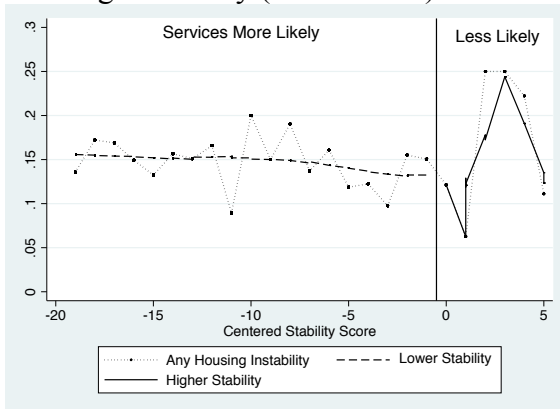


**Figure 4.15.** Outcomes Around Higher Stability Score Cutoff: Rates of Any Housing Instability (50% of data)
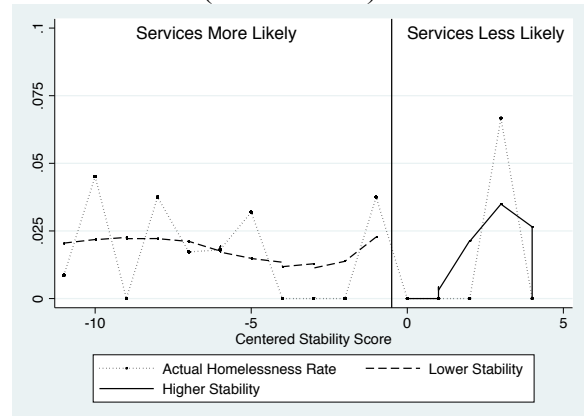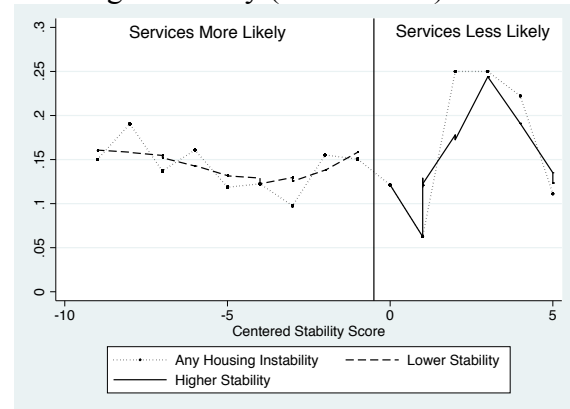


## Regression Discontinuity Design (RDD) - Analysis

As stated in the introduction, the current chapter uses a combination of parametric and non-parametric FRD to examine the effectiveness of HPRP to reduce rates of homelessness in Alameda County, CA (Imbens & Lemieux, 2008). To estimate the effectiveness of treatment, some methodological issues should be addressed. First, the goal of carrying out the FRD is to obtain an unbiased estimate of the effect of treatment for individuals who received treatment. However, the figures showing outcomes around the cutoff scores show visualizations for those who were intended to be treated, without

111

consideration of whether or not they actually received treatment. As Figure 4.2 showed,

many applicants in the comparison group received treatment even though they scored into

the comparison side of the cutoff. Conversely, some applicants failed to take up treatment

when they were assigned to the treatment group. Simply comparing groups who did and

did not receive HPRP thus confounds the causal effect of treatment with the effects of

whatever factors in addition to the instability score influenced whether applicants

received treatment. The Local Average Treatment Effect (LATE) – or the causal effect of

treatment – can be retrieved using an instrumental variable approach (Jacob, Zhu,

Somers, & Bloom, 2012).

The FRD is estimated with an instrumental variable framework, using the scoring

variable, an indicator of whether the applicant should be assigned to treatment based

solely on the cutoff score, and an endogenous treatment variable (whether applicants

actually received services or not). This estimation is carried out with two-stage least

squares regression (Trochim, 1984; 2001). The FRD is calculated with the following two

equations, borrowed from Jacob, Zhu, Somers, and Bloom (2012):

First stage: $\quad\quad\quad\quad\quad\quad T_i = \alpha_1 + \gamma_0 C_i + f_1(s_i) + \epsilon_i$
Second stage: $\quad\quad\quad\quad\quad Y_i = \alpha + \beta_0 \hat{T}_i + f_2(s_i) + \mu_i$

Where:
$T_i = 1$ if an individual receives treatment, and 0 otherwise;
$C_i = 1$ if an individual should be assigned to treatment based on the cutoff score, and 0
otherwise;
$s_i$ = centered scoring variable for each individual (the primary covariate);
$f_1(s_i)$ = the association between the scoring variable and actual treatment receipt for an
individual;
$\epsilon_i$ = random error in the first stage;
$Y_i$ = the outcome for an individual;
$f_2(s_i)$ = the association between the scoring variable and the outcome for an individual;
$\mu_i$ = random error in the second stage.

The equations use ordinary least squares (OLS) regressions to estimate each stage. The predicted value of treatment ($\hat{T}$) is used in the second-stage regression to estimate the effect of treatment for each outcome. Stata's *ivregress* command, specified for two-stage least squares regressions, corrects the standard errors in the second equation (Nichols, 2007).[25]

In addition to the parametric estimation, the current study estimates non-parametric, local linear regressions with different bandwidth choices as recommended in the literature (Imbens and Lemieux, 2008). Using Stata's *rd* command, LATEs are obtained using an optimal bandwidth based on the data (Imbens, & Kalyanaraman, 2012). The optimal bandwidth varies with each outcome, which inhibits a formal comparison with the parametric estimates. Additionally, to investigate the robustness of the effects across bandwidth choice, effects are obtained for the following percentages of the optimal bandwidth around each of the cut-off scores: 30, 50, 150, and 200. All models were estimated with and without baseline covariates and, separately, with risk scores from an efficiency analysis (chapter 3). The addition of these variables did little to the results of the RDD models substantively, although confidence intervals became slightly narrower compared to analyses without covariates. The analyses in the results section are presented without additional covariates in the models.

---

[25] Each of the outcomes in the current study is binary, which would normally call for a regression better suited for dichotomous outcomes (e.g., logit, probit). However, for two-stage least squares analyses, linear regressions achieve consistent estimates of the average effect of treatment, even for binary outcomes (Angrist & Pischke, 2009; Nichols, 2011). Thus the current study estimates a linear probability model based on the equations presented above.

**Results**

The parametric two-stage least-squares analyses suggested that HPRP might have substantial effects on rates of some forms homelessness, but confidence bounds were wide, so results were at best marginally significant. Table 4.2 reports estimates of the effect of treatment (modeled as an instrumented variable), including the scoring variable (modeled as an exogenous variable). Regression discontinuity is especially focused on observations around the cutoff score. To examine effects as distant observations further from the cutoff are increasingly ignored, four models are estimated for each outcome around the cutoff scores. The first model includes all of the data from the lowest observed stability score and stops at the upper cutoff score. The second model contains 80% of observations around the lower cutoff score. The third model includes 50% of observations around the lower cutoff score. The fourth model contains 20% of observations around the lower cutoff score. Accordingly, the models focus increasingly on the cutoff score and exclude distant cases from the analyses.

The models that include all of the data or 20% of the data represent the most inclusive and least inclusive parametric analyses, whereas the models that include 80% or 50% of the data contain the same cases presented graphically in the methods section. The same strategy was used for the upper cutoff score, with the lower cutoff score representing the lower boundary of the data and the highest observed stability score as its upper boundary. If treatment reduced the probability of housing instability, the effects would be negative for both halves of Table 4.2.

The LATE for threatened homelessness approached significance in the expected direction (-.31, $p < .10$) for the smallest sample of cases around the low stability cutoff.

114

Thus, nearly a third of applicants excluded from HPRP by their cutoff score were

estimated to experience homelessness, compared to almost no applicants whose score led

them to receive HPRP.    For any housing instability, the effect of treatment was nearly as

large and again approached significance (-.29, p < .10) for 20% of cases around the low

stability cutoff. For literal homelessness, the effect was in the expected direction around

the low stability cutoff, and large enough to be meaningful (-.17, p < .15). For all three

outcomes, estimated effects were largest close to the cutoff. As expected, the LATE was

not significant around the high stability cutoff. However, estimated effects were in the

unexpected direction.

**Table 4.2.** Two-stage Least-squares Analysis: Regression Coefficients for the Local
Average Treatment Effects of HPRP

| N = 2761 | Low Stability Cutoff | | | | High Stability Cutoff | | | |
|---|---|---|---|---|---|---|---|---|
| | All | 80% | 50% | 20% | All | 80% | 50% | 20% |
| **Literal Homelessness** **n =60 (2.2%)** | | | | | | | | |
| Treatment Instrument | -.01 (.04) | -.01 (.04) | -.03 (.06) | -.17 (.11) | .00 (.03) | .01 (.05) | .00 (.00) | .10 (.07) |
| Scoring Variable | .00 (.00) | .00 (.00) | .00 (.00) | .02 (.01) | .00 (.00) | .00 (.00) | .00 (.00) | .01 (.01) |
| **Homeless or** **Imminent Home Loss** **n =246 (8.9%)** | | | | | | | | |
| Treatment Instrument | -.06 (.07) | -.06 (.08) | -.11 (.09) | -.31 (.17)~ | -.05 (.07) | -.05 (.07) | -.05 (.08) | .19 (.13) |
| Scoring Variable | .00 (.00) | .00 (.00) | .00 (.00) | .03 (.02) | .00 (.00) | .00 (.00) | .00 (.00) | .03 (.01)~ |
| **Any Housing** **Instability** **n =389 (14.1%)** | | | | | | | | |
| Treatment Instrument | -.07 (.08) | -.07 (.10) | -.12 (.11) | -.29 (.19)~ | -.03 (.08) | -.05 (.09) | -.05 (.10) | .18 (.52) |
| Scoring Variable | .00 (.00) | .00 (.00) | .00 (.00) | .03 (.02) | .00 (.00) | .00 (.00) | .00 (.00) | .03 (.01)~ |

n is number of applicants who experienced the outcome subsequently
SEs in parentheses
~ significant at p < .10

In addition to the parametric investigations, the current study uses local linear

regressions to examine additional evidence of treatment effectiveness. For the non-

parametric analysis, multiple models are fitted to the data with various bandwidths for

each model. The *rd* command in Stata selects an appropriate bandwidth based on the data

(for a detailed explanation of bandwidth selection, see Imbens, & Kalyanaraman, 2012). For the FRD, a local Wald estimator represents the estimated LATE. The Wald estimator is a ratio of the estimated discontinuity (i.e., the estimated local mean difference) for the outcome divided by the estimated discontinuity of the treatment variable.

Findings for local linear regressions are similar to those from the parametric analysis. The Wald estimators (i.e., LATEs) were non-significant (p < .05), but approached significance in the expected direction around the low stability cutoff score. Conversely, the Wald estimators around the high stability cutoff score were in the opposite direction – individuals who received treatment would have higher expected rates of the unwanted outcome, if the models had reached significance.

Table 4.3 shows that the discontinuities in treatment (under the treatment column) were all significant in the expected direction. For example, the difference in the estimated average rate of treatment was .42 for applicants immediately on the right side of the low stability cutoff score compared to applicants immediately on the left for the optimal bandwidth associated with Literal Homelessness (in this case, 2.3 points on the stability scale). The differences in estimated effects of treatment varied across outcomes, because different optimal bandwidths were selected for each outcome (see graphs 16 through 21 for exact optimal bandwidth values across outcomes). The best estimates of effects at the optimal bandwidth (Wald estimates) were quite similar to those from the parametric analysis (-.19, -.32, -.32), but the confidence bounds were broader and no results approached significance. None of the discontinuities in outcomes (under the outcome column) were significant at $p < .05$. Literal homelessness approached significance (p < .10) around the high stability cutoff score, albeit in the unexpected direction.

116

**Table 4.3.** Wald Estimates (Effects of Treatment at Cutoff), Outcome Discontinuities (Difference in Conditional Means of Outcomes at Cutoff), and Treatment Discontinuities (Difference in Conditional Means of Treatment at Cutoff) of Local Linear Regression at Optimal Bandwidths with Standard Errors as well as 95% and 90% Confidence Intervals

| N = 2761 | Low Stability Cutoff Score | | | High Stability Cutoff Score | | |
|---|---|---|---|---|---|---|
| | Wald Estimate | Outcome Discontinuity | Treatment Discontinuity | Wald Estimate[a] | Outcome Discontinuity | Treatment Discontinuity |
| **Literal Homelessness** | | | | | | |
| Coefficient | -.19 (.15) | -.08 (.06) | .42 (.13)*** | .17 (.13) | -.05 (.03) ~ | -.32 (.16)* |
| 95% CI | -.48, .10 | -.19, .03 | .15, .68 | -.09, .43 | -.12, .01 | -.63, -.02 |
| 90% CI | -.43, .06 | -.18, .02 | .21, .63 | -.04, .38 | -.10, .00 | -.58, -.06 |
| **Threatened Homelessness** | | | | | | |
| Coefficient | -.32 (.22) | -.12 (.08) | .38 (.11)*** | .38 (.29) | -.13 (.09) | -.36 (.12)** |
| 95% CI | -.74, .10 | -.28, .04 | .17, .59 | -.20, .94 | -.31, .04 | -.59, -.14 |
| 90% CI | -.68, .04 | -.25, .01 | .19, .56 | -.10, .86 | -.28, .02 | -.56, -.16 |
| **Any Housing Instability** | | | | | | |
| Coefficient | -.32 (.23) | -.12 (.09) | .38 (.11)*** | .21 (.30) | -.07 (.11) | -.37 (.12)** |
| 95% CI | -.78, .13 | -.30, .05 | .17, .59 | -.39, .81 | -.29, .13 | -.59, -.14 |
| 90% CI | -.70, .06 | -.27, .03 | .19, .56 | -.28, .70 | -.25, .11 | -.57, -.17 |

a The Wald estimates for the high-stability cutoff are positive, because the probability of experiencing the outcome and the probability of treatment both decline on the right side of the cutoff score. Thus the ratio of the two numbers yields a positive effect.
~$p<.10$, *$p < .05$, ** $p < .01$, *** $p < .001$

For the local linear regressions, bandwidth choice can alter estimates and the literature recommends sensitivity analyses with bandwidths that are larger and smaller than the optimal bandwidth (Jacob, Zhu, Somers, & Bloom, 2012). Accordingly, the current study examines LATEs and 95% confidence intervals at the optimal bandwidth and narrower and wider bandwidths to examine the robustness of the effect. To visualize the robustness of effects across multiple bandwidth choices, graphs of the estimates, 95% confidence intervals, and bandwidths at 30%, 50%, 150%, and 200% of the optimal bandwidth appear below (Figures 4.16 to 4.21).

Figure 4.16 shows the estimated effects of treatment for the optimal bandwidth (2.3 points, marked with dashed lines) along with estimates for alternative bandwidth

choices, above and below the optimal bandwidth. A bandwidth of 2.3 points suggests that

a local regression is fitted and adjusted at intervals of 2.3 points along the covariate of

interest, namely the stability score at intake. The graph shows that the LATE ranges from

-.1 to -.2, depending on the bandwidth choice. Further, the 95% confidence intervals

include zero, which is consistent with the lack of significance that I found from the

parametric analysis. However the treatment effect is in the expected direction and large

enough to be meaningful and it approaches significance (p < .10) for some of the

narrower bandwidth choices. Figures 4.16 through 4.18 tell similar stories about the

effectiveness of HPRP to reduce literal homelessness, threatened homelessness, and any

housing instability – although effects fail to reach significance, they are in the expected

direction.

**Figure 4.16**. Lower Stability: Optimal Bandwidth[26] of LATE for Literal Homelessness
and 95% CIs (with sensitivity analysis for bandwidths at 30%, 50%, 150%, and 200% of
optimal bandwidth)



---
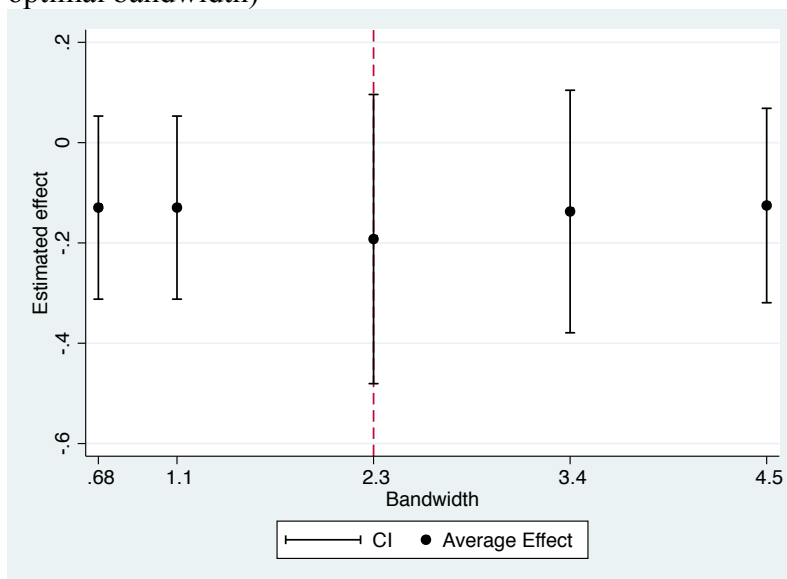
[26] Red dashed lines mark the position of the *optimal bandwidth* (Imbens, &
Kalyanaraman, 2012).

**Figure 4.17**. Lower Stability: Optimal Bandwidth of LATE for Threatened Homelessness and 95% CIs (with sensitivity analysis for bandwidths at 30%, 50%, 150%, and 200% of optimal bandwidth)



**Figure 4.18**. Lower Stability: Optimal Bandwidth of LATE for Any Housing Instability and 95% CIs (with sensitivity analysis for bandwidths at 30%, 50%, 150%, and 200% of optimal bandwidth)
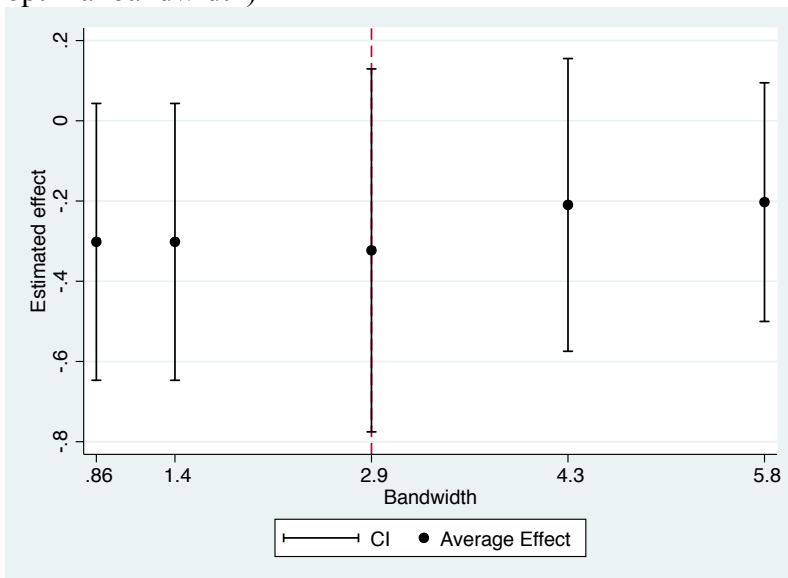


Figures 4.19 through 4.21 show parallel estimates of the effect of HPRP around the cutoff score associated with high housing stability. Crossing the cutoff was associated with significantly lower rates of treatment as expected. However, rates of the outcome

were also lower across the cutoff. Thus a positive (adverse) effect of treatment would be suggested if the Wald estimates were significant. In other words, treatment would seem to be associated with higher rates of housing instability. However, these findings should be interpreted with some caution. The graphical analyses in the Data section suggested that manipulation of service provision might have occurred around the high-stability cutoff score (see Figure 4.3). Accordingly, the validity of findings around this cutoff seems questionable.

**Figure 4.19**. Higher Stability: Optimal Bandwidth of LATE for Literal Homelessness and 95% CIs (with sensitivity analysis for bandwidths at 30%, 50%, 150%, and 200% of optimal bandwidth)
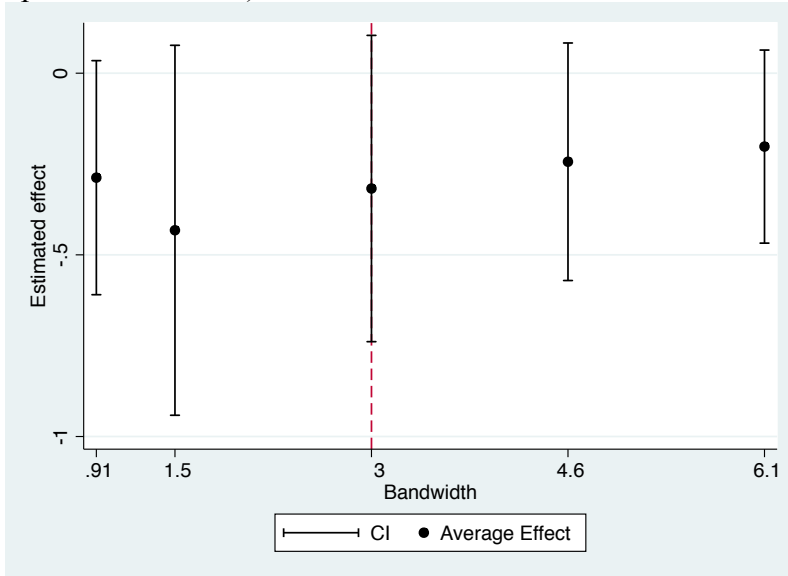
**Figure 4.20**. Higher Stability: Optimal Bandwidth of LATE for Threatened Homelessness and 95% CIs (with sensitivity analysis for bandwidths at 30%, 50%, 150%, and 200% of optimal bandwidth)
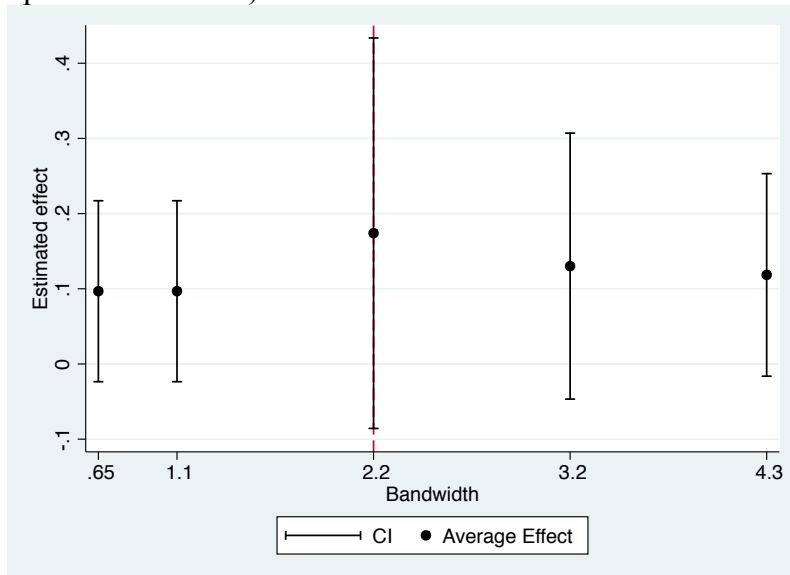


**Figure 4.21**. Higher Stability: Optimal Bandwidth of LATE for Any Housing Instability and 95% CIs (with sensitivity analysis for bandwidths at 30%, 50%, 150%, and 200% of optimal bandwidth)
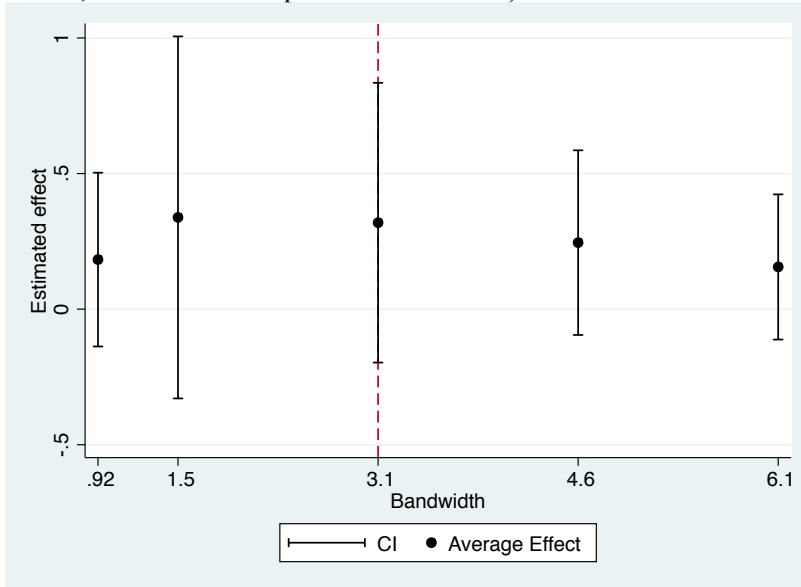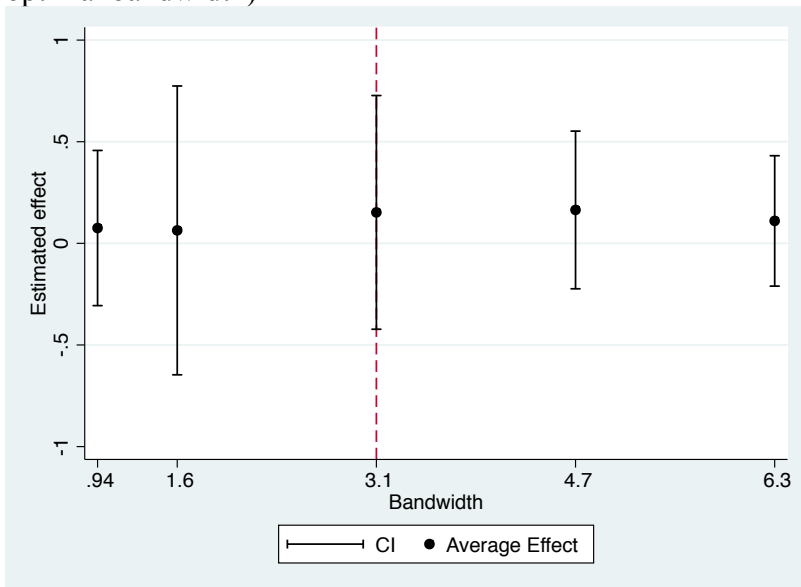
**Discussion**

Using regression discontinuity methods, this study found limited evidence that HPRP reduced homelessness rates in Alameda County, CA, in particular for applicants with very low housing stability ($p < 0.10$). The best estimate of the Local Average Treatment Effect (LATE) around the low stability cutoff is -.17 (90% CI: -.43, .06) for literal homelessness, -.32 (90% CI: -.68, .04) for threatened homelessness, and -.32 (90% CI: -.70, .06) for any housing instability. The design does not have power to rule out a null effect; neither does it have power to rule out a substantial effect.[27] Basic graphical analyses suggest discontinuities that are large enough to be meaningful in all three outcomes around the cutoff associated with lower housing stability, but parametric and non-parametric RDD analysis suggests that we cannot rule out chance as the explanation. Further, even these effects are not robust as we include cases further and further from the cutoff.  While results hinted at effectiveness for the cutoff associated with lower housing stability, no analyses provided evidence for effectiveness around the cutoff associated with high housing stability. If anything, results were in the opposite direction. The addition of covariates or an empirical risk score from previous analyses (Chapter 3) did little to the results of the RDD models substantively. The stability score that Alameda County used to determine service receipt was not related to literal homelessness but minimally related to the two more inclusive measures of threatened homelessness and housing instability.

The current study was limited in multiple ways. First, the limited sample size led to a lack of power, especially when smaller bandwidths of data were considered close to

---

[27] See Power Analysis in Appendix.

the cutoffs (Jacob et al., 2012). Power was also limited because the outcomes tended to be rare, especially literal homelessness (2.2% experienced the outcome). The best estimates of effects were large, but we cannot rule out that they were due to chance. Additionally, some assumptions of RDD seem to be violated around the cutoff associated with higher stability. In particular there was some evidence that some applicants around the cutoff associated with higher stability may have received scores that pushed them across the threshold favoring service receipt. This finding combined with the limited sample size above the higher cutoff (n = 104) compared to the sample size below the lower cutoff (n = 337) calls the validity of results around the cutoff associated with higher stability into question.

An additional limitation to this study is that the outcomes were treated as if they were uncensored. Given the limited follow-up time for some individuals (i.e., sometimes just over three months), some applicants may have become homeless after the study ended. However, previous studies of homelessness prevention found that applicants were at highest risk for homelessness soon after applying for services (Shinn et al., 2013, Chapter 2, Chapter 3). Nevertheless, future investigations might include longer periods of follow up (e.g., two or three years for all participants) so that all applicants have the same time period to potentially experience the outcome. The follow-up period is certainly not long enough to understand whether any effects of the temporary subsidies provided by HPRP are temporary or permanent.

A final limitation to this study is that some applicants failed to receive the intake survey when providers considered them to be poor candidates for HPRP. Reports from some providers suggested that applicants were not surveyed if they would have been

eligible for services but lacked the resources (especially income) necessary to avoid homelessness, even with the help of HPRP. If providers were correct that individuals would have been deemed eligible and would nevertheless have become homeless, the current study would have overestimated the effect of HPRP. Alternative assumptions about providers' accuracy could suggest a bias in the opposite direction. Overall, the elimination of some members of the population from the sample that was surveyed further calls the effectiveness findings from this study into question.

From a programmatic standpoint, findings suggest that Alameda County implemented HPRP without compromising the validity of the research design, at least around the lower-stability cutoff. For future studies, intake workers could record specific reasons for overriding the cutoffs. Because discontinuities in covariates failed to be apparent around the cutoffs, there is no evidence that intake workers granted services to specific groups. However, justifications from workers could improve future investigations, because evaluators could understand whether the scoring variable systematically excluded people with specific characteristics.

Improvement in the measurement of outcomes might provide clearer research findings and hence firmer guidelines for policy. According to data from the Annual Homeless Assessment Report (AHAR, 2013), Alameda County's HMIS captures almost 82% of shelter beds. This percentage is higher than percentages in many other areas, but Alameda County might do more to ensure that all shelter beds are included in their counts. Such an effort might lead to more accurate rates of the literal homelessness outcome measured in the current study. For example, the rate of literal homelessness for the current study (2.2%) was much lower than the rate of shelter entry for similar

populations in New York City (6.4% for individuals and 12.8% for families, albeit over longer periods than studied here), where the HMIS includes over 91% of shelter beds. A fully integrated HMIS would improve estimates of subsequent homelessness.

Measurements of the broader outcomes of threatened homelessness and any housing instability are also underestimates. These outcomes were recorded only if applicants for HPRP services (whether or not they received those services) called service providers again for additional assistance. Further, since applicants who received help might be more likely to seek additional help than those turned away initially, these measures could underestimate the effects of treatment. Observed effects would then be lower-bound estimates of actual effects.

Although the current study fails to find conclusive evidence that HPRP reduced rates of homelessness, RDD is an appropriate method to evaluate how well prevention programs work. Future studies of similar programs with larger samples – especially around the chosen cutoff(s) – and longer follow-up periods would yield clearer results (that is results with narrower confidence bounds). This study shows that prevention programs, like HPRP in Alameda County, can offer services in a way that is compatible with RDD analysis and with limited threats to validity. This is one of the first studies in the homelessness prevention literature to estimate effects using RDD. By using this design, the study included a counterfactual, which is vital to uncover how effective a program might be.

Despite the limited nature of the evidence for prevention in this study, there are two important implications for policy. First, in combination with studies of targeting of homelessness prevention efforts in Alameda County (Chapter 3) and elsewhere (Chapter

2; Shinn et al. 2013), the study suggests that prevention efforts are most likely to be effective if targeted at high-risk individuals with low prior housing stability. There is no evidence in any of these studies for widely used triage policies that exclude the highest-risk households from receiving assistance to prevent homelessness. Entry into homelessness was far from certain, even for broad definitions of homelessness and for households with the lowest housing stability scores.

Second, because the best estimates of the effects of HPRP were substantial, but poorly specified, collection of additional evidence is important. Other cities that can link HPRP assessments to data from Homelessness Management Information Systems could provide similar analyses. Even if each sample is relatively small and results inconclusive, given the low likelihood of people becoming homeless over short periods of time, the synthesis of evidence across studies, using meta-analysis or other tools, could provide clearer evidence about the effectiveness of efforts to prevent homelessness. HPRP is a relatively inexpensive program. If it has important effects, the policy implications would be enormous. Other studies should build on the initial step taken here to investigate how much homelessness prevention programs can reduce or eliminate the high rates of homelessness that plague the U.S.

# CHAPTER 5

## CONCLUSIONS AND CONTRIBUTION TO THE LITERATURE

### Homelessness Prevention: Effectiveness and Efficiency

This dissertation expands the understanding of effectiveness and efficiency of homelessness prevention programs with in-depth empirical analyses in Alameda County, CA and New York City, NY. The findings provide limited evidence that prevention programs can reduce entries into homelessness, and stronger evidence that programs can be made to work better by focusing efforts on individuals and families at higher risk. The studies also contribute to our understanding of the causes of homelessness, via the examination of particular risk factors in the two sites.

### Empirical Risk Models Increase Efficiency Compared to Intake Worker Judgments of Eligibility

Chapters 2 and 3 investigated efficiency – the extent to which services were targeted to those who most needed them.  Empirical models are useful in identifying relative risk. This is consistent with longstanding literature that supports empirical targeting models in multiple domains (Ægisdóttir et al., 2006; Dawes, Faust, & Meehl, 1989; Grove et al., 2000). In Alameda County, it was possible to target literal homelessness more efficiently than broader forms of homelessness; the two broader models do not differ significantly in terms of efficiency. The literal homelessness model for Alameda County falls between the efficiency of the two models in NYC – it is more

127

efficient than the model for families (especially at higher false alarm rates) and less efficient than the model for individuals (Chapter 3).

Remarkably, risk domains for homelessness tended to be consistent across populations (i.e., families and individuals) and location (i.e., New York City and Alameda County), although some survey questions differed substantially across sites. Previous homelessness was the predictor with the largest association with subsequent homelessness for all groups (Chapter 2; 3; Shinn et al., 2013). Perhaps people are less likely to avoid homelessness if they have already accepted an identity of homelessness from a previous experience than if they have never been without a home. Or perhaps the risk factors that precipitated the earlier episode persisted and continued to put people at heightened risk.

In addition to previous homelessness, risk factors associated with access to affordable housing and poverty that may be considered individual manifestations of structural risk were reliable predictors of subsequent homelessness across groups (Chapter 2; 3; Shinn et al., 2013). By way of contrast, individual vulnerabilities, such as mental illness, substance abuse, and domestic violence failed to be useful predictors of homelessness in any site, even though rates of these vulnerabilities varied across groups in the expected directions (Chapter 3). These findings contribute to our theoretical understanding of the origins of homelessness – individual vulnerabilities may operate through access to income and housing. Across groups, risk for homelessness was highest shortly after applying for services, although risk persisted for a year or more (Chapter 2; 3; Shinn et al., 2013).

Some risk factors differed, often because the programs asked different questions. For example, eviction was positively associated with homelessness across studies, but Alameda County asked applicants if they were being evicted from public housing, and this indicator was more reliable than simply having an eviction history (Chapter 3). Further, verbal eviction was more reliable than formal written eviction for individuals in NYC (Chapter 2).

In the future, programs might tailor their assessment procedures to include reliable predictors from other sites to examine their reliability in new locations. Findings across studies suggest that risk models may generalize across location (Chapter 3). This suggestion stems from the finding that structural causes manifesting as individual risk factors consistently predicted homelessness across sites (Chapter 3). Of course, risk models should be examined for fit with particular locations. Both communities studied here had lower homeownership rates, lower rates of rental vacancies, higher unemployment rates, and higher poverty rates than national averages (Chapter 3). As more locations investigate the efficiency of their targeting models, the degree to which models generalize across sites will become clearer.

**For the Narrowest Definitions of Homelessness, Programs Appear to Be Most Effective for those at Highest Risk**

Chapter 4 investigated the effectiveness of a modest prevention program in Alameda County, CA, and found limited evidence of large effects. For those with low housing stability, homeliness prevention services were protective, but effects were not statistically reliable (Chapter 4). The best estimate of the Local Average Treatment Effect (LATE) around the low stability cutoff is -.17 (90% CI: -.43, .06) for literal

homelessness, -.32 (90% CI: -.68, .04) for threatened homelessness, and -.32 (90% CI: -.70, .06) for any housing instability. The design does not have power to rule out a null effect; neither does it have power to rule out a substantial effect. This finding is supported by evidence that applicants at the highest risk of homelessness who received services experienced lower subsequent shelter-entry rates than those who were denied services in New York City, but services were not associated with lower shelter entry rates for low risk applicants  (Chapter 2; Shinn et al., 2013). For applicants in Alameda County, the outcome of literal homelessness exhibited a similar trend (Chapter 3). For those with lower risk (i.e., higher stability), services failed to be associated with different rates of subsequent homelessness – if anything, services may have led to increased rates of subsequent homelessness for this group (Chapter 4). However, for broader definitions of homelessness, providing services at any level of risk seemed to be associated with lower rates of these outcomes (Chapter 3).

One of the most substantial difficulties with investigating effectiveness is finding a suitable counterfactual, and Chapter 4 was the first known homelessness investigation to employ a regression discontinuity design (RDD). Such a design mirrors a randomized controlled trial around a cutoff score, and thus, includes a counterfactual. RDD studies are challenged with multiple threats to validity, but Alameda County did not seem to compromise the design, especially for those with the lowest rates of housing stability (Chapter 4).

Two primary challenges to the current dissertation likely affected the lack of reliability for the effectiveness investigation. First, while RDD is an appropriate design to uncover the Local Average Treatment Effect, RDD requires much larger sample sizes

than randomized controlled trials, because only participants with scores close to the cutoff that determines receipt of services are included in analyses. With sample sizes that were likely too small in the current dissertation, the confidence intervals were too broad to lead to reliable conclusions. Second, the outcome of subsequent literal homelessness was rare, which increased the difficulty of finding an effect. At least some of the infrequency of the measured outcome was due to a homelessness management information system (HMIS) that failed to incorporate all shelter beds (Chapter 4) and to short follow-up periods for some applicants. Nevertheless, the best estimate of the effect of prevention on homelessness was substantial – especially for the broadest definitions of homelessness – so it is important to the field to get a more reliable fix on how much prevention helps. Because other sites are unlikely to have much larger samples of high-risk applicants to work with, it will be important to aggregate results across sites (see recommendations in the Future Research section below). For now, homelessness prevention in Alameda County remains promising, but unproven.

Outcomes of prevention for broader definitions of homelessness (i.e., threatened homelessness and any housing instability) in Alameda County also appeared promising but not proven. In addition to the difficulties already cited for measures of literal homelessness, the broader outcomes contained additional biases, because applicants could report these outcomes only by returning to providers for additional services. Thus rates of each outcome were likely underestimates and, to the extent that applicants who received services at the outset were more likely to return to ask for additional help, estimates of intervention effects were likely conservative (Chapter 4).

**Policy Recommendations**

Empirical models are vital to efficiently offer prevention services to those at highest risk (Chapter 2; 3; Shinn et al., 2013). Accordingly, homelessness policy should support investigations of efficiency by including the guidance and funding necessary for rigorous evaluations. Such support should include technical assistance for local providers, and funding for ongoing examinations of model efficiency. Further, policy should emphasize the importance of sharing and integrating results across locations, and as evidence accumulates, offer guidelines for prevention based on the most recent and rigorous empirical risk models.

The nation's largest homelessness prevention effort, the Homelessness Prevention and Rapid Re-housing Program (HPRP), encouraged communities to use a triage design (Chapter 4), and none of the empirical investigations here support such an approach (Chapters 2; 3; 4). Instead, evidence from the current dissertation suggests that programs should focus on the highest risk applicants for subsequent shelter entry or literal homelessness (Chapter 2; 3). On the other hand, prevention seems to be associated with decreased rates of broader forms of homelessness and housing instability, regardless of risk level (Chapter 3). However, with no indication that prevention effectively reduced rates for those at low risk for homelessness (defined in all three ways) in the regression discontinuity chapter (Chapter 4), the apparent benefits of prevention for broader forms of homelessness and housing instability may not be concentrated at the high-risk end of the spectrum.

Previous homelessness, a lack of affordable housing, and poverty appeared to be the primary risk factors across NYC and Alameda County (Chapter 2; 3) that pushed

people into homelessness. Each of these risk factors has ties to policy. First, people who experience previous homelessness appear to be at highest risk of subsequent homelessness across populations and location. Accordingly, policy should provide additional support for those with previous bouts of homelessness so that they might escape cycling back into the shelter system. Such supports would not only increase the ability of the most vulnerable people to avoid homelessness, but also save public funds by avoiding repeated expensive shelter stays.

Second, a lack of affordable housing is apparent in NYC and Alameda County (Chapter 3), and such problems are often found in housing markets across cities. When a lack of affordable housing coincides with high poverty rates, the most vulnerable populations are left without access to suitable housing (Chapter 1). Housing vouchers and social welfare initiatives (e.g., TANF) are examples of policy that might be expanded or tailored to those at highest risk of becoming homeless. Efforts to combat both the lack of affordable housing and high rates of poverty would likely help to keep the most vulnerable people housed. In this sense, the findings of this study support the theorized *population* and *high-risk framework* (Apicello, 2010), where interventions would take place at multiple levels simultaneously.

**Prevention Recommendations**

In contrast to macro-level policy recommendations, the current subsection discusses recommendations for local prevention programs. Across studies, efficiency was substantially improved by using empirical targeting models compared to intake worker determinations of eligibility for services (Chapter 2; 3). Accordingly, this dissertation recommends using and evaluating empirical models to target services to applicants of

homelessness prevention programs. Further, for cities that share ecological characteristics with New York and Alameda County, local prevention providers might adopt a version of the targeting models compared in Chapter 3. For localities with more benign housing markets, more research is needed to see whether risk models are transferrable (see discussion on future research below).

To examine effectiveness and efficiency in additional sites, prevention programs should attempt to offer services in a way that remains true to the goals of the intervention while allowing evaluations to be conducted. This dissertation found that Alameda County offered services to the majority of applicants (83%) without compromising the validity of a regression discontinuity design for those with the lowest housing stability according to the County's rating system (Chapter 4). Other studies have also encouraged evaluation teams to investigate the effectiveness of their programs (e.g., HomeBase). Where resources are limited, a lottery (allowing a randomized control trial) is potentially a fair way to allocate scarce resources. Alternatively, quasi-experimental designs, especially regression discontinuity designs, may be more ethically defensible and lower cost alternatives to randomized control trials, although they require substantially larger sample sizes.

In addition to offering services in ways that permit rigorous evaluations with suitable counterfactuals, programs might increase the coverage of Homelessness Management Information Systems and consider integrating data from multiple social service agencies, not just those that are part of the homeless service system (Chapter 4). Homelessness outcomes occur rarely, and evaluators would obtain more reliable and valid estimates if locations captured outcome occurrences accurately. While obtaining

accurate estimates of shelter use might be the first step in capturing homelessness outcomes, designing a HMIS to capture broader outcomes of homelessness and housing instability would provide additional information for individuals and families who are insufficiently housed but have avoided shelter entry.

Some locations might be most interested in targeting services to those at highest risk of shelter entry because it is expensive. Other locations with greater access to resources might expand their prevention efforts to include broader definitions of homelessness. The current study found applicants who experienced shelter entry or literal homelessness shared more similarities with each other across location than with applicants who experienced threatened homelessness or any housing instability in the same location (Chapter 3). Accordingly, providers who are interested in investigating broader definitions of homelessness should keep in mind that risk factors may differ for those who experience literal homelessness compared to those with other types of instability.

In general, locations that seek to examine the effectiveness and efficiency of their programs should encourage intake workers to collect data thoroughly. Intake workers may be more interested in providing services than collecting accurate data, but poor data quality can be detrimental to the clients that service providers seek to assist. High rates of missing data can lead to problems that range from complicated analyses (e.g., multiple imputation) to biased results. Perhaps, supervisors and managers can share results from other investigations with their staff to highlight what high-quality data collection might mean for their clients. Additionally, researchers and evaluators might interact directly with program staff to incentivize adequate data collection.

Finally, prevention programs should publicize results so that they can learn best practices from one another. Local prevention programs can examine the effectiveness and efficiency of their models, and the current dissertation and previous research suggest that efficiency will improve with empirical models compared to current worker decisions (Chapter 2; 3; Shinn et al., 2013).

**Future Research**

Above all other findings, the current dissertation points out the need for future research to examine effectiveness and efficiency across locations. Effectiveness studies of homelessness prevention require a counterfactual to attribute differences in homelessness rates to treatment (Chapter 4). Future studies should strive for designs that include comparison groups that are similar as possible to the treatment group.

For designs that require large sample sizes, research results that seem substantial but that lack the power necessary for reliability can become a part of meta-analyses of homelessness prevention (Chapter 4). Accordingly, evaluators should pursue rigorous investigations and publicize results even when they may lack the ability to detect an effect. Additionally, further research is needed to test the effectiveness of particular kinds of interventions. Neither HomeBase in New York City nor HPRP in Alameda County provided detailed descriptions of exactly what they did for different applicants when they provided prevention services. Documenting programs (and any services available to comparison groups) is critical to learning from study results. Further, literature reviews of empirical studies are needed to illustrate the current state of effectiveness investigations.

The efficiency findings for similarities and differences across groups in the current dissertation highlight exciting possibilities for future research. First, as additional

jurisdictions incorporate empirical risk models, researchers should examine the extent to which models are generalizable (Chapter 3). Such investigations should expand beyond urban settings with limited access to affordable housing. Instead, prevention programs in rural areas as well as in cities with more benign housing markets should investigate empirical models for generalizability to their locations. Additional investigations of efficiency in different cities could go far to assist researchers and service providers to better understand how interactions of structural causes of homelessness with individual vulnerabilities differ across locations.

This dissertation begins to establish empirical evidence for the effectiveness and efficiency of modest prevention services. However, the findings are just a beginning, and they appear to leave more questions than definite answers. Although unreliable, effects of prevention were encouraging for those who had low housing stability. Additional studies and meta-analyses that combine results from multiple sites will allow researchers to specify the effects of prevention programs more precisely. In terms of efficiency, the results from studies in two housing markets with limited access to affordable housing suggest that structural issues are the driving forces of homelessness (Chapter 3). Future research is necessary to examine similarities and differences for these findings across housing markets.

The importance of individual manifestations of structural factors in predicting homelessness lends support to the *population and high-risk framework*. This study, and others (Messeri, O'Flaherty, & Goodman, 2011; Rolston, Geyer, & Locke, 2013; Shinn et al., 2013) suggest that prevention programs that provide services to individuals can counteract structural risk. Nevertheless, prevention programs might do better not only to

provide immediate prevention for individuals and families but also to combat the

structural forces that lead to high homelessness rates in the first place.

# REFERENCES

Ægisdóttir, S., White, MJ., Spengler, PM., Maugherman, AS., Anderson, LA., Cook, RS., Nichols, CN., Lampropoulos, GK., Walker, BS., Cohen, G., Rush, JD. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist. 34*(3), 341-382.

Alesina, A., & Glaeser, E. L. (2004). *Fighting poverty in the U.S. and Europe: A world of difference*. Oxford: Oxford University Press.

Apicello, J. (2010). A paradigm shift in housing and homeless services: Applying the population and high-risk framework to preventing homelessness. *The Open Health Services and Policy Journal, 3,* 41-52.

Banks, S., McHugo, G., Williams, V., Drake, R., & Shinn, M. (2002). A prospective meta-analytic approach in a multisite study of homelessness prevention, *New Directions for Evaluation, 94*, 45-60.

Barnett, J., Burand, B., DeBower, J., Danso, K., Han, L., Lynn, K. (2011). *Family homelessness in Hennepin County: A mixed methods examination of the rapid exit program and patterns of shelter mobility*. University of Minnesota: Humphrey School of Public Affairs.

Bassuk, E., Volk, K., & Olivet, J. (2010). A framework for developing supports and services for families experiencing homelessness. *The Open Health Services and Policy Journal, 3,* 34-40.

Blank, R. M. (2010) The New American Model of Work-Conditioned Public Support, in:

J. Alber and N. Gilbert (eds.) United in Diversity: Comparing Social Models in

Europe and America (New York: Oxford University Press), pp. 176–198.

Bloom, H., Michalopoulos, C., & Hill, C. (2005). Using experiments to assess

nonexperimental comparison-group methods for measuring program effects. In H.

Bloom (Ed.), *Learning more from social experiments: Evolving analytic*

*approaches* (173-235). New York: Russell Sage Foundation.

Blumer, H. (1958). Race prejudice as a sense of group position. *Pacific Sociological*

*Review, 1*, 3–7.

Bobo, L. (1999). Prejudice as group position: Microfoundations of a sociological

approach to racism and race relations. *Journal of Social Issues, 55*(3), 445-472.

Bronfenbrenner, U. (1979). The ecology of human development: Experiments by nature

and design. London: Harvard University Press.

Burt, M. (1991). Causes of the growth of homelessness during the 1980s. *Housing Policy*

*Debate, 2*, 903–936.

Burt, M. (2001). Homeless families, singles, and others: Findings form the 1996 National

survey of homeless assistance providers and clients. *Housing Policy*

*Debate.12*(4), 737-780.

Burt, M., Pearson, C., & Montgomery, A. (2007). Community-wide strategies for

preventing homelessness: Recent evidence. *Journal of Primary Prevention, 28*,

213-228.

Byrne, T., Munley, E., Fargo, J., Montgomery, A., & Culhane, D. (2012). New

    perspectives on community-level determinates of homelessness. *Journal of Urban*

    *Affairs,* DOI: 10.1111/j.1467-9906.2012.00643.x

Cappelleri, J., Darlington, R., & Trochim, W. (1994). Power analysis of cutoff-based

    randomized clinical trials. *Evaluation Review, 18,* 141-152.

Caton, CL., Dominguez, B., Schanzer, B., Hasin, DS., Shrout, PE., & Felix, A. (2005)

    Risk factors for long-term homelessness: Findings from a longitudinal study of

    first-time homeless single adults. *American Journal of Public Health. 95*, 1753-

    1759.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots.

    *Journal of the American Statistical Association, 74*, 829-836.

Coalition for the Homeless (2013). Basic facts about homelessness: New York City Data

    and Charts. Retrieved from: http://www.coalitionforthehomeless.org/pages/basic-

    facts-about-homelessness-new-york-city-data-and-charts

Coie, J., Watt, N., West, S., Hawkins, D., Asarnow, J., Markman, H., Ramey, S., Shure,

    M., & Long, B. (1993). The science of prevention. *American Psychologist, 48*,

    1013–1022.

Coldwell, C., & Bender, W. (2007). The effectiveness of assertive community treatment

    for homeless populations with severe mental illness: A meta-analysis. *American*

    *Journal of Psychiatry, 164*(3), 393-399.

Cook, T. (2008). "Waiting for life to arrive": A history of the regression-discontinuity

    design in psychology, statistics, and economics. *Journal of Econometrics, 142,* 2,

    636-654.

Cook, T., & Campbell, D. (1979). *Quasi-Experimentation: Design & analysis issues for field settings.* Boston: Houghton Mifflin Company.

Cornish, F., & Campbell, C. (2009). The social conditions for successful peer education: A comparison of two HIV prevention programs run by sex workers in India and South Africa. *American Journal of Community Psychology, 44,* 123-135.

Cox, D.R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society, Series B, 34*, 187-202.

Cragg, M., & O'Flaherty, B. (1999). Do homeless shelter conditions determine shelter population? The case of the Dinkins deluge. *Journal of Urban Economics, 46*(3), 377-415.

Crane, M., & Warnes, M. (2000). Evictions and prolonged homelessness. *Housing Studies, 15*(5), 757-773.

Culhane, D., Metraux, S., & Byrne, T. (2011). A prevention-centered approach to homelessness assistance: A paradigm shift? *Housing Policy Debate, 21*(2), 295-315.

Culhane D., Metraux S., Byrne T., Steno M., & Bainbridge J. (2013). The age structure of contemporary homelessness: Evidence and implications for public policy. *Analyses of Social Issues and Public Policy. 13*, 1-17. National Center on Homelessness among Veterans.

Culhane, D., Metraux, S., & Hadley, T. (2002). Public service reductions associated with placement of homeless persons with severe mental illness in supportive housing. *Housing Policy Debate, 13*(1), 107-163.

Culhane, D. Park, J., & Metraux, S. (2011a). The Patterns and Costs of Services Use among Homeless Families. *Journal of Community Psychology.*

Czischke, D., & Gruis, V. (2007). Managing social rental housing in the EU in a changing policy environment: Towards a comparative study. *Comparative Housing Policy,* Workshop in Dublin, 1-33.

Dana, J., & Dawes, RM. (2004). The superiority of simple alternatives to regression for social science predictions. *Journal of Educational and Behavioral Statistics, 29*(3), 317-331.

Dasinger L., & Speiglman, R. (2007). Homelessness prevention: the effect of a shallow rent subsidy program on housing outcomes among people with HIV or AIDS. *AIDS Behavior, 11,* 128-139.

Dawes, RM. (1979) The robust beauty of improper linear models in decision making. *Am. Psychol, 34*(7), 571-582.

Dawes, RM., & Corrigan, B. (1974). Linear models in decision making. *Psychol. Bull. 81*(2), 95-106.

Dawes, RM., Faust, D., Meehl, PE. (1989). Clinical versus actuarial judgment. *Science. 243*(4899), 1668-1674.

Dobson, R., & McNeill, J. (2011). Review article: Homelessness and housing support services: Rationales and policies under New Labour. *Social Policy & Society, 10*(4), 581-589.

Drake, R., McHugo, G., Xie, H., Fox, M,. Packard, J., & Helmstetter, B. (2006). Ten-year recovery outcomes for clients with co-occuring schizophrenia and substance use disorders. *Schizophrenia Bulletin, 32*(3), 464-473.

Drake, R., Yovetich, N., Bebout, R., Harris, M., & McHugo, G. (1997). Integrated

    treatment for dually diagnosed adults. *Journal of Nervous and Mental Disease,*

    *185,* 298-305.

Early, D. (1998). The role of subsidized housing in reducing homelessness: An empirical

    investigation using micro-data. *Journal of Policy Analysis and Management, 17,*

    687-696.

Early, D. W., & Olsen, E. O. (2002). Subsidized housing, emergency shelters, and

    homelessness: An empirical investigation using data from the 1990 Census.

    *Advances in Economic Analysis and Policy, 2*(1), 2–34.

Early, D. (2004). The determinants of homelessness and the targeting of housing

    assistance. *Journal of Urban Economics 55*, 195–214.

Edgar, B. (2006). *National strategy to prevent and tackle homelessness: The pathway to a*

    *permanent home*. European Commission: Social Inclusion.

Fallis, G. (2010). Progressive housing policy in the 21st century: A contrarian view.

    *Journal of Sociology & Social Welfare, 37*(4), 173-199.

FEANTSA. (2005). ETHOS – European typology on homelessness and housing exlusion.

    Retrieved from http://www.feantsa.org/code/EN/pg.asp?Page=484

Goldberger, A. (1972). *Selection bias in evaluating treatment effects: Some formal*

    *illustrations.* Working Paper, Economics Department, University of Wisconsin.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual*

    *Review of Psychology, 60*, 549-576.

Graham, J. W., Cumsille, P. E., & Elek-Fisk, E. (2003). Methods for handling missing data. In J. A. Schinka & W. F. Velicer (Eds.). Research Methods in Psychology (pp. 87-114). Volume 2 of Handbook of Psychology (I. B. Weiner, Editor-in-Chief). New York: John Wiley & Sons.

Graham, J. W., & Hofer, S. M. (2000). Multiple imputation in multivariate research. In T. D. Little, K. U. Schnabel, & J. Baumert, (Eds.), Modeling longitudinal and multiple-group data: Practical issues, applied approaches, and specific examples. (201-218). Hillsdale, NJ: Erlbaum.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prev. Sci.8*, 206-213.

Greenberg, G., Hoblyn, J., Seibyl, C., & Rosenheck, R. (2006). Housing outcomes for hospitalized homeless veterans. *Journal of Health Care for the Poor and Underserved.17*, 425-440.

Grove, WM., Zald, DH., Lebow, BS., Snitz, BE., Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment. 12*(1), 19-30.

Gubits, D., Khadduri, J., & Turnham, J. (2009). *Housing patterns of low income families with children: Further analysis of data from the study of the effects of housing vouchers on welfare families.* Cambridge, MA: Joint Center for Housing Studies.

Gutierrez, O., & Friedman, D. (2005). Managing Project Expectations in Human Services Information Systems Implementations: The Case of Homeless Management Information Systems. *International Journal of Project Management, 23,* 513-523.

Hess, R. (2006). HomeBase homelessness prevention services. Retrieved from: http://www.innovations.harvard.edu/awards.html?id=52611

Honig, M. & Filer, R. (1993). Causes of intercity variation in homelessness. *American Economic Review, 83,* 248-255.

Hopper, K. (1991). Homelessness old and new: The matter of definition. *Housing Policy Debate, 2*, 757–813.

Hopper, K., Jost, J., Hay, T., Welber, S., & Haugland, G. (1997). Homelessness, severe mental illness, and the institutional circuit. *Psychiatric Services, 48*(5), 659-665.

Hopper, K., Shinn, M., Laska, E., Meisner, M., & Wanderling, J. (2008). Estimating numbers of unsheltered homeless people through plant-capture and postcount survey methods. *American Journal of Public Health, 98*(8), 1-5.

Hudson C., & Vissing Y. (2010). The geography of adult homelessness in the US: validation of state and county estimates. *Health & Place.16*, 828-837.

Imbens, G., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics, 142*, 615-635.

Imbens, G., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies, 79,* 933-959.

Israel, N., Toro, P., & Ouellette, N. (2010). Changes in the composition of the homeless population: 1992-2002. *American Journal of Community Psychology, 46*, 49-59.

Jencks, C. (1995). *The Homeless*. Cambridge, MA: Harvard University Press.

Khadduri, J. (2008). *Housing vouchers are critical for ending family homelessness*. Washington, DC: National Alliance to End Homelessness.

Kleinbaum, D.G., & Klein, M. (2005). *Survival Analysis: A Self-Learning Text.* New

     York: Springer.

Koegel, P., Burnam, M. A., & Baumohl, J. (1996). *The causes of homelessness.*

     *Homelessness in America, 24.*

Lee, B., Price-Spratlen, T., &Kanan, J. (2003). Determinants of homelessness in

     metropolitan areas. *Journal of Urban Affairs, 25,* 335–356.

Lee, B., Tyler, K., & Wright, J. (2010). The new homelessness revisited. *Annual Review*

     *of Sociology, 36,* 501-521.

Lee, D., & Lemieux, T. (2010). Regression discontinuity designs in economics, *Journal*

     *of Economic Literature, 48,* 281-355.

Lee, H., & Munk, T. (2008). *Using Regression Discontinuity Design for Program*

     *Evaluation,* Section on Survey Research Methods, Joint Statistical Meetings.

     Denver CO: American Statistical Association.

Levine, M., & Levine, A. (1970). *A social history of helping services: Clinic, court,*

     *school, and community.* New York: Appleton Century Crofts.

Link, B., & Phelan, J. (2001). Conceptualizing stigma. *Annual Review of Sociology, 27,*

     363-385.

Luke, D.A. (1993). Charting the Process of Change: A primer on survival analysis.

     *American Journal of Community Psychology, 21*(2), 203-246.

McChesney, K. (1990). Family homelessness: A systemic problem. *Journal of Social*

     *Issues, 46*(4), 191-205.

McFate, K., Lawson, R., & Wilson, W. (1995). Poverty, inequality and the future of

     social policy: Western states in the new world order. New York: Russell Sage

     Foundation.

McNeill, J. (2010). Employability pathways and perceptions of 'work' amongst single

     homeless and vulnerably housed people. *Social Policy & Society, 10*(4), 571-580.

Messeri, P., O'Flaherty, D., & Goodman, S. (2011). *Can homelessness be prevented?*

     *Evidence from New York City's HomeBase program.* Retrieved from:

     http://jagiellonia.econ.columbia.edu/~bo2/research/homebase.pdf

Mrazek, P., & Haggerty, R. (1994). *Reducing risks for mental disorders: Frontiers for*

     *preventive intervention research.* Institute of Medicine. National Academy Press,

     Washington DC.

NAEH. (2012). Changes in the HUD definition of "homeless."

National Institute on Drug Abuse (NIDA) (2012). *Drug facts: Nationwide trends.* US

     Department of Health and Human Services. NIH, Washington DC. Retrieved

     from: http://www.drugabuse.gov/sites/default/files/nationwide_0.pdf

National Surveys on Drug Use and Health (NSDUH), 2011 (revised October 2013). The

     Substance Abuse and Mental Health Services Administration (SAMHSA).

     Retrieved on 01-15-2014 from: http://www.samhsa.gov/data/spotlight/spot111-

     adults-mental-illness-substance-use-disorder.pdf .

Nelson, G. B., & Prilleltensky, I. (2005). Community psychology: In pursuit of liberation

     and well-being. Basingstoke: Palgrave Macmillan.

Nichols, A. (2011). Causal inference for binary regression. In Stata conference Chicago,

     (version June 14, 2011).

Nolan, C., ten Broeke, C., Magee, M., & Burt, M. (2005). *The family permanent supportive housing initiative: Family history and experiences in supportive housing.* Washington D.C.: Urban Institute.

Nooe, R., & Patterson, D. (2010). The ecology of homelessness. *Journal of Human Behavior in the Social Environment, 20,* 105-152.

O'Connell, M., Kasprow, W., & Rosenheck, R. (2008). Rates and risk factors for homelessness after successful housing in a sample of formerly homeless veterans. *Psychiatric Services, 59*(3), 268-275.

O'Flaherty B. (2004). Wrong person or wrong place: For homelessness, the conjunction is what matters. *Journal of Housing Economics, 13*, 1-15.

Office of the Deputy Prime Minister (2013). Live tables on homelessness. https://www.gov.uk/government/statistical-data-sets/live-tables-on-homelessness

Padgett D., Stanhope V., Henwood B., Stefancic A. (2011). Substance use outcomes among homeless clients with serious mental illness: Comparing Housing First with Treatment First programs. *Community Mental Health Journal, 2,* 227-232.

Paquette, K. (2011). Current statistics on the prevalence and characteristics of people experiencing homelessness in the United States. Substance Abuse and Mental Health Services Administration. Retrieved on 01-15-2014 from: http://homeless.samhsa.gov/ResourceFiles/hrc_factsheet.pdf .

Pearson, C., Montgomery, A., & Locke, G. (2009). Housing stability among homeless individuals with serious mental illness participating in housing first programs. *Journal of Community Psychology, 37*(3) 404-417.

Perlman, J., & Parvensky, J. (2006). *Denver housing first collaborative: Cost benefit analysis and program outcomes report*. Denver, CO: Colorado Coalition for the Homeless.

Quigley, J., Raphael, S., & Smolensky, E. (2002). Homeless in America, Homeless in California. *Working Papers, Berkeley Program on Housing and Urban Policy, Institute of Business and Economic Research*, UC Berkeley.

Rafferty, A., & Wiggan, J. (2011). Choice and welfare reform: Lone parents' decision making around paid work and family life. *Journal of Social Policy, 40*(2), 275-293.

Raphael, S. (2010). Housing market regulation and homelessness. In I. G. Ellen, & B. O'Flaherty (Eds.), *How to house the homeless* (pp. 110–140). New York: Russell Sage Foundation.

Rolston, H., Geyer, J., & Locke, G. (2013). *Evaluation of the HomeBase Community Prevention Program.* Bethesda, MD: Abt Associates Inc.

Rosenbaum, P., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*(1), 41–55.

Rossi, P. H. (1994). Troubling families: family homelessness in America. *American Behavioral Scientist*, *37*(3), 342+.

Rossi, P. H., Lipsey, M., & Freeman, H. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: Sage Publications.

Rubin, D.B. (1987) *Multiple Imputation for Nonresponse in Surveys*. J. Wiley & Sons, New York.

Sadowski, L., Kee, R., Van der Weele, T., & Buchanan, D. (2009). Effect of a housing
and case management program on emergency department visits and
hospitalizations among chronically ill homeless adults: A randomized trial.
*JAMA, 301*(17), 1771-1778.

Sard, B. (2009). Number of homeless families climbing due to recession: Recovery
package should include new housing vouchers and other measures to prevent
homelessness. Washington DC: Center on Budget and Policy Priorities.

Schafer, J.L. (1999). NORM: Multiple imputation of incomplete multivariate data under
a normal model, version 2. Software for Windows 95/98/NT, available from
http://www.stat.psu.edu/~jls/misoftwa.html.

Schafer. J. L. & Graham, J. W. (2002). Missing data: Our view of the state of the art.
*Psychological Methods, 7*(2), 147-177.

Schochet, P. (2009). Statistical power for regression discontinuity designs in education
evaluations. *Journal of Educational and Behavioral Statistics,* 34(2), 238-266.

Schwartz, A. (2010). *Housing Policy in the United States.* New York: Russell Sage
Foundation.

Sclar, E. (1990). Homelessness and housing policy: A game of musical chairs. *American
Journal of Public Health, 80*(9), 1039-1040.

Scott, D. (1979). On optimal and data-based histograms. *Biometrika, 66*(3), 605-610.

Shier, M., Jones, M., & Graham, J. (2012). Employment difficulties experienced by
employed homeless people: Labor market factors that contribute to and maintain
homelessness. *Journal of Poverty, 16*, 27-47.

Shinn, M. (1992). Homelessness: What is a psychologist to do? *American Journal of Community Psychology, 20*(1), 1-24.

Shinn, M. (2007). International homelessness: Policy, socio-cultural, and individual perspectives. *Journal of Social Issues, 63*, 659–679.

Shinn, M. (2010). Homelessness, poverty and social exclusion in the United States and Europe. *European Journal of Homelessness, 4,* 19-44.

Shinn, M., Baumohl, J., & Hopper, K. (2001). The prevention of homelessness revisited, *Analyses of Social Issues and Public Policy,* 95-127.

Shinn, M., & Greer, A.L. (2011). The European consensus conference on homelessness-Kudos and some cautions, to Europe. *European Journal of Homelessness. 5*, 183-190.

Shinn, M., Greer, A. L., Bainbridge, J., Kwon, J., & Zuiderveen, S. (2013). Efficient targeting of homelessness prevention services for families. *American Journal of Public Health, 103*(S2), *S324- S330.*

Shinn, M., Weitzman, BC. (1994). You can't eliminate homelessness without housing. *American Behavioral Scientist. 37*(3), 435-442.

Shinn, M., Weitzman, B., Stojanovic, D., Knickman, J., Jimenez, L., & Duchon, L. (1998). Predictors of homelessness among families in New York City: from shelter request to housing stability. *Am J Public Health. 87,* 256-262.

Sinharay, S., Stern, HS., & Russell, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods.6*(4), 317-329.

Smeeding, T.M. (2005). Government programs and social outcomes: The United States in comparative perspective. In A. Auerbach, D. Card, & J. Quigley (Eds.), *Poverty, Public Policy, and the Distribution of Income.* New York, NY: Russell Sage Foundation.

Spellman, B., Khadduri, J., Sokol, B., & Leopold, J. (2010). Costs associated with first-time homelessness for families and individuals. Washington D.C.

Stretch, J., & Kreuger, L. (1993). A social-epidemiological five year cohort study of homeless families: A public/private venture policy analysis using applied computer technology. *Computers in Human Services, 9*(3-4), 209-230.

Swets, J. (1996). *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers.* Hillsdale, NJ: Lawrence Erlbaum.

The United States Conference of Mayors. (2008). Hunger and homelessness survey: A status report on hunger and homelessness in America's cities: A 25-city survey. Washington, D. C.

Toro, P. A. (2007). Toward an international understanding of homelessness. *Journal of Social Issues, 63*, 461–481.

Toro, P., Trickett, E., Wall, D., & Salem, D. (1991). Homelessness in the United States: An ecological perspective. *American Psychologist, 46*(11), 1208-1218.

Tsai, J., Kasprow, W., & Rosenheck, R. (2011). Exiting homelessness without a voucher: A comparison of independently housed and other homeless veterans. *Psychological Services, 8*(2), 114-122.

Tsemberis, S., Gulcur, L., & Nakae, M. (2004). Housing First, consumer choice, and

harm reduction for homeless individuals with a dual diagnosis. *American Journal
of Public Health, 94*(4), 651-656.

UK Census (2013). Ethnicity and National Identity in England and Wales 2011.

http://www.ons.gov.uk/ons/rel/census/2011-census/key-statistics-for-local-
authorities-in-england-and-wales/rpt-ethnicity.html

U.S. Census (2011). American Community Survey Data. Retrieved on 01-15-2014 from:

http://factfinder2.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t .

U.S. Conference of Mayors. (2011). *Status Report on Hunger and Homelessness.*

Retrieved from: http://usmayors.org/pressreleases/uploads/2011-hhreport.pdf .

U.S. Department of Housing and Urban Development. (2011). *HPRP Eligibility
Determination and Documentation.* Retrieved from:

https://www.onecpd.info/resources/documents/hprp_eligibilityanddocumentationg
uidance.pdf.

U.S. Dept. of Housing and Urban Development. *The 2011 Annual Housing Assessment
Report to Congress.* Washington, DC. 2012a.

U.S. Dept. of Housing and Urban Development. *Volume I of the 2012 Annual Homeless
Assessment Report.* Washington, DC. 2012b.

U.S. Dept. of Housing and Urban Development. *Sheltered Homeless Persons in New
York City (NY) 10/1/2010-09/30/2011.* Washington, DC. 2012c.

US Dept. of Housing and Urban Development. *The 2013 Annual Homeless Assessment
Report (AHAR) to Congress: Part 1 Point-in-Time Estimates of Homelessness.*
Washington, DC. 2013a.

US Dept. of Housing and Urban Development. *The 2012 Annual Homeless Assessment Report (AHAR) to Congress: 2012 AHAR HMIS Estimates of Homelessness.* Washington, DC. 2013b.

Wayman, J. C. (2003). Multiple Imputation for Missing Data: What is it and how can I use it? Paper presented at the 2003 Annual Meeting of the American Educational Research Association, Chicago, IL.

Willse, C. (2010). Neo-liberal biopolitics and the invention of chronic homelessness. *Economy and Society 39*(2), 155-184.

Witte, P. (2012). *The State of Homelessness in America in America 2012: A research report on homelessness.* Washington, DC: National Alliance to End Homelessness.

Wong, Y., Culhane, D., & Kuhn, R. (1997). Predictors of exit and reentry among family shelter users in New York City. *Social Services Review, 71*(3), 441-462.

Wood, M., Turnham, J., & Mills, G. (2008). Housing affordability and family well-being: Results from the housing voucher evaluation. *Housing Policy Debate, 19*(2), 367-412.

Zlotnick, C., Robertson, M., & Lahiff, M. (1999). Getting off the streets: Economic resources and residential exits from homelessness. *Journal of Community Psychology, 27,* 209-224.

# APPENDIX CHAPTER 2

## NOTES ON MULTIPLE IMPUTATION

The following questions and responses are intended to provide information about the rationale and procedure for carrying out Multiple Imputation (MI). A detailed description of MI is beyond the scope of this appendix, but a thorough foundation for MI can be found in multiple sources (e.g., Graham et al., 2003; Rubin, 1987).

**What is Multiple Imputation?**

Multiple imputation is a technique that predicts missing values for a variable by using information from non-missing values across other variables (Wayman, 2003). The MI technique can estimate a value for each missing cell in a data set and replicates this method for a specified number of data sets. Multiple data sets are needed because final estimates should include variability to account for the uncertainty of the missing value. Thus, with MI, each missing cell in a data set will have multiple estimates across a number of imputed data sets.

Rubin's (1987) rules for combining datasets suggest that the intended analysis should be carried out on each data set, and the results should be combined across data sets. First, statistical programs save the estimates and standard errors for each dataset across analyses. Then, the estimates are averaged across datasets. However, standard errors must be corrected for *within imputation* variance as well as *between imputation* variance (for formulae, see Rubin, 1987). The resulting estimates and standard errors make use of all available information, rather than eliminating cases that may contain only one (or more) missing value across variables. As with most quantitative tools, there are multiple assumptions about when to properly use MI, and while the intention of this

156

appendix is not to outline all assumptions, MI has been shown to be remarkably tolerant to violations of assumptions and even across forms of missingness that might not be random (Graham, 2009; Wayman, 2003).

**Why is multiple imputation needed?**

The MI procedure has been shown to reduce bias that arises from other ways of dealing with missing data (e.g., list-wise deletion, mean matching, single imputation), especially with substantial proportions of missing data (Graham & Hofer, 2000; Schafer & Graham, 2002). Naïve techniques that eliminate missing data, such as list-wise deletion, assume that the missing cases are missing completely at random (MCAR). However, missingness can also arise from data that are missing at random (MAR) (i.e., known variables account for the missingness) and missing not at random (MNAR) (i.e., unknown variables account for the missingness). Thus, a blanket elimination of cases that are anything other than MCAR can lead to biased results when an underlying reason for missingness is ignored. Even when MCAR cases are lost due to list-wise deletion, elimination will lead to a loss of power, although results will likely be unbiased (Wayman, 2003).

**How many data sets should be imputed?**

The number of data sets that one should impute remains an issue of debate, but much of the literature suggests that 1) more data sets yield less biased results, and 2) high rates of missingness call for many imputations (Graham et al., 2007). While some statisticians suggest that very few imputations are enough to obtain excellent results (e.g., Schaffer, 1999), the current study imputes 50 data sets for the analyses in Chapters 2 and 3.

# APPENDIX CHAPTER 3

## DESCRIPTIVE STATISTICS BY FAMILY STATUS

| Variables | Individuals (n = 1261) Percent or Mean | Families (n = 1500) Percent or Mean |
|---|---|---|
| **Subsequent Outcome Variables** | | |
| *Literal Homelessness* | 2.54 | 1.80 |
| *Homeless or Imminently Losing Housing* | 8.96 | 8.93 |
| *Any Housing Instability* | 13.56 | 14.73 |
| **Baseline Covariates** | | |
| ***Demographics*** | | |
| *Female* | 58.21 | 85.87*** |
| *African American* | 70.90 | 65.93** |
| *Hispanic* | 10.79 | 19.40*** |
| *Age* | 42.97 | 36.15*** |
| *Veteran* | 10.31 | 2.53*** |
| *Married* | 11.42 | 17.13*** |
| *Pregnant* | N/A | 4.47 |
| ***Human Capital*** | | |
| *Employed Part-Time* | 16.97 | 18.73 |
| *Unemployed* | 57.34 | 49.53*** |
| *Receiving Non-Cash Benefits* | 59.64 | 73.47*** |
| *<30% AMI* | 85.01 | 81.07** |
| *High school Diploma* | 82.24 | 75.60*** |
| ***Housing Conditions*** | | |
| *Leaseholder History* | 72.72 | 84.47*** |
| *Doubled up* | 17.05 | 17.80 |
| *Eviction History* | 11.10 | 14.93** |
| *Being Evicted from Public Housing* | 5.95 | 8.73** |
| *Rent > 40% Income* | 85.41 | 80.40** |
| *Outstanding Debt > $1000* | 35.92 | 42.87*** |
| ***Disability/Criminal Justice History*** | | |
| *Disabled* | 46.15 | 31.93*** |
| *Unhealthy Compared to Peers* | 30.69 | 22.07*** |
| *Criminal Justice History* | 19.03 | 14.07*** |
| ***Interpersonal Discord*** | | |
| *Domestic Violence History* | 25.30 | 35.47*** |
| *Escaping Domestic Violence Currently* | 3.57 | 8.93*** |
| ***Shelter History*** | | |
| *Homeless >once last year* | 26.01 | 15.47*** |
| *Previously Homeless (Self Report)* | 36.16 | 27.07*** |
| *Previously Homeless (HMIS Record)* | 21.33 | 9.67*** |
| *Previous Night Homeless (Self Report)* | 27.52 | 15.53*** |

*** $p < .001$, ** $p < .01$, * $p < .05$ for chi-squared analyses or two-tailed t-tests across groups

# APPENDIX CHAPTER 4

## COVARIATE DISCONTINUITIES

| Variables | Difference in rates of crossing cutoff score | P-Value |
|---|---|---|
| **Baseline Covariates** | | |
| *Employment and Income Potential* | | |
| Unemployed | .01 | .98 |
| High school Diploma | .33 | .48 |
| Receiving Non-Cash Benefits | .51 | .30 |
| *Financial Status* | | |
| <30% AMI | .39 | .43 |
| Rent > 40% Income | .04 | .93 |
| Outstanding Debt > $1000 | -.20 | .68 |
| *Housing and Legal* | | |
| Eviction History | -.23 | .44 |
| Leaseholder History | -.05 | .84 |
| Previously Homeless (Self Report) | -.22 | .57 |
| Previous Night Homeless (Self Report) | -.36 | .32 |
| Criminal Justice History | .13 | .54 |
| *Additional Covariates* | | |
| Female | .25 | .56 |
| African American | -.28 | .49 |
| Hispanic | .36 | .31 |
| *Age (mean)* | *-6.68* | *.53* |
| Veteran | -.14 | .53 |
| Married | .37 | .20 |
| Pregnant | -.12 | .48 |
| Escaping Domestic Violence Currently | .04 | .56 |
| Domestic Violence History | -.15 | .75 |
| Living Doubled up | .16 | .64 |
| Disabled | -.03 | .94 |
| Unhealthy Compared to Peers | .49 | .17 |

# APPENDIX CHAPTER 4

# POWER ANALYSIS

One difficulty with the RD design is the much larger sample sizes required compared to the randomized controlled trial (RCT) to have enough statistical power to show an effect (Schochet, 2009). In fact, if the score variable were normally distributed with the cut-point at the center of the distribution, the sample for a RD design would have to be 2.75 times larger than for a RCT to have the same precision (Goldberger, 1972). Because of this challenge with the RD design, some researchers have focused on expanding the range around cut-off values to increase power (Cappelleri, Darlington, & Trochim, 1994).

Lee and Munk (2008) offer the formula for determining the sample size of a parametric RD based on the desired minimum detectable (standardized) effect size (MDES), proportion of participants assigned to treatment, $R^2$ of the model, significance level, and power.

$$n = \frac{(1 - R_M^2)(z_{1-\alpha} - z_\beta)^2}{\{M^2 P(1 - P)(1 - R_T^2)\}}$$

"where M is the MDES, $z_\alpha$ and $z_{1-\beta}$ are normal $100\alpha$ –th and $100(1-\beta)$-th percentiles, respectively." (p. 1679) $R_M^2$ is the r-squared statistic for the model. $R_T^2$ is the squared correlation between treatment and score variable (.09 in the current study; the correlation was .3 for 10 points above and below each cut point). P is the proportion of participants assigned to treatment (83% in the current study). A power analysis for the non-parametric investigation was not found in the literature. However, one study addresses the bandwidth size (or area of inclusion) in detail (Imbens & Kalyanaraman

2012). Because different sizes of bandwidths include more or fewer cases, a power

analysis should include considerations of both sample sizes and power.

Two meta-analyses offer anticipated effect sizes. The first meta-analysis found an

average treatment effect of .37 (95% CI .18-.55) for community interventions across six

randomized controlled trials with various homelessness outcomes (Coldwell & Bender,

2007). The second study, a prospective meta-analysis of eight sites, found an average

effect size of .22 (95% CI .09-.34) for an outcome of housing stability (Banks, McHugo,

Williams, Drake, & Shinn, 2002). To align with the effects found in previous analyses, I

provide a range of power analyses for effect sizes from .09-.55.

**Table 1**. Sample Sizes for .8 Power

| R-squared:<br>MDES | .01 | .1 | .2 | .3 |
|---|---|---|---|---|
| .09 | 5855 | 5323 | 4731 | 4140 |
| .2 | 1186 | 1078 | 958 | 838 |
| .3 | 527 | 479 | 426 | 373 |
| .4 | 297 | 269 | 240 | 210 |
| .5 | 190 | 172 | 154 | 134 |
| .55 | 157 | 143 | 127 | 111 |

Note: Sample sizes were always rounded up

A related formula for determining power is given as:

$$1 - \beta = 1 - P\left\{Z < \left(z_{1-\alpha} - M\sqrt{\frac{nP(1-P)(1-R_T^2)}{1-R_M^2}}\right)\right\}.$$

If I assume a sample size of 500 around each cut point, table 1 shows the power for

variations of effect sizes and r-squared values.

**Table 2**. Power for Various R-squared Values and Effect Sizes (n = 500)

| R-squared: | .01 | .1 | .2 | .3 |
|---|---|---|---|---|
| MDES | | | | |
| .09 | .18 | .19 | .20 | .22 |
| .2 | .49 | .52 | .56 | .61 |
| .3 | .78 | .81 | .85 | .89 |
| .4 | .94 | .96 | .97 | .99 |
| .5 | .99 | 1 | 1 | 1 |
| .55 | 1 | 1 | 1 | 1 |

Note: Power is rounded to the nearest hundredth

In the absence of a power analysis for the non-parametric RD design[28], this power

analysis for a parametric RD design provides the best approximation of power for various

sample sizes, effect sizes, and r-squared values. To have power of .8 or higher, table 1

shows the sample sizes necessary for differing effect sizes and r-squared values. For a

modest to large effect size (>.2) and with a desired power of .8, the current sample size (n

= 2,726) falls within the range required to detect an effect. However, for the smallest

effect sizes found in homelessness literature (.09), the current sample would be too small

to have sufficient power to detect an effect. Additionally, more and more of the sample

will be excluded as the bandwidths of the non-parametric RD design grow narrower.

Table 2 shows that if the sample size were held at 500 (for a somewhat large bandwidth

around a cut point), the effect would have to be .3 or greater to be detected with a

---

[28] This precision estimate is for the full parametric models only. For the non-parametric
models, current research does not offer a reliable way to estimate precision (Jacob et al.,
2012). The problem with estimating precision for non-parametric models is that
observations around the cutoff are weighted so heavily that distant observations are often
ignored. Thus non-parametric models are inherently less precise than parametric models.

minimum of .8 power. Further, if the effect were .3, the r-squared value would have to be

larger than .01. Overall, the sample size seems sufficient to detect large effects.