

Cityscape

*A Journal of Policy
Development and Research*

HOUSING DISCRIMINATION TODAY
VOLUME 17, NUMBER 3 • 2015



PD&R



Managing Editor: Mark D. Shroder
Associate Editor: Michelle P. Matuga

Advisory Board

Dolores Acevedo-Garcia
Brandeis University

Ira Goldstein
The Reinvestment Fund

Richard K. Green
University of Southern California

Mark Joseph
Case Western Reserve University

Matthew E. Kahn
University of California, Los Angeles

C. Theodore Koebel
Virginia Tech

Jens Ludwig
University of Chicago

Mary Pattillo
Northwestern University

Carolina Reid
University of California

Patrick Sharkey
New York University



PD&R

Cityscape

*A Journal of Policy
Development and Research*

HOUSING DISCRIMINATION TODAY
VOLUME 17, NUMBER 3 • 2015

U.S. Department of Housing and Urban Development
Office of Policy Development and Research

The goal of *Cityscape* is to bring high-quality original research on housing and community development issues to scholars, government officials, and practitioners. *Cityscape* is open to all relevant disciplines, including architecture, consumer research, demography, economics, engineering, ethnography, finance, geography, law, planning, political science, public policy, regional science, sociology, statistics, and urban studies.

Cityscape is published three times a year by the Office of Policy Development and Research (PD&R) of the U.S. Department of Housing and Urban Development (HUD). Subscriptions are available at no charge and single copies at a nominal fee. The journal is also available on line at huduser.gov/periodicals/cityscape.html.

PD&R welcomes submissions to the Refereed Papers section of the journal. Our referee process is double blind and timely, and our referees are highly qualified. The managing editor will also respond to authors who submit outlines of proposed papers regarding the suitability of those proposals for inclusion in *Cityscape*. Send manuscripts or outlines to cityscape@hud.gov.

Opinions expressed in the articles are those of the authors and do not necessarily reflect the views and policies of HUD or the U.S. government.

Visit PD&R's website, huduser.gov, to find this report and others sponsored by PD&R. Other services of HUD USER, PD&R's Research and Information Service, include listservs, special interest and bimonthly publications (best practices, significant studies from other sources), access to public use databases, and a hotline (1-800-245-2691) for help with accessing the information you need.

Contents

Symposium

Housing Discrimination Today 1

Guest Editors: Margery Austin Turner and Judson James

Guest Editors' Introduction

Discrimination as an Object of Measurement 3

What Have We Learned From Paired Testing in Housing Markets? 15

by Sun Jung Oh and John Yinger

Housing Discrimination Among Available Housing Units in 2012: Do Paired-Testing Studies Understate Housing Discrimination? 61

by Rob Pitingolo and Stephen L. Ross

Changing Contexts and New Directions for the Use of Testing 87

by Fred Freiberg and Gregory D. Squires

Targeting Disability Discrimination: Findings and Reflections From the National Study on Housing Discrimination Against People Who Are Deaf and People Who Use Wheelchairs 103

by Claudia L. Aranda

Other Protected Classes: Extending Estimates of Housing Discrimination 123

by Margery Austin Turner

Commentary: Some Thoughts on Field Experiments on Housing Discrimination From a European View 137

by Ali M. Ahmed

Commentary: Housing Discrimination Research in the 21st Century 143

by Samantha Friedman

Commentary: Expanding the Fair Housing Testing Landscape 151

by James Perry

Commentary: Testing Benefits Housing Providers and the Industry 155

by Fred Underwood

Refereed Papers 159

What Happens to Housing Assistance Leavers? 161

by Robin E. Smith, Susan J. Popkin, Taz George, and Jennifer Comey

Do the GSEs Meet the Credit Needs of Underserved Communities of Color? 193

by Michela Zonta

Departments 219

Data Shop

Measuring Neighborhood Opportunity With AFFH Data 221
by Brent D. Mast

Graphic Detail

Civil Unrest and Marginalization in Baltimore 231
by John C. Huggins

Industrial Revolution

Rural America: Perceptions of Residential Energy Retrofits 233
by Nathan Barry

Foreign Exchange

Preparing Our Housing for the Transition to a Post-Baby Boom World: Reflections on
Japan's May 26, 2015 Vacant Housing Law 239
by Peter Manda

SpAM

Predicting Local Crime Clusters Using (Multinomial) Logistic Regression 249
by Martin A. Andresen

Evaluation Tradecraft

Fair Housing Testing: Selecting, Training, and Managing an Effective Tester Pool 263
by Claudia L. Aranda and Sarale H. Sewell

Symposium

Housing Discrimination Today

Guest Editors: Margery Austin Turner and Judson James

Guest Editors' Introduction

Discrimination as an Object of Measurement

Margery Austin Turner
Urban Institute

Judson James
U.S. Department of Housing and Urban Development

The views expressed in this introduction are those of the guest editors and do not represent the official positions or policies of the Urban Institute, the Office of Policy Development and Research, the U.S. Department of Housing and Urban Development, or the U.S. government.

Introduction

For nearly four decades, the Office of Policy Development and Research (PD&R) in the U.S. Department of Housing and Urban Development (HUD) has sponsored large-scale, paired-testing studies to rigorously measure the incidence and forms of housing discrimination in metropolitan areas nationwide. As we approach the 50th anniversary of passage of federal fair housing protections, this issue of *Cityscape* offers a comprehensive review of HUD's paired-testing research, which first focused exclusively on discrimination against African-American homeseekers, and, most recently, expanded to measure discrimination based on disability, sexual preference and gender identity, family composition, and housing voucher reciprocity.

Origins of Paired Testing as a Research Tool

Paired testing as a research tool developed out of its initial use in helping to illustrate the practice of housing discrimination against racial minorities. As Sun Jung Oh and John Yinger document in more detail in appendix B of their article (Oh and Yinger, 2015) in this symposium, the initial step was conducting community audits during the 1950s in which testers were paired in their housing inquiries and the results compared. The testing was referred to as “audits”—as measures of community practice.

Some of the most basic techniques of paired testing evolved early in these audits. They included matched inquiries for the same type of unit addressed to the same housing provider. Tester selection and protocols were designed to reduce any differences in the income and housing

requirements presented to the housing provider, making the only readily apparent difference between the two testers the characteristic being tested. The treatment experienced by each tester was separately and systematically recorded immediately afterward.

The use of paired testing expanded in the 1960s as a tool for exposing discrimination in the housing market and for enforcing new federal fair housing protections of the Fair Housing Act of 1968 and related state laws. For purposes of enforcement of housing discrimination laws, paired testing had the important advantage of providing clear evidence in a form that best suits the manner that trial courts prefer to proceed to decisions; that is, a specific set of facts about well-defined interactions between specific individuals. General patterns, while providing credibility, are less important in litigation than the facts of a specific case.

The case law developed around this early enforcement activity helped establish the legal standing of testers and facilitated the launching of large-scale studies for research purposes without the hazard of legal challenges. The key legal issue is the claim by housing providers that testing constitutes *entrapment* (that is, an active solicitation of illegal behavior) and is a burden on the normal business activities of the housing provider. The response to these claims is that neither burden nor entrapment occurs because—

- The housing rental and sales agents involved are carrying out a normal business activity in responding to one of many potential clients.
- The Supreme Court has ruled that such testing does not impose an inappropriate burden on the activities of rental and sales agents.¹
- Congress has conferred on all “persons” a legal right to truthful information about available housing, 42 U.S.C. § 3604(d), a right made enforceable through the creation of an explicit cause of action in § 812(a) of the Act, 42 U.S.C. § 3612(a).
- The tester passively responds to and records the information received about a potential housing opportunity advertised by the sales or rental agent. He or she does not ask any questions except those appropriate to an initial exploratory inquiry.

Testing for research purposes and testing for enforcement purposes have common roots and complementary objectives, even though they have distinctive objectives and differences in methodology. The research audit can demonstrate to a broader community the scale of the discriminatory behaviors and their consequences, which serves to build support for enforcement activity and for parallel education and outreach on discriminatory practices. Together, research audits and enforcement testing built up the body of practical knowledge of how to conduct tests, including the production of “how-to” manuals. The research on the extent and forms of housing discrimination can inform the targeting of enforcement efforts.

Research (audit) testing and enforcement testing differ, however, in three important respects.

(1) Whereas research testing is concerned with representative samples, consistent protocols, and closed-ended reporting to maximize the scientific value and generalizability of its findings, enforcement testing is focused on documenting a specific interaction in a manner that is likely to produce

¹ *Havens Realty Corp. v. Coleman*, 455 U.S. 363, 373 (1982).

a basis for successful litigation. (2) Research testing has to focus on objectivity and consistent procedures to be credible to a broader public as evidence about the general patterns of housing discrimination, whereas enforcement testing procedures can be adaptable in building the strongest legal case in a specific set of circumstances. (3) To maintain its claim to objectivity, the results of research or audit testing are not directly available for enforcement purposes and individual testers do not have access to the results of their own tests and cannot pursue enforcement actions on the basis of that specific experience, a significant conflict with the purposes of enforcement testing. Research findings can indirectly prompt pattern and practice investigations, however, by both governmental and private enforcement agencies.²

The Office of Policy Development and Research (PD&R) of the U.S. Department of Housing and Urban Development took the lead in expanding the role of paired testing in housing discrimination research with the initiation in 1977 of the Housing Market Practices Survey (HMPS), the first national audit of racial discrimination in housing sales and rentals. The lead role of PD&R was based on the authority given by Section 808(e) of the Civil Rights Act of 1968.

The Secretary of Housing and Urban Development shall (1) make studies with respect to the nature and extent of discriminatory practices in representative communities, urban, suburban, and rural, throughout the United States....³

Under this authority, PD&R began what became a series of periodic national audits of housing discrimination practices against racial and ethnic minorities. Other paired-testing studies directed toward discriminatory housing practices affecting other protected classes under the Fair Housing Act of 1988 also followed, along with related housing discrimination research, including examination of the performance of private fair housing enforcement agencies funded by HUD grants.

The HMPS had two major objectives: (1) measuring the nature and extent of housing discrimination against African-American homeseekers in American metropolitan housing markets, and (2) determining what factors, including the enforcement of housing civil rights legislation, influence the observed discrimination against African-American homeseekers. Data collection for HMPS was conducted in the spring of 1977 and the final report was published in May 1979. The study executed a total of 3,264 paired tests in 40 metropolitan areas, by far the largest paired-testing operation to that date. An extensive selection-and-training process of testers and a pretest of testing protocols (tester instructions, and so on) preceded the testing. Of the 40 test sites, 5 were audited more intensively with 200 paired tests each, between two and three times the rate at the other sites, to examine some potentially relevant factors in more detail. Testers recorded a number of the specific responses by the sales or rental agent—including initial unit availability, number of units available, courtesy of treatment received, and information requested and volunteered. It is important to note that HMPS established a structural format and a set of procedures to be built on and refined in later paired-testing research studies.

² More specifically, in recent audit studies, the research organization has turned over all individual test results to HUD, on the understanding (written into the contractual agreement) that information about specific housing providers cannot be used as evidence in litigation but can be used to target further (enforcement) testing.

³ 42 U.S.C. § 3608.

The HMPS findings provided systematic evidence on the widespread housing discrimination against African-American individuals in metropolitan areas throughout the United States. In addition, HMPS demonstrated the value of the paired-testing methodology and the survey's substantive findings supported further developments: the Fair Housing Act of 1988; HUD funding of enforcement activity by both private and state enforcement agencies; and funding for further housing discrimination research by PD&R.

Evolution of Research Testing

Over the three and a half decades since the pioneering HMPS was completed, the paired-testing methodology has been continuously adapted and refined to measure different dimensions of housing discrimination and to respond to evolving housing market practices. This powerful research tool has also been extended to quantify the extent and forms of discrimination in other domains, with particular attention to discrimination in hiring.

National Estimates of Racial and Ethnic Discrimination in Housing Rentals and Sales

HUD has funded a national paired-testing study of discrimination against minority homebuyers and renters once each decade. The 1979 HMPS found high levels of discrimination in both rental and sales markets (Wienk et al., 1979). At that time, it was not uncommon for African-American homeseekers to be told that no homes or apartments were available to them or to be denied an opportunity to meet with a rental or sales agent. The 1989 Housing Discrimination Study (HDS1989) measured discrimination against Hispanic and African-American homeseekers, and it again found high levels of discriminatory treatment in both rental and sales markets nationwide. That study concluded that overall levels of discrimination against African-American homeseekers had not changed significantly since 1977, although its forms were changing to become more subtle and less easily detectable (Turner, Struyk, and Yinger, 1991).

Roughly a decade later, the HDS2000 again found statistically significant levels of discrimination against African-American, Hispanic, and Asian homeseekers (Turner and Ross, 2003a, 2003b; Turner et al., 2002). That study was explicitly designed to measure changes in discrimination and concluded that, between 1989 and 2000, the overall incidence of discrimination against African-American homeseekers declined in both rental and sales markets nationwide. The incidence of discrimination against Hispanic homebuyers also declined, but no significant change occurred for Hispanic renters (Turner et al., 2002).⁴ Finally, the most recent national study, HDS2012, found that African-American, Hispanic, and Asian homeseekers are generally just as likely as equally qualified White homeseekers to get an appointment and learn about at least one available housing unit but that minority homeseekers are told about and shown fewer homes and apartments than White homeseekers (Turner et al., 2012).

⁴ HDS2000 also tested for discrimination against Native American renters (not homebuyers) in a small selection of metropolitan areas. This exploratory effort documented very high levels of discrimination, including the types of "door slamming" discrimination that African-American homeseekers experienced in the 1977 study.

Each of these major studies built on its predecessors but updated or refined the basic paired-testing methodology in an effort to better respond to evolving market conditions. Of particular importance are: (1) changes in methods for constructing a representative sample of housing providers to test, (2) documenting differences in treatment during telephone inquiries and during in-person visits with housing providers, and (3) assessing the racial or ethnic identifiability of the testers and implications for differential treatment.

Identifying a Representative Sample of Housing Providers

The core objective of a paired-testing study is to observe the relative treatment that housing providers offer White and minority homeseekers in the private market. Because these studies measure provider behavior, one would ideally draw a representative sample of rental and sales providers in which an individual's (or firm's) probability of selection reflects his or her share of available housing units.

All four national, HUD-funded testing studies have approximated this objective by drawing random samples of advertisements for rental and sales housing to represent the universe of housing units—and housing providers—in the marketplace. The 1977 HMPs drew a one-time sample of advertisements from the major newspaper in each metropolitan area where testing was conducted. This sample was then used to identify housing providers that testers visited to inquire about available homes or apartments. This approach was modified in the 1989 HDS to draw a fresh sample of advertisements each week—again from major metropolitan newspapers. Testers then referred explicitly to the house or apartment in the sampled advertisement when they visited housing providers and made their inquiries about availability. This new approach ensured that both minority and White testers were conveying the same initial signals about the type of housing they were seeking and that the housing providers they visited had suitable units available.

HDS2000 also created weekly ad samples, but it drew on multiple advertising vehicles, identified by local fair housing organizations as important applied sources of information for homeseekers. In addition, because evidence suggested that online advertising was becoming increasingly common, the 2000 study began to experiment with the use of online sources. For example, the online versions of some newspapers' advertising sections were used in place of the print version. Finally, in response to concerns that some housing providers might try to exclude minority homeseekers by not advertising available units in publicly accessible venues, HDS2000 explored strategies for supplementing its ad samples with information drawn from community newspapers, fliers, and foreign language newspapers. This exploratory effort did not find evidence that patterns of discrimination varied across ad sources, but it did not completely address the important issue (discussed in depth by Fred Freiberg and Gregory D. Squires in this symposium) that discrimination may be more prevalent among housing providers that avoid advertising altogether (Freiberg and Squires, 2015).

By 2010, major metropolitan newspapers were no longer a primary source of advertisements for rental or sales housing. Instead, most housing providers and homeseekers appeared to rely on online sources, including Craigslist, Apartments.com, and Zillow. Major sources of online advertisements vary across metropolitan areas. Therefore, HDS2012 relied entirely on online ad sources, continuously drawing fresh ad samples from a rotating list of sources tailored to each metropolitan

area in the study sample. The 2012 study also implemented a two-stage approach to ensure that the geographic distribution of sampled advertisements corresponded to the geographic distribution of rental and homeowner housing within each metropolitan area (Turner et al., 2014).

Measuring Differential Treatment at the Telephone Inquiry Stage

The first three national paired-testing studies reported differences in treatment that occurred during in-person visits to housing providers. Testers often made phone calls in advance of these visits to secure an appointment, but the results of these calls were not systematically recorded or analyzed. In effect, differences in treatment were not recorded or reported until both testers had appeared in person at the housing provider's office. The rationale for this approach was that researchers could not be sure housing providers had identified each tester's race or ethnicity until they had seen them in person. But critics argued that housing providers might be screening out minority customers at the phone stage, based on their perceptions of callers' race or ethnicity. Therefore, HDS2012 developed new protocols to record any differential treatment that may occur before in-person visits.

Identifiability of a Tester's Race or Ethnicity

When homeseekers call to make an appointment, the housing provider might or might not accurately identify their race or ethnicity. Even when homeseekers meet in person with housing providers, it is not certain that their race or ethnicity is accurately identified. In HDS2012, a team of coders assessed the race/ethnicity of each tester based on reading the tester's name and listening to a recording of his or her speech—the information available to an agent over the phone. A parallel assessment was conducted based on name, speech, and a photograph—the information available to an agent during an in-person meeting. This assessment made it possible to address the question of whether minority testers who are identifiable are more likely to experience discrimination.

Discrimination Against Other Groups of Homeseekers

In recent years, HUD has supported the extension of the paired-testing methodology to measure the extent and forms discrimination against other potentially vulnerable groups of homeseekers. These extensions have generally been achieved by first conducting a small-scale pilot effort to assess feasibility and to test revised protocols and measures before launching a full-scale national study. To date, this two-stage process has been or is being applied to measure discrimination against renters with disabilities; lesbian, gay, and transgender renters; renter families with children; and renters participating in the Housing Choice Voucher program.

Interpreting the Results From Paired-Testing Studies

Over the course of a long history of paired-testing research, scholars have debated and refined the statistical measures used to report and interpret the results. A major topic of debate has been whether to highlight “gross” or “net” measures of differential treatment. *Gross measures* report the share of all tests in which the White tester is favored over the minority tester—the most straightforward indicator of adverse treatment based on race or ethnicity. Although gross measures of differential treatment are easily understandable, most researchers believe, however, that they generally

overstate the frequency of systematic discrimination. In any paired-testing study, some tests show treatment that favors the minority tester over the White tester (for at least some indicators). These tests could reflect either systematic reverse discrimination or the effects of random, nondiscriminatory influences. For example, another customer may have rented the advertised apartment between two testers' visits, or the real estate agent may have been in a rush to get home and therefore showed her late-afternoon customer fewer available homes).

This reality has led many researchers to construct *net measures*, which report the proportion of White-favored treatment minus the proportion of minority-favored treatment (along with corresponding measures of statistical significance). To the extent that minority testers are systematically favored over White testers in some share of housing inquiries (reverse discrimination), the net measure will understate the incidence of discrimination against minority testers, so it is thought to provide a *lower bound estimate* of systematic discrimination in favor of White homeseekers.⁵

Analyses over the past 25 years strongly suggest that gross measures include substantial random differences in treatment, and that net measures more accurately reflect the systematic disadvantages that minority homeseekers face. One important source of evidence on this issue is a small sample of three-part tests conducted as part of HDS2000. In these tests (conducted in two metropolitan areas), the in-person visit by a White tester was followed by two visits by a minority tester, or the in-person visit by a minority tester was followed by two visits by a White tester, all following the same protocols. Comparing the treatment of the two same-race testers provides a direct estimate of random (not race- or ethnicity-based) differential treatment. This exploratory triad testing effort suggested that most, if not all, minority-favored treatment is random; it provides no convincing evidence that minority-favored treatment systematically exceeds differences in the treatment of same-race testers (see Turner and Ross, 2003a).

A second major measurement challenge for paired-testing researchers involves the definition of composite measures that summarize the results across the multiple forms of treatment typically captured in a paired test. Three basic approaches have been applied over the years, all of which have significant limitations: (1) cumulative measures, (2) consistency measures, and (3) hierarchical measures. The examples that follow assume a White versus an ethnic and racial minority discrimination test, but they would apply in the same manner to any comparison of testers from a protected class under discrimination statutes with any control group of testers.

1. **Cumulative measures** report the share of tests in which the White tester was favored over the minority tester on *any* of several measures. For example, if for a given test both testers got an appointment and both were told the advertised apartment was available, but the White tester was shown more available apartments, the test would be classified overall as White favored. If in the same test, however, the minority tester was quoted more favorable terms for the same advertised unit, that test would also have to be classified as minority favored. If random factors are contributing to some of the observed differences in treatment, cumulative measures essentially magnify their impact, yielding very high estimates of both White-favored and minority-favored treatment.

⁵ For outcomes from a test that can be measured in amounts (such as the number of units recommended or the monthly rent quoted), the net measure reflects the average degree of differential treatment experienced by minority testers relative to White testers, providing a measure of the severity of discrimination.

2. **Consistency measures** focus on the extent to which the White and minority testers are consistently favored over their counterparts in an effort to produce a measure of systematic treatment in favor of White testers that would not be affected by random factors. For example, for a given set of outcome measures, a finding of *consistent adverse treatment favoring White homeseekers* is made when a White tester receives preferential treatment on one or more measures while the minority tester fails to receive preferential treatment on any measure. Instances in which both testers receive preferential treatment for one or more outcomes (or when both receive the same treatment) have a consistency measure of *neutral*. The consistency measures attempt to isolate tests in which one can be reasonably sure that systematic preference exists for White homeseekers. A nontrivial share of tests, however, show consistent treatment in favor of the *minority* tester—more than one would expect if all randomness were eliminated.
3. **Hierarchical measures** are designed to give the most weight in determining differential treatment to the more important treatment items. Not all measured treatment items are of equal importance. Suppose that a White homeseeker is shown more available housing units than his or her minority counterpart, but the minority homeseeker is given more complete information about rental terms and conditions (application and other fees, rent, security deposit) but not told about the incentive program. A hierarchical measure would classify this as a White-favored test, because the White was favored on the more consequential treatment indicator. Again, however, if a large number of treatment indicators are incorporated into a hierarchical composite, then random factors can inflate the share of tests classified as either White favored or minority favored.

The most recent national paired-testing study, HDS2012, adopted a new approach to summarize findings across the many treatment indicators. Instead of trying to define a single, “headline” measure of discrimination, HDS2012 reported seven key measures that, taken together, provide a rounded picture of both the incidence and the severity of differential treatment over the natural course of a test. It then combined these into two overall measures that summarize the severity of adverse treatment across the various stages of the test—(1) the average difference in the number of homes recommended to White and minority homeseekers and (2) the average difference in the number of homes shown. These summary measures were selected to reflect the most consequential forms of differential treatment observed in the 2012 study.

Goals and Contents of This Symposium

This collection of articles and commentaries takes stock of the current state of paired-testing as a tool for rigorously measuring housing market discrimination, highlighting both its important accomplishments and the challenges moving forward. Leaders from a variety of disciplines in the field of housing discrimination research have tackled a wide range of topics and approaches to several vital questions: How did we get here? Where are we now? Where do we go from here? So, although some of the assembled articles present research evidence, others focus on design and methodological issues, and others offer ideas for the further evolution of the paired-testing methodology. We hope that, read as a whole, the collection offers readers a well-rounded—and provocative—picture of the current state of research in this area.

Oh and Yinger (2015) review the evidence about discrimination obtained from in-person paired testing in housing markets, with an emphasis on the major national studies of racial and ethnic discrimination. They review the testing methodology and then the research results on a variety of discrimination measures. Finally, they examine the linkage between paired testing and public policy, including passage of significant legislation and funding strategies for fair housing enforcement. Oh and Yinger also provide two useful appendixes; the first summarizes other relevant studies and the second summarizes the origin of housing audits.

Rob Pitingolo and Stephen L. Ross tackle the technically difficult and substantively important issue of the degree to which paired testing underestimates the degree of housing discrimination (Pitingolo and Ross, 2015). The constraints on the paired-testing methodology in scope and depth of the housing market transactions it can examine make its results a lower bound estimate of likely discrimination. Just how much lower the estimate is a subject of some concern and debate.

Freiberg and Squires (2015) take on the question of where research on racial and ethnic discrimination in housing should go next. How have changes in the housing market and housing provider practices begun to limit the application of paired testing and what are the alternatives? They raise a series of challenges to those who wish to use paired testing in future research on housing discrimination.

Claudia Aranda summarizes a very recently finished study of housing discrimination against people who are deaf or hard of hearing and people who use wheelchairs. This study extends housing discrimination research to a protected class (people who have physical disabilities), which had received only limited attention up to now. In addition to presenting the study findings, Aranda (2015) reviews some novel issues in the application of paired testing to this important protected class of homeseekers. The article also describes the nuts and bolts of conducting a national audit to illustrate the method in more detail.

Margery Austin Turner reviews a set of pilot studies currently under way that extend paired testing to other protected classes and offers a set of key design questions that future studies must tackle as they seek to apply the paired testing more broadly (Turner, 2015).

The concluding section of this issue presents commentary on the articles from a variety of perspectives. Ali M. Ahmed represents international scholarship on housing discrimination research; Samantha Friedman represents active academic research on housing discrimination; James Perry represents the private fair housing organizations, which are very active in the enforcement of fair housing laws; and Fred Underwood provides a perspective from the housing industry. Together, their diversity of experience and outlook can broaden the discussion, raise additional questions and concerns, and pose additional suggestions for further inquiry (Ahmed, 2015; Friedman, 2015; Perry, 2015; Underwood, 2015).

Acknowledgments

The symposium has greatly benefited from the input of the numerous reviewers who contributed their expertise and time to very thoughtful and extremely helpful critiques of earlier drafts of the articles described in this Guest Editors' Introduction. Their assistance was in the best spirit of the cooperative joint enterprise that successful inquiry in a field of knowledge must be, and the co-editors of the symposium greatly appreciate their contribution.

Guest Editors

Margery Austin Turner is Senior Vice President for Program Planning and Management at the Urban Institute.

Judson James is a social science analyst in the Office of Policy Development and Research at the U.S. Department of Housing and Urban Development.

References

- Ahmed, Ali M. 2015. "Commentary: Some Thoughts on Field Experiments on Housing Discrimination From a European View," *Cityscape* 17 (3): 137–142.
- Aranda, Claudia L. 2015. "Targeting Disability Discrimination: Findings and Reflections From the National Study on Housing Discrimination Against People Who Are Deaf and People Who Use Wheelchairs," *Cityscape* 17 (3): 103–121.
- Freiberg, Fred, and Gregory D. Squires. 2015. "Changing Contexts and New Directions for the Use of Testing," *Cityscape* 17 (3): 87–101.
- Friedman, Samantha. 2015. "Commentary: Housing Discrimination Research in the 21st Century," *Cityscape* 17 (3): 143–149.
- Oh, Sun Jung, and John Yinger. 2015. "What Have We Learned From Paired Testing in Housing Markets?" *Cityscape* 17 (3): 15–59.
- Perry, James. 2015. "Commentary: Expanding the Fair Housing Testing Landscape," *Cityscape* 17 (3): 151–153.
- Pitingolo, Rob, and Stephen L. Ross. 2015. "Housing Discrimination Among Available Housing Units in 2012: Do Paired-Testing Studies Understate Housing Discrimination?" *Cityscape* 17 (3): 61–85.
- Turner, Margery Austin. 2015. "Other Protected Classes: Extending Estimates of Housing Discrimination," *Cityscape* 17 (3): 123–136.
- Turner, Margery, and Stephen Ross. 2003a. *Discrimination in Metropolitan Housing Markets: Phase 2—Asians and Pacific Islanders*. Washington, DC: U.S. Department of Housing and Urban Development.
- . 2003b. *Discrimination in Metropolitan Housing Markets: Phase 3—Native Americans*. Washington, DC: U.S. Department of Housing and Urban Development.
- Turner, Margery, Stephen Ross, George Galster, and John Yinger. 2002. *Discrimination in Metropolitan Housing Markets: National Results From Phase I of HDS2000*. Washington, DC: U.S. Department of Housing and Urban Development.
- Turner, Margery Austin, Rob Santos, Diane K. Levy, Doug Wissoker, Claudia Aranda, and Rob Pitingolo. 2014. *Housing Discrimination Against Racial and Ethnic Minorities 2012*. Washington, DC: U.S. Department of Housing and Urban Development.

Turner, Margery, Raymond Struyk, and John Yinger. 1991. *Housing Discrimination Study: Synthesis*. Washington, DC: U.S. Department of Housing and Urban Development.

Underwood, Fred. 2015. "Commentary: Testing Benefits Housing Providers and the Industry," *Cityscape* 17 (3): 155–157.

Wienk, Ronald E., Clifford E. Reid, John C. Simonson, and Frederick J. Eggers. 1979. *Measuring Discrimination in American Housing Markets: The Housing Market Practices Survey*. Washington, DC: U.S. Department of Housing and Urban Development.

What Have We Learned From Paired Testing in Housing Markets?

Sun Jung Oh
Korea Labor Institute

John Yinger
Syracuse University

Abstract

Fair housing audits or tests, which compare the way housing agents treat equally qualified homeseekers in different racial or ethnic groups, are an important tool both for enforcing fair housing laws and for studying discriminatory behavior in housing markets. This article explains the features of two types of housing audits: in-person paired audits and correspondence audits, which are usually conducted over the Internet. In addition, this article reviews evidence provided by audit studies about the extent of housing discrimination. The studies reviewed include four national studies in the United States based on in-person audits and many studies based on correspondence audits in the United States and in several European countries. This article also reviews audit-based evidence about the causes of discrimination in housing markets. Despite variation in methods, sample sizes, and locations, audit studies consistently find evidence of statistically significant discrimination against homeseekers who belong to a historically disadvantaged racial or ethnic group. The 2012 national audit study found, for example, that the share of audits in which a White homebuyer was shown more available houses than an equally qualified Black homebuyer was 9 percentage points higher than the share in which the Black homebuyer was shown more houses than his or her White counterpart. In the United States, housing discrimination against Black and Hispanic homeseekers appears to have declined in some types of agent behavior, such as whether the advertised unit is shown to a customer, but to have increased in others, such as steering Black and Hispanic homeseekers toward minority neighborhoods. This article also discusses the past use and continued importance of fair housing audits as a fair-housing-enforcement tool.

Introduction

In-person paired testing is a methodology explicitly designed to observe differential treatment of equally qualified homeseekers in different groups—that is, to observe discrimination. Testing is also called auditing; we use the two terms as synonyms. Paired testing in housing markets was first conducted in the 1950s by partnerships between scholars and community groups. This type of testing then gained prominence during the 1960s and 1970s as localities, states, and the federal government passed fair housing legislation, private fair housing groups refined testing methods for enforcement purposes, and scholars discovered that this method could be used to study discriminatory behavior. Building on this foundation, testing has yielded extensive information on the nature, extent, and causes of discrimination in housing; it has, by documenting discrimination, provided influential support for fair housing legislation, such as the 1988 Fair Housing Amendments Act (FHAA); and it has been used extensively as a fair housing enforcement tool by private fair housing groups and by governmental civil rights agencies. In addition, the recent development of testing methods using the Internet, usually called correspondence tests or correspondence audits, has resulted in a large number of studies of housing discrimination in many different countries.

This article begins with a detailed review of the testing method and of evidence about discrimination against African-American, Hispanic-American, and Asian-American homeseekers obtained from in-person paired testing in housing markets. This review is followed by an exploration of studies based on e-mail audits (some of which apply to other countries), an exploration of the link between paired testing and fair housing policy, and a brief review of the use of paired testing in some other markets. The focus is on the use of audits (tests) for research purposes.¹ The next section examines testing methodology by reviewing the basics of paired testing, discussing audits that do not involve face-to-face contact (called correspondence audits), explaining how testing results can be used to study the causes of discrimination, and describing key methodological issues in the four national housing audit studies. Results from paired-testing studies in the housing market are presented in the subsequent section. To be specific, this section explores the incidence of discrimination, trends in discrimination, results concerning racial and ethnic steering, and evidence about the causes of discrimination. Appendix A provides further evidence from smaller paired-testing studies and from correspondence audits. The final section addresses paired testing and public policy, focusing on the link between paired testing and fair housing policy, but also providing a brief review of paired-testing research in markets other than the housing market. Appendix B describes the origins of fair housing audits.

Testing Methodology

The testing method can be used both to measure the incidence of discrimination and to test hypotheses about discrimination's causes. This section explains the methodology of both in-person paired testing and correspondence audits, shows how these methods can be used to test hypotheses about the causes of discrimination, and introduces the four national in-person audit studies in the United States.

¹ See Smith (1994) and Schwemm (2014) for details of enforcement audits.

The Basics of In-Person Paired Testing

In-person paired-testing research involves six main steps. First, auditors are selected. Each auditor must be capable of playing the role of a typical homeseeker and not have unusual traits that might influence his or her treatment in the housing market relative to the auditor with whom he or she is paired.

Second, auditors are trained about the role they should play during an audit. In most cases, they are instructed to inquire about an advertised unit and then to ask for additional suggestions from the housing provider. In some audit studies, the audits are *blind*, in the sense that the auditors are told only that they are helping with a study about the marketing of housing and are not told that they have a partner or that the study is investigating discrimination. In the 1989 and 2000 Housing Discrimination Studies, however, the training was not blind in this sense. Auditors were told the purpose of the study and were trained to provide information as accurately as possible, and, to the extent possible, managers were given protocols to check on the accuracy of the information provided. In these studies, auditors in different racial or ethnic groups were trained together to ensure that they received the same training. Blind auditing is not appropriate in these studies because auditors can observe the group composition of the trainees; without an explanation for this composition, auditors might make their own guesses about the purpose of the study.²

Third, a sample of available housing units is randomly drawn, usually from the major local newspaper. In some audit studies, some neighborhoods are oversampled or the sample from the major newspaper is supplemented with other sources, such as community newspapers. Each sampled unit then becomes the basis for one audit.

Fourth, auditors are matched for each test with one member from a historically disadvantaged group.³ Paired testers are assigned income and other household traits that make them equally qualified for the sampled advertised unit about which they are inquiring.⁴ Even if the auditors are the same, the assigned income and household traits vary from one audit to the next to match the associated advertised unit. Teammates are assigned similar incomes and other traits for a given audit so that differences in these traits do not lead to differences in treatment. Because housing market transactions are relatively simple and because the people marketing housing do not usually

² The auditor training manual for HDS1989 began by saying, “Thank you for helping in this study of housing discrimination in the nation’s sales and rental housing markets. Your role as an ‘auditor’ is absolutely critical to the success of the study. Your activities will provide the raw material from which others will be able to make assessments about the nature and extent of practices of housing discrimination by members of the housing industry” (Urban Institute, 1991: Annex 4, p. 1). Ross and Turner (2005: 174) explained another reason for avoiding blind audits in the early national studies: “In 1989, many minority testers experienced blatant discriminatory treatment. For example, one African American tester saw the real estate agent visibly react as the tester got out of the car; the agent then jumped in his/her car and quickly drove off. Many testers returned from their visits upset and angry at the treatment they had received. In fact, tester training for the 2000 study explicitly prepared testers so that they would not overreact to such treatment and invalidate the test.”

³ In most cases, we refer to the “White” auditor and the “minority” auditor. An audit could, of course, also be conducted with men and women in teams, some of which consist of two White people. Three-person teams are sometimes used in some enforcement audits, but because of their added expense are rarely used to research (see appendix A).

⁴ In enforcement audits, the auditor from the disadvantaged group is usually given slightly better qualifications. In a research audit study, random assignment of income is sufficient to avoid bias, but in the national audit studies, the auditor from the disadvantaged group was always given a slightly higher income.

ask potential customers about very many traits, the qualifications of audit teammates are almost identical. It follows that differences in the way teammates are treated can be attributed to random factors or to discrimination. Because membership in a historically disadvantaged group cannot be randomly assigned, this approach cannot fully rule out the possibility that some unassigned trait influences treatment, thereby biasing estimates of discrimination up or down; however, good management makes this outcome unlikely. As discussed in the section “The Housing Discrimination Studies,” some audit studies have also collected information on auditors’ actual traits, such as their income and education, to see if these traits affect measures of discrimination.⁵

Fifth, audit teammates separately contact the housing agent associated with one of the selected advertisements and attempt to schedule a visit. The initial contacts are completed during a short period, but not so short as to be suspicious to the agent. In most studies, the order of the visits is randomized. The visit, if it occurs, then follows the script that the auditors learned in training, with inquiries about the advertised unit and similar units.⁶ Auditors are generally encouraged to learn about and visit as many units as possible, while not stating preferences (beyond an interest in the advertised unit and units similar to it) that would guide this process.

Sixth, and finally, after an audit is complete, each audit teammate is asked to record what he or she was told and how he or she was treated. These audit forms provide information on the number of houses or apartments shown to each auditor and also on many other aspects of housing agent behavior. Audit teammates have no contact with each other during an audit and they fill out their audit survey forms independently. Most audit studies then schedule debriefing sessions in which an audit manager reviews these forms with each auditor to ensure that all information on the forms is accurate.

Unlike alternative approaches that look for signals of discrimination in housing prices, housing quality, homeownership, and segregation patterns, in-person paired testing provides direct measures of discrimination by comparing the outcomes of equally qualified White and minority testers. Moreover, paired testing makes it possible to examine the multiple, complex forms that discrimination can take by observing many types of housing agent behavior. This methodology yields a powerful narrative concerning the way people in different groups are treated. This narrative adds credibility to findings of discrimination in research, policy, and court settings. In addition, the results of the paired testing can shed light on the causes of discrimination because they provide information on the circumstances in which discrimination occurs. This article returns to research on the causes of discrimination in the section “Testing Hypotheses About the Causes of Discrimination.”

One important feature of paired audits is that some of the unobservable factors are shared between audit teammates. This type of unobservable factor does not lead to bias in estimates of

⁵ Some disagreement among scholars remains about the importance of traits that are not matched in the audit design. Heckman and Siegelman (1993) argue that they could be an important source of bias (in an unknown direction), whereas Yinger (1993) argues that they are unlikely to be an important source of bias. Controlling for auditors’ actual traits in the HDS2000, one possible unobservable in previous studies, has little impact on the results but does not rule out the possibility of bias from other factors.

⁶ In the first national audit study (Wienk et al., 1979), the auditors asked about the type of unit and general location defined by the selected advertised unit but did not ask about the advertised unit specifically.

discrimination, but if not accounted for it does lead to an upward bias in standard errors. The studies discussed in the section “The Results of Paired-Testing Studies” provide a variety of ways to avoid this type of bias. This issue also arises in the following section on correspondence audits.

A New Development: Correspondence Audits

One disadvantage of paired testing is that it is expensive; a large management structure must be created and auditors must be hired, trained, sent into the field, and debriefed. To address this practical problem, scholars have developed an alternative method, called a correspondence audit, that is less expensive and more precise but addresses a narrower set of questions. This methodology is based on e-mail inquiries instead of visits and relies on names instead of personal contact to convey race or ethnicity.⁷ In addition, correspondence audits record the housing agent’s response to an e-mail instead of his or her in-person treatment of an auditor. The rapid growth in the use of the Internet for marketing housing has made this an appealing strategy.

Correspondence audits for research purposes have focused on rental housing, usually based on advertisements posted on a particular website, such as craigslist. Unlike in-person paired testing, correspondence audits can literally assign race or ethnicity randomly. The audit managers write several versions of an audit e-mail and then randomly select a version and a group membership for each inquiry. This randomization eliminates potential bias from unobserved differences between White and minority homeseekers. This sharper identification strategy comes at a cost, however, because correspondence audits can address only a relatively narrow set of questions concerning housing agents’ initial responses to an inquiry.

One distinction between in-person paired testing and correspondence audits is that correspondence audits do not have to rely on pairing; that is, a housing agent need not receive two e-mails, one each from a White homeseeker and a minority homeseeker. With a one-e-mail approach and random assignment of racial or ethnic identity, discrimination is the difference between the average treatment of e-mails with a White identity minus the average treatment of the e-mails with the identity of a racial or ethnic minority.⁸ This single-inquiry strategy lowers the possibility of detection, because housing agents do not receive two somewhat-similar inquiries in a relatively short time span. This strategy also raises the standard errors of discrimination estimates for a given sample size, because unobservable factors shared by teammates cannot be removed. Because correspondence audits are relatively cheap, however, the problem of high standard errors can be addressed by expanding the sample size.⁹

⁷ Some early correspondence audits used phone calls instead of e-mails. In this case, minority status is conveyed both through the auditor’s name and through his or her accent. This approach, like standard paired testing, cannot randomize group membership—at least not if a person’s accent is part of the study design.

⁸ In principle, in-person audits do not have to rely on pairing, either. Audit managers could randomly select the group membership for the single inquiry associated with each advertisement in the sample. This approach would lose the narrative power of a two-person audit, however, and it would require a much larger sample size. To the best of our knowledge, no in-person audit study has followed this strategy. Although this approach would preserve the lack of correlation between auditors and the circumstances they encounter, it would not eliminate the potential bias from a correlation between group membership and unobserved auditor traits that influence treatment in the housing market.

⁹ Another disadvantage of nonpaired audits is that they do not yield gross measures of discrimination (defined in the next section).

Testing Hypotheses About the Causes of Discrimination

Many scholars have used audits to study the causes of discrimination. In the audit context, a hypothesis about a cause of discrimination is stated as a situation in which discrimination is more likely to occur. The hypothesis is then tested by determining whether discrimination is higher or more likely during audits in which that situation arises. Three main hypotheses have appeared in the literature: (1) the agent-prejudice hypothesis, (2) the customer-prejudice hypothesis, and (3) the statistical-discrimination hypothesis. These hypotheses, which are not mutually exclusive, are briefly described here; existing empirical tests of these hypotheses are discussed in the section “The Causes of Discrimination.”

The agent-prejudice hypothesis states that discrimination may occur because real estate agents have strong personal biases against minority homeseekers. Because agent prejudice is not directly observed, studies have tested this hypothesis using variables that are known to be associated with prejudice. These variables include the race, age, and gender of the agent and the gender or marital status of the auditor. Studies have found, for example, that White prejudice is higher among men than among women (Schuman, Steeh, and Bobo, 1985; Schuman et al., 1997) and increases with age (Schuman and Bobo, 1988), so the agent-prejudice hypothesis predicts that discrimination will be higher if the agent is male or older.

The customer-prejudice hypothesis states that agents may avoid renting to minority customers to protect their actual or potential business with prejudiced White customers.¹⁰ This hypothesis predicts that agents discriminate more against a minority customer if some of the customer’s characteristics are particularly likely to upset their prejudiced White customers and certain types of customers are more likely to be racially prejudiced. Such characteristics may include low educational level of White property owners (Schuman et al., 1997) and also a low household income and a large number of children of minority homeseekers (Schuman, Steeh, and Bobo, 1985). The customer-prejudice hypothesis also predicts more discrimination when the agent’s office is small, with a smaller client base; the office is in a White neighborhood; or the advertised unit is in a largely owner-occupied neighborhood.

The statistical-discrimination hypothesis states that discrimination occurs when agents treat people in different groups differently because they believe that group membership is correlated with unobserved characteristics that affect the profitability of their actions.¹¹ In the rental housing market, for example, a rental agent may use customers’ race or ethnicity as a signal about their preferences for housing type, neighbors, or both or about constraints that are related to the probability of a transaction. The statistical-discrimination hypothesis also predicts that discrimination against minority

¹⁰ Several scholars (Galster, 1990c, 1987; Yinger, 1986) have pointed out that this hypothesis breaks down when racial neighborhood transition from White to Black is imminent or under way. In this case, real estate brokers may be able to maximize turnover, and hence commissions, by selling to Black households.

¹¹ Several scholars have identified more specific hypotheses in this category. Galster (1990c, 1987) and Newberger (1989) argue that agents may or may not show houses to Black homeseekers in some neighborhoods because of “anticipated discrimination” against them by White homesellers or mortgage lenders. Galster (1990c) also argues that agents may discriminate on the basis of their beliefs about what customers in different groups prefer. Ondrich, Ross, and Yinger (2003) argue that agents may have stereotypes about the financial capabilities of people in certain groups that lead to intergroup differences in treatment despite the equal qualifications of audit teammates.

homeseekers will decrease with the minority population in a neighborhood if real estate agents believe that minority homeseekers prefer living where minority residents are concentrated. In addition, this hypothesis predicts that discrimination increases with the value of houses if real estate agents believe that minority individuals have a relatively high probability of financial difficulties.

The Housing Discrimination Studies

The largest paired-testing studies in the United States are the Housing Market Practices Survey (HMPS) in 1977 and the three Housing Discrimination Studies (HDS1989, HDS2000, and HDS2012) sponsored by the U.S. Department of Housing and Urban Development (HUD).¹² These studies were designed to yield statistically reliable national estimates of discrimination against certain racial and ethnic groups in urban housing markets. Exhibit 1 presents a few common features of the four studies (including race or ethnicity tested, scale, and locations where tests were conducted) along with some of their differences. In addition, several scholars have conducted smaller scale in-person audit studies, and numerous e-mail correspondence audits have been conducted during the past two decades.

Outcomes measured in paired-testing studies indicate the incidence and severity of unfavorable treatment that minority homeseekers experience. Unfavorable treatment may arise, for example, in the probability of being told the advertised units are available; the probability of making an appointment when inquiring about advertised units; the probability of at least one in-person visit to an available apartment; the number of apartments suggested or shown; the characteristics of apartments shown or inspected; and the terms and conditions of the lease, such as the rent, the security deposit, or the lease length.

Two types of discrimination measures have appeared in the literature: (1) gross measures and (2) net measures. Gross measures indicate the share of all audits in which the White auditor is favored over his or her minority teammate. Although gross measures are easily understandable, they may overstate the frequency of systematic discrimination because nondiscriminatory random events are responsible for some portion of observed treatment. A White auditor might appear to be favored, for example, because she went first and the apartment was rented before her minority teammate arrived. Net measures of discrimination are the proportion of audits in which the White auditor is favored minus the proportion of audits in which the minority auditor is favored. The net measure indicates the disadvantage minority homeseekers face in the housing market relative to White homeseekers. The net measure provides a lower bound estimate of systematic discrimination in favor of White homeseekers. To the extent that minority homeseekers are systematically favored over White homeseekers in some share of housing inquiries, such as inquiries concerning apartments in largely minority neighborhoods, the net measure will understate the incidence of discrimination against minority homeseekers. Although the audit design makes it possible to calculate reasonable net and gross measures based on sample proportions, more precise measures of discrimination can be obtained with more advanced statistical procedures (see Ondrich, Ross, and Yinger, 2003; Ondrich, Stricker, and Yinger, 1999, 1998; Page, 1995; Zhao, 2005).

¹² The HMPS, the first national audit study of housing market discrimination, was conducted by the National Committee Against Discrimination in Housing. The Urban Institute conducted the three Housing Discrimination Studies.

Exhibit 1

Comparison of National Paired-Testing Studies

Study	Race/Ethnicity Tested	Scale	Location	Features
HMPS1977	Black	3,264 tests	40 metropolitan areas	Individual real estate agents and apartment rental complexes were randomly selected from only one sample of newspaper advertisements for each metropolitan area. Auditors did not explicitly ask for the advertised unit. Black auditors always preceded White auditors in rental audits, but White auditors always preceded Black auditors in sales audits.
HDS1989	Black Hispanic	3,800 tests	25 metropolitan areas	Each audit began with a request for a specific, advertised unit randomly selected from the most recent Sunday newspaper (“anchoring” audit). It was the first study that measured racial and ethnic steering. For the purpose of steering analysis, auditors were instructed to ask about the availability of other homes similar in size and prices to the advertised unit. Order of initial call was randomized.
HDS2000	Black Hispanic Asian Native American	4,600 tests	23 metropolitan areas	Like HDS1989, the sample of housing units was randomly selected from the Sunday classified advertisements of major metropolitan newspapers. HDS2000 also used geographic oversampling and supplemental samples from secondary newspapers for areas that were underrepresented in the newspaper advertisements. In addition, it recorded some auditors’ actual characteristics, such as income and education. Testers made appointment calls for sales and rental tests, and the order of initial call was randomized. On sales tests, testers were not to mention the advertised home during this call and were also to refrain from providing their personal and financial information. Testers inquired about the availability of the advertised housing unit that prompted their visit and about similar units.
HDS2012	Black Hispanic Asian	8,047 tests	28 metropolitan areas	Testers attempted to make appointments for in-person visits by telephone or e-mail. Order of initial contact was randomized. On sales tests, testers were not to mention the advertised home during telephone conversation or e-mail. If making an appointment was successful, testers used the in-person visit to learn about available homes or apartments. Testers inquired to view the home that was advertised. If told about at least one available housing unit, testers sought to inspect homes or apartments.

HDS = Housing Discrimination Study, HMPS = Housing Market Practices Survey.

The Results of Paired-Testing Studies

This section first presents the findings from four nationwide paired-testing studies that began in the late 1970s and were sponsored by HUD. In addition, this section reviews the results of studies based

on the data from the four national studies and of other studies that conducted paired tests. It also reviews the evidence on trends in housing discrimination based on the four HUD-sponsored studies.

This section presents the results of sales and rental audit studies separately. The results for these two markets are not strictly comparable.¹³ Buying a house is a more complex procedure than renting an apartment, and less of the buying process can be examined than that of the renting process. The complexity of the sales market provides many opportunities for discriminatory treatment, not all of which can be examined. Moreover, real estate agents earn their incomes from commissions. Prospective Black buyers may receive systematically different treatment or service than White buyers—but may still receive service. Audit studies must therefore be careful to look for differences in the services provided, not just differences in whether service was provided at all. In addition, the results of audit studies should not be interpreted as comprehensive measures of discrimination but instead as measures of discrimination in key types of agent behavior. Moreover, the types of behavior that can be observed may not be the same in the sales and rental markets.

The Incidence of Discrimination

This section presents findings from HDS2012, which was conducted in 28 metropolitan areas to measure discrimination against minority home renters and buyers in 2012. Based on overall measures of differential treatment for renters, White renters experience more favorable treatment than equally qualified Black renters in 28.4 percent of inquiries compared with 19.6 percent in which Black renters are favored (Turner et al., 2013). White renters similarly experience more favorable treatment than equally qualified Hispanic renters in 28.9 percent of inquiries compared with 18.9 percent in which Hispanic renters are favored (Turner et al., 2013). In the sales tests, White homebuyers experience more favorable treatment than equally qualified Black homebuyers in 40.7 percent of inquiries compared with 30.9 percent in which Black homebuyers are favored (Turner et al., 2013).

Exhibit 2 presents the results (net measures) of HDS2012 for the eight types of auditor treatment that are similarly measured for both rental and sales tests. The top two panels present the rental and sales test results for Black homeseekers, and the bottom two panels present the rental and sales results for Hispanic homeseekers. The left panel presents the outcomes for which White testers were favored (defined as statistically significant net measures), and the right panel presents the outcomes for which no discrimination was detected or Black testers were favored. A few interesting patterns of housing discrimination emerge.

First, when comparing the results of rental tests with sales tests for Black homeseekers, it is clear that the magnitude of discrimination against Black homeseekers is higher in sales tests than in rental tests (for the outcomes a, b, c, and d). For instance, Black homeseekers are told about fewer available units (outcome a) than White homeseekers in 13.4 percent of inquiries in the sales tests compared with 9.0 percent of inquiries of the rental tests. Also, Black homeseekers are told about 0.5 fewer units available than White homeseekers (outcome b) in the sales tests compared with 0.2 fewer units in the rental tests. Finally, Black homeseekers are shown fewer units than White homeseekers (outcome c) in 9.3 percent of inquiries of the sales tests compared with 2.8 percent of inquiries of the rental

¹³ This point was clearly explained in Wienk et al. (1979: 175–176).

Exhibit 2

Results of HDS2012

		White Favored Against Black	Neither Favored or Black Favored
Rental	a.	Told about more available units (9.0%*)	f. Only one tester able to make appointment (0.4%)
	b.	Average number of units available per visit (0.20*)	g. Only one tester told units available (0.9%)
	c.	Shown more units (2.8%*)	h. Level of agent helpfulness (- 0.03)
	d.	Average number of units shown (0.04*)	
	e.	Average rent (- \$4*)	
Sales	a.	Told about more available units (13.4%*)	e. Average price (- \$4,012)
	b.	Average number of units available per visit (0.50*)	g. Only one tester told units available (2.1%)
	c.	Shown more units (9.3%*)	h. Level of agent helpfulness (0.12)
	d.	Average number of units shown (0.30*)	
	f.	Only one tester able to make appointment (2.4%*)	
		White Favored Against Hispanic	Neither Favored or Hispanic Favored
Rental	a.	Told about more available units (12.8%*)	b. Average number of units available per visit (- 0.22*)
	c.	Shown more units (6.0%*)	f. Only one tester able to make appointment (0.2%)
	d.	Average number of units shown (0.07*)	h. Level of agent helpfulness (0.02)
	e.	Average rent (- \$6*)	a. Told about more available units (2.3%)
	g.	Only one tester told units available (1.8%*)	b. Average number of units available per visit (0.28)
Sales			c. Shown more units (2.0%)
			d. Average number of units shown (0.10)
			e. Average price (- \$5,621)
			f. Only one tester able to make appointment (0.4%)
			g. Only one tester told units available (- 0.2%)
			h. Level of agent helpfulness (0.08)

HDS = Housing Discrimination Study.

* Indicates statistical significance at the 90-, 95-, or 99-percent level.

Note: Net measures are presented in parentheses.

Source: Estimates are from exhibits IV-1 and IV-14 of Turner et al. (2013)

tests. In terms of the number of units shown (outcome d), Black homeseekers are shown 0.30 fewer units than White homeseekers in the sales tests compared with 0.04 fewer units in the rental tests.

Second, in rental markets, Hispanic renters experience more discrimination than Black renters for four outcomes that exhibit discrimination (outcomes a, c, d, and e). For instance, Hispanic renters are told about fewer available units than are White renters (outcome a) in 12.8 percent of inquiries, whereas Black renters are told about fewer available units in 9.0 percent of inquiries. In addition, Hispanic renters are shown fewer units than equally qualified White renters (outcome c) in 6.0 percent of inquiries compared with 2.8 percent of Black renters' inquiries. On average, Hispanic renters are shown 0.07 fewer units than White renters (outcome d), but Black renters are shown 0.04 fewer units than White renters. Agents also quote slightly higher rents (\$6 per month on average) to Hispanic renters than White renters, and agents quote \$4 higher rents to Black renters than White renters. In contrast with the results of the rental tests, however, Hispanic homebuyers are as favored as White homebuyers, whereas White homebuyers are favored over Black homebuyers in five out of the eight outcomes.

Third, minority homeseekers are rarely denied appointments (outcome f), and when both White and minority testers meet with an agent in person, they are rarely told that no unit is available (outcome g). Compared with comparable White customers, however, Black homebuyers are slightly more likely to be denied an in-person appointment (in 2.4 percent of inquiries), and Hispanic renters are slightly more likely to be told that no homes or apartments are available (in 1.8 percent of inquiries). Overall levels of agent helpfulness to White and minority homeseekers are not significantly different.

In addition to HDS2012, several smaller scale audits have been conducted in individual cities in the United States and in European countries. These studies traditionally used in-person paired tests, but recent studies generally use e-mail correspondence tests. Most correspondence tests determine whether ethnically linked names influence the probability that the agent responds to an inquiry or allows the homeseeker to make an appointment.

Appendix A presents the summary results for other testing studies of housing.¹⁴ This appendix focuses on the studies involving more than 100 tests. Although the location, time period, and minority groups in these studies are quite varied, the studies consistently found that, for various outcomes, discrimination against racial and ethnic minority homeseekers is a common feature of housing markets in many countries. A number of studies also found discrimination against immigrants and people with low socioeconomic status.

Trends in Discrimination

Tracking discrimination over time can help determine how successful antidiscrimination interventions have been. This section presents the trends in rental and sales discrimination using the results of the four national studies. The four national studies provide a reasonable approximation to national trends in housing discrimination because they were conducted about 10 years apart using a similar methodology.¹⁵ Exhibits 3 and 4 summarize the findings of discriminatory treatment for Black and Hispanic homeseekers, based on types of behavior that are consistently measured in all four national studies.¹⁶ This section presents the results of rental tests and sales tests separately. The incidence of discrimination is drawn from the final reports of HMPS1977, HDS1989, HDS2000 and HDS2012.¹⁷ Two common outcomes presented for both rental and sales tests are (1) whether the agent told only the White tester that the advertised unit was available and (2) whether the agent showed the White tester more units. For sales tests, this section presents an additional

¹⁴ For a review of earlier audit literature, see Yinger (1987).

¹⁵ Because they used similar research methods, HDS1989 and HDS2000 provide particularly clear measures of changes in discrimination between 1989 and 2000. Comparisons across other national studies are less precise. To incorporate changes in housing search practices, for example, HDS2012 used e-mail inquiries to make appointments with housing agents. This issue was recognized by HUD: "Although tracking trends in the incidence of discrimination is also important, HUD placed higher priority on accurately capturing current market practices than on precisely measuring change over time" (HDS2012, Goals for the 2012 Housing Discrimination Study: 2). Moreover, HMPS1977, unlike all subsequent national studies, instructed auditors to ask about a type of house, not a specific house.

¹⁶ The outcomes presented in exhibits 3 and 4 that are similarly measured for the four national studies are different from the outcomes presented in exhibit 2.

¹⁷ The final reports of the four national studies are Wienk et al. (1979) for HMPS1977; Turner, Struyk, and Yinger (1991) for HDS1989; Turner et al. (2002) for HDS2000; and Turner et al. (2013) for HDS2012.

Exhibit 3

Results of National Rental Tests

Minority	Study	Advertised Unit Available		Inspected More Units	
		White Favored (%)	Net Measure (%)	White Favored (%)	Net Measure (%)
Black	HMPS1977	30	19*	27	6*
	HDS1989	19	7*	34	19*
	HDS2000	12	4*	23	7*
	HDS2012	5	0	18	3*
Hispanic	HDS1989	17	9*	27	10*
	HDS2000	12	7*	21	6*
	HDS2012	5	3*	21	8*
Asian	HDS2000	7	0	13	-5
	HDS2012	5	2	22	6*

HDS = Housing Discrimination Study, HMPS = Housing Market Practices Survey.
 * Indicates statistical significance for net measures at the 90-, 95-, or 99-percent level.
 Note: Gross estimates (percent White favored) are by definition statistically significant.
 Sources: Estimates of HMPS1977 are from table 2 of Wienk et al. (1979); estimates of HDS1989 and HDS2000 (except for the Asian minority group) are from exhibits 3-1 and 3-2 of Turner et al. (2002); Asian estimates of HDS2000 and estimates of HDS2012 are from exhibit V-1 of Turner et al. (2013)

Exhibit 4

Results of National Sales Tests

Minority	Study	Advertised Unit Available		Inspected More Units		Help With Financing Offered	
		White Favored (%)	Net Measure (%)	White Favored (%)	Net Measure (%)	White Favored (%)	Net Measure (%)
Black	HMPS1977	21	10*	38	10*	—	—
	HDS1989	10	4*	24	12*	21	5*
	HDS2000	16	1	43	12*	19	1
	HDS2012	13	-1	37	9*	19	2
	HDS1989	9	4*	27	13*	19	-1
Hispanic	HDS2000	12	-3	36	-2	22	12*
	HDS2012	13	0	34	1	16	1
Asian	HDS2000	16	1	46	14	29	15
	HDS2012	15	3	38	10*	19	8*

HDS = Housing Discrimination Study, HMPS = Housing Market Practices Survey.
 * Indicates statistical significance for net measures at the 90-, 95-, or 99-percent level.
 Note: Gross estimates (percent White favored) are by definition statistically significant.
 Sources: Estimates of HMPS1977 are from table 25 of Wienk et al. (1979); estimates of HDS1989 and HDS2000 (except for the Asian minority group) are from exhibits 3-11, 3-12, 3-14, 3-17, 3-18, and 3-20 of Turner et al. (2002); Asian estimates of HDS2000 and estimates of HDS2012 are from exhibit V-2 of Turner et al. (2013)

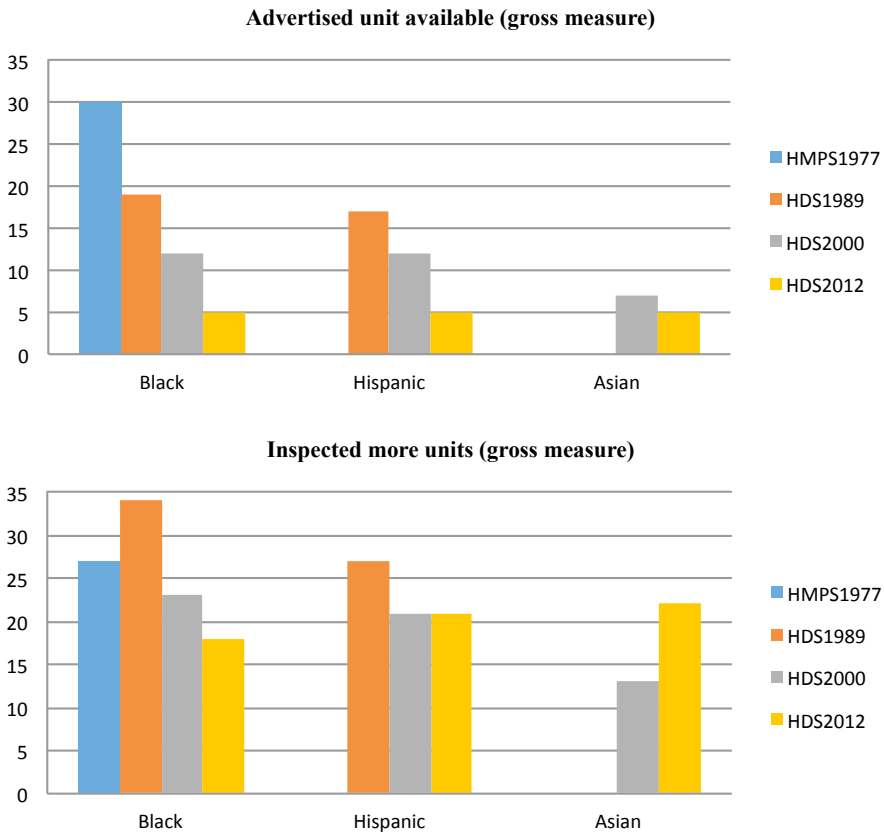
outcome—whether the real estate agent offered help with financing. To be more specific, this outcome indicates whether the agent had a general discussion with the homeseeker about the mortgage process or offered to provide a mortgage prequalification for a maximum loan amount. The net and gross measures consistently show that minority homeseekers receive less favorable treatments than White homeseekers in both the rental and sales housing markets (Turner, Struyk, and Yinger, 1991; Turner et al., 2013; Turner et al., 2002; Wienk et al., 1979).

Exhibits 3 and 4 show that, based on gross measures, racial or ethnic minority homeseekers had a 5- to 30-percent lower probability than White homeseekers of being told that the advertised unit was available. Moreover, minority homeseekers inspected fewer housing units than did their White teammates from 13 to 46 percent of the time. In sales tests, minority homebuyers had 16 to 29 percent lower probability of receiving financial help. The net measures indicate substantially lower levels of discrimination in both outcomes. Indeed, several of the results are not significantly different from zero. Nevertheless, the net measure indicates significant discrimination, as high as 19 percent, in several other cases. Exhibits 3 and 4 also show that the incidence of discrimination tends to be somewhat higher against African-American homeseekers¹⁸ than against Hispanic homeseekers. Moreover, gross measures of discrimination tend to be higher in sales tests than in rental tests, with the notable exception of the availability of the advertised unit in 1977.

Exhibits 5 and 6 present the trends in rental and sales discrimination, respectively, using gross measures. In general, housing discrimination on the outcomes in exhibits 3 and 4 has declined

Exhibit 5

Trends in Rental Discrimination

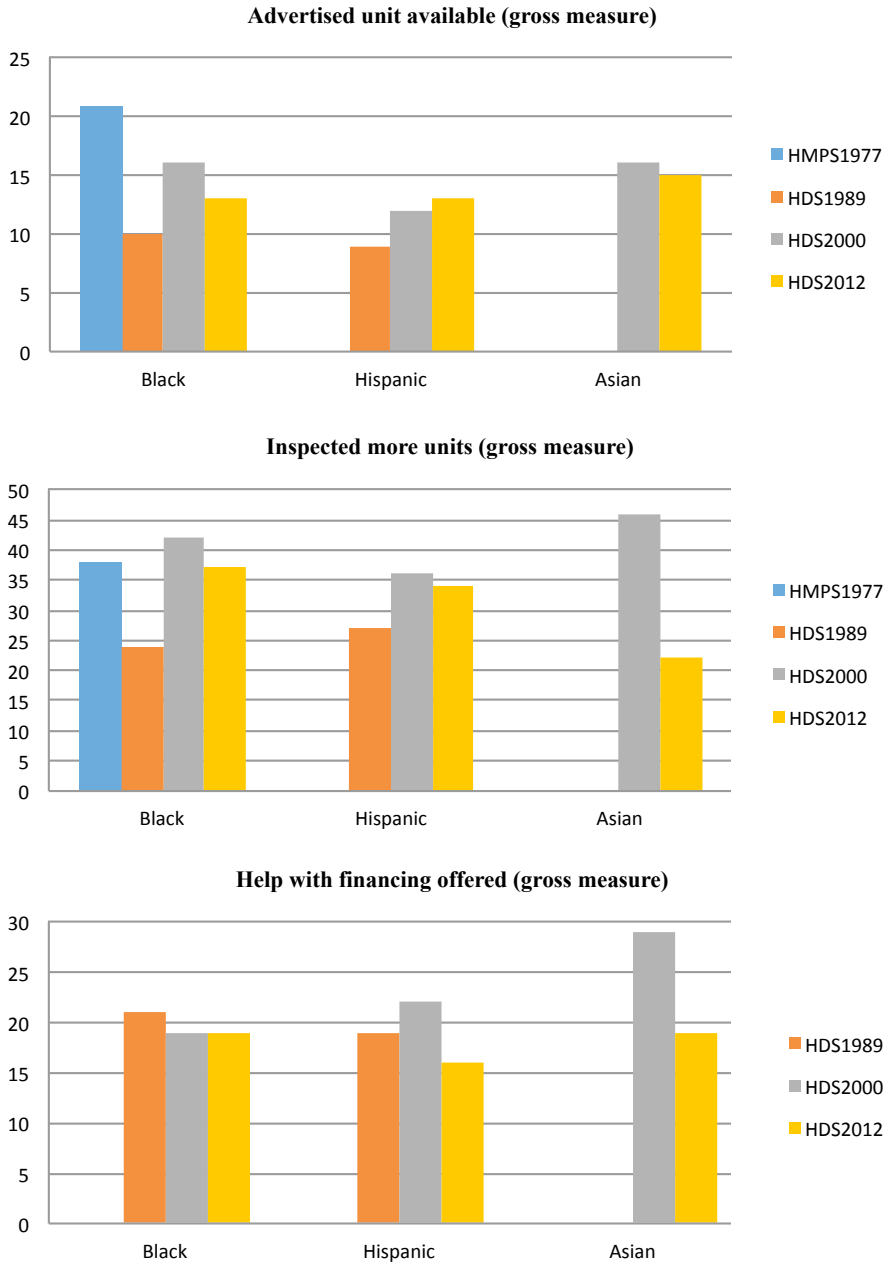


HDS = Housing Discrimination Study. HMPS = Housing Market Practices Survey.

¹⁸ This article uses *African-American* and *Black* as synonyms.

Exhibit 6

Trends in Sales Discrimination



HDS = Housing Discrimination Study, HMPS = Housing Market Practices Survey.

over time. In 1977, Black homeseekers were frequently denied access to advertised units that were available to equally qualified White homeseekers. For instance, one in three Black renters and one in every five Black homebuyers were told that there were no homes available in 1977 (Wienk et al., 1979). In 2012, however, minority renters or homebuyers who called to inquire about advertised homes or apartments were rarely denied appointments that their White counterparts were able to make (Turner et al., 2013). The decline in discrimination is more apparent in rental tests than sales tests and is larger for Black homeseekers than for Hispanic or Asian homeseekers. The decline also differs across outcomes. In both the rental and sales tests, for example, differential treatment in the number of inspected units has not declined very much. In the rental tests, discrimination against Asian homeseekers has increased for the same outcome. Moreover, for the financial-help-offered outcome, no clear evidence indicates that the discrimination has declined over time. Although the most blatant forms of housing discrimination (such as refusing to show the advertised unit) have declined since the first national audit study in 1977, housing opportunities for minority homeseekers are still limited in significant ways.

Racial and Ethnic Steering

Steering occurs when the characteristics of the neighborhoods in which a homeseeker is shown houses depend on the homeseeker's race or ethnicity. Black homeseekers, for example, may be steered away from affluent, predominantly White neighborhoods and instead offered housing in neighborhoods where the residents are largely Black, integrated, relatively poor, or a combination of the three, and White homeseekers may be steered away from neighborhoods where a significant number of Black families reside.¹⁹ This outcome could reflect the customer-prejudice hypothesis (if agents are trying to avoid upsetting their White customers) or the statistical-discrimination hypothesis (if agents are trying to please customers based on stereotypes about their preferences). Steering is difficult for individual homebuyers to detect. One central objective of HDS was to measure steering, which contributes to residential segregation. Because it is time consuming and expensive for auditors to visit a large number of houses, HDS auditors were instructed to obtain the addresses of as many houses as possible, by asking the agent to recommend houses that they might visit together at another time or that the auditor might drive by to determine their suitability (Turner, Mikelsons, and Edwards, 1990). Steering analysis compares the average characteristics of neighborhoods where houses were shown or recommended to minority and White auditors.²⁰

Exhibit 7 presents estimates of steering from the three HDS studies for houses recommended and houses inspected.²¹ Racial steering is defined to exist if, compared to the White auditor in the same audit, the minority auditor is recommended or shown houses in neighborhoods where the percentage of the population that is White is lower. As exhibit 7 illustrates, each HDS found evidence of steering. The gross estimates of steering in this exhibit range from 4 to 26 percent, and the net measures for both houses recommended and houses inspected are statistically significant for

¹⁹ The HDS2012 report used seven summary outcome measures: (1) differential denial of in-person meeting, (2) differential denial of available units, (3) differential number of units recommended, (4) differential number of units shown, (5) differences between testers in agent helpfulness, (6) differential rent or sales price, and (7) differential neighborhood racial/ethnic composition.

²⁰ Steering can be also analyzed at various geographical levels such as municipalities and school districts (see Galster and Godfrey, 2005).

²¹ See Galster (1990b) for a review of steering results from earlier small-scale tests.

Black homeseekers in 2000 and 2012. The net measure for houses inspected is also significant for Hispanic homeseekers in 2000. Exhibit 8 illustrates the trends in steering based on the incidence of steering in exhibit 7. This chart shows that the incidence of steering has become larger over time. These results indicate that steering plays a role in the overall pattern of unfavorable treatment in the housing market. Despite the clear evidence of steering, the HDS studies also found that the composition of neighborhoods recommended to minority homebuyers is similar to the composition of those recommended to equally qualified White buyers (Turner, Struyk, and Yinger, 1991; Turner, Ross, and Galster, 2002; Turner et al., 2013). This apparent contradiction arises because the differences in neighborhood ethnic composition between teammates are small in magnitude; most of the houses shown and recommended to both minority and majority homeseekers were located in predominantly White neighborhoods. Houses for sale in minority-integrated neighborhoods are underrepresented among advertisements in major metropolitan newspapers (Galster, Freiberg, and Houk, 1987; Newburger, 1995; Turner, 1992), and, consequently, these neighborhoods are underrepresented in the HDS sample. Thus, results of the HDS studies reflect the incidence of steering in only one segment of the market.

Several other studies also shed some light on steering. Using HDS2000 data, Ondrich, Ross, and Yinger (2003) found less discrimination in suburban integrated areas than in White areas. Using HDS1989 and HDS2000 data, Galster and Godfrey (2005) found that both Black and Hispanic customers have a significant chance of encountering steering in 2000, particularly in the form of negative comments about minority neighborhoods. According to Galster and Godfrey (2005), Black customers are more likely to encounter steering in 2000 than in 1989.

Exhibit 7

Steering Evidence From Housing Discrimination Studies

Minority	Study	Houses Recommended in Whiter Tracts		Houses Inspected in Whiter Tracts	
		White Favored (%)	Net Measure (%)	White Favored (%)	Net Measure (%)
Black	HDS1989	6	-6	4	-2
	HDS2000	16	4*	11	4*
	HDS2012	25	8*	21	5*
		Non-Hispanic White Favored (%)	Net Measure (%)	Non-Hispanic White Favored (%)	Net Measure (%)
Hispanic	HDS1989	12	4	7	2
	HDS2000	17	2	15	5*
	HDS2012	23	2	26	5

HDS = Housing Discrimination Study.

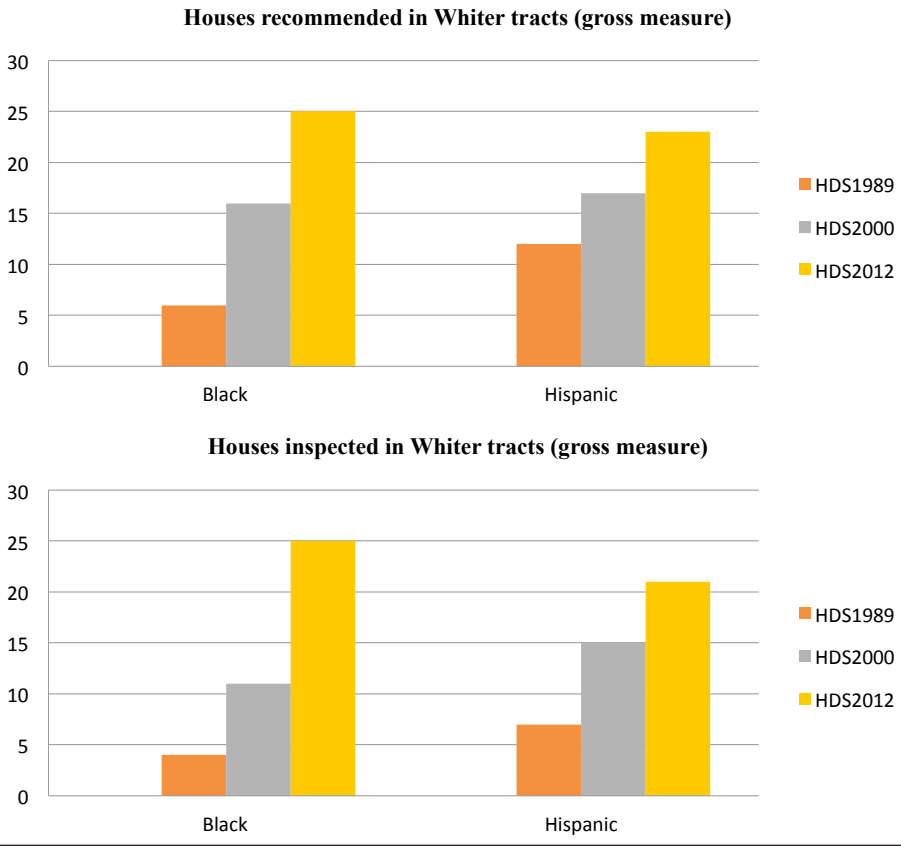
* Indicates statistical significance for net measures at the 90-, 95-, or 99-percent level.

Notes: The HDS1989 report on steering (Turner, Mikelsons, and Edwards, 1990) considered only a difference in percent White of more than 5 percentage points as discrimination, so the estimates are drawn from the HDS2000 report, which used the same measurement for its analysis of both HDS1989 and HDS2000. Gross estimates (percent White favored) are by definition statistically significant. Statistical significance of HDS1989 net measures is not available due to the lack of data.

Sources: Estimates of HDS1989 and HDS2000 are from exhibits 3-13 and 3-19 of Turner et al. (2002); estimates of HDS2012 are from exhibits IV-19 and IV-24 of Turner et al. (2013)

Exhibit 8

Trends in Steering



HDS = Housing Discrimination Study.

The Causes of Discrimination

Exhibit 9 summarizes how audit characteristics would affect discrimination based on the three hypotheses about the cause of discrimination discussed in the section “Testing Hypotheses About the Causes of Discrimination” and the findings of eight studies presented in the following paragraphs.²² The characteristics tested for each hypothesis are not always mutually exclusive, and different hypotheses may predict the opposite effect of the same characteristic on discrimination. In addition, several audit characteristics interact with other characteristics in a complex way, and some predictions of the agent-prejudice and customer-prejudice hypotheses cannot be separated. For instance, either housing agents or their White customers may have stronger prejudice against younger minority homeseekers than against older minority homeseekers (Choi, Ondrich, and Yinger, 2005).

²² See also table 3 of Zhao (2005) and exhibit 1 of Choi, Ondrich, and Yinger (2008).

Exhibit 9

Predictions and Findings of Discrimination by Causal Hypotheses

Characteristics	Agent-Prejudice Hypothesis	Customer-Prejudice Hypothesis	Statistical-Discrimination Hypothesis
Agent	Minority agent of the same race (-) (Choi, Ondrich, and Yinger, 2008, 2005; Zhao, 2005; Zhao, Ondrich, and Yinger, 2006) Age of agent (+) (Choi, Ondrich, and Yinger, 2005; Ondrich, Stricker, and Yinger, 1999) Male agent (+) (Choi, Ondrich, and Yinger, 2005)	Size of real estate agency (-) (Ondrich, Ross, and Yinger, 2003; Ondrich, Stricker, and Yinger, 1999, 1998) Use of multi real estate listing service (-) (Zhao, Ondrich, and Yinger, 2006)	Size of real estate agency (-) (Ondrich, Ross, and Yinger, 2003; Ondrich, Stricker, and Yinger, 1999, 1998)
Auditor	Age of auditor (-) (Ondrich, Stricker, and Yinger, 1999) Male auditor (+) (Choi, Ondrich, and Yinger, 2005; Zhao, Ondrich, and Yinger, 2006)	Age of auditor (-) (Ondrich, Stricker, and Yinger, 1999) Male auditor (+) (Choi, Ondrich, and Yinger, 2005; Zhao, Ondrich, and Yinger, 2006) Assigned marital status (-) (Ondrich, Ross, and Yinger, 2003) Assigned income (-) (Page, 1995; Zhao, 2005)	Assigned income (-) (Page, 1995; Zhao, 2005) Asked housing value (+) (Page, 1995)
Neighborhood		Percent owner-occupied house (+) (Ondrich, Stricker, and Yinger, 1999; Zhao, 2005) Percent White residents before tipping (+) (Ondrich, Ross, and Yinger, 2003; Ondrich, Stricker, and Yinger, 1999; Page, 1995; Zhao, Ondrich, and Yinger, 2006)	Percent White residents before tipping (+) (Ondrich, Ross, and Yinger, 2003; Ondrich, Stricker, and Yinger, 1999; Page, 1995; Zhao, Ondrich, and Yinger, 2006) Neighborhood housing value (+) (Choi, Ondrich, and Yinger, 2005; Ondrich, Ross, and Yinger, 2003; Zhao, 2005)

(+) Indicates that discrimination against a minority is positively correlated with the factor (for indicator variables, more discrimination if the factor is applicable).

(-) Indicates that discrimination against a minority is negatively correlated with the factor (for indicator variables, less discrimination if the factor is applicable).

Note: The findings that support the predictions of each hypothesis are based on the advertised-unit-available/inspected, similar-unit-inspected, or the number-of-houses-shown outcomes of Black-White audits.

Based on those predictions, eight studies examined the causes of housing discrimination using HDS data (Choi, Ondrich, and Yinger, 2008, 2005; Ondrich, Ross, and Yinger, 2003; Ondrich, Stricker, and Yinger, 1999, 1998; Page, 1995; Zhao, 2005; Zhao, Ondrich, and Yinger, 2006).²³ Except for Ondrich, Ross, and Yinger (2003), these studies use an audit pair as the unit of analysis and use an audit fixed-effects model to control for the fact that audit teammates share values of unobservable variables (Yinger, 1986).²⁴ In addition, based on multivariate analysis, HDS2012 examined potential contributions of audit characteristics to differences in the number of housing units shown.²⁵

Several results of the eight studies support the agent-prejudice hypothesis. Unless otherwise indicated, this article focuses on the results of Black-White discrimination.²⁶ For the advertised-unit-inspected or similar-unit-inspected outcome, results show more discrimination by older agents (Choi, Ondrich, and Yinger, 2005; Ondrich, Stricker, and Yinger, 1999), less discrimination by female agents (Choi, Ondrich, and Yinger, 2005), and less discrimination against female auditors (Choi, Ondrich, and Yinger, 2005; Zhao, Ondrich, and Yinger, 2006; HDS2012) and older auditors (Ondrich, Stricker, and Yinger, 1999). Results concerning the effect of the agent's race on discrimination are inconsistent. Ondrich, Stricker, and Yinger (1999, 1998) found more discrimination against Black homeseekers when the agent is Black, whereas Zhao (2005), Choi, Ondrich, and Yinger (2008, 2005), Zhao, Ondrich, and Yinger (2006), and HDS2012 (only sales tests) found less discrimination when the agent is Black. In the sales market, however, Black agents are rare. According to HDS2012, testers met with a Black agent in only 5 percent of the sales tests.

The customer-prejudice hypothesis suggests that discrimination is likely to increase with the assigned income of auditors, the percent owner-occupied housing units in the neighborhood, and the share of White residents in the White-majority neighborhood. Some results support these predictions. For the advertised-unit-inspected outcome and the number-of-houses-shown outcome, Black homebuyers face less discrimination in the neighborhoods with a significant share of Black residents (Zhao, 2005) but more discrimination in the neighborhoods with a higher percentage of owner-occupied housing units (Ondrich, Stricker, and Yinger, 1999; Zhao, 2005). Page (1995) and Zhao (2005) found that Black homeseekers with higher incomes encounter less discrimination in the number of houses shown; however, the effect of having a white-collar job or higher education on discrimination is unclear. Ahmed, Andersson, and Hammarstedt (2010); Bosch, Carnero, and Farre (2010); and Carlsson and Eriksson (2014) found no effect of positive information on discrimination, but Baldini and Federici (2011) and Hanson and Hawley (2011) found that minority homeseekers receive more e-mail responses when they reveal positive information. Finally,

²³ Choi, Ondrich, and Yinger (2008, 2005) and Ondrich, Stricker, and Yinger (1999) examined discrimination in rental housing; Ondrich, Ross, and Yinger (2003), Ondrich, Stricker, and Yinger (1998), Zhao (2005), and Zhao, Ondrich, and Yinger (2006) examined discrimination in sales housing; and Page (1995) examined both.

²⁴ A significant advance of Ondrich, Ross, and Yinger (2003) is that, based on a random-effect multinomial logit model, they used a housing unit as a unit of analysis to avoid an endogeneity problem in explanatory variables that are influenced by agent choices. In addition, Choi, Ondrich, and Yinger (2005), Zhao (2005), and Zhao, Ondrich, and Yinger (2006), using HDS2000 data, accounted for auditors' actual characteristics to overcome potential bias from unmatched characteristics of audit pairs.

²⁵ These multivariate analyses are based on ordinary least squares regressions using "tests in which both teammates met with an agent" (HDS2012: 34). Because of this sample-selection strategy, the results of these analyses provide little insight into hypotheses about the causes of discrimination.

²⁶ Less consistent patterns emerge when comparing results across different minority groups.

Ondrich, Ross, and Yinger (2003) found that real estate agents discriminate more against higher income Black customers, and Carlsson and Eriksson (2013) found that racial discrimination is higher for ethnic minority homeseekers with a high-skill job than those with a low-skill job.

Some studies also found more discrimination against Black homeseekers in high-value neighborhoods (Choi, Ondrich, and Yinger, 2005; Ondrich, Ross, and Yinger, 2003). For the advertised-unit-inspected outcome in sales audits, Ondrich, Stricker, and Yinger (1999) found that discrimination decreases as the ratio of assigned auditor income to housing value increases. Several results related to *tipping*, defined as the rapid exit of White residents from a neighborhood once the minority composition of the neighborhood reaches a certain point, also support the customer-prejudice hypothesis. Page (1995) and Ondrich, Stricker, and Yinger (1999) found that discrimination increases as the percentage of minority representation approaches a neighborhood tipping point and decreases when the percentage exceeds the tipping point. Moreover, Ondrich, Ross, and Yinger (2003) found that discrimination is relatively high in central city integrated areas, which are the ones most likely to be threatened with tipping. Zhao, Ondrich, and Yinger (2006) also found that Black homebuyers are less likely to encounter discrimination in Hispanic neighborhoods.

The customer-prejudice hypothesis also predicts that larger real estate agencies, which are less dependent on a particular neighborhood for their business, are less likely to discriminate. Ondrich, Stricker, and Yinger (1998) found that discrimination is less likely in larger agencies for the advertised-unit-inspected outcome in rental audits, and Ondrich, Ross, and Yinger (2003) and Ondrich, Stricker, and Yinger (1999) found similar results in sales audits. By contrast, Choi, Ondrich, and Yinger (2008) and HDS2012 found the opposite result: larger agencies discriminate more against minority customers than do smaller agencies (for the advertised-unit-inspected, the number-of-units-inspected, or rental-incentive-provided outcomes). Finally, Zhao, Ondrich, and Yinger (2006) found that real estate agents who use the Internet, which allows them to steer Black customers away from prejudiced White neighborhoods, are less likely to discriminate in terms of the number of units recommended.

Ondrich, Ross, and Yinger (2003) found strong evidence for statistical discrimination. They found that agents' marketing efforts increase with asking price for White homeseekers but not for Black homeseekers, an implication that is consistent with the hypothesis that agents practice statistical discrimination based on a preconception about the ability of Black customers to purchase expensive homes. They also found that Black customers, but not White customers, are shown units that are cheaper than the advertised unit that is the basis for the audit. Ondrich, Ross, and Yinger (2003) and Zhao (2005) also found that real estate agents discriminate more in neighborhoods with higher house values, even controlling for the value of the houses being shown. In addition, Page (1995) found that Black homeseekers encounter more discrimination when inquiring about more expensive houses, and a similar finding appears in Choi, Ondrich, and Yinger (2005). Choi, Ondrich, and Yinger (2005) found a positive, significant coefficient for median house value in the number-of-units-inspected regression but concluded that this result is also consistent with the customer-prejudice hypothesis. Ondrich, Stricker, and Yinger (1998), however, found that discrimination is not significantly different when the advertised unit is in an integrated or a White neighborhood. Page (1995) and Ondrich, Ross, and Yinger (2003) found less discrimination in integrated areas than in White areas in the sales markets, which is consistent with the view that

agents try to maximize the chances of a successful match by making race-based assumptions about a customer's preference. Choi, Ondrich, and Yinger (2008), however, found no evidence to support the statistical-discrimination hypothesis in the Black-White audits.

Ewens, Tomlin, and Wang (2014) designed correspondence audits of rental housing to look for landlord prejudice and statistical discrimination. If landlord prejudice is at work, they argued, then discrimination should be lower in largely Black than in largely White neighborhoods. Their results do not support this prediction. They also argued that landlords practicing statistical discrimination will find explicit signals about the “quality” of an applicant from a particular racial or ethnic group to be more believable if they have more experience dealing with tenants from that group. If so, landlords will respond more favorably to positive information (such as an indication that the tenant is a nonsmoker or has a desirable job) about a Black applicant in a largely Black neighborhood than in a largely White neighborhood. Their empirical results support this prediction and, therefore, support the conclusion that some landlords practice statistical discrimination.

Finally, as Page (1995) recognized, Yinger (1995) shows that the level of discrimination depends on a broker's opportunities to discriminate, defined as his access to available housing units. With controls for the opportunity to discriminate (an agent's available units), Yinger (1995) found evidence that real estate brokers discriminate to protect their business with prejudiced White customers and on the basis of stereotypes about Black and Hispanic customers. These results indicate that the causes of discrimination in rental and sales housing are complex. The strongest results from Ondrich, Ross, and Yinger (2003) support statistical discrimination, but both the prejudice of agents and their responses to the prejudice of their White customers also appear to be at work in some cases.

Housing discrimination is a complex social phenomenon, and its causes may differ over time and place. Existing studies provide some evidence to support the hypotheses that agent and customer prejudice can lead to discrimination, but this type of evidence does not appear in most audit studies. These findings suggest that these hypotheses cannot fully explain the amount of discrimination observed in audit studies, although, to some degree, they may also reflect the limitations of the hypothesis tests that are possible with audit data. Two studies that appear to have relatively compelling methods, Ondrich, Ross, and Yinger (2003) and Ewens, Tomlin, and Wang (2014), both found strong evidence that housing discrimination is sometimes based on a housing provider's perceptions about the likelihood of a successful transaction with customers from different racial or ethnic groups, which is a form of statistical discrimination. These findings should be of great interest to policymakers. Because statistical discrimination arises as an illegal way for a housing provider to maximize profits based on stereotypes, enforcement agencies need to use audits and other methods to ensure that the costs of discrimination are higher than the benefits—at least for housing providers who might otherwise break the law.

Paired Testing and Public Policy

Studies that measure the incidence of discrimination and provide evidence about discrimination's causes obviously are relevant for fair housing policy. This section explores the links between paired testing and fair-housing-enforcement activities and also briefly surveys the use of the paired-testing method in markets other than housing.

Paired Testing and Fair Housing Enforcement

The audit methodology and fair housing enforcement have evolved together and are connected to each other in three important ways. The foundation of the first connection is a provision in the 1968 Fair Housing Act that gives private, nonprofit fair housing agencies legal standing to bring court cases against alleged discriminators. Fair housing audits were developed in the 1950s by partnerships between community groups and scholars who wanted to highlight the extent of discrimination (see appendix B). After local, state, and federal fair housing laws were implemented, starting in 1958 in New York City, however, community groups interested in combatting discrimination quickly figured out that audits could also be used for enforcement purposes.²⁷ An enforcement audit by a private agency typically begins with a complaint about a given housing provider. The agency then conducts one or more audits to determine whether the alleged discrimination exists. When the audits are carefully conducted and the legal requirements for a fair housing lawsuit are met, audit evidence, even from a single audit, can provide compelling evidence about the existence of discrimination and, if it exists, about the form it takes.²⁸ In recognition of the important role that private organizations played in combatting discrimination, Congress passed the Fair Housing Initiatives Program (FHIP) under the Housing and Community Development Act of 1987, which became effective in 1988. This program, which provides federal funding for the auditing and other activities of these agencies, became permanent in 1993.

As of 2011, 98 private nonprofit agencies were engaged in fair housing enforcement.²⁹ In 2006 alone, these agencies conducted more than 5,000 tests (Temkin, McCracken, and Liban, 2011). Since shortly after the Fair Housing Act was passed, these agencies have used tests to establish discrimination and to obtain settlements in hundreds of cases.³⁰ In almost all cases, fair housing organizations obtain injunctive relief that includes a change in behavior, policies, or both; training to prevent future discrimination; and monitoring to ensure compliance with the Fair Housing Act. Some examples follow.

- In 2013 and 2014, based on complaints, the Fair Housing Justice Center (FHJC) conducted audits in rental housing in the Woodlawn neighborhood of the Bronx (J.J.A Holding Corporation). The FHJC and three African-American testers alleged that J.J.A Holdings engaged in racially discriminatory rental practices: Among other things, an agent told African-American testers that no apartments were available while showing apartments to White testers on the same day. In 2015, J.J.A Holdings agreed to change rental practices and pay the plaintiffs \$200,000 for damages and attorneys' fees (Gorman, 2015).
- In 2013, Latino homeseekers did not receive a rental application from apartment managers of Bailey Properties in Arkansas (or did so after significant delay), whereas prospective White

²⁷ Private fair housing groups also learned to use audits to shed light on the nature and extent of discrimination in their service area. Galster (1990a) reviewed 71 audit studies that private fair housing groups conducted in the 1980s.

²⁸ The typical prima facie elements for establishing a fair housing case are (1) the plaintiff was eligible for the unit available, (2) the plaintiff was denied the unit or the housing provider refused to negotiate, (3) the plaintiff is a member of a legally protected category of persons, and (4) the housing opportunity remained available (*Smith v. Anchor Building Corp.* 8th Circuit 1976. 536 F2d 231).

²⁹ FHIP also provides some funds for fair lending enforcement activities by these agencies.

³⁰ See Freiberg (1993), Fudge (2014), Schwemm (2014, 1992), and Yinger (1995).

renters promptly received documents. The Arkansas Fair Housing Commission confirmed the discrimination against Hispanic renters based on six correspondence audits conducted by the National Fair Housing Alliance (NFHA, 2014).

- In 2009, an African-American couple was told no units were available at Geneva Terrace in La Crosse, Wisconsin, but their White friend who called the rental office later was told that units were available. The couple called the office again in 15 minutes but was again told that no units were available. This refusal to rent to Black homeseekers was confirmed in two audits by Metropolitan Milwaukee Fair Housing Council. The couple received \$47,500 in damages, and the owners of Geneva Terrace were required to complete fair housing training (HUD, 2014).

The second connection between audits and fair housing enforcement is that the results of research audits and enforcement audits have provided powerful evidence in support of continued or expanded fair housing enforcement. Most importantly, perhaps, the audit results from HMPS were highlighted in congressional testimony about the FHAA, which greatly expanded the federal government's powers to enforce fair housing laws.

Testimony by HUD's general council,³¹ for example, began by citing the HMPS results but also explained that "HUD staff extrapolated from those findings to conclude that about 2 million instances of housing discrimination occurred every year. Experts in statistical methodology may quibble over that extrapolation, but even if the estimate is wrong by half, it is nonetheless staggering and, to put it mildly, deeply disconcerting." Testimony before the same subcommittee by the director of the Kentucky Commission on Human Rights³² cited not only the HMPS results for his state but also results from audits conducted by his agency in 1977 and 1985. In addition, several articles in professional journals cited evidence from HMPS in building a case for passage of the FHAA (James and Crow, 1986; Rice, 1984).³³ This influence of research audits on enforcement policies is well summarized in the article by Freiberg (1993: 230), which states that, "Data showing widespread patterns of unlawful housing discrimination [from HMPS1977 and HDS1989] understandably evoke a response from well-meaning policymakers for more vigorous enforcement of fair housing laws."

The HMPS results also appear to have been influential in the 1987 passage of FHIP and in the 1984 passage of the Fair Housing Assistance Program (FHAP), which is discussed in the following paragraphs. See, for example, the testimony on FHIP by HUD's general council,³⁴ which includes the quotation in the previous paragraph from his FHAA testimony. This hearing concerned

³¹ Knapp, John J. 1986b. Statement. In *Hearing Before the Subcommittee on Civil and Constitutional Rights of the Committee on the Judiciary, House of Representatives, Ninety-Ninth Congress, Second Session, on H.R. 4119, Fair Housing Amendments Act*, Serial No. 120, July 17. Washington, DC: U.S. Government Printing Office: 73–99.

³² Martin, Galen. 1986b. Statement. In *Hearing Before the Subcommittee on Housing and Urban Affairs of the Committee on Banking, Housing, and Urban Affairs, United States Senate, Ninety-Ninth Congress, Second Session, on Proposed Guidelines for Fair Housing Testing Included in Section 602 of S. 2507*, July 18. Washington, DC: U.S. Government Printing Office: 122–172.

³³ The James and Crow article (1986) was also incorporated into the record of the congressional hearings on FHAA in 1986.

³⁴ Knapp, John J. 1986a. Statement. In *Hearing Before the Subcommittee on Housing and Urban Affairs of the Committee on Banking, Housing, and Urban Affairs, United States Senate, Ninety-Ninth Congress, Second Session, on Proposed Guidelines for Fair Housing Testing Included in Section 602 of S. 2507*, June 18. Washington, DC: U.S. Government Printing Office: 2–10.

amendments to restrict testing with FHIP funds. At that hearing, Martin³⁵ explicitly argued that testing by private fair housing groups in Kentucky had been very effective and that there was a “need for funding especially for private fair housing groups... without the restriction of the amendments.”

In more recent years, many of the scholarly publications presenting results from HDS or other audit studies conclude by citing the need for continued enforcement of fair housing laws, including actions by HUD and the U.S. Department of Justice (DOJ), and continued funding for FHIP and FHAP. See, for example, Galster and Godfrey (2005), Ross and Turner (2005), Yinger (1995), and Zhao, Ondrich, and Yinger (2006).³⁶

The third audit-enforcement connection is that, shortly after the passage of fair housing laws, audits became a standard tool in the efforts of governmental fair housing enforcement agencies at all levels of government. Audits provide compelling evidence about discriminatory behavior, and a well-publicized audit program may encourage housing providers to be more careful to meet their obligations under the Fair Housing Act.³⁷ Although HUD and DOJ did not conduct their own enforcement audits for many years after the passage of the Fair Housing Act, DOJ relied on audits conducted by private fair housing groups as early as 1972 (Lee, 1999; Schwemm, 1992). Moreover, the Comptroller General of the United States (CGUS) reported in 1978 that “HUD does use testing data developed by local fair housing organization” (CGUS, 1978: 26). At the state and local levels, FHAP provides funding to governmental fair housing agencies with antidiscrimination legislation that is substantially equivalent to federal law. Some of this funding is used to contract with private nonprofit fair housing organizations to carry out audits. At the federal level, FHAA in 1988 expanded the enforcement powers of DOJ and HUD, and explicitly gave HUD the power to investigate cases of possible discrimination using audits and other techniques, with or without a complaint from a homeseeker.

DOJ started the Fair Housing Testing Program in 1992. Based on its experience with this program, DOJ has come to the conclusion that “testing can be a valuable tool to investigate housing market practices and to document illegal housing discrimination” (DOJ, 2014). Since it started its testing program, DOJ has filed 98 pattern and practice testing cases with evidence directly generated from the Fair Housing Testing Program (DOJ, 2014). The vast majority of testing cases filed to date are based on testing evidence that involved allegations of agents misrepresenting the availability of rental units or offering different terms and conditions based on race, national origin, familial status, or a combination of the three. From the 96 resolved cases, DOJ has recovered more than \$12.9 million, including more than \$2.3 million in civil penalties and more than \$10.5 million in other damages (DOJ, 2014). Most of these cases also call for changes in the defendants’ behavior to prevent discrimination in the future. For example—

- In response to a complaint that a corporate owner and leasing agent discriminated based on race, in 2013, DOJ conducted a series of three tests at Baldwin Commons in Pittsburgh,

³⁵ Martin, Galen. 1986a. Statement. In *Hearing Before the Subcommittee on Housing and Urban Affairs of the Committee on Banking, Housing, and Urban Affairs, United States Senate, Ninety-Ninth Congress, Second Session, on Proposed Guidelines for Fair Housing Testing Included in Section 602 of S. 2507*, June 18. Washington, DC: U.S. Government Printing Office: 17–27.

³⁶ Many of these scholars also point out that audits could increase the effectiveness of fair housing enforcement efforts by helping enforcement officials identify the circumstances under which discrimination is most likely to occur.

³⁷ Ross and Galster (2007) provide some preliminary evidence that an active fair housing enforcement program deters housing providers from engaging in discrimination.

Pennsylvania. The tests found that White testers were shown apartments and were offered the opportunity to rent them but Black testers were told that the same apartments were unavailable. The court entered a consent decree in *United States v. S-2 Properties, Inc.* (2014), and the defendants will pay a civil penalty to the United States of \$15,000, develop and maintain nondiscrimination housing policies, and attend fair housing training (DOJ, 2014).

- Based on the complaint that these defendants discriminated against Black customers, paired tests were conducted by DOJ. These tests found that Somali testers were told to make appointments to see apartments, whereas White testers were shown apartments when they walked in. The consent decree in *United States v. Highland Management Group, Inc.* (2013) contains injunctive relief and civil penalties of \$30,000 (DOJ, 2014).

Despite the powers given to it by FHAA, HUD does not frequently conduct fair housing audits itself. Instead, HUD has undertaken enforcement actions in partnership with organizations that conduct audits. In 2012, for example, HUD brought a case against Peachtree Apartments in Clanton, Alabama, on the grounds that the owners discriminated against tenants based on national origin. The Central Alabama Fair Housing Center conducted audits and found that Peachtree Apartments required prospective Hispanic tenants to provide documentation of their immigration status while not asking the same of non-Hispanic individuals. As a result of HUD's actions, the owner of these apartments voluntarily entered into a settlement agreement that requires nondiscriminatory admission policies, a plan to market housing opportunities to populations with limited English proficiency, and the provision of translation services and fair housing training to its employees and contractors (HUD, 2012).

Overall, therefore, audits have become a crucial tool in the fair housing enforcement system, and audit results have provided support for improvements in and continued support for this system. Although the evidence reviewed in this article indicates that some key forms of housing discrimination have declined over time, this evidence also indicates that a significant amount of housing discrimination remains and that a few forms of discrimination have actually increased. Adjustments in the nature or location of audits may be called for, such as an increase in the use of correspondence audits for enforcement purposes. It is clear, however, that housing discrimination has by no means gone away and that fair housing audits for both research and enforcement will be needed in the future.

Paired Testing in Other Markets

Audits are sufficiently advanced to conduct national-level tests in the areas of housing sales and rentals. In-person audits, in which individuals are matched for all relevant characteristics other than the one that is expected to lead to discrimination (for example, race or ethnicity), have also been used in several other markets, including entry-level hiring, inquiries about home mortgages, house insurance, car sales, and selected areas of public accommodations such as taxi service.³⁸ The first in-person and correspondence audits to measure hiring discrimination were conducted in Britain (Daniel, 1968; Jowell and Prescott-Clarke, 1970). In the United States, the Urban Institute conducted the first in-person audits of hiring discrimination against Hispanic men applying for entry-level jobs in Chicago and San Diego in 1989 (Cross et al., 1989). Several in-person and correspondence tests of hiring discrimination against minority groups have also been conducted in the

³⁸ See Pager (2007) and Pager and Shepherd (2008) for a review of the literature.

United States and European countries since then (Bendick, Jackson, and Reinoso, 1994; Bendick et al., 1991; Bertrand and Mullainathan, 2004; Bursell, 2007; Carlsson and Rooth, 2012, 2007; Goldberg, Mourinho, and Kulke, 1995; Kaas and Manger, 2011; Nunley et al., 2014; Pager, 2003; Pager, Western, and Bonikowski, 2009; Turner, Fix, and Struyk, 1991; Wood et al., 2009).³⁹ In addition, audits have been used to examine discrimination in automobile sales (Ayres, 1991; Ayres and Siegelman, 1995), taxicab service (Ridley, Bayton, and Outtz, 1989), home mortgage inquiries (Galster, 1993; Turner and Skidmore, 1999), homeowners' insurance (Wissoker, Zimmermann, and Galster, 1998), and the provision of medical care (Schulman et al., 1999).

Audits have also been used to study discrimination in shopping (Gneezy and List, 2004; Gneezy, List, and Price, 2012; List, 2004; Zussman, 2013) and beverage service (Perry, 2005). The economic costs of discrimination in these everyday commercial transactions are undoubtedly smaller than the costs of discrimination in employment or housing, but these costs, in the form of higher prices, additional waiting time or hassle, or psychological issues, may be significant. Correspondence audits were recently used for studying certain types of commercial transactions over the Internet (Doleac and Stein, 2013; Nunley, Owens, and Howard, 2011). The possibilities for the use of this method to study discrimination have certainly not been exhausted.

Conclusion

This article reviews the results of audit studies in housing markets. The audit methodology has been widely used in the United States and many European countries to measure the incidence of discrimination in housing markets. The number of audits conducted and the types of behavior examined vary significantly across the studies that we have reviewed. Some early studies conducted fewer than 100 tests in a single city, for example, whereas the 2012 Housing Discrimination Study conducted more than 8,000 tests in 28 metropolitan areas. With the rapid growth in the use of the Internet for marketing housing, many recent audit studies have used e-mail correspondence. This approach has been widely used, for example, to study housing discrimination in European countries when immigration has introduced new ethnic divisions. Despite their variation in methodology and social context, housing audit studies consistently find that racial and ethnic minority homeseekers experience unfavorable treatment compared with racial and ethnic majority homeseekers. In terms of trends, the four national audit studies in the United States found that housing discrimination had declined over time in some important types of agent behavior, such as making an advertised apartment available to a customer. Discrimination against Black and Hispanic homeseekers has not declined very much in some other types of agent behavior, however, and the steering of Black homeseekers away from White neighborhoods appears to have increased over time.

Paired audits offer a uniquely effective tool for directly observing differential treatment of equally qualified homeseekers. Because of their narrative power, these audits provide compelling evidence about discrimination for educating the public, for influencing fair housing policy, and for providing evidence in court. Nevertheless, audits also have some limitations. First, in-person audits are expensive and difficult to manage. Second, audits also observe only the marketing phase of a transaction and may miss discrimination that occurs in housing advertisements or after price negotiations begin,

³⁹ See Riach and Rich (2002) and Rich (2014) for comprehensive reviews of audit studies.

in the search for a mortgage, or, in rental markets, when the terms of the lease are specified. Third, audit studies are based on a sample of advertisements, usually from major metropolitan newspapers or ad-listing websites. These advertisements may not correspond to the actual housing experience of minority groups, who may use other means of identifying available housing or who may not be qualified for a share of this advertised housing. The discrimination actually experienced by minority homeseekers could therefore be higher or lower than the discrimination measured by an audit study. Evidence from HDS2000 addressed these issues by conducting some audits based on posted advertisements and by oversampling advertisements in neighborhoods with a high minority concentration. These steps did not lead to significant changes in measures of discrimination (see Turner et al., 2002).⁴⁰

Despite these limitations, in-person paired testing is still a valuable tool for scholars and public officials who want to shed light on discrimination. Such testing is a proven method that can observe discrimination in many types of behavior involved in a housing market transaction. Even in the Internet age, important components of any such transaction involve face-to-face contact and the resulting possibility of discrimination. Of course, changes in housing markets have also opened the door to correspondence audits, which are less expensive and more precise, but which cannot examine nearly as many types of behavior. Further investigation into the best circumstances for using each of these methods would certainly be warranted.

Another possibility for future research is to combine audits with other types of data and research methods. Linking audit results with survey evidence on prejudice and discrimination could be quite valuable, for example. This type of linkage would make it possible to ask a variety of new questions about discrimination: Does variation across locations in perceived discrimination correspond with variation in discrimination measured with audits? Do audit-based measures of discrimination against a minority group increase in neighborhoods where surveys find White people with relatively high levels of prejudice against that group? A related possibility is to administer surveys to the landlords or real estate brokers involved in an audit study. This step would make it possible to ask whether landlords or housing agents with relatively high prejudice are more likely to be the ones that exhibited discriminatory behavior during the audit study.

Discrimination in housing markets has certainly evolved over the years, and discrimination in some types of housing agent behavior has declined. Nevertheless, paired testing has shown that significant discrimination remains in several important types of agent behavior. Gross measures indicate continuing discrimination in the number of units shown to a customer and offers to help a customer find financing, for example, and net measures signal ongoing discrimination in the number of apartments inspected and in racial steering. As long as this type of behavior continues to occur, paired testing and the new methods that are its descendants will be valuable tools both for scholars who want to measure discrimination and understand its causes and for fair housing enforcement officials who want to protect the housing rights of minority households.

⁴⁰ It is also possible to adjust discrimination measures so that they correspond to the actual income distribution of the minority group under study. Yinger (1995) found that this type of calculation has little impact on the results of HDS1989. Some other research suggests, however, that less discrimination occurs when testers reveal that they have a more professional job (Baldini and Federici, 2011; Ewens, Tomlin, and Wang, 2014; Hanson and Hawley, 2011). This possibility might be more common in actual searches than in audit studies. Minority homeseekers also might avoid housing agents who are known to discriminate. This type of behavior imposes a cost on these households, of course, but it also might lower measured discrimination. For further discussion of this issue, see Ross and Yinger (2006).

Appendix A. Results of Housing Discrimination Audit Studies

Exhibit A-1

Results of Housing Discrimination Audit Studies (1 of 6)

Authors	Data/Methodology ^a	Scale	Other Factors Considered With Race/Ethnicity and Gender	Location, Period, and Market Examined	Main Findings
Ahmed, Andersson, and Hammarstedt (2010)	E-mail correspondence tests Probit model	1,032 units (not paired: one e-mail per unit)	Information on employment, education, marital status, age, smoking behavior, and rent payment issues	Sweden 2008 Rental tests	Arabic males have 15 to 23 percentage points and 17 to 20 percentage points lower probability of receiving a call back and being invited to further contacts than do Swedish males. Providing additional information about themselves increases the probability of receiving a call back, being invited to further contacts, and being invited to showing by 12 to 19 percent for both Swedish and Arabic males. Increasing the amount of information in the application will not reduce discrimination in the housing market.
Ahmed and Hammarstedt (2008)	E-mail correspondence tests Probit model	500 units (500 x 3 = 1,500 e-mails)		Sweden 2007 Rental tests	Arabic males have 21 to 26 and 7 percentage points lower probability of being invited to further contacts and being invited to a showing than do Swedish males, respectively. Swedish males are almost 13 percentage points less likely to be invited to a flat showing than are Swedish females.
Andersson, Jakobsson, and Kotsadam (2012)	E-mail correspondence tests Probit model	950 units (not paired: one e-mail per unit)	Socioeconomic status using occupation	Norway 2009–2010 Rental tests	Arabic renters have 13 percentage points lower probability of receiving a positive response compared with Norwegian renters. For all ethnicities, the probability of receiving a positive response is lower by about 7 percentage points for males than females and by 7 percentage points for warehouse workers than economists.

Exhibit A-1

Results of Housing Discrimination Audit Studies (2 of 6)

Authors	Data/Methodology ^a	Scale	Other Factors Considered With Race/Ethnicity and Gender	Location, Period, and Market Examined	Main Findings
Baldini and Federici (2011)	E-mail correspondence tests Probit model	3,676 units (not paired: one e-mail per unit)	Socioeconomic status using information on occupation and marital status	Italy 2010 Rental tests	Arabic renters have 22 percentage points lower probability of receiving a positive response than native Italian renters. East European renters have 16 percentage points lower probability. The magnitude of discrimination is greater for men (24 percentage points) than for women (15 percentage points). Providing information reduces discrimination for foreign names, especially for males.
Bosch, Carnero, and Farre (2010)	E-mail correspondence tests Probit, linear probability, and unit fixed-effects models Telephone correspondence tests	1,809 units (4,709 e-mails: 2 to 4 e-mails per unit) 201 units (201 x 2 = 402 calls)	Socioeconomic status using occupation	Spain 2009 Rental tests Spain 2008 Rental tests	Moroccan immigrants are 13 to 18 percentage points less likely to receive a response than are native Spanish renters. Discrimination is much higher for Moroccan males (22 percentage points) than for Moroccan females (10 percentage points). Revealing positive information about occupation increases the contacts by about 6 to 8 percentage points. Positive information reduces discrimination only for males. Moroccan renters have around a 10-percentage point lower response rate, and Moroccan males are substantially more discriminated than Moroccan females (15 versus 7 percentage points). Discrimination against Hispanic renters is of a similar magnitude.
Bovenkerk et al. (1979)	In-person tests	135 units (135 x 3 = 405 visits)		France 1976 Rental tests	Black renters experience net discrimination of 31.9 percent compared with native French renters, but there was no discrimination against Portuguese renters.

Exhibit A-1

Results of Housing Discrimination Audit Studies (3 of 6)

Authors	Data/Methodology^a	Scale	Other Factors Considered With Race/Ethnicity and Gender	Location, Period, and Market Examined	Main Findings
Carlsson and Ericksson (2013)	E-mail correspondence tests Linear probability model	5,143 units (not paired: one e-mail per unit)	Socioeconomic status using occupation	United Kingdom 2011–2012 Shared housing (for example, roommate and housemate)	Ethnic minority home-seekers have 13 percentage points lower probability of being invited to an apartment showing than native British (White) renters. Ethnic minority home-seekers with high-skill jobs face 10 percentage points more discrimination than ethnic minority home-seekers with low-skill jobs. Discrimination against ethnic minority home-seekers is highest among Arabic home-seekers (compared with Eastern European, Indian, and African home-seekers).
Carlsson and Ericksson (2014)	E-mail correspondence tests Linear probability model	5,827 units (not paired: one e-mail per unit)	Employment status, age, and socioeconomic status using occupation	Sweden 2010–2011 Rental tests	Arabic males and females have 11 and 7 percentage points lower probability of being invited to an apartment showing than do Swedish males. Unemployed applicants are more discriminated than employed applicants, but including employment status does not change the magnitude of ethnic discrimination. The study finds no evidence of age discrimination.
Carpusor and Loges (2006)	E-mail correspondence tests	1,115 units (not paired: one e-mail per unit)		Los Angeles, California 2003 Rental tests	African-American and Arabic renters receive 33 percentage points and 23 percentage points lower positive e-mail responses (being told that the unit was available) than White renters, respectively.
Ewens, Tomlin, and Wang (2014)	E-mail correspondence tests	14,237 units (not paired: one e-mail per unit)		34 U.S. cities Late 2000s (year not specified) Rental tests (one bedroom, studio only)	African-American renters have 8 to 9 percentage points lower positive responses (being told that the unit is available) than White renters. Revealing positive information about the socioeconomic status (using information on nonsmoking behavior, a respectable occupation, and credit rating) increases the positive response rate, but it does not narrow the racial gap in the response rate.

Exhibit A-1

Results of Housing Discrimination Audit Studies (4 of 6)

Authors	Data/Methodology ^a	Scale	Other Factors Considered With Race/Ethnicity and Gender	Location, Period, and Market Examined	Main Findings
Feins and Bratt (1983)	In-person tests	156 rental and 118 sales tests		Boston, Massachusetts (seven neighborhoods) 1981 Rental and sales tests Steering	African-American homebuyers have 27 percent [37 percent] and 15 percent [21 percent] lower probability of being told, when asked if an advertised unit was available in rental and sales markets. The study finds no strong significant differences in the racial composition of neighborhoods in which White and African-American auditors were shown houses. This steering result was based on the association of the housing characteristics and auditors' race.
Hanson and Hawley (2011)	E-mail correspondence tests Probit model	4,728 tests (4,728 x 2 = 9,456 e-mails)	Socioeconomic status using the prose quality of e-mails	10 U.S. cities 2009 Rental tests	African-American renters are treated less favorably than White renters by landlords. Landlords reply faster, reply with an e-mail that is longer to inquiries made, make formal greetings, and use polite language when replying to e-mail inquiries from a White homebuyer.
Hanson and Santas (2014)	E-mail correspondence tests	3,072 tests (3,072 x 2 = 6,144 e-mails)	Socioeconomic status using the prose quality of e-mails	21 large U.S. metropolitan areas 2011 Rental tests	Assimilated Hispanic-American renters experience little discrimination, but recent Hispanic immigrants receive 2.9 percent lower response rates than White renters. When e-mail prose quality of the immigrants is low, discrimination doubles to 5.8 percent for nonresponse outcomes, and the incidence of discrimination is 6.9 percent for positive-response outcomes.
Hogan and Berry (2011)	E-mail correspondence tests Fixed-effects logit model	1,124 tests (1,124 x 5 = 5,620 e-mails)		Toronto, Canada 2007 Rental tests	Arabic homebuyers face discrimination in 12 percent of experiments. The level of discrimination is modest but significant for Asian men (7 percent), Black homebuyers (5 percent), and Arabic women (5 percent). "Opportunity denying" discrimination (exclusion through nonresponse) was 10 times as common as "opportunity diminishing" discrimination (for example, additional rental conditions).

Exhibit A-1

Results of Housing Discrimination Audit Studies (5 of 6)

Authors	Data/Methodology ^a	Scale	Other Factors Considered With Race/Ethnicity and Gender	Location, Period, and Market Examined	Main Findings
James, Cummings, and Tynan (1984)	In-person tests	253 tests		Denver, Colorado 1982 Rental and sales tests Steering	African-American and Hispanic homebuyers have 1 to 11 percentage points higher probability of being told that either advertised or others similar houses are not available than White homebuyers (using net measures). Discrimination against minority homebuyers was higher in rental markets than sales markets. Minority homebuyers were offered less assistance with financing arrangements. In general, Hispanic homebuyers experienced higher discrimination than African-American homebuyers. Discrimination in sales tests was evident in White neighborhoods, and discrimination against minority renters was evident in minority neighborhoods.
Massey and Lundy (2001)	Telephone correspondence tests Logit and OLS models	79 units (79 x 6 = 474 calls)	Socioeconomic status using racially distinctive English (with racially neutral names)	Philadelphia, Pennsylvania 1999 Rental tests	African-American renters with high social class have 1 to 5 percentage points lower probability of being told that a unit is available than do White renters. The gap for African-American renters with low social class is 21 to 23 percentage points. Males have 4 to 10 percentage points higher probability of being told that a unit is available than do females, and African-American renters with low social class have 16 to 22 percentage points lower probability than do African-American renters with high social class.
McIntosh and Smith (1974)	In-person tests	178 tests (137 sales; 41 rental)		United Kingdom (five regions) 1967 and 1973 Rental and sales tests	Much lower levels of net discrimination against ethnic minority home seekers (West Indian, Indian, Pakistani, and Greek individuals) were found in 1973 than in 1967.

Exhibit A-1

Results of Housing Discrimination Audit Studies (6 of 6)

Authors	Data/Methodology ^a	Scale	Other Factors Considered With Race/Ethnicity and Gender	Location, Period, and Market Examined	Main Findings
Pearce (1979)	In-person tests	97 tests		Detroit, Michigan 1974–1975 Sales tests Steering	African-American homeseekers have 44 percent [46 percent] lower probability of being shown houses by real estate agents. African-American homeseekers were shown houses in slightly higher percent Black census tracts and in communities with lower house values than were White homeseekers.
Purnell, Idsardi, and Baugh (1999)	Telephone correspondence tests	989 calls		San Francisco Bay Area, California 1997 Rental tests	African-American renters have 8 to 41 percentage points lower rates of having a confirmed appointment to view an apartment in the White-majority areas but 3 to 22 percentage points higher rate of having an appointment in Black-majority areas. Mexican-American renters have 7 to 17 percentage points lower rates of making an appointment than do African-American renters in all the areas.
Roychoudhury and Goodman (1992)	In-person tests Ordered probit model	568 tests		Detroit, Michigan 1980–1990 Rental tests	For each additional apartment available to an agent, the probability of discrimination against an African-American auditor increases by 0.50 for the number of units withheld and by 0.58 for the number of units inspected.
Roychoudhury and Goodman (1996)	In-person tests Ordered probit and bivariate probit model	319 tests		Detroit, Michigan 1980–1990 Sales tests Steering	Discrimination against African-American homebuyers was substantially less frequent among African-American housing agents and more frequent among older agents. An audit-by-audit comparison suggests that a considerable number of White homeseekers were steered toward predominantly White and more affluent neighborhoods.

OLS = ordinary least squares.

^a Methodology indicates any statistical analysis other than computing gross and net measures or the differences-in-means tests.

Note: Gross measures are presented in brackets.

Appendix B. The Origins of Fair Housing Audits

Fair housing audits have been developed over the years by scholars, private fair housing groups, and governmental enforcement officials—often in partnerships.

An extensive library and Internet search by the authors indicates that the first published reference to fair housing audits appeared in the *New York Times* in 1956 (Rowland, 1956). This article describes audits conducted by the Committee on Civil Rights in East Manhattan (CCRM) in 1953 and 1954. These audits are also discussed in Frost (1958) and McEntire (1960).⁴¹ This study “was supervised by a group of social scientists associated with the sponsoring Committee [CCRM]” (McEntire, 1960: 240n).⁴² According to Frost (1958: 69), “CCRM was not seeking to duplicate the work of established organizations in the field of civil rights. Rather, it hoped to give to a varied group of individuals a more intensive experience as well as an opportunity to work cooperatively in producing changes. The technique selected for achieving this was that of the community audit.”

In the CCRM audits, a Black person or a Black couple “would visit the office of a designated real estate broker and inquire about the availability of apartments of a specified type. After the minority tester left, the control tester (white) would proceed to the same office with an identical set of apartment specifications... [D]iscriminatory practice was to be the difference in treatment accorded the two testers” (Frost, 1958: 71). Testers were selected and trained, and “in November [1953] and March [1954] a total of 27 firms were visited and tested” (Frost, 1958: 72). In 22 of the 27 cases, discrimination against African-American renters was found. In addition, the CCRM conducted 17 more tests in May 1954 based on advertisements in the Sunday *New York Times*. Discrimination was found in 10 of the tests.

McEntire (1960: 239) also reports on an audit study in Los Angeles in 1955, which appears to be the first fair housing audit study of the sales market. In this study, “a white couple, representing themselves as possible house buyers, called on twelve real estate brokers doing business in a new residential area of 12,000 homes, chiefly FHA- and VA-financed. The couple was followed after a brief interval by a Negro, also purporting to be looking for a house to buy.” Discrimination was found in every case. The report on this study was co-authored by James H. Kirk, a professor at Loyola University of Los Angeles.⁴³

⁴¹ These housing audits built on CCRM’s experience with restaurant audits, which were conducted as far back as 1950. These restaurant audits, like many of the housing audits that followed, involved a partnership between scholars and CCRM. According to Selltiz (1955), these scholars included Kenneth Clark (City College of New York), Dan Dodson (New York University), Herbert Hyman (Columbia University), Patricia Kendall (Columbia University), Sophia M. Robison (New York School of Social Work, now the Columbia University School of Social Work), and Claire Selltiz (New York University). Jou (2014) provides a detailed history of the restaurant audits.

⁴² Although many scholars are listed in the preface to McEntire (1960), the particular social scientists who contributed to this audit study are not identified. It seems likely that Claire Selltiz was involved, because she was both a “technical consultant” to the CCRM restaurant audits (Selltiz, 1955: 19) and was listed in McEntire (1960: xi) as someone who prepared “Research memoranda on sociopsychological aspects of housing and minority groups.”

⁴³ The affiliation of the co-author, Lane D. Spane, is not indicated. This study appears to be one in “a series of related studies prepared by cooperating social scientists and other experts” upon which McEntire (1960: x) is “largely based.”

The authors' research indicates that the first use of enforcement testing by a private fair housing group was in Brooklyn, New York, in 1960 (Purnell, 2013). The first legislation in the country banning discrimination in private housing became effective in New York City in 1958, and the New York branch of the Congress of Racial Equality (CORE) decided to help enforce this act by conducting tests and, if necessary, giving the results to the relevant enforcement agency: the New York City Commission on Intergroup Relations. The first documented case of this strategy occurred in August 1960, when the New York CORE helped a Black family who was told that a Brooklyn apartment they wanted was no longer available. During the next 5 days, White testers were told over the phone that the apartment was still available, but testers who identified themselves as Black were told that it had been rented. When the apartment was subsequently advertised in the *New York Times*, the Black family was told once again that the apartment was no longer available, whereas a White tester who visited the rental agent's office was told that he could rent the apartment. These techniques were then picked up by the Brooklyn branch of CORE. According to Purnell (2013), "Over the next year and a half, the chapter's housing activists also improved on this basic model... and helped scores of African Americans move into apartments and homes in mostly white areas of Brooklyn."

The first documented case of enforcement tests by private fair housing groups that were not linked to complaints comes from Chicago in the mid-1960s. The Coordinating Council of Community Organizations (CCCO) was formed in Chicago in April 1962; its first focus was on school segregation in the city (Cohen and Taylor, 2000). Chicago passed a fair housing ordinance in 1963, and at some point CCCO initiated a testing program. This program came to light in 1966 when Martin Luther King, Jr., and the Southern Christian Leadership Conference joined with the CCCO to initiate the Chicago Freedom Movement. As part of its efforts to combat discrimination, "The Freedom Movement had been sending testers into Gage Park," one of the White neighborhoods in Chicago, and by the time this organization began a series of marches in White neighborhoods in July 1966, it "had already documented 121 cases of racial discrimination" (Cohen and Taylor, 2000: 392). An all-night vigil was held "at F.H. Halvorsen Realty in Gage Park... because, according to recent testing, it repeatedly discriminated against black applicants" (Cohen and Taylor, 2000: 392).

In August 1966, Chicago Mayor Richard Daley invited the participants in the Freedom Movement to a housing summit. "The movement also embarked... on a pre-summit campaign of real estate-agent testing. As expected, blacks were lied to about the availability of housing in white neighborhoods and turned away" (Cohen and Taylor, 2000: 400). Moreover, this testing program "collected enough evidence to file seventy-four discrimination complaints against sixteen real estate brokers. Equally important, the testing gave them fresh evidence going into the summit that the problem of housing discrimination was real, and that the city's Fair Housing Ordinance of 1963 was not being enforced" (Cohen and Taylor, 2000: 400).

The passage of the Fair Housing Act and the U.S. Supreme Court's resurrection of the Civil Rights Act of 1866, both of which occurred in 1968, greatly expanded the opportunity for audits to be used as an enforcement tool. Most importantly, FHA gave private fair housing groups the standing to sue alleged discriminators. As a result, the use of audits by private fair housing groups quickly spread. The first testing-based case to appear in federal court relied on evidence from tests conducted in 1968 in Brown County, Ohio (Schwemm, 1992). As pointed out by Yinger

(1995: 20), “how-to manuals for conducting audits were widely available” by the early 1970s. See, for example, Kovar (1974), Leadership Council for Metropolitan Open Communities (1975), or Murphy (1972). An assessment of the early use of audits by private fair housing groups is provided by Freiberg (1993). Governmental civil rights enforcement agencies eventually also started using audits, and they became a crucial tool in fair housing cases (Schwemm, 1992). In 1982, the U.S. Supreme Court upheld the use of audits as a fair housing enforcement tool.⁴⁴

Scholars also recognized the power of audits to uncover discriminatory behavior, and more audit studies began to appear as early as the 1960s. Housing audits were conducted in Great Britain, for example, starting in 1967 (see Daniel, 1968, and McIntosh and Smith, 1974).⁴⁵ In the United States, a study published in 1971 (Johnson, Porter, and Mateljan, 1971) used audits to examine discrimination against Black and Mexican-American homeseekers in a Southern California city. Another 1971 audit study, which examined the behavior of real estate brokers and landlords in Akron, Ohio, was “devised by the author [a professor at Kent State University] as part of the ongoing research program of the Fair Housing Contact Service of Akron, a voluntary open housing group” (Saltman, 1975: 41).⁴⁶ A study conducted in Detroit in 1974 and 1975 used audits to examine racial steering by real estate brokers (Pearce, 1979).⁴⁷ In 1977, scholars at the U.S. Department of Housing and Urban Development (HUD) directed the first nationwide audit study (Wienk et al., 1979).⁴⁸

The federal government did not develop its own testing programs for many years after the passage of the Fair Housing Act. Indeed, as late as 1978, the Comptroller General of the United States (CGUS) observed that “HUD officials . . . are reluctant to use testing because some people view it as harassment. Some officials also question its legality, but we were told by HUD’s General Counsel that testing is legal” (CGUS, 1978: 26). Despite this reluctance to conduct testing, however, the CGUS (1978: 26) also reported that “HUD does use testing data developed by local fair housing organizations.” Moreover, according to Lee (1999: 48n, endnote 21), “In the early years of its enforcement efforts, the [Justice] Department often relied upon testing evidence provided to it by local fair housing groups.” To be specific, “the first reported case in which tester evidence was used in a suit brought by the Attorney General” involved tests conducted in March 1970 (Schwemm, 1992: 40).

Although the audit technique is used for both enforcement and research purposes, the requirements for these two applications are not exactly the same. Over time, private fair housing groups and governmental agencies refined the use of this tool for enforcement purposes and scholars developed new measures of discrimination, new statistical procedures, and ways to use audit results for testing hypotheses about discriminatory behavior. Moreover, many scholars based their research on data from audits originally designed for enforcement purposes. These developments are reviewed in the text of this article.

⁴⁴ *Havens Realty Corp. v. Coleman*, 455 U.S. 363, 373 (1982).

⁴⁵ Moreover, our research indicates that the first appearance of correspondence audits was in a study of racial discrimination in employment in Britain (Jowell and Prescott-Clarke, 1970).

⁴⁶ Dr. Saltman was one of the founders of the Fair Housing Contact Service in 1965. See Walbeck (1974).

⁴⁷ At the time this study was published, Dr. Pearce was a professor at the University of Illinois at Chicago Circle.

⁴⁸ HUD’s partner for this study was the National Committee against Discrimination in Housing, which carried out the audits. See Wienk et al. (1979).

Acknowledgments

The authors thank two anonymous referees for their helpful comments.

Authors

Sun Jung Oh is a research fellow at the Korea Labor Institute.

John Yinger is a professor of economics and public administration at the Maxwell School, Syracuse University.

References

- Ahmed, Ali M., Lina Andersson, and Mats Hammarstedt. 2010. "Can Discrimination in the Housing Market Be Reduced by Increasing the Information About the Applicants?" *Land Economics* 86 (1): 79–90.
- Ahmed, Ali M., and Mats Hammarstedt. 2008. "Discrimination in the Rental Housing Market: A Field Experiment on the Internet," *Journal of Urban Economics* 64 (2): 362–372.
- Andersson, Lisa, Niklas Jakobsson, and Andreas Kotsadam. 2012. "A Field Experiment of Discrimination in the Norwegian Housing Market: Gender, Class, and Ethnicity," *Land Economics* 88 (2): 233–240.
- Ayres, Ian. 1991. "Fair Driving: Gender and Race Discrimination in Retail Car Negotiations," *Harvard Law Review* 104 (4): 817–872.
- Ayres, Ian, and Peter Siegelman. 1995. "Race and Gender Discrimination in Bargaining for a New Car," *American Economic Review* 85 (3): 304–321.
- Baldini, Massimo, and Marta Federici. 2011. "Ethnic Discrimination in the Italian Housing Market," *Journal of Housing Economics* 20 (1): 1–14.
- Bendick, Marc, Charles W. Jackson, and Victor A. Reinoso. 1994. "Measuring Employment Discrimination Through Controlled Experiments," *Review of Black Political Economy* 23 (1): 25–48.
- Bendick, Marc, Charles W. Jackson, Victor A. Reinoso, and Laura E. Hodges. 1991. "Discrimination Against Latino Job Applicants: A Controlled Experiment," *Human Resource Management* 30 (4): 469–484.
- Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review* 94 (4): 991–1013.
- Bosch, Mariano, Angeles M. Carnero, and L dia Farr . 2010. "Information and Discrimination in the Rental Housing Market: Evidence From a Field Experiment," *Regional Science and Urban Economics* 40 (1): 11–19.

Bovenkerk, Frank, Benjamin Kilborne, Francois Raveau, and David Smith. 1979. "Comparative Aspects of Research on Discrimination Against Non-White Citizens in Great Britain, France and the Netherlands." In *Problems in International Comparative Research in the Social Sciences*, edited by Jan Berting, Felix Geyer, and Ray Jurkovich. Oxford, United Kingdom: Pergamon Press: 105–124.

Bursell, Moa. 2007. What's in a Name? A Field Experiment Test for the Existence of Ethnic Discrimination in the Hiring Process. Working paper 7. Stockholm, Sweden: Stockholm University, Linnaeus Center for Integration Studies (SULCIS). http://www.su.se/polopoly_fs/1.55389.1321514449/SULCISWP2007_7.pdf.

Carlsson, Magnus, and Stefan Ericksson. 2014. "Discrimination in the Rental Market for Apartments," *Journal of Housing Economics* 23: 41–54.

———. 2013. *Ethnic Discrimination in the Market for Shared Housing*. Kalmar, Sweden: Linnaeus University Centre for Labour Market and Discrimination Studies.

Carlsson, Magnus, and Dan-Olof Rooth. 2012. "Revealing Taste-Based Discrimination in Hiring: A Correspondence Testing Experiment With Geographic Variation," *Applied Economics Letters* 19 (18): 1861–1864.

———. 2007. "Evidence of Ethnic Discrimination in the Swedish Labor Market Using Experimental Data," *Labour Economics* 14 (4): 716–729.

Carpusor, Adrian G., and William E. Loges. 2006. "Rental Discrimination and Ethnicity in Names," *Journal of Applied Social Psychology* 36 (4): 934–952.

Choi, Seok Joon, Jan Ondrich, and John Yinger. 2008. "Changes in Rental Housing Discrimination Since 1989," *Cityscape* 10 (2): 301–330.

———. 2005. "Do Rental Agents Discriminate Against Minority Customers? Evidence From the 2000 Housing Discrimination Study," *Journal of Housing Economics* 14 (1): 1–26.

Cohen, Adams, and Elizabeth Taylor. 2000. *Pharaoh: Mayor Richard J. Daley: His Battle for Chicago and the Nation*. Boston: Back Bay Books.

Comptroller General of the United States (CGUS). 1978. "Stronger Federal Enforcements Needed to Uphold Fair Housing Laws." CED-78-21; B-164855. <http://www.gao.gov/assets/130/121301.pdf>.

Cross, Harry, Genevieve Kenney, Jane Mell, and Wendy Zimmerman. 1989. *Differential Treatment of Hispanic and Anglo Job Seekers: Hiring Practices in Two Cities*. Washington, DC: Urban Institute.

Daniel, W. 1968. *Racial Discrimination in England*. Middlesex, United Kingdom: Penguin Books.

Doleac, Jennifer L., and Luke C.D. Stein. 2013. "The Visible Hand: Race and Online Market Outcomes," *Economic Journal* 123 (572): F469–F492.

Ewens, Michael, Bryan Tomlin, and Liang Choon Wang. 2014. "Statistical Discrimination or Prejudice? A Large Sample Field Experiment," *Review of Economics and Statistics* 96 (1): 119–134.

Feins, Judith D., and Rachael G. Bratt. 1983. "Barred in Boston: Racial Discrimination in Housing," *Journal of the American Planning Association* 49 (3): 344–355.

Freiberg, Fred. 1993. "Promoting Residential Integration: The Role of Private Fair Housing Groups." In *Housing Markets and Residential Mobility*, edited by G. Thomas Kingsley and Margery Austin Turner. Washington, DC: Urban Institute: 219–242.

Frost, Olivia. 1958. "CCRM—Grass Roots in the Big City," *The Crisis* (February): 69–80. http://books.google.com.pr/books?id=71cEAAAAMBAJ&pg=PA67&lpg=PA67&dq=Frost+The+Crisis+1958&source=bl&ots=XnDRWFSQlc&sig=wzJJDvAUuPh3Q6YBD1RGNW2HhSM&hl=en&sa=X&ei=X6lOU-7CDq3p0QG6xYG4AQ&redir_esc=y#v=onepage&q=Frost%20The%20Crisis%201958&f=false.

Fudge, Keith. 2014. "Fair Housing Organizations Use Testing To Expose Discrimination," *Evidence Matters* (Spring/Summer): 16–26.

Galster, George. 1993. "The Use of Testers in Investigating Discrimination in Mortgage Lending and Insurance." In *Clear and Convincing Evidence: Measurement of Discrimination in America*, edited by Michael E. Fix and Raymond J. Struyk. Washington, DC: Urban Institute Press: 287–334.

———. 1990a. "Racial Discrimination in Housing Markets in the 1980s: A Review of the Audit Evidence," *Journal of Planning Education and Research* 9 (3):165–175.

———. 1990b. "Racial Steering in Urban Housing Markets: A Review of the Audit Evidence," *Review of Black Political Economy* 18 (3): 105–129.

———. 1990c. "Racial Steering by Real Estate Agents: Mechanisms and Motivations," *Review of Black Political Economy* 19 (1): 39–63.

———. 1987. "The Ecology of Racial Discrimination in Housing: An Exploratory Model," *Urban Affairs Quarterly* 23 (1): 84–107.

Galster, George, Fred Freiberg, and Diane L. Houk. 1987. "Racial Differentials in Real Estate Advertising Practices: An Exploratory Case Study," *Journal of Urban Affairs* 9 (3): 199–215.

Galster, George, and Erin Godfrey. 2005. "By Words and Deeds: Racial Steering by Real Estate Agents in the US in 2000," *Journal of the American Planning Association* 71 (3): 251–268.

Gneezy, Uri, and John A. List. 2004. Are the Disabled Discriminated Against in Product Markets? Evidence From Field Experiments. Working paper. Pittsburgh, PA: American Economic Association. https://www.aeaweb.org/annual_mtg_papers/2005/0108_0800_1303.pdf.

Gneezy, Uri, John List, and Michael K. Price. 2012. Toward an Understanding of Why People Discriminate: Evidence From a Series of Natural Field Experiments. NBER Working Paper 17855. Cambridge, MA: National Bureau of Economic Research. <http://www.nber.org/papers/w17855>.

Goldberg, Andreas, Dora Mourinho, and Ursula Kulke. 1995. *Labour Market Discrimination Against Foreign Workers in Germany*. International Migration Papers 7. Geneva, Switzerland: International Labour Organization. http://www.ilo.org/global/topics/labour-migration/publications/WCMS_201039/lang--en/index.htm.

Gorman, Elizabeth. 2015. "Opening Acts: March 12, 2015," *Fair Housing Justice Center Newsletters*. <http://www.fairhousingjustice.org/2015/03/12/opening-acts-march-12-2015/>.

Hanson, Andrew, and Zackary Hawley. 2011. "Do Landlords Discriminate in the Rental Housing Market? Evidence From an Internet Field Experiment in U.S. Cities," *Journal of Urban Economics* 70 (2): 99–114.

Hanson, Andrew, Zackary Hawley, and Aryn Taylor. 2011. "Subtle Discrimination in the Rental Housing Market: Evidence From E-mail Correspondence With Landlords," *Journal of Housing Economics* 20 (4): 276–284.

Hanson, Andrew, and Michael Santas. 2014. "Field Experiment Tests for Discrimination Against Hispanics in the U.S. Rental Housing Market," *Southern Economic Journal* 81 (1): 135–167.

Heckman, James J., and Peter Siegelman. 1993. "The Urban Institute Audit Studies: Their Methods and Findings." In *Clear and Convincing Evidence: Measurement of Discrimination in America*, edited by Michael E. Fix and Raymond J. Struyk. Washington, DC: Urban Institute Press: 187–258.

Hogan, Bernie, and Brent Berry. 2011. "Racial and Ethnic Biases in Rental Housing: An Audit Study of Online Apartment Listings," *City & Community* 10 (4): 351–372.

James, Franklin J., and Eleanor G. Crow. 1986. "Stronger Administrative Enforcement of Federal Housing Laws? Lessons From the States," *The Journal of Intergroup Relations* 13 (4): 3–27.

James, Franklin J., Betty L. McCummings, and Eileen A. Tynan. 1984. *Minorities in the Sunbelt*. New Brunswick, NJ: Rutgers University Press.

Johnson, Daniel A., Richard J. Porter, and Patricia L. Mateljan. 1971. "Racial Discrimination in Apartment Rentals," *Journal of Applied Social Psychology* 1 (4): 364–377.

Jou, Chin. 2014. "Neither Welcomed, Nor Refused: Race and Restaurants in Postwar New York City," *Journal of Urban History* 40 (2): 232–251.

Jowell, Roger, and Patricia Prescott-Clarke. 1970. "Racial Discrimination and White-Collar Workers in Britain," *Race* 11 (4): 397–417.

Kaas, Leo, and Christian Manger. 2011. "Ethnic Discrimination in Germany's Labour Market: A Field Experiment," *German Economic Review* 13 (1): 1–20.

Kovar, Leopold J. 1974. *Auditing Real Estate Practices: A Manual*. Philadelphia: National Neighbors.

Leadership Council for Metropolitan Open Communities. 1975. *Investigation and Auditing in Fair Housing Cases*. Chicago: Leadership Council for Metropolitan Open Communities.

Lee, Bill Lann. 1999. "An Issue of Public Importance: The Justice Department's Enforcement of the Fair Housing Act," *Cityscape* 4 (3): 35–56.

List, John A. 2004. "The Nature and Extent of Discrimination in the Marketplace: Evidence From the Field," *The Quarterly Journal of Economics* 119 (1): 49–89.

Massey, Douglas S., and Garvey Lundy. 2001. "Use of Black English and Racial Discrimination in Urban Housing Markets," *Urban Affairs Review* 36 (4): 452–469.

McEntire, Davis. 1960. *Residence and Race*. Berkeley, CA: University of California Press.

McIntosh, Neil, and David John Smith. 1974. *The Extent of Racial Discrimination, Political and Economic Planning Broadsheet No. 547*. London, United Kingdom: Political and Economic Planning.

Murphy, J.K. 1972. *Audit Handbook: Procedures for Determining the Extent of Racial Discrimination in Apartment Rentals*. Palo Alto, CA: Midpeninsula Citizens for Fair Housing.

National Fair Housing Alliance (NFHA). 2014. *Arkansas Fair Housing Commission Confirms Discrimination Against Prospective Latino Tenants*. Washington, DC: National Fair Housing Alliance. <http://www.nationalfairhousing.org/>.

Newburger, Harriet B. 1995. "Sources of Difference in Information Used by Black and White Housing Seekers: An Exploratory Analysis," *Urban Studies* 32 (3): 445–470.

———. 1989. "Discrimination by a Profit-Maximizing Broker in Response to White Prejudice," *Journal of Urban Economics* 26 (1): 1–19.

Nunley, John M., Mark F. Owens, and R. Stephen Howard. 2011. "The Effects of Information and Competition on Racial Discrimination: Evidence From a Field Experiment," *Journal of Economic Behavior and Organization* 80 (3): 670–679.

Nunley, John M., Adam Pugh, Nicholas Romero, and Richard Alan Seals. 2014. *An Examination of Racial Discrimination in the Labor Market for Recent College Graduates: Estimates From the Field*. Working paper no. auwp2014-06. Auburn, AL: Auburn University, Department of Economics. <http://cla.auburn.edu/econwp/Archives/2014/2014-06.pdf>.

Ondrich, Jan, Stephen Ross, and John Yinger. 2003. "Now You See It, Now You Don't: Why Do Real Estate Agents Withhold Available Houses From Black Customers?" *Review of Economics and Statistics* 85 (4): 854–873.

- Ondrich, Jan, Alex Stricker, and John Yinger. 1999. "Do Landlords Discriminate? The Incidence and Causes of Racial Discrimination in Rental Housing Markets," *Journal of Housing Economics* 8 (3): 185–204.
- . 1998. "Do Real Estate Brokers Choose to Discriminate? Evidence From the 1989 Housing Discrimination Study," *Southern Economic Journal* 64 (4): 880–901.
- Page, Marianne. 1995. "Racial and Ethnic Discrimination in Urban Housing Markets: Evidence From a Recent Audit Study," *Journal of Urban Economics* 38 (2): 183–206.
- Pager, Devah. 2007. "The Use of Field Experiments for Studies of Employment Discrimination: Contributions, Critiques, and Directions for the Future," *The Annals of the American Academy of Political and Social Science* 609 (1): 104–133.
- . 2003. "The Mark of a Criminal Record," *American Journal of Sociology* 108 (5): 937–975.
- Pager, Devah, and Hana Shepherd. 2008. "The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets," *Annual Review of Sociology* 34: 181–209.
- Pager, Devah, Bruce Western, and Bart Bonikowski. 2009. "Discrimination in a Low-Wage Labor Market: A Field Experiment," *American Sociological Review* 74 (5): 777–799.
- Pearce, Diana M. 1979. "Gatekeepers and Homeseekers: Institutional Patterns in Racial Steering," *Social Problems* 26 (3): 322–342.
- Perry, James. 2005. *An Audit of Discriminatory Treatment of African-American Males by Bourbon Street Nightclubs and Bars*. New Orleans: Greater New Orleans Fair Housing Action Center.
- Purnell, Brian. 2013. *Fighting Jim Crow in the County of Kings: The Congress of Racial Equality in Brooklyn*. Lexington, KY: University Press of Kentucky.
- Purnell, Thomas, William Idsardi, and John Baugh. 1999. "Perceptual and Phonetic Experiments on American English Dialect Identification," *Journal of Language and Social Psychology* 18 (1): 10–30.
- Riach, P.A., and J. Rich. 2002. "Field Experiments of Discrimination in the Market Place," *The Economic Journal* 112 (483): F480–F518.
- Rice, Willy E. 1984. "Judicial Enforcement of Fair Housing Laws: An Analysis of Some Unexamined Problems That the Fair Housing Amendments Act of 1983 Would Eliminate," *Howard Law Journal* 27 (1): 227–274.
- Rich, Judith. 2014. What Do Field Experiments of Discrimination in Markets Tell Us? A Meta Analysis of Studies Conducted Since 2000. IZA Discussion paper no. 8584. Bonn, Germany: Forschungsinstitut zur Zukunft der Arbeit. <http://ftp.iza.org/dp8584.pdf>.

Ridley, Stanley E., James A. Bayton, and Janice Hamilton Outtz. 1989. *Taxi Service in the District of Columbia: Is It Influenced by Patron's Race and Destination?* Washington, DC: Lawyer's Committee for Civil Rights Under the Law.

Ross, Stephen L., and George C. Galster. 2007. "Fair Housing Enforcement and Changes in Discrimination Between 1989 and 2000: An Exploratory Study." In *Fragile Rights Within Cities*, edited by John Goering. Plymouth, United Kingdom: Rowman and Littlefield: 177–202.

Ross, Stephen L., and Margery A. Turner. 2005. "Housing Discrimination in Metropolitan America: Explaining Changes Between 1989 and 2000," *Social Problems* 52 (2): 152–180.

Ross, Stephen L., and John Yinger. 2006. "Uncovering Discrimination: A Comparison of the Methods Used by Scholars and Civil Rights Enforcement Officials," *American Law and Economics Review* 8 (3): 562–614.

Rowland, Stanley, Jr. 1956. "Race Bias Easing in Churches Here: Many Congregations Employ 'Open Door' Policy—Housing Discrimination Persists," *New York Times*, February 12: 1.

Roychoudhury, Canopy, and Allen C. Goodman. 1996. "Evidence of Racial Discrimination in Different Dimensions of Owner-Occupied Housing Search," *Real Estate Economics* 24 (2): 161–178.

———. 1992. "An Ordered Probit Model for Estimating Racial Discrimination Through Fair Housing Audits," *Journal of Housing Economics* 2 (4): 358–373.

Saltman, Juliet. 1975. "Implementing Open Housing Laws Through Social Action," *Journal of Applied Behavioral Science* 11 (1): 39–61.

Schulman, Kevin A., Jesse A. Berlin, Willam Harless, Jon F. Kerner, Shyrl Sistrunk, Bernard J. Gersh, Ross Dubé, Christopher K. Taleghani, Jennifer E. Burke, Sankey Williams, John M. Eisenberg, William Ayers, and José J. Escarce. 1999. "The Effect of Race and Sex on Physicians' Recommendations for Cardiac Catheterization," *New England Journal of Medicine* 340 (8): 618–626.

Schuman, Howard, and Lawrence Bobo. 1988. "Survey-Based Experiments on White Racial Attitudes Toward Residential Integration," *American Journal of Sociology* 94 (2): 273–299.

Schuman, Howard, Charlotte Steeh, and Lawrence Bobo. 1985. *Racial Attitudes in America*. Cambridge, MA: Harvard University Press.

Schuman, Howard, Charlotte Steeh, Lawrence Bobo, and Maria Krysan. 1997. *Racial Attitudes in America: Trends and Interpretations*. Cambridge, MA: Harvard University Press.

Schwemm, Robert G. 2014. *Housing Discrimination: Law and Litigation*. Eagan, MN: Thomson Reuters.

———. 1992. *Fair Housing Cases Involving Testing: A Legal Review of Reported Federal Court Decisions 1968–1991*. Washington, DC: Urban Institute.

Selltiz, Claire. 1955. "The Use of Survey Methods in a Citizens Campaign Against Discrimination," *Human Organization* 14 (3): 19–25.

Smith, Shanna L. 1994. "The National Fair Housing Alliance at Work." In *Residential Apartheid: The American Legacy*, edited by Robert D. Bullard, J. Eugene Grigsby III, and Charles Lee. Los Angeles: CAAS Publications: 237–256.

Temkin, Kenneth, Tracy McCracken, and Veralee Liban. 2011. *Study of the Fair Housing Initiatives Program*. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research. http://www.huduser.org/Publications/pdf/FHIP_2011.pdf.

Turner, Margery A. 1992. "Discrimination in Urban Housing Markets: Lessons From Fair Housing Audits," *Housing Policy Debate* 3 (2): 183–215.

Turner, Margery A., Michael Fix, and Raymond J. Struyk. 1991. *Opportunities Denied, Opportunities Diminished: Racial Discrimination in Hiring*. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research.

Turner, Margery A., Maris Mikelsons, and John Edwards. 1990. *Analysis of Steering in the Housing Discrimination Study*. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research.

Turner, Margery A., Stephen L. Ross, George C. Galster, and John Yinger. 2002. *Discrimination in Metropolitan Housing Markets: National Results From Phase 1 HDS 2000*. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research.

Turner, Margery A., Rob Santos, Diane K. Levy, Doug Wissoker, Claudia Aranda, and Rob Pitingolo. 2013. *Housing Discrimination Against Racial and Ethnic Minorities 2012*. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research.

Turner, Margery A., and Felicity Skidmore. 1999. *Mortgage Lending Discrimination: A Review of Existing Evidence*. Washington, DC: Urban Institute.

Turner, Margery A., Raymond Struyk, and John Yinger. 1991. *Housing Discrimination Study: Synthesis*. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research.

Urban Institute. 1991. *Housing Discrimination Study. Methodology and Data Documentation*. UI Report 3864-06 A. Washington, DC: Urban Institute.

U.S. Department of Housing and Urban Development (HUD). 2014. "Fair Housing Enforcement Organizations Use Testing To Expose Discrimination," *Evidence Matters: Transforming Knowledge Into Housing and Community Development Policy* Spring/Summer 2014. <http://www.huduser.org/portal/periodicals/em/spring14/highlight3.html>.

———. 2012. “HUD and Owner, Manager of Alabama Apartment Complex Settle Allegations of Discrimination Against Prospective Hispanic Tenants.” Press release, July 13. <http://archives.hud.gov/local/al/news/pr2012-07-13.cfm>.

U.S. Department of Justice (DOJ). 2014. “Fair Housing Testing Program.” http://www.justice.gov/crt/about/hce/housing_testing.php.

Walbeck, Jack. 1974. “Fair Housing Service Founder Saltman up for ‘Woman of Year’,” *Daily Kent Stater* 18 (79; March 13): 13. <http://dks.library.kent.edu/cgi-bin/kentstate?a=d&d=ks19740313-01.2.45>.

Wienk, Ronald E., Clifford E. Reid, John C. Simonson, and Frederick J. Eggers. 1979. *Measuring Racial Discrimination in American Housing Markets: The Housing Market Practices Survey*. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research, Division of Evaluation.

Wissoker, Douglas A., Wendy Zimmermann, and George C. Galster. 1998. *Testing for Discrimination in Home Insurance*. Washington, DC: Urban Institute.

Wood, Martin, Jon Hales, Susan Purdon, Tanja Sejersen, and Oliver Hayllar. 2009. *A Test for Racial Discrimination in Recruitment Practice in British Cities*. Department for Work and Pensions Research Report No. 607. London, United Kingdom: National Centre for Social Research. <http://www.natcen.ac.uk/media/20541/test-for-racial-discrimination.pdf>.

Yinger, John. 1995. *Closed Doors, Opportunities Lost: The Continuing Costs of Housing Discrimination*. New York: Russell Sage Foundation.

———. 1993. “Access Denied, Access Constrained: Results and Implications of the 1989 Housing Discrimination Study.” In *Clear and Convincing Evidence: Measurement of Discrimination in America*, edited by Michael E. Fix and Raymond J. Struyk. Washington, DC: Urban Institute Press: 69–112.

———. 1987. “The Racial Dimension of Urban Housing Markets in the 1980s.” In *Divided Neighborhoods: Changing Patterns of Racial Segregation*, edited by Gary A. Tobin. Newbury Park, CA: SAGE Publications: 43–67.

———. 1986. “Measuring Racial Discrimination With Fair Housing Audits: Caught in the Act,” *American Economic Review* 76: 881–893.

Zhao, Bo. 2005. “Does the Number of Houses a Broker Shows Depend on the Homeseeker’s Race?” *Journal of Urban Economics* 57 (1): 128–147.

Zhao, Bo, Jan Ondrich, and John Yinger. 2006. “Why Do Real Estate Brokers Continue to Discriminate? Evidence From the 2000 Housing Discrimination Study,” *Journal of Urban Economics* 59 (3): 394–419.

Zussman, Asaf. 2013. “Ethnic Discrimination: Lessons From the Israeli Online Market for Used Cars,” *The Economic Journal* 123 (572): F433–F468.

Housing Discrimination Among Available Housing Units in 2012: Do Paired-Testing Studies Understate Housing Discrimination?

Rob Pitingolo
Urban Institute

Stephen L. Ross
The University of Connecticut

Abstract

The 2000 Housing Discrimination Study (HDS2000) documented substantial declines in discrimination between HDS1989 and HDS2000, and the most recent study (HDS2012) tends to mirror HDS2000 in its findings. The results of HDS2000 led to considerable debate about whether paired-testing studies of the type conducted in HDS2000 understate the extent of housing discrimination. Using data from HDS2012 and earlier evidence, this article considers three of the significant concerns raised regarding paired-testing studies of housing discrimination: (1) exclusion of minority homeseekers during the process of setting up appointments, (2) the net measure of adverse treatment understating discrimination because some housing units are systematically not shown to White testers, and (3) the use of metropolitanwide advertisements that may systematically underrepresent neighborhoods where discrimination is higher. HDS2012 directly addresses the first concern, finding at most very low levels of discrimination in obtaining an appointment over the phone. The evidence for the second concern is mixed. Steering persisted against both Black and Asian homeseekers in owner-occupied housing. On the other hand, the levels of equal treatment in HDS2012 in terms of basic access were quite high, leaving little room for the systematic exclusion of White homeseekers from specific housing units. Further, three-person tests in HDS2000 involving same-race pairs did not suggest that the net measure was biased. To partially address the third concern, this article conducts a new empirical analysis in which we measure the availability of rental and owner-occupied housing in each broad neighborhood represented in HDS2012 and reweight the tests to represent the spatial availability of housing across each metropolitan site. Although the reweighting substantially changed the weights on individual tests, the average attributes of the neighborhoods represented by those tests experienced only modest changes from reweighting, and the estimated measures of adverse treatment were unchanged.

Introduction

In 2012, the U.S. Department of Housing and Urban Development launched the fourth major nationwide Housing Discrimination Study (HDS2012), with the goal of measuring housing discrimination in rental and owner-occupied housing for Black, Hispanic, and Asian homeseekers. The first major study, conducted in 1977,¹ found high levels of discrimination against Black homeseekers, including frequent occurrences of overt exclusion in which Black homeseekers could not complete or even schedule an appointment with real estate agents and often were summarily told that no housing was available (Weink et al., 1979). This study was followed by a nationwide study in 1989² of discrimination against Black and Hispanic homeseekers. This study was the first to document high levels of discrimination against Hispanic homeseekers, and it found no evidence of a decline in the measured levels of discrimination against Black homeseekers since HMPS1977, although most overt acts of exclusion, such as “door slamming,” had declined considerably (Turner, Struyk, and Yinger, 1991; Yinger, 1995). The next study, conducted in 2000,³ was designed to provide estimates of the changes in housing discrimination that Black and Hispanic homeseekers faced nationwide (Turner et al., 2002); a second phase in 2002 provided national estimates for Asian homeseekers (Turner and Ross, 2003a). Although meaningful levels of discrimination were detected for all three groups tested in both markets, HDS2000 found substantial declines in discrimination against Black homeseekers in the rental market and against Black and Hispanic homeseekers in the sales market, and the testers’ narratives suggested a dramatic improvement in the environment that Black and Hispanic homeseekers experienced during their housing search (Ross and Turner, 2005). In 2012, the most recent nationwide housing discrimination study—HDS2012—continued to find persistent, but in many cases modest, levels of discrimination against Black, Hispanic, and Asian homeseekers that appear comparable to the levels detected during 2000 and 2002 (Turner et al., 2013).

The substantial declines in discrimination observed between HDS1989 and HDS2000 led to a considerable debate about whether paired-testing studies of the type conducted in HDS2000 understate the extent of housing discrimination. Paired-testing studies involve sending a White and a minority tester to the same establishment to make the same market inquiry. By their design, testing studies in housing are naturally limited to the portion of the housing transaction that can be observed either during the preapplication process in rental housing or in the housing search assistance provided by a real estate agent in sales market, and so they may miss substantial discrimination later in the process. Even at this early stage, however, several concerns have arisen that suggest that paired tests might understate discrimination. The first and most straightforward concern is that discrimination may take place when the tester attempts to schedule an appointment, because the agent may be able to identify the tester’s race or ethnicity over phone, and so perhaps the most discriminatory rental or sales agents were never tested in HDS2000 (Baugh, 2007; Massey and Lundy, 2001). Second, the most conservative measure of discrimination, the net measure of adverse treatment, is calculated by subtracting the share of tests that favor minority homeseekers from the share of tests that favor White homeseekers, under the assumption that the share of minority-favored tests provides a proxy

¹ The Housing Market Practices Survey (HMPS1977).

² The Housing Discrimination Study (HDS1989).

³ The Housing Discrimination Study (HDS2000).

for the frequency with which differences in test outcomes arise for random reasons. This net measure will be too small if agents systematically steer White homeseekers away from particular units (Ross, 2002). The gross measure, which is the share of White-favored tests, has often been referenced to address this concern, and gross measures of discrimination are typically much larger than net measures. Even in a world with no discrimination, however, the gross measure could be quite large due to random differences in the circumstances of the testers' visits. Finally, paired-testing studies of housing discrimination have almost always tested for discrimination using a sample of marketwide advertisements for housing as the entry point into the housing market, but such studies might understate discrimination if the agents who tend to discriminate also tend not to advertise available housing or tend not to advertise housing in "protected" neighborhoods (Yinger, 1995).

Another article in this symposium summarizes the HDS2012 methodology and presents the study's core findings. This article reviews the findings and reanalyzes the data to consider three significant concerns raised regarding paired-testing studies of housing discrimination: (1) the exclusion of minority homeseekers during the process of setting up appointments, (2) the net measure of adverse treatment understating discrimination because some housing units are systematically not shown to White testers, and (3) the use of metropolitanwide advertisements that may systematically underrepresent neighborhoods where discrimination is higher. Our discussion mostly focuses on the net measure of adverse treatment in which one of the three concerns is that the net measure may understate discrimination relative to the gross measure. In the main body of the article, we first present the basic estimates of treatment patterns from HDS2012 rental tests and then present the sales test results. HDS2012 directly addresses the first concern raised previously in that the ability to make an appointment over the phone is tracked and measures of adverse treatment are developed based on this outcome. In rental markets, no differences are observed in the frequency of White-favored and minority-favored tests (net measure) on the ability to make an appointment for any group. In sales markets, statistically significant differences are observed for White-Black tests in the ability to make an appointment, but these differences are small—less than 3 percentage points—and no differences were observed for Anglo-Hispanic⁴ or White-Asian tests (Turner et al., 2013). The study findings suggest that the inability of minority testers to obtain an appointment is not a major source of bias in measuring housing discrimination.

The evidence on the second concern—bias in the net measure—is mixed but, in our opinion, tends to support the use of the net measure. In HDS2012, the likelihood of equal treatment was quite high for both obtaining an appointment and the availability of housing. Therefore, for obtaining an appointment and, in the case of rental housing, for availability, there was virtually no room for any systematic favoring of minority testers, and the observed net measures are quite small due to the high rate of equal treatment. On the other hand, HDS2012 found larger net differences in the likelihood of the White tester versus the minority tester having either more units available or inspecting more units for all three groups in the rental tests and for Black and Asian testers in the sales tests. Further, these measures have substantial room for differences between the net and gross measures, especially in sales tests in which the gross measure is often 30 to 40 percentage points higher than the net measure.

⁴ *Anglo* traditionally is used to refer to non-Hispanic White testers in the context of tests involving Hispanic testers, even though all tests use non-Hispanic White testers as the majority group.

Therefore, we turn to earlier studies to further address the second question. Ondrich, Ross, and Yinger (2000), using data from HDS1989, found direct evidence that White homeseekers are sometimes systematically favored, but only for one of the many measures of treatment considered. HDS2000 found evidence that steering against Black homeseekers increased between HDS1989 and HDS2000 (Ross and Turner, 2005) and the incidence of steering against Black and Asian homeseekers in HDS2012 was similar in magnitude (Turner et al., 2013), but the incidence of steering in all cases was far too small to explain the large observed differences between net and gross measures for number of units available or number of units inspected. Finally, a pilot study of three-person tests in which two of the three testers were the same race was conducted at two sites in 2002. In those tests, randomness was assessed by comparing same-race pairs, and that pilot study did not find any evidence that the net measure understated discrimination (Turner and Ross, 2003b).

The final concern that paired-testing studies of this type understate discrimination is that they are forced to rely on publically available, marketwide advertisements, either in the newspaper or on the Internet. In HDS2012, the advertisement selection process for testing is designed to represent the stock of rental and owner-occupied housing in each metropolitan area. The resulting tests, however, may understate discrimination either if discriminatory landlords are not represented in metropolitanwide advertising sources or if the stock of housing is not representative of the current market for available housing. At present, little direct evidence exists on this last concern. The one exception is a limited sampling of alternative neighborhood-level advertisements at the larger sites of HDS2000 (Turner and Ross, 2003b). The level of adverse treatment in the alternative sample of White-Black rental tests was not systematically higher than the traditional newspaper sampling based on estimates in HDS2000, but the alternative sample of White-Black sales tests did exhibit substantially higher levels of adverse treatment on the availability and inspection of units.

To partially address this final concern, we use data from the 2011 American Community Survey (ACS) to reweight the tests conducted in HDS2012 and so more accurately represent the population of available housing units. We specifically identify new residents in rental (or owner-occupied) housing within the past year as evidence that a rental (owner-occupied) housing unit turned over and was available for rent (sale) during that year. Using this proxy for available rental or owner-occupied housing, we estimate the number of available rental and owner-occupied housing units in each Public Use Microdata Area (PUMA) at each metropolitan site for HDS2012. For both rental and sales tests, we compare these shares with the shares of tests for each PUMA in each site and then develop weights that give higher weights to tests in PUMAs with more available rental housing. We use these weights to develop measures of adverse treatment that are representative of treatment in the sample of available housing units, as opposed to the housing stock or the sample of advertised units. We conduct this analysis for White-Black, Anglo-Hispanic, and White-Asian rental tests over the seven main measures⁵ of adverse treatment in the rental market and for four key measures⁶ for the owner-occupied market that were presented in HDS2012. This exercise

⁵ The seven main measures are (1) tester(s) able to make an appointment, (2) tester(s) told any units available, (3) one tester told about more units than partner, (4) average number of units available (per visit), (5) average rent, (6) one tester inspected more units than partner, and (7) average number of units inspected (per visit).

⁶ The four key measures are (1) tester(s) able to make an appointment, (2) tester(s) told any units available, (3) one tester told about more units than partner, and (4) average number of units available (per visit).

naturally cannot address concerns that landlords who intend to discriminate strategically choose not to advertise housing in metropolitanwide venues, nor can it capture variation in adverse treatment that arises at lower levels of geography.

Our analysis of HDS2012 suggests that estimates of housing discrimination are unaffected by the broad geographic distribution of tests across PUMAs within metropolitan areas. We find virtually no difference between the measures of adverse treatment from HDS2012 and alternative measures using the weights created for this article. The descriptive statistics suggest that weights vary dramatically across PUMAs and within metropolitan sites. We also measure, however, the average neighborhood characteristics of the metropolitan site using census tract measures of average housing price, median income, share Black, share Hispanic, and share owner-occupied using HDS2012 tests with and without our new weights. Across the six samples and the many neighborhood variables, we find very little differences for rental tests and modest differences for sales tests, with the sales tests in HDS2012 underrepresenting high-income, lower minority share neighborhoods.

Methods

HDS2012 examines housing discrimination through the lens of the available housing that appears in paid, metropolitanwide advertisements. In the case of HDS2012, the rental advertisements appear on line in craigslist, apartments.com, rent.com, and similar websites; sales advertisements were drawn from sites like zillow.com. Tests were based on advertisements that were selected with probabilities based on the stock of available housing in each market. In this section, we describe our approach for reweighting the data to represent the population of available housing during the year of the study; in the results section, we discuss the broader set of evidence available from HDS2012.

HDS2012 uses a two-stage sampling process in which, first, metropolitan areas are selected with probability associated with their minority population, and then advertisements are selected within each site. Metropolitan areas were organized into subsets or strata based on their populations of each minority group. The areas with the largest representation of each minority group were selected with certainty for the tests associated with that group, and then a set number of metropolitan areas were randomly selected from each of the remaining strata. The original weights were created solely to address the sampling of metropolitan area sites. Each test i in each metropolitan area s is then assigned a weight that is equal to the inverse of the selection probability (P_s) times the inverse of the number of tests in the site (N_s), or—

$$W_{is} = \frac{1}{P_s N_s} , \tag{1}$$

so that each metropolitan area has a total weight over all tests equal to the inverse of the selection probability. Note that the selection probability for a site is proportional to the site's share of the minority group being tested in that particular stratum.

Our weights are designed to leave the total weight associated with a site unchanged but to allow tests that must represent a larger number of available units geographically to have higher weights. We specifically exploit a lower level of geography within each metropolitan area—the PUMA—and develop separate weights for all tests in each PUMA. We create this weight by dividing the number

of tests in PUMA p (n_{ps}) by the number of estimated available units in PUMA p (a_{ps}) and then scaling the weight so that the total weight associated with a metropolitan area is unchanged. To be specific, the new weight is described by—

$$\tilde{W}_{ips} = (W_{is}N_s) \frac{a_{ps}}{n_{ps} A_s} = \frac{a_{ps}}{n_{ps} P_s A_s} \quad (2)$$

where the first term in the expression is the original weight associated with the metropolitan area, the second term is the ratio of number of tests to number of available units in a PUMA, and the third term is simply the sum of the second term over all tests in a metropolitan area, or—

$$A_s = \sum_{i=1}^{N_s} \frac{a_{ps}}{n_{ps}} \quad (3)$$

The logic behind this equation is as follows.

1. Each test receives a weight based on its PUMA, which is the ratio of number of available units divided by number of tests.
2. This ratio is divided by the sum of the ratio over all tests in the site so that the sum of weights for each site is 1.
3. The resulting number is then multiplied by the original weight assigned to each site, 1 over the probability of selection.

Adverse treatment is then defined using the traditional net measure,

$$Net[y_{iw}, y_{im} | \tilde{W}_{ips}] = Fr[y_{iw} = 1, y_{im} = 0 | \tilde{W}_{ips}] - Fr[y_{iw} = 0, y_{im} = 1 | \tilde{W}_{ips}] \quad (4)$$

for discrete outcomes, or—

$$Net[y_{iw}, y_{im} | \tilde{W}_{ips}] = Mean[y_{iw} | \tilde{W}_{ips}] - Mean[y_{im} | \tilde{W}_{ips}] \quad (5)$$

for continuous measures of treatment, where Fr is the empirical frequency and both the means and the frequencies are weighted based on \tilde{W}_{ips} .

Data

Our analysis begins with the data arising from HDS2012. As noted earlier, HDS2012 begins by selecting a sample of sites from specific strata. The details of site selection are in Turner et al. (2013). Exhibit 1 presents the selected sites by their strata. The four sites in the large minority strata are selected with certainty for all three sets of tests. The next set of sites is selected with certainty for White-Black and Anglo-Hispanic tests, and the final set of four sites is selected with certainty only for Anglo-Hispanic tests. The last three strata are sites selected for their Black, Hispanic, or Asian representation only. To economize on administrative overhead associated with adding additional sites, however, all groups had sites selected randomly from every other stratum based on their own group's representation in each site. This strategy maximized the number of sites in which multiple groups were tested. White-Asian tests were conducted in three of the sites selected with certainty for both the Black and Hispanic and the Hispanic strata, and White-Black tests were conducted in three of the sites selected with certainty for the Hispanic stratum. As stated previously, the subset of sites selected randomly for one group from another group's stratum was selected based on its own group's representation.

Exhibit 1

HDS Site Selection and Test Count

Strata	FIPS Codes: Site Names	Rental Tests			Sales Tests		
		Black	Hispanic	Asian	Black	Hispanic	Asian
Large minority	5600: New York, NY-Northeastern NJ	135	135	60	43	43	49
	1600: Chicago-Gary-Lake, IL	135	135	60	43	43	49
	4480: Los Angeles-Long Beach, CA	135	135	60	43	43	49
	3360: Houston-Brazoria, TX	135	135	60	43	43	49
Large Black and Hispanic	8840: Washington, DC/MD/VA	135	46	46	43	43	49
	6160: Philadelphia, PA/NJ	135	46	46	43	43	49
	520: Atlanta, GA	135	46	46	43	43	49
	2160: Detroit, MI	135	46	NA	43	43	NA
Large Hispanic	5000: Miami-Hialeah, FL	47	135	NA	43	43	NA
	6780: Riverside-San Bernardino, CA	47	135	47	43	43	49
	1920: Dallas-Fort Worth, TX	47	135	47	43	43	49
	7320: San Diego, CA	NA	135	47	NA	43	49
Black	5605: Newark, NJ	44	44	46	43	43	49
	8280: Tampa-St. Petersburg-Clearwater, FL	44	44	46	43	43	49
	3760: Kansas City, MO/KS	44	44	46	43	43	49
	3120: Greensboro-Winston Salem-High Point, NC	44	44	NA	43	43	NA
Hispanic	1680: Cleveland, OH	44	44	46	43	43	49
	720: Baltimore, MD	44	44	46	43	43	49
	6760: Richmond-Petersburg, VA	44	NA	NA	43	NA	NA
	4482: Orange County, CA	44	44	46	43	43	49
Asian	7240: San Antonio, TX	44	44	46	43	43	49
	1921: Fort Worth-Arlington, TX	44	44	46	43	43	49
	200: Albuquerque, NM	44	44	46	43	43	49
	7400: San Jose, CA	NA	44	47	NA	43	49
Small Black	1120: Boston, MA	44	44	47	43	43	49
	7600: Seattle-Everett, WA	44	44	47	43	43	49
	5604: Middlesex-Somerset-Hunterdon, NJ	44	44	47	43	43	49
	1760: Columbia, SC	44	NA	47	43	NA	49
Total tests		1,881	1,880	1,121	1,118	1,118	1,127

FIPS = Federal Information Processing Standard. HDS = Housing Discrimination Study. NA = no tests conducted.

We then turn to the public use microdata of the 2011 ACS to identify the degree of turnover in rental and owner-occupied housing within metropolitan areas. The public use microdata identifies the location of housing units within PUMAs that are estimated to contain a minimum of 100,000 individuals. The ACS is sent to nearly 3 million addresses each year, and so it constitutes the only survey of housing with a sufficient number of observations to accurately characterize the population and housing below the metropolitan area level in all U.S. metropolitan areas. Whereas the census requires 5 years of the ACS to develop statistics at the census tract level, PUMAs are 10 to 25 times larger than a typical census tract, and so 1 year of the ACS should be sufficient for our purposes. To identify available rental or owner-occupied housing, we use the moved-last-year variable in the ACS to identify all housing units for which every resident of that unit lived at a different address in the preceding year. ACS sampling weights are then used to construct the estimated number of available units in each PUMA by tenure.

Exhibit 1 also presents the number of tests conducted at each site. It is notable that, in the rental tests, the very largest metropolitan areas are oversampled to obtain site-specific estimates. Such oversampling is not conducted in the sales tests, and, as a result, in the very largest metropolitan areas, the total number of tests provides very limited coverage across the PUMAs. Exhibit 2 presents the number of PUMAs in each site and the average number of tests per PUMA for the selected sites. In New York, Chicago, and Los Angeles, the number of sales tests per PUMA is always less than one, but it is more than two for the Black and Hispanic rental tests. The larger samples of tests required for site-specific estimates were also conducted for several additional large Black and Hispanic rental sites providing better coverage across those sites. Although less extensive, additional Asian rental tests were also conducted for the very largest Asian sites.

It is important to note that the small number of tests is more a concern about the general results arising from HDS2012, as opposed to a concern about the exercise conducted in this article. If systematic variation in adverse treatment exists across locations, HDS2012 measures will suffer from spatial error or noise arising from the small number of tests in each site. To be specific, if variation in discrimination across neighborhoods exists within metropolitan areas, the area estimates on which the national estimates are based may have considerable measurement error because the number of tests per site is too small to accurately cover the many distinct regions or neighborhoods in each site. The corrected weights are based on turnover in a broad sample of rental housing units in each PUMA, and so they provide a quite accurate indication of the relative turnover in each location. The number of tests in each PUMA is equal to the number of observations in each PUMA by definition. Therefore, the new weights provide a very accurate mapping from the information generated by the tests in each site to a population of available rental housing units, with one exception discussed in the next paragraph. Therefore, the implication of the small number of tests is not about bias, but rather that the reweighting corrects for two problems: (1) the potential systematic undersampling of some PUMAs relative to the amount of available rental housing and (2) the random spatial error or noise added to HDS2012 measures due to the relatively sparse number of tests across each site.

The one exception that creates bias in the proposed weighting scheme is that some PUMAs in large metropolitan statistical areas may have no tests, and so those PUMAs must be ignored in any measure of adverse treatment. Of course, discrimination in those PUMAs was also omitted

Exhibit 2

Site PUMAs and Tests per PUMA

Strata	FIPS Codes: Site Names	PUMA	Rental Tests			Sales Tests		
			Black	Hispanic	Asian	Black	Hispanic	Asian
Large minority	5600: New York, NY-Northeastern NJ	64	2.109375	2.109375	0.9375	0.671875	0.671875	0.765625
	1600: Chicago-Gary-Lake, IL	59	2.288136	2.288136	1.016949	0.728814	0.728814	0.830508
	4480: Los Angeles-Long Beach, CA	67	2.014925	2.014925	0.895522	0.641791	0.641791	0.731343
	3360: Houston-Brazoria, TX	32	4.21875	4.21875	1.875	1.34375	1.34375	1.53125
	8840: Washington, DC/MID/VA	33	4.090909	1.393939	1.393939	1.30303	1.30303	1.484848
	6160: Philadelphia, PA/NJ	37	3.648649	1.243243	1.243243	1.162162	1.162162	1.324324
Large Black and Hispanic	520: Atlanta, GA	32	4.21875	1.4375	1.4375	1.34375	1.34375	1.53125
	2160: Detroit, MI	32	4.21875	1.4375	NA	1.34375	1.34375	NA
	5000: Miami-Hialeah, FL	19	2.473684	7.105263	NA	2.263158	2.263158	NA
	6780: Riverside-San Bernardino, CA	20	2.35	6.75	2.35	2.15	2.15	2.45
	1920: Dallas-Fort Worth, TX	23	2.043478	5.869565	2.043478	1.869565	1.869565	2.130435
	7320: San Diego, CA	16	NA	8.4375	2.9375	NA	2.6875	3.0625
Black	5605: Newark, NJ	16	2.75	2.75	2.875	2.6875	2.6875	3.0625
	8280: Tampa-St. Petersburg-Clearwater, FL	20	2.2	2.2	2.3	2.15	2.15	2.45
	3760: Kansas City, MO/KS	13	3.384615	3.384615	3.538462	3.307692	3.307692	3.769231
	3120: Greensboro-Winston Salem-High Point, NC	9	4.888889	4.888889	NA	4.777778	4.777778	NA
	1680: Cleveland, OH	19	2.315789	2.315789	2.421053	2.263158	2.263158	2.578947
	720: Baltimore, MD	22	2	2	2.090909	1.954545	1.954545	2.227273
Hispanic	6760: Richmond-Petersburg, VA	5	8.8	NA	NA	8.6	NA	NA
	4482: Orange County, CA	17	2.588235	2.588235	2.705882	2.529412	2.529412	2.882353
	7240: San Antonio, TX	12	3.666667	3.666667	3.833333	3.583333	3.583333	4.083333
	1921: Fort Worth-Arlington, TX	12	3.666667	3.666667	3.833333	3.583333	3.583333	4.083333
	200: Albuquerque, NM	6	7.333333	7.333333	7.666667	7.166667	7.166667	8.166667
	7400: San Jose, CA	14	NA	3.142857	3.357143	NA	3.071429	3.5
Asian	1120: Boston, MA	33	1.333333	1.333333	1.424242	1.30303	1.30303	1.484848
	7600: Seattle-Everett, WA	20	2.2	2.2	2.35	2.15	2.15	2.45
	5604: Middlesex-Somerset-Hunterdon, NJ	8	5.5	5.5	5.875	5.375	5.375	6.125
	1760: Columbia, SC	4	11	NA	NA	10.75	NA	NA

FIPS = Federal Information Processing Standard. NA = no tests conducted. PUMA = Public Use Microdata Area.

by definition from the traditional estimates. Exhibit 3 presents the total number of tests for each site by rental (column 1) and by sales (column 3) and also the number of tests used in calculating specific treatment variables. To be specific, most of the treatments considered by HDS2012 are observed only if both testers make it relatively far into the process, which happens more frequently for rental tests. For example, whether the advertised unit is inspected conditional on units being available is an important treatment variable that is observed for approximately 85 percent of rental tests but for only between 60 and 70 percent of sales tests. Finally, columns 2 and 4 show the share of PUMAs for each set of treatments that can be included in the revised measures of adverse treatment because at least one test reached this stage of the process in that PUMA. For White-Black and Anglo-Hispanic rental tests, the samples of tests always cover at least 70 percent of the PUMAs in the sample of sites. The initial percentage for White-Asian rental tests is smaller, at 60 percent, but it never falls below 55 percent. For sales tests, weights provide coverage of less than 50 percent of the PUMAs for many of the treatment variables. Therefore, we conduct this exercise only for the four treatment indicators from the sales tests that are calculated for either the entire sample of tests or the sample of tests in which both testers were able to meet with an agent. For these four treatments, at least 50 percent of the PUMAs have tests in all samples.

Exhibit 3

Test Subsamples and Within-Site Representativeness

	Rental Tests		Sales Tests	
	Sample Size	Percent of PUMAs	Sample Size	Percent of PUMAs
White-Black Tests				
Full sample	2,009	74.8	1,244	69.4
If able to meet with an agent	1,813	72.4	1,072	57.4
If available units recommended	1,710	71.8	800	50.9
If unit inspected			441	35.4
Anglo-Hispanic Tests				
Full sample	1,986	76.1	1,193	70.6
If able to meet with an agent	1,775	73.8	1,043	58.7
If available units recommended	1,654	72.6	737	48.6
If unit inspected			432	33.3
White-Asian Tests				
Full sample	1,150	60.1	1,170	63.3
If able to meet with an agent	1,037	57.2	1,047	52.3
If available units recommended	968	55.9	799	47.1
If unit inspected			463	33.7

PUMA = Public Use Microdata Area.

Note: The panels represent the groups being tested, the sample size is the number of tests for each group in each market, and the percent of PUMAs is the fraction of PUMAs that contain at least one test.

Results

Exhibits 4 through 6 present the rental market estimates of adverse treatment of Black, Hispanic, and Asian testers, respectively, relative to their White counterparts. The first panel of each table presents the original HDS2012 estimates, and the second panel presents the estimates reweighted to represent

the availability of rental housing throughout each site. The rows in each panel represent the key treatment variables beginning with whether testers were able to make an appointment. If both testers were able to make an appointment, the following treatment variables are considered: whether the tester was told about units available, whether the tester was told about more available units than the other tester, and the number of units available. If both testers learn about available units, the final set of treatment variables examined are average rent, whether one tester inspected more units than the other, and the number of units inspected. The first column identifies the fraction of tests in which either both testers received favorable treatment or learned about or saw the same number of units. The next two columns identify the fraction of tests in which either the majority or minority tester was treated favorably, and the fourth column presents the differences in those two columns, or the net measure of adverse treatment. The final column presents the confidence with which the net measure can be reported as differing from zero, indicating evidence of discrimination.

In the first panel, we find no significant differences for any minority group in the likelihood of obtaining an appointment, which is significant, because this treatment was not captured in HDS2000, and the inability to obtain an appointment represented a potential source of bias in many earlier housing discrimination studies. The differences for rental tests shown in exhibits 4, 5, and 6 are always less than 0.5 percent and never significant. The second important observation to draw from the first panels in exhibits 4, 5, and 6 is that moderate improvements continue in the number of testers receiving equal treatment in the rental market. The share of tests in which both testers obtained an appointment and, if an appointment was obtained, the share of tests for which rental housing was available to both testers is always about 95 percent, which leaves very little room for differential treatment of any kind. Therefore, it is unlikely that net measures in this area are understated because minority testers are sometimes systematically favored. These changes represent substantial improvements for Black and Hispanic testers in the likelihood of equal treatment relative to HDS2000. In HDS2000, 20 percent of White-Black tests had differences between testers in the availability of the advertised unit and almost 30 percent of White-Black tests had differences in the availability of similar units; for Anglo-Hispanic tests, the share of tests with differences were 17 and 24 percent of tests for advertised and similar units, respectively (Ross and Turner, 2005). For White-Asian tests, the percent of tests with differences in HDS2000 were 15 and 22 percent of tests. Although the incidence of equal treatment is not as high on the number of units available or inspected, Black and Hispanic testers still show notable improvements, with Black testers' frequency of equal treatment on number of units available and inspected rising from 49 to 54 and from 60 to 69, respectively, and Hispanic testers' frequency rising from 50 to 65 and from 66 to 69, respectively.

Nonetheless, significant levels of discrimination remain for all three groups. Minority testers are more likely to be told about fewer available units or to inspect fewer units than their White counterparts, with net differences of 9.0, 12.8, and 8.8 percentage points on availability and of 2.8, 6.0, and 5.5 percentage points on inspection, respectively, for Black, Hispanic, and Asian testers. These HDS2012 differences compare to HDS2000 differences for Black, Hispanic, and Asian testers, respectively, in number of available units of 6.2, 8.9, and 3.9 and in inspected units of 6.8, 6.1, and -4.8 (Asian favored). Observed discrimination in the rental market appears to be somewhat higher in HDS2012 in terms of the number of housing units available. Finally, in HDS2012, Black and Hispanic testers are quoted slightly higher rents than their White counterparts, \$4 and \$6 per month difference respectively.

Exhibit 4

White-Black Measures of Adverse Treatment in Rental Housing

	Both	White	Black	Difference	p-Value
Rental market treatment measures (original weights)					
Appointment					
Information and availability	94.6%	1.0%	0.6%	0.4%	0.26
Tester(s) able to make an appointment <i>If able to meet with an agent:</i>					
Tester(s) told any units available	95.0%	3.0%	2.1%	0.9%	0.20
One tester told about more units than partner	54.1%	27.5%	18.5%	9.0%	0.00
Average number of units available (per visit) <i>If available units recommended:</i>					
Average rent		\$1,122	\$1,126	-\$4	0.06
One tester inspected more units than partner	69.1%	16.9%	14.1%	2.8%	0.03
Average number of units inspected (per visit)		1.28	1.23	0.04	0.04
Inspections					
Rental market treatment measures (available units)					
Appointment					
Information and availability	93.7%	1.0%	0.3%	0.6%	0.06
Tester(s) able to make an appointment <i>If able to meet with an agent:</i>					
Tester(s) told any units available	95.4%	2.8%	1.8%	1.0%	0.22
One tester told about more units than partner	55.7%	26.9%	17.5%	9.4%	0.00
Average number of units available (per visit) <i>If available units recommended:</i>					
Average rent		1.78	1.59	0.19	0.00
One tester inspected more units than partner	69.0%	16.1%	14.9%	1.2%	0.64
Average number of units inspected (per visit)		1.24	1.21	0.03	0.35

Notes: The first panel presents the results from the 2012 Housing Discrimination Study report, and the second panel presents results that are representative of available owner-occupied housing. The column labeled "Both" contains the share of tests in which equal treatment occurred. The column labeled by race or ethnicity represents either the fraction of tests in which that group was favored or the average of a continuous variable. The column labeled "Difference" is the difference between the preceding two columns. The column labeled "p-Value" represents the statistical significance of this difference.

Exhibit 5

Anglo-Hispanic Measures of Adverse Treatment in Rental Housing

	Both	Anglo	Hispanic	Difference	p-Value
Rental market treatment measures (original weights)					
Appointment					
Information and availability	96.1%	0.4%	0.2%	0.2%	0.52
Tester(s) able to make an appointment <i>If able to meet with an agent:</i>					
Tester(s) told any units available	94.6%	3.6%	1.8%	1.8%	0.04
One tester told about more units than partner	54.7%	29.1%	16.2%	12.8%	0.00
Average number of units available (per visit)		1.82	1.60	0.21	0.00
<i>If available units recommended:</i>					
Average rent		\$1,291	\$1,297	-\$6	0.01
One tester inspected more units than partner	68.1%	18.9%	12.9%	6.0%	0.00
Average number of units inspected (per visit)		1.40	1.33	0.07	0.00
Inspections					
Rental market treatment measures (available units)					
Appointment					
Information and availability	95.9%	0.4%	0.2%	0.3%	0.46
Tester(s) able to make an appointment <i>If able to meet with an agent:</i>					
Tester(s) told any units available	94.9%	3.4%	1.7%	1.7%	0.03
One tester told about more units than partner	56.1%	28.3%	15.6%	12.7%	0.00
Average number of units available (per visit)		1.81	1.58	0.23	0.00
<i>If available units recommended:</i>					
Average rent		\$1,277	\$1,281	-\$4	0.08
One tester inspected more units than partner	68.5%	18.9%	12.6%	6.4%	0.00
Average number of units inspected (per visit)		1.40	1.31	0.08	0.00

Notes: The first panel presents the results from the 2012 Housing Discrimination Study report, and the second panel presents results that are representative of available owner-occupied housing. The column labeled "Both" contains the share of tests in which equal treatment occurred. The column labeled by race or ethnicity represents either the fraction of tests in which that group was favored or the average of a continuous variable. The column labeled "Difference" is the difference between the preceding two columns. The column labeled "p-Value" represents the statistical significance of this difference.

Exhibit 6

White-Asian Measures of Adverse Treatment in Rental Housing

	Both	White	Asian	Difference	p-Value
Rental market treatment measures (original weights)					
Appointment					
Information and availability					
Tester(s) able to make an appointment <i>If able to meet with an agent:</i>	96.0%	0.4%	0.1%	0.4%	0.35
Tester(s) told any units available	94.9%	2.7%	2.4%	0.3%	0.77
One tester told about more units than partner	54.9%	27.0%	18.1%	8.8%	0.01
Average number of units available (per visit) <i>If available units recommended:</i>		1.79	1.63	0.17	0.00
Average rent		\$1,391	\$1,389	\$2	0.71
One tester inspected more units than partner	64.8%	20.4%	14.9%	5.5%	0.03
Average number of units inspected (per visit)		1.44	1.36	0.08	0.01
Inspections					
Rental market treatment measures (available units)					
Appointment					
Information and availability					
Tester(s) able to make an appointment <i>If able to meet with an agent:</i>	96.4%	0.3%	0.1%	0.3%	0.34
Tester(s) told any units available	95.1%	2.8%	2.0%	0.8%	0.25
One tester told about more units than partner	55.5%	25.8%	18.7%	7.1%	0.06
Average number of units available (per visit) <i>If available units recommended:</i>		1.75	1.63	0.12	0.02
Average rent		\$1,376	\$1,373	\$3	0.27
One tester inspected more units than partner	65.7%	19.5%	14.8%	4.7%	0.15
Average number of units inspected (per visit)		1.44	1.37	0.06	0.16

Notes: The first panel presents the results from the 2012 Housing Discrimination Study report, and the second panel presents results that are representative of available owner-occupied housing. The column labeled "Both" contains the share of tests in which equal treatment occurred. The column labeled by race or ethnicity represents either the fraction of tests in which that group was favored or the average of a continuous variable. The column labeled "Difference" is the difference between the preceding two columns. The column labeled "p-Value" represents the statistical significance of this difference.

The calculations in the second panel of exhibits 4 through 6 show virtually no systematic difference between the estimates using the original weights and the results using weights based on within metropolitan turnover or the availability of rental housing. For example, the net measure for which a tester was told about more available units rises from 9.0 to 9.4 percentage points, falls from 12.8 to 12.7 percentage points, and falls from 8.8 to 7.1 percentage points for Black, Hispanic, and Asian testers, respectively. For being shown more units, the net measure falls from 2.8 to 1.2 percentage points, rises from 6.0 to 6.4 percentage points, and falls from 5.5 to 4.7 percentage points for these three groups. All these changes are substantially smaller than the standard errors associated with the estimates of net adverse treatment. We find no evidence that PUMAs that were undertested in HDS2012 relative to the amount of available rental housing have systematically higher levels of adverse treatment against minority individuals who are seeking rental housing.

Exhibits 7 through 9 present the sales market estimates of adverse treatment of Black, Hispanic, and Asian testers, respectively. As in exhibits 4 through 6, the top panel of each table presents the traditional estimates, and the bottom panel presents the reweighted estimates. The first rows in each panel present estimates for whether testers were able to make an appointment. If both testers were able to make an appointment, the following treatment variables are considered: whether the tester was told about units available, whether the tester was told about more available units than the other tester, and the number of units available. In order, the columns present the fraction of tests in which either both testers received favorable treatment or learned about or saw the same number of units, the fraction of tests in which either the majority or minority tester was treated favorably, and the differences in those two columns. The final column presents the confidence with which the net measure can be reported as differing from zero, indicating evidence of discrimination.

As in the rental market, differences in the likelihood of obtaining an appointment could have had, at most, modest impacts on the measured incidence of discrimination on other treatments. For White-Black tests, the net measure for obtaining an appointment is 2.4 percent and is statistically significant. Even if all these landlords discriminated on the key variables such as being told about more homes (net of 12.4 percent of tests) or inspecting more homes (net of 9.3 percent of tests), however, this finding would imply relatively small increases in the measured incidence of discrimination on White-Black sales tests (for example, raising net measures to 14.8 and 11.7 percent). The net differences in obtaining an appointment are substantially smaller and statistically insignificant for the Anglo-Hispanic and White-Asian sales tests.

The net differences in adverse treatment on availability and number of units are typically insignificant. Net differences in whether the tester saw or inspected at least one unit are insignificant for all three groups, and net differences in whether the White tester saw more units is significant only for the White-Black tests. The incidence of equal treatment on the availability of housing, however, is significantly lower—below 85 percent for all three groups—in the sales market as compared with the rental market. Therefore, this market has more room for gross differences in adverse treatment and for the possibility that the net measure understates discrimination because White testers are systematically not told about housing that is being made available to minority testers. Further, although not shown in the exhibits, net differences also exist for Black and Asian testers in terms of being steered away from neighborhoods with higher shares of White residents, of 5.0 and 5.9 percentage points. On the other hand, the differences in steering are much smaller than the fraction

Exhibit 7

White-Black Measures of Adverse Treatment in Owner-Occupied Housing

	Both	White	Black	Difference	p-Value
Sales market treatment measures (original weights)					
Appointment					
Tester(s) able to make an appointment	88.3%	3.4%	1.1%	2.4%	0.01
<i>If able to meet with an agent:</i>					
Tester(s) told any units available	78.9%	9.6%	7.4%	2.2%	0.14
One tester told about more units than partner	21.2%	46.1%	32.6%	13.5%	0.00
Average number of units available (per visit)		3.41	2.90	0.51	0.00
Rental market treatment measures (available units)					
Appointment					
Tester(s) able to make an appointment	89.2%	3.8%	0.6%	3.2%	0.02
<i>If able to meet with an agent:</i>					
Tester(s) told any units available	79.7%	9.0%	7.1%	1.8%	0.35
One tester told about more units than partner	22.1%	46.0%	31.9%	14.1%	0.01
Average number of units available (per visit)		3.33	2.86	0.46	0.06

Notes: The first panel presents the results from the 2012 Housing Discrimination Study report, and the second panel presents results that are representative of available owner-occupied housing. The column labeled "Both" contains the share of tests in which equal treatment occurred. The column labeled by race or ethnicity represents either the fraction of tests in which that group was favored or the average of a continuous variable. The column labeled "Difference" is the difference between the preceding two columns. The column labeled "p-Value" represents the statistical significance of this difference.

Exhibit 8

Anglo-Hispanic Measures of Adverse Treatment in Owner-Occupied Housing

	Both	Anglo	Hispanic	Difference	p-Value
Sales market treatment measures (original weights)					
Appointments	92.8%	2.2%	1.8%	0.4%	0.68
Information and availability					
Tester(s) able to make an appointment <i>If able to meet with an agent:</i>					
Tester(s) told any units available	76.7%	8.8%	9.0%	- 0.2%	0.93
One tester told about more units than partner	22.4%	39.9%	37.7%	2.3%	0.63
Average number of units available (per visit)		3.04	2.76	0.28	0.21
Rental market treatment measures (available units)					
Appointments	94.9%	1.7%	0.9%	0.8%	0.32
Information and availability					
Tester(s) able to make an appointment <i>If able to meet with an agent:</i>					
Tester(s) told any units available	77.0%	8.6%	9.3%	- 0.7%	0.84
One tester told about more units than partner	22.8%	39.8%	37.4%	2.4%	0.61
Average number of units available (per visit)		2.93	2.68	0.25	0.22

Notes: The first panel presents the results from the 2012 Housing Discrimination Study report, and the second panel presents results that are representative of available owner-occupied housing. The column labeled "Both" contains the share of tests in which equal treatment occurred. The column labeled by race or ethnicity represents either the fraction of tests in which that group was favored or the average of a continuous variable. The column labeled "Difference" is the difference between the preceding two columns. The column labeled "p-Value" represents the statistical significance of this difference.

Exhibit 9

White-Asian Measures of Adverse Treatment in Owner-Occupied Housing

	Both	White	Asian	Difference	p-Value
Sales market treatment measures (original weights)					
Appointment					
Tester(s) able to make an appointment	92.2%	1.6%	2.6%	- 1.0%	0.56
<i>If able to meet with an agent:</i>					
Tester(s) told any units available	79.5%	9.1%	7.5%	1.6%	0.51
One tester told about more units than partner	20.9%	44.1%	35.0%	9.2%	0.13
Average number of units available (per visit)		3.36	2.81	0.55	0.08
Rental market treatment measures (available units)					
Appointment					
Tester(s) able to make an appointment	92.4%	1.1%	2.7%	- 1.6%	0.40
<i>If able to meet with an agent:</i>					
Tester(s) told any units available	80.6%	8.7%	6.9%	1.9%	0.53
One tester told about more units than partner	21.6%	44.9%	33.5%	11.3%	0.09
Average number of units available (per visit)		3.34	2.68	0.66	0.06

Notes: The first panel presents the results from the 2012 Housing Discrimination Study report, and the second panel presents results that are representative of available owner-occupied housing. The column labeled "Both" contains the share of tests in which equal treatment occurred. The column labeled by race or ethnicity represents either the fraction of tests in which that group was favored or the average of a continuous variable. The column labeled "Difference" is the difference between the preceding two columns. The column labeled "p-Value" represents the statistical significance of this difference.

of White-favored tests in terms of number of units available and inspected, which ranges between 40 and 46 percent. Therefore, a substantial fraction of minority-favored tests on availability and inspection likely arises due to random differences in the circumstances of the tester's visit.

The calculations in the second panel of exhibits 7 through 9 show virtually no systematic differences between the estimates using the original weights and the results using the revised weights. The net measures for appointment and having a unit available for the three samples remain consistently small. The net measure for which tester was told about more available units rises from 13.5 to 14.1 percentage points, rises from 2.3 to 2.4 percentage points, and rises from 9.2 to 11.3 percentage points for Black, Hispanic, and Asian testers, respectively. Only the change for Asian testers is appreciable in magnitude, and those estimates are very noisy and statistically insignificant, even though the point estimate of net adverse treatment is about 10 percent.

The distribution of weights is illustrated by presenting the distribution of the ratio of the new weights to the original weights. Because the old weights are constant for all tests in a metropolitan area, this ratio illustrates the level of variation in weights within each site. The results for rental and owner-occupied housing are shown in exhibits 10 and 11, respectively. The three panels present the distribution of the within-metropolitan-area weights for each of the three groups, in order, White-Black, Anglo-Hispanic, and White-Asian tests. The rows in each panel represent the weights for various subsamples, because the number of tests varies across the treatment variables and, as a result, the weights vary across the variables. The first row is the full sample for which we observe whether testers were able to meet with an agent, the second row is the subsample in which both testers were able to meet with an agent and we observe the availability of units, and the third (rental only) is the subsample in which units are available for both testers and we learn about treatments such as rent and ability to inspect a unit. The columns present, in order, the minimum, 25th percentile, median, 75th percentile, and maximum ratios. All sets of weights in both exhibits show substantial variation, with the 75th percentile weights being more than double the 25th percentile weights (almost three times for sales tests), and so the weights contributed substantial information

Exhibit 10

Ratio of Available Rental Units Weights to Original Weights

	Minimum	25th Percentile	Median	75th Percentile	Maximum
White-Black rental tests					
Full sample	0.13	0.58	0.78	1.13	13.60
If able to meet with an agent	0.16	0.57	0.78	1.15	9.04
If available units recommended	0.15	0.58	0.78	1.13	8.77
Anglo-Hispanic rental tests					
Full sample	0.17	0.56	0.77	1.19	13.53
If able to meet with an agent	0.14	0.55	0.74	1.23	12.75
If available units recommended	0.14	0.56	0.77	1.24	12.23
White-Asian rental tests					
Full sample	0.14	0.55	0.80	1.19	7.30
If able to meet with an agent	0.14	0.56	0.80	1.20	7.34
If available units recommended	0.14	0.55	0.80	1.22	6.69

Note: The table presents the descriptive statistics over the sample of tests for the ratio of the weights based on available rental units divided by the original site weights.

Exhibit 11

Ratio of Available Owner-Occupied Units Weights to Original Weights

	Minimum	25th Percentile	Median	75th Percentile	Maximum
White-Black sales tests					
Full sample	0.08	0.44	0.75	1.20	11.88
If able to meet with an agent	0.17	0.46	0.76	1.21	10.28
Anglo-Hispanic sales tests					
Full sample	0.11	0.45	0.73	1.26	9.00
If able to meet with an agent	0.10	0.47	0.71	1.26	8.24
White-Asian sales tests					
Full sample	0.07	0.45	0.72	1.20	9.53
If able to meet with an agent	0.08	0.45	0.74	1.20	9.93

Note: The table presents the descriptive statistics over the sample of tests for the ratio of the weights based on available owner-occupied housing units divided by the original site weights.

that could have led to large changes in the measures of adverse treatment. For all groups, the weights are skewed toward a small number of tests with relatively large weights. This skewness in weights raises some concerns about results being driven by outliers, but the largest weights are never much more than 10 and the tests with the largest weights never represent much more than 1 to 2 percent of the sample by weight (less than 1 percent for rental tests).

Exhibits 12 and 13 attempt to shed additional light on why the effects of the weights on estimated adverse treatment are so small. The first panel of each table presents the average of several key census tract variables over all tests, using both the original weights and the weights based on turnover or availability, and the second panel presents the ratio of the averages based on the original and revised weights. Each subpanel shows the averages for the total sample of tests; the subsample in which both testers had an appointment; and, in the case of rental housing in exhibit 12, the subsample in which both testers were told units were available. The five tract attributes considered are median income, median housing value, share Black, share Hispanic, and share households that are owner-occupants. For rental housing in exhibit 12, the differences in tract exposure are relatively modest for all variables considered, and the ratios of the tract exposure means are usually less than 3 percent and never more than 6 percent away from 1. We find considerable variation within sites in the weights, but the variation is approximately orthogonal to the attributes of the neighborhoods in which the tests are located. As a result, even if adverse treatment were higher in some areas of each metropolitan area, the differences between the distribution of tests and the distribution of available housing appear to be relatively close to random.

On the other hand, for owner-occupied housing in exhibit 13, the effect on neighborhood attributes is more substantial. Reweighting raises average median income of the census tracts by between 4 and 10 percent, decreases average percent Black by between 11 and 19 percent, and decreases average percent Hispanic by 7 to 10 percent. The largest changes arise for the sample of Anglo-Hispanic tests. Looking back at exhibits 7, 8, and 9, the changes in the net measure are somewhat larger for the owner-occupied sample than for the rental sample, but they are still modest and nonsystematic, with some measures of adverse treatment increasing and others decreasing. Even with the larger changes in the neighborhood composition for the sales tests, we still find no evidence of a systematic bias away

Exhibit 12

Weighted Rental Sample Exposure to Census Tract Attributes

Tract Attributes	Median Income (\$)		Housing Value (\$)		Percent Black		Percent Hispanic		Percent Owner-Occupied	
	Original	Available	Original	Available	Original	Available	Original	Available	Original	Available
White-Black rental tests										
Full sample	53,360	52,422	281,910	278,238	23.8	24.3	19.9	20.6	44.5	44.0
If able to meet with an agent	53,085	52,407	278,297	272,499	24.8	25.5	19.2	19.5	43.8	43.8
If available units recommended	53,353	52,541	279,093	272,955	24.8	25.6	19.2	19.5	43.8	43.6
Anglo-Hispanic rental tests										
Full sample	56,203	54,538	362,289	348,432	12.8	13.5	32.1	33.2	40.3	40.7
If able to meet with an agent	56,441	54,710	364,421	349,824	12.8	13.6	31.8	32.9	40.5	40.8
If available units recommended	56,372	54,813	363,492	350,284	12.9	13.7	31.9	32.9	40.5	40.7
White-Asian rental tests										
Full sample	57,884	57,233	392,525	388,644	14.0	14.3	25.7	26.9	40.2	39.2
If able to meet with an agent	58,239	57,342	395,049	390,068	13.9	14.3	25.6	26.8	40.2	39.0
If available units recommended	58,031	57,506	393,923	390,318	14.0	14.4	25.8	26.9	40.1	39.0
Ratio of Exposure Using Original Weights to Exposure Using Revised Weights										
White-Black rental tests										
Full sample	1.018		1.013		0.978		0.968		1.012	
If able to meet with an agent	1.013		1.021		0.972		0.985		1.002	
If available units recommended	1.015		1.022		0.971		0.982		1.004	
Anglo-Hispanic rental tests										
Full sample	1.031		1.040		0.949		0.969		0.990	
If able to meet with an agent	1.032		1.042		0.938		0.968		0.992	
If available units recommended	1.028		1.038		0.942		0.970		0.995	
White-Asian rental tests										
Full sample	1.011		1.010		0.983		0.957		1.025	
If able to meet with an agent	1.016		1.013		0.975		0.954		1.030	
If available units recommended	1.009		1.009		0.971		0.957		1.028	

Notes: The first panel presents the average of census tract attributes over all tests and various subsamples of tests weighted by either the original site weights and by the weights designed to represent the population of available rental housing. The second panel presents the average for the original weights divided by the average for the revised weights.

Exhibit 13

Weighted Owner-Occupied Sample Exposure to Census Tract Attributes

Tract Attributes	Median Income (\$)		Housing Value (\$)		Percent Black		Percent Hispanic		Percent Owner-Occupied	
	Original	Available	Original	Available	Original	Available	Original	Available	Original	Available
White-Black sales tests										
Full sample	76,770	72,858	313,702	303,986	10.5	12.5	12.3	13.2	66.6	63.6
If able to meet with an agent	76,556	73,357	313,162	299,029	12.1	13.6	11.9	12.7	65.8	64.2
Anglo-Hispanic sales tests										
Full sample	77,893	70,538	360,335	339,145	6.82	8.39	23.24	25.79	63.89	60.22
If able to meet with an agent	78,482	72,002	368,335	348,691	6.91	8.31	22.72	24.82	63.82	60.51
White-Asian sales tests										
Full sample	83,980	79,173	438,500	416,136	6.37	7.86	16.77	18.14	61.91	59.03
If able to meet with an agent	84,863	79,187	446,565	418,975	6.40	7.77	16.93	18.36	61.96	58.66
Ratio of Exposure Using Original Weights to Exposure Using Revised Weights										
White-Black sales tests										
Full sample	1.054		1.032		0.841		0.930		1.047	
If able to meet with an agent	1.044		1.047		0.884		0.936		1.024	
Anglo-Hispanic sales tests										
Full sample	1.104		1.062		0.813		0.901		1.061	
If able to meet with an agent	1.090		1.056		0.832		0.916		1.055	
White-Asian sales tests										
Full sample	1.061		1.054		0.810		0.925		1.049	
If able to meet with an agent	1.072		1.066		0.824		0.922		1.056	

Notes: The first panel presents the average of census tract attributes over all tests and various subsamples of tests weighted by either the original site weights and by the weights designed to represent the population of available owner-occupied housing. The second panel presents the average for the original weights divided by the average for the revised weights.

from detecting discrimination against minority homebuyers. In practice, this lack of evidence implies that adverse treatment in the sales market was not systematically higher in the high-income, lower share minority submarkets that were underrepresented by the paired tests conducted in HDS2012.

Conclusion

This article discusses three major concerns that have been raised about paired-testing studies by individuals who question whether these studies understate the level of discrimination. The first concern that rental and sales agents who intend to discriminate will filter out minority homeseekers during initial phone calls has been addressed directly by HDS2012. This recent study found no differences for rental tests and only small White-Black differences for sales tests in the likelihood of obtaining an appointment with both testers obtaining an appointment in the vast majority of cases. As a result, exclusion at the appointment stage could have, at most, only a very modest effect on estimates of discrimination at later stages of the tests.

The second concern is that net measures of adverse treatment understate discrimination because some cases of favorable treatment of minority homeseekers might arise from discriminatory behavior. The continued evidence of steering in the sales market supports these concerns. The incidence of steering, however, is quite small, typically 5 percentage points, relative to the 30 or more percentage point differences between net and gross measures of adverse treatment on having more available housing or inspecting more housing units. Whereas the net measure might modestly understate discrimination, the gross measure likely dramatically overstates discrimination. In addition, the only direct evidence on this question comes from the use of three-person or triad tests in HDS2000. That analysis found the same rate of unequal treatment between same-race testers and testers of different races, suggesting no bias in net measures.

The third and final concern discussed is that the sampling of housing units from metropolitanwide advertisement sources may miss or underrepresent housing units or neighborhoods where discrimination is especially high. If landlords or real estate agents who intend to discriminate simply do not advertise housing in metropolitanwide sources, then such discrimination cannot be detected using the information from paired tests based on such metropolitanwide sources. Further, as discussed earlier, evidence from HDS2000 on housing units advertised in nontraditional sources suggests that this practice might be a concern for the sales market. An important, unanswered question is whether increased reliance on the Internet for marketing housing may have changed the importance of these nontraditional sources since HDS2000.

On the other hand, paired-testing studies could face bias because discrimination on available housing may be higher in regions or neighborhoods of the metropolitan areas that are underrepresented by advertisement-based sampling. The data from such studies can be reweighted to represent the average level of adverse treatment for available housing if the study contains sufficient numbers of tests in each site to provide broad geographic coverage across the site. This article has conducted such a reweighting but finds only limited evidence of systematic underrepresentation of certain neighborhoods in each metropolitan area and no evidence of bias in the measures of adverse treatment. A key caveat to this conclusion is that we cannot rule out differences in treatment that operate at lower levels of geography, such as census tracts or block groups.

Finally, for rental housing, it is important to acknowledge that paired-testing studies cannot detect adverse treatment that arises much later in the rental process. For example, a landlord may treat all potential tenants the same until he or she accepts a formal application and runs a credit check on applicants, but then rent only to White applicants. Such behavior might be a rational response of discriminatory landlords in the face of federally and locally funded fair housing enforcement actions. In fact, Galster and Ross (2007) found that rental discrimination against Black homeseekers between HDS1989 and HDS2000 fell the most in metropolitan areas with the highest levels federally funded enforcement. It is impossible to know whether these enforcement actions reduced discrimination or just pushed discrimination until later in the rental process, and large national paired-testing studies are not the appropriate vehicle for investigating this possibility.

Acknowledgments

The authors thank Margery Austin Turner and Judson James for their detailed comments. All remaining errors are the authors' own.

Authors

Rob Pitingolo is a research associate in the Metropolitan Housing and Communities Policy Center at the Urban Institute.

Stephen L. Ross is a professor of economics at the University of Connecticut.

References

- Baugh, John. 2007. "Attitudes towards Variation and Ear-Witness Testimony: Linguistic Profiling and Voice Discrimination in the Quest for Fair Housing and Fair Lending." In *Sociolinguistic Variation: Theory, Methods, and Applications*, edited by Robert Bayley and Ceil Lucas. Cambridge, United Kingdom: Cambridge University Press: 338–348.
- Galster, George, and Stephen L. Ross. 2007. "Fair Housing Enforcement and Changes in Discrimination Between 1989 and 2000." In *Fragile Rights within Cities: Government, Housing, and Fairness*, edited by John Goering. Lanham, MD: Rowman & Littlefield: 178–202.
- Massey, Douglas S., and Garvey Lundy. 2001. "Use of Black English and Racial Discrimination in Urban Housing Markets: New Methods and Findings," *Urban Affairs Review* 36 (4): 452–469.
- Ondrich, Jan, Stephen L. Ross, and John Yinger. 2000. "How Common Is Housing Discrimination? Improving on Traditional Measures," *Journal of Urban Economics* 47 (3): 470–500.
- Ross, Stephen L. 2002. "Paired Testing and the 2000 Housing Discrimination Study." In *Measuring Housing Discrimination in a National Study: Report of a Workshop*, edited by Stephen E. Feinberg and Angela Williams Foster. Washington, DC: National Academy of Sciences, National Research Council.

Ross, Stephen L., and Margery A. Turner. 2005. "Housing Discrimination in Metropolitan America: Explaining Changes between 1989 and 2000," *Social Problems* 52 (2): 152–180.

Turner, Margery A., and Stephen L. Ross. 2003a. *Discrimination in Metropolitan Housing Markets: Phase 2—Asians and Pacific Islanders*. Washington, DC: U.S. Department of Housing and Urban Development.

———. 2003b. *Discrimination in Metropolitan Housing Markets: Phase I—Supplement*. Washington, DC: U.S. Department of Housing and Urban Development.

Turner, Margery A., Stephen L. Ross, George Galster, and John Yinger. 2002. *Discrimination in Metropolitan Housing Markets: Phase 1*. Washington, DC: U.S. Department of Housing and Urban Development.

Turner, Margery A., Rob Santos, Diane K. Levy, Doug Wissoker, Claudia Aranda, and Rob Pitin-golo. 2013. *Housing Discrimination Against Racial and Ethnic Minorities 2012*. Washington, DC: U.S. Department of Housing and Urban Development.

Turner, Margery, Raymond Struyk, and John Yinger. 1991. *Housing Discrimination Study Synthesis*. Washington, DC: U.S. Department of Housing and Urban Development.

Wienk, Ronald E., Clifford E. Reid, John C. Simonson, and Frederick J. Eggers. 1979. *Measuring Discrimination in American Housing Markets: The Housing Market Practices Survey*. Washington, DC: U.S. Department of Housing and Urban Development.

Yinger, John. 1995. *Closed Doors, Opportunities Lost: The Continuing Costs of Housing Discrimination*. New York: Russell Sage Foundation.

Changing Contexts and New Directions for the Use of Testing

Fred Freiberg

Fair Housing Justice Center

Gregory D. Squires

George Washington University

Abstract

For decades, testing has been an effective investigative tool for documenting housing discrimination in fair housing enforcement efforts and scholarly research. This article discusses evidence gathered from recent testing investigations in the New York City region and how many violators of fair housing laws have tailored their practices to elude detection. Some changes in housing provider practices portend serious challenges for researchers and enforcement practitioners who have traditionally relied only on paired testing methodologies to identify discriminatory housing practices. In view of these changes, we offer guidance on preliminary steps that might develop credible testing approaches for the purpose of investigating or studying contemporary housing market practices. We provide some recommendations for structural changes and suggest new directions for both research and enforcement organizations. We submit that efforts to eliminate discrimination from our nation's housing markets would be greatly enhanced if we better understood housing provider practices and the changing nature of housing discrimination.

Introduction

Testing has long been a powerful instrument for documenting housing discrimination. It is a valuable research method for understanding housing market practices and the varied experiences of particular groups of homeseekers. In the fair housing enforcement context, testing has proved to be the single most effective investigative tool for collecting evidence of illegal housing discrimination.

Testing faces limitations in both research and enforcement. It conversely has the potential to be more widely used in segments of the housing market where it has not been employed and in ways that are not always considered by researchers and enforcement practitioners. This article

offers some lessons from previous testing that apply to conventional rental and sales testing and also to the new frontiers where testing might be applied more widely and effectively.

In the research context, paired testing has been used extensively to study race and national origin discrimination in the nation's housing markets (Turner et al., 2012). Testing has been used to study other forms of housing discrimination; for example, disability and sexual orientation (Friedman et al., 2013). Paired testing has also been used to identify discriminatory practices in other markets; for example, mortgage lending, homeowners insurance, employment, restaurants, hotels, and taxicab services (Fix and Struyk, 1992; Pager, 2007; Smith and Cloud, 1997; Turner et al., 2013).

Since the passage of the federal Fair Housing Act 47 years ago, paired testing has been used to gather evidence of illegal housing discrimination. More than four decades of legal challenges to discriminatory housing practices based on testing evidence have led to many changes in housing-provider policies and practices. One outcome of these changes, for certain, has been greater compliance.

Mounting evidence, however, from recent testing investigations and fair housing litigation indicates that some housing providers, those intent on violating fair housing laws, have become adept at disguising or altering their practices in a way that effectively reduces their chances of being detected by researchers, government enforcement agencies, and, most importantly, ordinary consumers. New and more subtle forms of discrimination have been identified, leading Douglas Massey (2005) to conclude that racial discrimination in housing has become a "moving target."

In view of these changes, along with changes in the housing market in general and particularly in the way housing and housing-related services are provided, this article explores how testing might be more effectively used in both research and enforcement contexts to identify housing discrimination. We offer some guidance about preliminary steps that might be taken to develop credible testing approaches to investigate or study housing market practices in segments of the housing market that have received less attention since the passage of fair housing laws and to more effectively use testing in rental and sales markets, where this tool has long been employed. Finally, we point to some recommendations for structural changes and new directions for both research and enforcement organizations that suggest how testing might be used to simultaneously advance our knowledge about discriminatory housing practices while seeking to eradicate these practices from our nation's housing markets. Perhaps the most important lesson is the need to better understand the context in which housing is provided and discrimination occurs when developing testing programs for either research or enforcement.¹

¹ Testing, by definition, is a covert activity and, to control the process, testers are often assigned personal, financial, and homeseeking characteristics that are not their own. Testers are trained and deployed to simulate or replicate consumer behavior in order to gather information and capture observations about the ordinary business practices of housing providers. In this sense, testers are proxies for ordinary consumers. Testers generally do not possess any specialized expertise about housing market practices. Testers follow directions, adhere to assigned characteristics, carry out their assignments, and report on their test experiences in an accurate, complete, and unbiased manner. Most testing, although we hasten to emphasize not all testing, is focused on obtaining observations about housing practices and the treatment of people during the preapplication stage of a housing transaction. The person responsible for supervising the testers is the principal investigator or test coordinator. Testing can be used to obtain information and observations about the policies and practices of housing providers and compare them against the requirements of fair housing laws. Testing frequently provides a comparison that may indicate whether similarly qualified populations are receiving equal treatment and equal access to housing without regard to their race or some other protected characteristic under fair housing laws. While not detailed in this article, we note that the architecture applied to the type of paired testing used in social science research is often very different from the protocols used to conduct testing investigations for the purpose of enforcing fair housing laws.

Contemporary Housing Discrimination: Lessons From New York City

The Fair Housing Justice Center (FHJC) is a regional civil rights organization that conducts testing throughout New York City and seven surrounding New York counties. Since 2010, systemic testing investigations conducted by FHJC have resulted in the filing of numerous fair housing lawsuits. The testing evidence in recent FHJC cases provides insights and reveals some interesting characteristics about the nature of contemporary housing discrimination.

For instance, it is clear that some housing providers take steps to avoid or minimize contact with unwanted populations by not advertising at all or by selectively advertising the available housing in ways that will reach only certain populations.² FHJC recently completed a testing investigation involving a landlord who controls hundreds of rental units in a predominantly White Bronx neighborhood. The landlord refrained from advertising available apartments, instead relying entirely on referrals from existing White tenants to fill vacancies. This system made it less likely that the rental manager would ever have contact with African-American applicants, virtually assuring that his buildings would remain predominately White. Through an intricately designed testing investigation that arranged for African-American testers to have contact with the rental manager, FHJC was able to observe the manager informing African-American testers that no apartments were available but showing available apartments to White testers. Apart from misrepresenting the availability of apartments to African-American testers, the manager also confided to a White tester that the landlord does not advertise available apartments because “if you run ads, you get all kinds of things.” Instead, the company relied on its mostly White tenant population to locate and refer prospective applicants to fill vacant apartments.

In recent years, FHJC has identified many housing providers who never publicly advertise or selectively advertise available rental units. Although housing providers may have many reasons for using fewer public sources to reach prospective renters or buyers, discrimination is more likely to occur when providers restrict knowledge of, or access to, available housing by limiting advertising primarily to favored populations. When testing investigations compel these same housing providers to have contact with testers of different races, FHJC frequently finds differential treatment in the form of misrepresentations about the availability of apartments or quotes of higher rents and security deposits to African-American and Latino testers.

An abundance of evidence also indicates that some violators disguise their discriminatory contact with a friendly disposition, polite conversation, and good manners. Another recent FHJC testing case involved a New York City landlord that controls a thousand rental apartments that, again, were never publicly advertised. Multiple tests revealed that the rental manager at one of the buildings tested was conversant, friendly, and encouraging when approached by African-American testers while all the time maintaining that no apartments were available. The same agent acted in a more businesslike and less conversant fashion toward White testers who visited the apartment building, merely telling them about and showing them available apartments. The congenial

² The authors recognize that implicit bias can also affect housing market practices, but our focus here is on those individuals and businesses that are intentionally evading the law.

conduct of the agent camouflaged the fact that he was lying to African-American prospective renters. Whereas a paired-testing approach was effectively employed to document the discriminatory practices in this case, it is unlikely that this housing provider would ever have been sampled in a research study, given the lack of advertised apartments, and it is equally unlikely that an African-American consumer would have filed a housing discrimination complaint, given the exceedingly friendly demeanor of the rental manager. Other violators may inform all prospective homeseekers about a set of stringent requirements, qualifications, or procedures for renting an apartment or buying a home, and then, as applicants express stronger interest and have additional contact, an agent may offer to waive, change, or reduce the requirements for the more “desirable” applicants. Applying what appear to be facially neutral policies in an unequal manner can exclude or “disqualify” unwanted populations while maintaining the outward appearance of a fair process. In several recent rental cases, FHJC sent out matched paired testers, one White and one African-American tester, both posing as part of married households. Agents initially told both testers that an application was pending on the only available apartment. When testers returned with their tester spouses of the same race, however, the facts changed. The White testers were told (once the agent could see that the spouse was also White) that the application was no longer pending and that multiple apartments were now available for rent, but the housing remained at all times unavailable to the African-American testers. Recent enforcement testing suggests that contacts by testers to housing providers, as part of initial visits by matched paired testers, may not always capture the housing-provider practices in a way that adequately discloses or confirms whether fair and equal treatment is being provided. Additional contact between the testers and the housing provider may be needed to assess whether all applicants are ultimately being afforded equal access and equal treatment, including the same terms and conditions.

Some critics have argued that multiple contacts between testers and housing providers raise potential ethical concerns because agent time is consumed with additional deceptive inquiries. Courts have understood, however, that it is frequently difficult to develop proof in housing discrimination cases and that evidence provided by testers is valuable, if not, indispensable. One court described the situation as follows.

It is surely regrettable that testers must mislead commercial landlords and homeowners as to their real intentions to rent or buy housing. Nonetheless; we have long recognized that this requirement of deception was a relatively small price to pay to defeat racial discrimination. The evidence produced by testers benefits unbiased landlords by quickly dispelling false claims of discrimination and is a major resource in society’s continuing struggle to eliminate the subtle but deadly poison of racial discrimination.³

Still other facially neutral policies may be adopted and enforced in an apparently neutral manner, but in a way that effectively excludes populations based on race or national origin. Two housing cooperatives (co-ops) containing more than 1,000 detached homes in the Bronx maintained their predominately White neighborhoods by simply requiring that any prospective buyers provide three written references from existing shareholder residents. FHJC was successful in obtaining evidence of alleged discrimination by these developments after conducting only one paired test in which

³ *Richardson v. Howard*, 712 F.2d 319 (7th Cir. 1983).

an African-American couple and a White couple had dozens of contacts during a 2-month period with a real estate agent who had specialized in selling homes in these two co-op developments for more than four decades. The testing investigation confirmed how the requirement was being applied at each of the co-ops and how it was being used to unfairly advantage White homebuyers and discriminate against African-American homebuyers. A subsequent lawsuit resulted in a real estate broker having to surrender her license, eliminating the three-shareholder-reference requirement at both developments, and implementing activities to ensure future compliance with fair housing laws.

Other recent FHJC investigations disclosed that some Section 8 rental assistance programs operated by White suburban communities maintained policies that were masquerading as “residency preferences” that favored current residents, but that were effectively operated as illegal, discriminatory residency requirements.⁴ These policies excluded racial minorities from participating in the programs. The combination of testing (and not always paired testing) coupled with public document requests enabled FHJC to elicit valuable information and unravel *how* or *why* the stated policies were being applied in a discriminatory manner. In one instance, a White tester posing as a nonresident called a Section 8 rental assistance program operated by a suburban town to inquire about obtaining a voucher. The town’s website and management plan stated that residents received first preference but that nonresidents would receive lower priority on the waiting list. The White tester was told that a preference was given to current residents, but she was also told that she might want to consider moving to the town so that she could apply to the program and receive the higher preference. The employee followed up by sending the White tester an application in the mail. When African-American and Latino testers posing as nonresidents inquired about the possibility of obtaining a voucher, they were openly discouraged from adding their names to the waiting list and were not provided an application. Instead, the minority testers were told to apply to housing authorities in the communities where they resided, despite the fact that the waiting lists for those Section 8 programs had been closed for some time.

A testing investigation can often be helpful in cases that initially appear to involve only allegations that certain policies or practices have a disparate impact. For example, for cases in which a racially homogeneous community has adopted a residency preference for more benign reasons, the preference may unlawfully restrict access to housing and reinforce patterns of residential segregation. In some of those cases, implementation of these policies may not simply be a matter of impact. A carefully designed testing investigation can often yield additional insights and information that may have probative value and occasionally provide evidence of intentional discrimination.

By engaging in linguistic or other types of profiling to screen inquiries from prospective home-seekers, providing deceptive or misleading information to prospective applicants, or using third parties to selectively screen prospective applicants, some housing providers manage to continue their discriminatory practices with little concern that their exclusionary practices will be exposed or, more importantly, that housing discrimination complaints will be filed. For example, FHJC recently documented that a landlord in a predominately White Westchester County suburb was lying to African-American testers about apartment availabilities and falsely representing that he was

⁴ The authors acknowledge that residency preferences may be benign in some situations, but in other circumstances they may be exclusionary and involve intentional discrimination.

just a “worker” and not the person responsible for renting out apartments. In another recent FHJC testing case, an agent for a landlord in Queens not only consistently lied to African-American testers about apartment availabilities, but also he provided a fictitious name to the African-American testers, while providing White testers with his real name.

We know that unconscious or implicit bias can infect a housing market transaction at any stage. Given that some violators are now more sophisticated and adroit at eluding detection in addition to the role that unconscious or implicit bias can play, what do these changes portend for the use of paired testing by researchers or enforcement agencies? The surreptitious practices described in the previous paragraphs have important lessons for both fair housing research and enforcement.

Implications for Research and Enforcement

For future research into housing market practices, what are the implications? First, when discriminating housing providers who collectively control access to thousands of housing units elect not to advertise available units in newspapers or online search websites so they can avoid unwanted populations based on race or national origin, it follows that they would never be tested in paired-testing studies that sample only advertised units. Second, if the nature of housing discrimination has changed to the point at which the conduct is not readily apparent or initially revealed in early contacts with testers, it follows that a standardized or “one-size-fits-all” approach to paired testing may not be capable of detecting some of the most pernicious discriminatory conduct. For the reasons stated, these realities and changes in housing-provider conduct raise serious questions about the efficacy and usefulness of conducting future national paired-testing studies, similar to those that the U.S. Department of Housing and Urban Development (HUD) has conducted every decade to measure the level of disparate treatment on the basis of race and national origin. Instead, we believe these changes argue for other types of research, including some that employ testing as a research method, that may advance our knowledge about housing discrimination.

Likewise, if government enforcement agencies continue to rely primarily on consumer complaints to identify illegal housing discrimination, it follows that many of these housing providers will never become the object of any enforcement action. The realities of contemporary housing discrimination based on race and national origin strongly suggest that a predominantly complaint-responsive approach to enforcing fair housing laws is inadequate. New enforcement priorities are needed—those that place a greater emphasis on proactive testing to uncover systemic discrimination. An examination of demographic data and other publicly available information has enabled fair housing organizations to more strategically use scarce testing resources and identify violators with much greater precision. Testing organizations have also demonstrated that systemic testing investigations are capable of pulling back the curtain and illuminating some of the more subtle or furtive discriminatory practices that are not always detected when ordinary matched paired testing is conducted.

The ability to obtain a more complete picture of housing market practices and housing-provider conduct may depend on the quality, sequence, timing, and extent of the contacts and interaction between testers and housing providers. For instance, traditional paired testing that has been used in major research studies involves testers having an initial contact with a housing provider to inquire about and view available housing. As previously described, however, experienced

enforcement-testing practitioners are learning that some housing providers may provide relatively equal treatment during this initial contact. Thus, research solely focused on an initial inquiry may not be able to capture the types of differences in treatment that might be observed after multiple contacts between testers and housing providers. The changes we describe in housing-provider practices may suggest that housing discrimination is less obvious or more difficult to observe than it once was, but not necessarily less prevalent. As Krysan et al. (2011: 23) concluded, “the door is not slammed in the face of the minority home seeker so much as it is flung wide open for the white tester.” Perhaps a more appropriate image would be that of a “revolving door” as unsuspecting homeseekers are too often politely and courteously escorted into, out of, and ultimately away from the desired housing. The stealth-like character of contemporary housing discrimination means that many homeseekers have virtually no way to know that they are being unlawfully discriminated against in housing. As a result of these changes, most enforcement-testing practitioners rarely limit themselves to using only simple paired tests to investigate housing discrimination. By designing tests that allow for more followup by testers, by having testers convey greater interest in the available housing, and by using new and more creative test structures, test coordinators are better able to uncover and document unlawful housing discrimination. Some enforcement-testing practitioners have even devised effective ways to employ testing to investigate claims of in-place discrimination involving harassment, provision of different services, reasonable accommodations, nonrenewals, and evictions.

In the area of sales and rentals, in which paired testing has been and still is effectively used, segments of housing markets and phases of housing transactions have not been widely investigated or studied through the use of testing. Investigating gated communities; condominiums and housing co-ops; supportive housing and other special needs housing; tax credit housing and government-assisted housing programs; assisted-living facilities, nursing homes, and continuing-care facilities; mortgage lending and appraisal practices; and other real estate-related services are just some of the areas in which we contend testing has not been used as extensively or effectively as it could be. The following section provides some guidance on preliminary steps that can be taken to apply this vital tool to studying or investigating other housing transactions or segments of the housing market about which far less is known; it also provides some guidance about how this tool can be used more effectively in markets where testing continues to be widely used.

Striking a Balance

We navigate a delicate balancing act in this article. We want to inform fair housing organizations, researchers, enforcement agencies, and policymakers about how testing might be more effectively employed to document contemporary housing discrimination. At the same time, we must avoid revealing minute details about investigative techniques and methods that are currently used to uncover unlawful discrimination to those who persist in violating the law. Disclosing specific testing techniques or approaches could cause violators to further refine, disguise, or conceal their practices so they can more easily circumvent detection. Effecting a change in housing-provider practices, of course, is an objective of fair housing law enforcement, but the change sought is compliance with the law and not further subterfuge that allows unlawful behavior to continue.

What we can and will do is describe some of the preliminary steps that might be taken before conducting any research or enforcement testing aimed at documenting discriminatory housing practices.

Expanding the Use of Testing

We know that changes in housing-provider practices and efforts to elude detection may explain why traditional paired testing fails to detect some discriminatory practices in the sales and rental markets. What accounts for the fact that less testing has been conducted in certain segments of the housing market (for example, gated communities; condominiums and housing co-ops; supportive and special needs housing; senior housing, assisted-living facilities, and nursing homes; mortgage lending; and home appraisals)? Resource limitations, of course, are a constant factor. The lack of information and training available to practitioners to learn how to test certain types of housing or housing services, the complexity or perceived level of difficulty of the testing, and the financial or human resources needed to implement the testing are all likely factors. The moving targets, of course, constitute another challenge. It is also the case that testing is not the only way or necessarily the best way in every circumstance to gather information about discriminatory housing practices. Because testing has a unique ability to shine a bright light on housing market practices and capture vital observations about how providers of housing and housing services treat consumers based on protected characteristics, however, it is often the most powerful investigative tool. The power of paired testing resides principally in the intuitive understanding that if similarly situated homeseekers who differ in only one respect (for example, race, ethnicity, or gender) are treated differently in the homeseeking process, it is fairly easy for jurors, judges, enforcement agencies, sophisticated analysts, and ordinary citizens to see that something is wrong. To overcome the limitations noted earlier (no advertising or selective advertising by housing providers, the need for multiple contacts between testers and housing providers, the need to employ alternative test structures to paired testing, and so on), however, some new and creative approaches to testing may be required. To design these new approaches, it is imperative that we better understand what is being tested and how consumers conventionally learn about and access the housing or housing services to be tested.

What Are We Testing?

Whether a research organization is planning to implement a study that employs testing as a data-gathering method to inform policy or an enforcement agency is preparing to conduct a systemic testing investigation to enforce compliance with the law, learning more about the type of housing, housing program, or housing service to be tested is a critical first step. Why is this important? In the final analysis, it is important that testers make requests and ask questions that might (1) credibly come from ordinary consumers and (2) elicit the vital information that enables one to compare the policies and practices of the entity being tested against a set of treatment variables (for research) or against the requirements of fair housing laws (for enforcement). Understanding more about how housing providers or housing programs and services operate can often provide important clues about where potential bias might be infecting or adversely affecting consumer transactions. A few examples follow.

- Examining advertising and marketing materials, reports, journal articles, newspaper and magazine articles, websites, census data, and other publicly available information can often be useful, depending on the type of entity to be tested. If the entity to be tested is a government agency (for example, housing authorities and tax credit allocation agencies), using open records laws to request key public documents can also provide insights that might inform the testing protocols.
- If the entity to be tested is regulated, licensed, or certified by local, state, or federal governments (for example, nursing homes, assisted-living facilities, continuing-care communities, condominium and co-op boards, mortgage brokers, and housing counseling agencies), identifying the rules and requirements that govern their operations can also be helpful in structuring any kind of testing investigation.
- In more complex testing (for example, nursing homes, mortgage lending, supportive housing, and appraisals), it may make sense to consult experts in the field who understand some of the nuances, trends, and factors that might need to be controlled in any testing investigation or study.
- One of the most obvious ways to gauge consumer activity is to talk with consumers who have had firsthand experience in searching for or using the type of housing or housing service to be tested.

Although any of the above sources can be beneficial, a very important caveat is in order. The information contained in marketing materials, reports, newspaper articles, and other written materials may not be true. The prescriptions of laws and regulations may not be followed by the entities to be tested. The opinions of experts may be dead wrong. Experiences shared by ordinary consumers, although useful, may not provide all the information that is needed to fashion a testing approach. Although consulting these sources is almost always a worthwhile exercise to guide an investigation or testing study, it is important to assess how valuable and accurate the information is by doing some exploratory testing before commencing any comprehensive testing study or investigation. There is no substitute for information obtained by people “on the ground” who have direct contact with the entity or entities to be tested.

Finding the Path

One of the most important tasks before conducting any testing study or investigation is to learn how consumers currently find out about the type of housing or services to be tested. In other words, *what is the path one needs to follow to inquire about the housing or service?*

A popular misconception is that if one is offering a commodity or service in an open market, it naturally follows that one would want to advertise and market that commodity or service as widely as possible, detailing the positive features and benefits, to gain a competitive advantage, generate demand, and attract consumers. This approach may well be how it works for department stores, car manufacturers, or restaurant chains, but it is not how it works in housing markets. For many housing consumers, a search involves a veritable maze that they must navigate, complete with trap doors, dead ends, circuitous routes, hidden compartments, misleading signs, and other types of barriers. The housing market is anything but open for many renters and buyers. Some members of the research community have suggested that we need to learn more about how various populations search for and locate housing. Although that information may be helpful, public policy might be

better served by research that examines the marketing and advertising practices that providers of housing and housing services use. Better understanding of the demand and supply sides would be informative, but it is the supply side that most needs to be demystified and better understood. That is, it is more important to further examine how housing providers provide their goods and services than how consumers shop for them. At any rate, finding the most direct path to an apartment or home can be a daunting and time-consuming task for any consumer.

For both research and enforcement purposes, the common goal is to learn more about how one actually inquires about the housing or service to be tested. Even after the initial background investigation has been conducted, however, including possibly some pretests (in a research context), or “scouting” or “advance reconnaissance” (in an enforcement context), complications may well persist, depending on the type of housing or housing service to be tested. For instance, assigning a tester to make an advance visit to a suburban apartment complex that has a clearly marked leasing office is fairly simple to accomplish. After one visit, the tester will likely be able to collect all kinds of useful information about the necessity for appointments, office hours, staffing, available housing, price ranges, and so on. This information can be extraordinarily helpful in developing a viable research design or investigative approach. Contrast that situation with a guarded and gated condominium community that never advertises available units and has no onsite sales office. In this situation, it may be necessary to have testers talk with a security guard, speak to existing residents who are coming or going from the development, meet with real estate agents who sell condos in the area, and/or talk with service workers and others who are coming and going from the complex. It may take multiple contacts and approaches before the path becomes clear about how consumers might learn of and inquire about available condos in this development. Knowing the path and knowing who to contact are critical to structuring any kind of credible testing approach, particularly when that path is well concealed or rarely traveled.

An essential first step to any testing involves determining how homeseekers identify and access the housing and housing services to be tested. For instance, if a common approach made by consumers to obtaining assisted living for aging relatives involves family members inquiring about housing on their behalf, then perhaps some type of proxy test may be an appropriate way to document how assisted-living facilities treat different types of homeseekers. Housing counselors in transitional housing agencies that serve homeless populations often call on behalf of their clients to look for permanent housing. Simulating these types of calls may elicit valuable information about housing-provider practices and identify discriminatory preferences. For any type of testing, finding out how consumers access the housing is the first step. This is particularly the case in areas of the housing market about which less is known, such as gated communities, tax credit housing, subsidized housing, special needs housing, home appraisal services, mortgage brokers, and various types of senior housing.

Experienced enforcement-testing practitioners understand that providers of housing and housing services are structured in many various ways so that a one-size-fits-all approach to testing may or may not be possible across an entire market. Enforcement-testing practitioners frequently vary and adapt their approaches accordingly from site to site to ensure that each testing approach is credible. Variations in how housing providers are structured and how they interact with consumers, however, may provide unique challenges for researchers who require a standardized testing methodology

to conduct tests across a large swath of housing providers. A significant amount of exploratory testing may be necessary to determine whether a singular testing approach can be replicated across an entire market and produce reliable information about housing discrimination.

Valuable information obtained from the exploratory testing along with the earlier background material collected may make it possible to design a test structure. Without disclosing all the specific options, it is important to underscore that traditional paired testing may not always be the only or best approach. Depending on the circumstances, enforcement practitioners have developed test structures over the years that employ one, two, three, four, or more testers to document housing discrimination; these approaches have yielded credible and objective evidence of illegal discrimination. The basic lesson is that far more needs to be known about the context in which the testing is to take place before the details of the testing protocol can be finalized.

Future Directions and Recommendations

First and foremost, researchers and enforcement-testing practitioners need to collaborate more. This is not to say that every research project will have enforcement benefits or that every enforcement investigation will yield new theoretical insights or contain policy implications. Researchers and enforcement practitioners both, however, would clearly benefit from regular dialogue and a cross-fertilization of ideas. Such convenings would make it possible for enforcement practitioners to share creative testing approaches that have been effectively employed to overcome and document some of the more evasive and deceptive practices that are being detected in local housing markets. Researchers are in a better position to decide if any of these newer testing methodologies could be replicated in fair housing research studies.

At least some members of the research community have been moving in this direction. Linguistic profiling, whereby non-White homeseekers are denied housing or treated differently based on the racial or ethnic identity associated with their voice, has been documented (Baugh, 2000; Fischer and Massey, 2004; Massey and Lundy, 2001). In a similar way, so called *cybersegregation*, in which non-White homeseekers are denied during online housing searches based on the racial or ethnic identity that is associated with their name, has been demonstrated (Ahmed and Hammarstedt, 2008; Hanson, Hawley, and Taylor, 2011). Both lines of research have opened up areas for enforcement actions. These two newly explored types of discrimination, reflecting again the moving target, indicate the importance of flexibility on the part of law enforcement agencies.

The findings of a 2002 Urban Institute study of mortgage lending practices that involved the use of paired testing to examine what happens to homebuyers of different races and national origins at the preapplication stage of a mortgage lending transaction blazed some new trails, and the findings yielded significant benefits for enforcement practitioners (Turner et al., 2002). Using the information from this study, FHJC adapted and refined the lender testing protocols for use in a recent enforcement investigation and, in 2015, brought the first federal lawsuit under the Fair Housing Act against a major bank based solely on testing evidence.⁵ Another positive development in the

⁵ *FHJC et al. v. M&T Bank et al.*, Case No. 15 Civ. 00779 (S.D.N.Y. 2015).

wake of the foreclosure crisis has been stepped-up enforcement of fair lending and other consumer protection laws in financial services (Consumer Financial Protection Bureau, 2014; Pratt, 2014).

We have not seen, however, any substantial testing or research in general on the nontraditional segments of the housing market noted above (for example, gated communities and homeowner associations; tax credit housing and subsidized housing programs; nursing homes, assisted-living facilities, and continuing-care facilities; home appraisal practices). We previously noted some of the reasons for this lack of research. These issues are complex and dynamic and these areas are difficult to access. Perhaps that is all the more reason why greater collaboration is needed between researchers and fair housing enforcement professionals. New areas of inquiry into these less explored areas might inform policy, open up avenues for expanded enforcement, or both.

The changes we described in housing-provider practices lead us to conclude that research methods used to measure the level of housing discrimination by sampling advertised housing and using traditional paired-testing techniques, as has been done in the past, are less likely to yield reliable or meaningful measures of differential treatment based on race or national origin today. Given the reality of limited funds for both research and enforcement, future testing should be strategically targeted to look at segments of the housing market we know less about by using a variety of testing approaches and techniques.

Future testing can be more informative if the supply side is targeted more than it has been in the past. That is, we need to learn more about how housing providers market their products and services. Consumer behavior and knowledge are important. Consumers, however, can more effectively protect their interests if they have a better understanding of how various actors in the housing market work. Evidence indicates, for example, that those who are better informed about fair housing laws are more likely to be supportive of stronger fair housing enforcement (Abravanel, 2002). Research is critical to understanding how housing markets work. Equally, if not more, important is the vital need to better protect the rights of consumers in the various housing and housing-related markets.

In 1968, when the federal Fair Housing Act was enacted, most housing discrimination based on race and national origin was still fairly overt. The use of testing made it possible for private civil rights organizations and researchers to document discriminatory practices. Legislative action, most notably the 1988 amendments to the Fair Housing Act, resulted in strengthened enforcement efforts. Decades of enforcement by private fair housing groups and increased involvement of government enforcement agencies since 1989 have led to significant changes in housing-provider practices, including greater compliance.

Some changes in provider practices, however, have not been as positive. Providers of housing and housing services who remained intent on violating fair housing laws became more sophisticated and adept at concealing their discriminatory activities from ordinary consumers. As recent evidence suggests, an almost stealth-like quality permeates contemporary housing discrimination, which is designed to elude detection by consumers and government enforcement agencies. If consumers are unaware that they are being discriminated against, it follows that they will not file complaints. If no complaints are filed, no government enforcement action will result. If no enforcement action takes place, discrimination continues. This pernicious cycle, fueled by changes in provider practices,

suggests that the current emphasis on a passive, complaint-driven approach to enforcement of fair housing laws by government is inadequate. Although a complaint-responsive mechanism must be preserved (and we suggest it could also be vastly improved if testing were more widely used by all enforcement agencies), a greater emphasis must be placed on conducting targeted, systemic testing investigations. A new fair housing enforcement paradigm is needed, one characterized by a more coordinated, proactive, strategically targeted, and better resourced approach with testing as a centerpiece. Whether government enforcement agencies develop their own internal testing capability as the Civil Rights Division of the U.S. Department of Justice did more than two decades ago, or whether these agencies contract with nonprofit fair housing testing programs that possess that testing capability, this investigative tool must be used more often and more effectively if, as a nation, we hope to make progress in eliminating housing discrimination. A starting point is a better understanding of how housing and housing-related services are provided and how discrimination occurs in today's (and tomorrow's) housing markets.

Testing Outside the Comfort Zone

Testing has been the single most powerful tool for documenting housing discrimination. We are learning, however, the limitations of using paired testing as a means of uncovering housing discrimination, even in many traditional rental and sales situations. Also, some segments of the housing market have not been subjected to significant testing, and these missed segments constitute another important shortcoming. In developing future testing studies, HUD should consider devoting resources to exploring those housing-provider practices about which less is known (for example, nursing homes, condominiums, and co-ops). Partnerships with enforcement practitioners to formulate innovative and credible testing approaches could enhance and strengthen HUD's research and enforcement efforts. Researchers and enforcement-testing practitioners need to go beyond what has emerged as fairly traditional approaches to testing. Exploring creative and effective testing approaches in the sales and rental markets and in less tested parts of the market could yield valuable observations about housing market practices. It will require that we leave the comfort zone of traditional paired testing and explore new applications and frontiers, increase opportunities for collaboration between researchers and enforcement-testing practitioners, and confront the challenges presented by structural variations and changes in housing-provider practices. Whether future testing is aimed at informing public policy, expanding our knowledge about the nature and effect of housing discrimination, or enforcing fair housing laws, the shared goals of researchers and enforcement practitioners should be to eliminate the invidious discrimination that too often infects housing market transactions, restricts access to housing opportunities, and reinforces segregation.

Authors

Fred Freiberg is Executive Director of the Fair Housing Justice Center in New York City.

Gregory D. Squires is a professor of sociology and of public policy and public administration at George Washington University.

References

- Abravanel, Martin D. 2002. "Public Knowledge of Fair Housing Law: Does It Protect Against Housing Discrimination?" *Housing Policy Debate* 13 (3): 469–504.
- Ahmed, Ali M., and Mats Hammarstedt. 2008. "Discrimination in the Housing Market: A Field Experiment on the Internet," *Journal of Urban Economics* 64 (2): 362–372.
- Baugh, John. 2000. *Beyond Ebonics: Linguistic Pride and Racial Prejudice*. New York: Oxford University Press.
- Consumer Financial Protection Bureau. 2014. *Semi-Annual Report of the Consumer Financial Protection Bureau*. Washington, DC: Consumer Financial Protection Bureau.
- Fischer, Mary J., and Douglas S. Massey. 2004. "The Ecology of Racial Discrimination," *City and Community* 3 (3): 221–241.
- Fix, Michael, and Raymond J. Struyk, eds. 1992. *Clear and Convincing Evidence: Measurement of Discrimination in America*. Washington, DC: Urban Institute Press.
- Friedman, Samantha, Angela Reynolds, Susan Scovill, Florence R. Brassier, Ron Campbell, and McKenzie Ballou. 2013. *An Estimate of Housing Discrimination Against Same-Sex Couples*. Washington, DC: U.S. Department of Housing and Urban Development.
- Hanson, Andrew, Zackary Hawley, and Aryn Taylor. 2011. "Subtle Discrimination in the Rental Housing Market: Evidence From e-mail Correspondence With Landlords," *Journal of Housing Economics* 20: 276–284.
- Krysan, Maria, Samantha Friedman, Gregory D. Squires, Jonathan D. Stringfield, and Katherine Moloney. 2011. "Subtle Cyber-Discrimination?. Not *If* but *How* Internet Housing Providers Respond Differently to Neil, Tyrone and Jorge." Population Association of America, Washington, D.C., March 31.
- Massey, Douglas S. 2005. "Racial Discrimination in Housing: A Moving Target," *Social Problems* 52 (2): 148–151.
- Massey, Douglas S., and Garvey Lundy. 2001. "Use of Black English and Racial Discrimination in Urban Housing Markets: New Methods and Findings," *Urban Affairs Review* 36 (4): 452–469.
- Pager, Devah. 2007. "The Use of Field Experiments for Studies of Employment Discrimination: Contributions, Critiques, and Directions for the Future." In *Race, Ethnicity, and Inequality in the U.S. Labor Market: Critical Issues in the New Millennium*, edited by George Wilson. *The Annals of the American Academy of Political and Social Science* 609: 104–133.
- Pratt, Sara. 2014. "Highlights of Fair Housing Enforcement by the U.S. Department of Justice and the U.S. Department of Housing and Urban Development." Presentation at the plenary session of the annual conference of the National Fair Housing Alliance, "Making Your ZIP Code = Opportunity," June 23.

Smith, Shanna L., and Cathy Cloud. 1997. "Documenting Discrimination by Homeowners Insurance Companies Through Testing." In *Insurance Redlining: Disinvestment, Reinvestment, and the Evolving Role of Financial Institutions*, edited by Gregory D. Squires. Washington, DC: Urban Institute Press: 97–117.

Turner, Margery Austin, Fred Freiberg, Erin B. Godfrey, Carla Herbig, Diane K. Levy, and Robin E. Smith. 2002. *All Other Things Being Equal: A Paired Testing Study of Mortgage Lending Institutions*. Washington, DC: Urban Institute.

Turner, Margery Austin, Rob Santos, Diane K. Levy, Doug Wissoker, Claudia Aranda, and Rob Pitingolo. 2013. *Housing Discrimination Against Racial and Ethnic Minorities 2012*. Washington, DC: Urban Institute.

Targeting Disability Discrimination: Findings and Reflections From the National Study on Housing Discrimination Against People Who Are Deaf and People Who Use Wheelchairs

Claudia L. Aranda
Urban Institute

Abstract

This article highlights the use of the paired-testing methodology to measure housing discrimination against people with disabilities, with special emphasis on the complexities of conducting disability-based testing. It presents findings from the U.S. Department of Housing and Urban Development-sponsored Housing Discrimination Study-Disabilities, which has produced the first national estimates of discrimination in the rental housing market against people who are deaf or hard of hearing and people who use wheelchairs. The results show that people who are deaf or hard of hearing face barriers during the homeseeking process, including communicating with housing providers and learning about available units. People who use wheelchairs face barriers at several points in the process, including finding accessible units, securing appointments with providers and being shown units, and receiving a definite response to their reasonable modification requests. In both studies, testers posed as well-qualified rental homeseekers with the same qualifications and needs. Future paired-testing studies could help produce additional evidence on the homeseeking experiences of people who are deaf or hard of hearing and people who use wheelchairs, including those who are less qualified. Additional testing studies should also focus on the compliance of the current housing stock to the design and construction requirements of the Fair Housing Act. Such findings could help provide critical information about where the need for accessible rental housing is greatest.

Introduction

The primary goal of the first national paired-testing study of housing discrimination against people who are deaf or hard of hearing and people who use wheelchairs, Housing Discrimination Study-Disabilities (HDS-Disabilities), was to produce national estimates of differential treatment in the rental market (Levy et al., 2014). Funded by the U.S. Department of Housing and Urban Development (HUD) and conducted by the Urban Institute, the study also measured the willingness of housing providers to approve reasonable modification requests made by prospective tenants who use a wheelchair.¹ Because HDS-Disabilities included two distinctly different populations, it was implemented as two separate studies, each with its own methodological, analytical, and practical complexities, many of which are discussed in this article. In both studies, testers posed as well-qualified rental homeseekers with the same qualifications and needs. From the perspective of the housing provider, the only difference between the two testers in a pair (who are matched on sex, race or ethnicity, and age) is their disability status. Testers with the same profiles should receive the same treatment; when housing providers offer different housing costs or terms, the tests provide direct evidence of discrimination. Overall, the findings of HDS-Disabilities highlight the challenges people with disabilities face when they search for a home; although they might not face higher costs, on average, than homeseekers without disabilities, they must contact more housing providers to find housing that meets their needs. The study helps provide important details about the experiences of homeseekers with disabilities, which strongly suggest directions for research, education, and advocacy.

HDS-Disabilities builds on the lessons of the 2005 pilot disabilities study, also funded by HUD and conducted by the Urban Institute, which explored the feasibility of using paired-testing to measure housing discrimination against people with disabilities (Turner et al., 2005).² The study findings showed that adverse treatment of people with disabilities occurs more often during the initial stages of housing searches than the adverse treatment of African-American or Hispanic renters (Turner et al., 2005). When housing providers accepted calls of testers who were deaf, the testers received significantly less information about the rental application process and fewer opportunities for followup than did comparable hearing testers who made telephone inquiries. Among people who used wheelchairs and visited rental properties to inquire about advertised units, the findings showed that they were just as likely as testers without disabilities to meet with a housing provider. People who used wheelchairs, however, were told about fewer available units than were testers without disabilities and received less information about the application process, although they were quoted lower fees than were comparable testers without disabilities (Turner et al., 2005). HDS-Disabilities ultimately drew heavily from the protocols and measures used in the pilot study

¹ A reasonable modification is a structural change made to existing premises, occupied or to be occupied by a person with a disability, in order to afford such person full enjoyment of the premises. Title VIII of the Civil Rights Act, 42 U.S.C. § 3604(f)(3)(A).

² The pilot study tested for discrimination against people who are deaf and hard of hearing and people who are in wheelchairs. The exploratory component of the study implemented various testing scenarios across different disabilities, including people who are blind, who are deaf or hard of hearing, or who have mental disabilities.

and from sampling³ and field implementation procedures from the more recent 2012 Housing Discrimination Study, which were updated to reflect changes in rental housing markets, housing search practices, and communication technologies (Levy et al., 2014).

Background

In 1988, the Fair Housing Act was amended to prohibit discrimination in the sale, rental, and financing of housing on the basis of a disability.⁴ The amendments, which were enacted in 1989, made it illegal for housing providers to refuse to rent or sell to people with disabilities; impose different qualification criteria; or require different fees, terms, or conditions. The Fair Housing Act requires that housing providers make reasonable accommodations in rules, policies, practices, or services when such accommodations may be necessary to afford a person with a disability the equal opportunity to use and enjoy a dwelling and when such changes do not create an undue burden for the housing provider.⁵ In addition, housing providers are required to allow people with disabilities to make reasonable structural modifications at their own cost. Further, the design and construction provisions of the Act require that certain new multifamily dwellings developed for first occupancy on or after March 13, 1991, meet specific accessibility standards.

Since the coverage of the Fair Housing Act was expanded to include people with disabilities, disability discrimination complaints have become the majority of those received by federal and local agencies. In 2013, nearly 54 percent of all fair housing investigations that HUD conducted and 53 percent of all fair housing investigations that local Fair Housing Assistance Program agencies throughout the United States initiated were based on allegations of disability discrimination (NFHA, 2014). The National Fair Housing Alliance also reported that 48 percent of the complaints that member organizations received in 2013 were brought on the basis of disability.

Although fair housing organizations and government agencies alike may be continuing to help preserve access to housing for people with disabilities, the need for accessible housing is expected to continue to rise as the number of older people in the country continues to increase. The exponential level of population growth is undeniable: In 2000, more than 35 million people age 65 or older were living in the United States, comprising 12 percent of the total population (Smith et al., 2012). By 2050, however, fully one-fifth (20 percent) of the U.S. population will be age 65 or older (Jacobsen et al., 2011). “Most of this increase will take place by 2030 as the last of the large baby-boom cohorts reaches age 65” (Jacobsen et al., 2011: 2). By 2040, the population age 80 or older is projected to be 28 million (JCHS, 2014). Smith, Rayer, and Smith (2008: 3) also wrote that “[s]ince disability rates rise with age, the aging of the population will bring large increases in the number of disabled persons.” To be more specific, by the age of 85, more than two-thirds of individuals have some type of disability (JCHS, 2014). It is also the case that many working-age people with disabilities are more likely to have low incomes than those without disabilities (She

³ Advertisements were harvested electronically from online sources that included apartments.com, rent.com, move.com, forrent.com, and craigslist.org (Levy et al., 2014).

⁴ Title VIII of the Civil Rights Act, 42 U.S.C. § 3604(f)(3)(A).

⁵ Title VIII of the Civil Rights Act, 42 U.S.C. § 3604(f)(3)(B).

and Livermore, 2009), which can limit their ability to save and purchase a home (Hoffman and Livermore, 2012), and which suggests that many people with disabilities will continue to depend on the rental market for accessible housing that meets their needs.

Complexities in Definition, Analysis, and Field Implementation

The application of the paired-testing methodology to the measurement of discrimination against people with disabilities presented distinct challenges (and choices) in the design, analysis, and field implementation phases of HDS-Disabilities. The first section of this article describes the design and analytical complexities of the study: Compared with the race and ethnicity HDSs, in which testers in a pair had one key characteristic between them that was different (that is, race or ethnicity), tester pairs in HDS-Disabilities had two fundamental differences between them. The specific characteristics at issue for both the deaf and hard of hearing and wheelchair components of the study are described in turn, as are additional, notable design features. The next section addresses the unique field implementation challenges of disabilities testing, which include the heightened risk for detection. The key findings of HDS-Disabilities are presented in the subsequent section followed by recommendations for next steps in research, housing-provider education and outreach, and advocacy.

Discrimination Against Homeseekers Who Are Deaf or Hard of Hearing

In the case of the deaf and hard of hearing study, in which testing was conducted remotely via telephone or the Internet, testers in a pair not only differed in their disability status (that is, one tester was deaf or hard of hearing and the other was hearing), but they also differed in the type of telecommunication relay service (TRS) they used to communicate with the housing provider. Hearing testers communicated via telephone and the deaf and hard of hearing testers used one of the three most commonly used TRS types: (1) Video Relay Service (VRS), (2) Internet Protocol Captioned Telephone Service (IP CTS), or (3) Internet Protocol Relay (IP Relay) Service. The project team hypothesized that the only way to remove the technology type variable from the tester pair would be to have control testers also using the same TRS as their deaf or hard of hearing counterpart. The idea was dismissed summarily, however, because hearing testers do not use TRSs to communicate with housing providers. The existence of the two variables in each pair was acknowledged as an inevitable difference between two people, one who is deaf and one who is hearing, communicating remotely with a housing provider. By including different technology types, the testing outcomes ultimately enabled the team to report on differences in treatment for each type.⁶

The three technology types that the deaf and hard of hearing testers used made up about 94 percent of TRS usage volume as of the beginning of the study (Rolka Loube Saltzer Associates, 2012). Data indicated that VRS was the most commonly used TRS overall and was also most used by people who are deaf (Rolka Loube Saltzer Associates, 2012). IP CTS was the second most commonly used TRS overall and was the service most used by people who are hard of hearing (Rolka Loube Saltzer Associates, 2012). After the research design was finalized, the relative usage volumes

⁶ See Levy et al. (2014), chapter IV.

for the three TRSs changed. IP CTS became the most used service, followed by VRS and IP Relay (Rolka Loube Saltzer Associates, 2014). Together, these three TRS types largely have supplanted the use of text-telephone (TTY) relay (Levy et al., 2014). The three technologies also differ in how people who are deaf or hard of hearing communicate their messages to call recipients, such as whether the use of an intermediary is necessary, and in the equipment needed to place calls.

When using VRS, a caller who is deaf and uses sign language places a call through a service and uses a video phone or web camera to communicate with a communication assistant (CA). When using VRS for HDS-Disabilities, the tester signed the message to be conveyed to the housing provider, while the CA spoke to the housing provider on the telephone. As the recipient spoke directly to the CA, the assistant signed the response to the tester through VRS. The use of sign language and speech enabled the tester and the housing provider to communicate at or near the pace of spoken language even though they did not have direct contact.

The IP Relay Service is similar to VRS but, instead of signing the message, the caller types it to the CA through an IP Relay website or text application. The CA telephones the recipient and speaks that message. After the recipient speaks a response directly to the CA, the assistant types the response to the caller. The IP Relay typed messages can take 2 to 5 seconds to appear, which can cause a slight delay in communication for both the tester and the housing provider.

Unlike the other two TRSs used in the study, IP CTS enables the user to have partial direct contact with a call recipient. A caller who is deaf or hard of hearing calls the recipient through a captioned telephone service or a captioned website and speaks directly to them. As the recipient speaks a response, a CA repeats the response to the caller and voice recognition technology creates the message in text through the captioned telephone or website. The communication delay associated with typing the call recipient's response can last from 7 to 10 seconds, which is the longest delay of the three technology types.⁷

On HDS-Disabilities, testers who were deaf conveyed their deaf status to the housing provider at the beginning of the contact to ensure the call recipient understood the call was from a person who was deaf.⁸ If the housing provider hung up after the first contact, testers were directed to ask the CA if she or he was able to convey that the call was from a person who is deaf before the call was disconnected. If the CA was not able to state as much, the tester made a second attempt to contact the housing provider. If the CA did convey the tester's deaf status before the hangup, the outcome was recorded and no subsequent attempt to contact the housing provider was made. HDS-Disabilities ultimately found that homeseekers who are deaf and use VRS (the technology type whose pace most closely aligns with that of spoken language) are more likely to have housing providers take their calls compared with deaf or hard of hearing testers who contact them using IP Relay or IP CTS. Although housing providers do not take all calls from homeseekers using VRS, housing providers are more likely to tell VRS users (compared with IP CTS or IP Relay users) whether units are available (Levy et al., 2014). It is important to note that, because the use of VRS relies on a person's ability to communicate in sign language, this technology type is not an option for everyone who relies on a TRS.

⁷ For detailed diagrams showing how the three TRSs function, see Levy et al. (2014), chapter II.

⁸ For the 2005 pilot disability study, the relay operator explained the TTY call to the recipient at the beginning of the call but otherwise did not explicitly convey that the caller was deaf. The pilot protocol for HDS-Disabilities was modified to ensure the recipient understood the caller was deaf.

Discrimination Against Homeseekers Who Use Wheelchairs

In the study on discrimination in the rental market against people who use wheelchairs, two testers in a pair also had two key differences between them: (1) their disability status (that is, one tester used a wheelchair and the control tester was ambulatory) and (2) the definition of housing that would meet their needs. For the ambulatory tester, *suitable units* were defined as “those that are within testers’ price range, are available when needed, and have at least the minimum number of bedrooms required for the testers’ (assigned) household” (Levy et al., 2014: xi). For the tester using a wheelchair, a *suitable unit* had to meet the same criteria but also had to be “accessible or modifiable to become accessible” (Levy et al., 2014: xi). Note that for HDS-Disabilities, *accessibility* was defined as “the ability of a tester who uses a wheelchair to access a building and access available units” (Levy et al., 2014: vi). This operational standard is not strictly equivalent to specific laws or regulations. This second difference between the two testers in a pair could, in and of itself, cause testers to have different outcomes, and it was essential to not confound the availability of accessible housing with differential treatment. The study design addressed this issue, in part, by having local project staff obtain additional information about sampled advertised housing to determine, to the extent possible, whether the building was configured so that a tester using a wheelchair could enter it. During this advance contact phase of the testing process, project staff used online visual tools, such as Google Earth; drove by the property; or, when necessary, asked the housing provider whether the building was accessible to someone with a stroller or who was temporarily on crutches. After the initial site evaluation was completed, only those advertisements for units in buildings believed to be accessible were used to create test assignments.

When testers who used wheelchairs made contact with housing providers over the telephone, they were required to disclose their use of a wheelchair or scooter to help reduce the likelihood that testers would arrive at a site and encounter an obstacle that would prevent the test from continuing. Although both the initial site evaluation and the telephone contact did help to eliminate housing from the sample that the testers in wheelchairs would be unable to enter, the process had its limitations, particularly because it relied on a housing provider’s understanding of a building’s accessibility features. As a result, during site visits, testers using wheelchairs still encountered barriers when they attempted to enter buildings and also specific units. For example, one agent was surprised to learn that the several steps at the entrance of an apartment complex would preclude a tester from entering the lobby. Because the building had no alternate, accessible entrance, the test could not proceed. After testers in wheelchairs entered a single-family home or multifamily building, they sometimes encountered obstacles such as narrow doorways and high thresholds when attempting to view particular rooms, specific apartments, or both.

During the design phase of the study, the project team determined, in consultation with the expert panel, that testers using wheelchairs should reflect the diversity of the population, which would also allow for analysis on whether the type of wheelchair has any effect on results. More than one-half of the HDS-Disabilities testers in a wheelchair used a power chair (52 percent), 44 percent of the testers used a manual chair, and only 4 percent used a scooter (Levy et al., 2014). Because testers used chairs and scooters that varied significantly in size, the chair or scooter could affect a tester’s ability to view specific units because of the structural limitations of the housing and also possibly the housing providers’ perceived likelihood that a chair may cause more “wear and tear”

or damage to a unit over time. The study findings showed that housing providers are somewhat less likely to share information about suitable units with homeseekers who use motorized wheelchairs or scooters than with those who use manual wheelchairs (Levy et al., 2014).⁹

For both components of HDS-Disabilities, the study reports key measures that reflect sequential milestones in the course of a test (Levy et al., 2014).¹⁰ When taken together, the measures provide a rounded picture of both the incidence and severity of differential treatment. In the case of the wheelchair study, this sequential approach to analysis also allowed for the comparison of available units to be undertaken only when a *suitable unit* was observed for a tester who uses a wheelchair; that is, because not all available units are accessible for a person who uses a wheelchair, a comparison of the number of units recommended or shown to both testers of a matched pair would likely overstate differences in treatment. In a similar way, a comparison of whether units could be inspected was undertaken only when a unit was observed (determined in part through the use of tester narratives) to be one that the tester who uses a wheelchair could inspect.

Requests for Reasonable Modifications

The wheelchair component of HDS-Disabilities also measured the unilateral outcomes for testers in wheelchairs who requested reasonable modifications (Levy et al., 2014). Only testers in wheelchairs made requests for modification; control testers did not. For tests in which *suitable units* were available for the tester using a wheelchair, the incidence of refusal to a reasonable modification was determined (Levy et al., 2014). The study protocol required testers who used wheelchairs to request an appointment even if they were told the property was not accessible and also to suggest that it might be possible to make a modification to the available housing. When one tester in a pair was unable to secure an appointment, the other tester proceeded with the site visit to collect observational data on building and unit accessibility and, in the case of the tester who used a wheelchair, data on responses to reasonable modification requests. Control testers were trained to notice stairs, thresholds, and other property features that might make a building, lobby area, or available housing units inaccessible to people who use wheelchairs. They documented any such observed features. Testers in wheelchairs were instructed to request up to three modifications, given what they encountered when they arrived at the test site. Because evidence suggests that the number of modification requests can affect the housing provider's willingness to agree, the number of requests was limited to three (Levy et al., 2014). The list of approved modification requests was reviewed and amended by an expert panel composed of researchers, advocates, and HUD staff members to include those modification requests that, in most instances, would be considered reasonable. In addition, testers in wheelchairs were directed to tell housing providers that the modifications would be made at the testers' expense. If questioned further by an agent, testers explained that they would restore a modified unit to its original condition on moving out.

⁹ Providers also are less likely to show units to people who have quadriplegia than to homeseekers who have paraplegia (Levy et al., 2014).

¹⁰ The Urban Institute implemented a similar sequential approach to summarize results in its employment discrimination studies and in Turner et al. (2013).

The list of modifications testers could request included—

- Lobby area and hallways.
 - Install a lever handle on the door.
 - Install an interior ramp to make elevators and hallways accessible from the lobby.
- Available and inspected units.
 - Lower thresholds in doorways over which rolling is difficult.
 - Install a lever handle on the door.
 - Reverse the swing of the entry door.
 - Lower the placement of light switches.
 - Reposition outlets.
 - Lower the placement of the thermostat.
 - Replace thick-pile carpeting with low-pile carpeting, tile, or hardwood flooring.
 - Replace a standard shower with a roll-in shower.
 - Install grab bars around the toilet or in the shower.
 - Remove the cabinet under the bathroom sink.
 - Lower the placement of kitchen cabinets.
 - Replace standard kitchen cabinet shelves with revolving or extending shelves.
 - Remove cabinets under the kitchen sink.

HDS-Disabilities found that when homeseekers who use a wheelchair asked whether modifications that would improve the accessibility of the available units were allowed, housing providers either fail to provide a clear response (21 percent) or explicitly deny (7 percent) more than one-fourth of the requests. Providers who do not provide a clear response say they do not know the answer, need to check with a supervisor, or simply do not offer a final response, which can limit the information a homeseeker has to make an informed and timely decision. Housing providers' modification approval rate varies by the type of request. Housing providers approve more than 80 percent of requests to install bathroom grab bars and lever door handles but approve fewer than 50 percent of requests to lower kitchen cabinets and replace carpets.

Tester Income Level

During the design phase of HDS-Disabilities, the project team also contemplated the income level testers should be assigned relative to the cost of the housing they sought. On previous HDSs, the testers have all been financial well qualified. By assigning lower incomes, testers' financial qualifications would more accurately reflect the income levels of people with disabilities. As Hoffman and

Livermore (2012) reported, in 2005, 1.1 to 1.4 million households with a working-age person with a disability had worst case housing needs, defined as “very low-income renters who do not receive government housing assistance and who either pay more than one-half of their income for rent, live in severely inadequate conditions, or both” (Nelson, 2008: 1). The project team determined, however, that by limiting the income levels of tester pairs, rather than allowing for them to reflect a range of income levels and relative housing options, the study’s focus would essentially shift. Rather than highlighting the experience of testers with disabilities, the study would detail the challenges of people with low income. Another consequence of assigning low incomes to tester pairs is that the study would have effectively excluded a significant segment of the rental housing stock—testers would not have income sufficient to consider many housing options. As a result, the project team determined testers on HDS-Disabilities would be assigned income consistent with past HDS practice, making testers well qualified for the housing about which they inquired.

Field Implementation Challenges of Testing for Disability Discrimination

The HDS-Disabilities field implementation team faced operational challenges, which were more significant than during previous testing studies because of the inherent complexity of testing for disability discrimination. Of the 29 local organizations that participated in HDS-Disabilities, more than one-third of the groups were centers for independent living or disability advocacy groups and the majority comprised fair housing groups. Most of the disability groups had no previous testing experience, and many of the fair housing organizations that had previously conducted disability testing had done very few tests or else had used actors or proxies representing disabled homeseekers, rather than using people with disabilities as testers. Regardless of the organizations’ previous experience or expertise, they all participated in a comprehensive training program, designed in consultation with the project’s expert advisors, to help prepare local project managers and test coordinators to skillfully coordinate tests while providing sufficient support to testers with disabilities.¹¹ Together, the extended field team organized in-person tester training sessions in each of the project sites. For the sessions with deaf and hard of hearing testers, organizations arranged for American Sign Language interpreters to attend trainings and also assist with practice tests. The training sessions for the wheelchair component were held at facilities that could accommodate testers with a variety of different wheelchairs (that is, manual, motorized or power chair, and scooters).¹² In addition to providing specific accommodations at the tester training sessions, the project implemented various modifications to the study’s protocols and procedures when they proved necessary. For example, because of the confidential nature of the work, previous HDSs have prohibited nonproject staff from attending training sessions and tester briefings and debriefings. Because some

¹¹ Local project managers and test coordinators attended in-person trainings conducted by Urban Institute staff and also participated in supplemental webinars.

¹² At some of the local tester training sessions, coordinators underestimated the amount of space needed to comfortably seat the number of attendees, particularly because of the varying sizes of wheelchairs. As a result, testers in wheelchairs were not able to move in and out of the room as easily as they might have in a larger venue. Any testing organization conducting training sessions with people in wheelchairs should carefully consider the size of the training facilities, the size of the office space, the dimension and placement of tables, and the availability and location of accessible parking and restrooms.

testers who used wheelchairs had personal aides who had to be present for training sessions or drove testers to site visits, however, the project established protocols (and a separate nondisclosure agreement) to delineate what aides could and could not do. Although they could help a tester enter a building where a site visit was being conducted, aides were prohibited from interacting with housing providers and accompanying a tester on the test. In addition, because notetaking was mandatory for all site visits, testers in wheelchairs with manual limitations were allowed to use other means to write down important information; some testers in wheelchairs used tablets or other electronic devices (which some found easier than writing by hand) or else were allowed to ask the housing provider to help them take notes. By outlining such protocol modifications and explicitly defining how they could be implemented, the project's field team helped ensure that procedures were consistently used among the 30 sites coordinating in-person tests.

Although the tests conducted for both components of HDS-Disabilities each had their own unique complexities, the tests completed for the wheelchair study were the most logistically difficult of the 14,000 paired-tests that the Urban Institute and its subcontractor testing organizations have completed since 2011. At the outset, informed in part by the experience of the 2005 disability pilot study, the project team anticipated that the most complicated operational challenges would fall into three main categories: (1) the recruitment and retention of testers, (2) the availability of reliable transportation, and (3) the risk of detection by housing providers. All three are factors in testing projects of any size, particularly those including in-person site visits; without capable, credible testers who have access to transportation, any study is doomed to failure. When a project's tester pool includes many people who use wheelchairs, however, completing a testing study while avoiding detection by the housing industry can be particularly complex.

The Recruitment and Retention of Testers

As with previous HDSs, HDS-Disabilities project staff expended considerable effort recruiting capable and committed testers who could be matched on sex, race or ethnicity, and age to compose suitable tester pairs. Local organizations worked to achieve specific targets for racial and ethnic groups based on metropolitan area census data provided by the Urban Institute. As with previous HDSs, the recruitment of Hispanic and Asian-American testers proved difficult; local project staff and the team based at the Urban Institute connected with Hispanic and Asian-American community groups and national umbrella organizations to try to increase the diversity of the tester pool. Ultimately, more than one-half of HDS-Disabilities testers were White (55 percent among testers in wheelchairs and 57 percent among control testers), one-fourth were Black (25 and 23 percent, respectively), and 16 percent of both testers in wheelchairs and control testers were Hispanic (Levy et al., 2014).

Even fair housing organizations with robust testing programs needed to recruit additional testers—both testers with disabilities and their tester matches—before the start of data collection. Many organizations expanded their typical tester recruitment efforts, forging new relationships with disability organizations, advocacy groups, and sports teams, such as wheelchair basketball and water polo. Even though the study's testing organizations were successful in reaching their overall recruitment goals, groups faced significant difficulties retaining testers. In previous studies, sites were affected by chronic underemployment of testers; when testers were offered permanent employment,

they left their short-term, part-time jobs as testers. This reality continued to prove true for the control testers on HDS-Disabilities. For the testers with disabilities, however, their ability to remain on the project was affected less by the lure of other employment than by two other distinct factors. First, many testers with disabilities receiving financial assistance through state or federal programs had to adhere to limits on the amount of additional income they could earn without having their benefits reduced for subsequent years. As a result, many testers had a strict, maximum number of test assignments they could accept, and, after they reached that limit, they left the project. Second, many of the testers who used wheelchairs had ongoing healthcare needs, which affected both their availability and their capacity to conduct site visits. During the course of data collection, a number of testers in wheelchairs left the project because they became seriously ill. As a result of the level of tester attrition, the project sites were forced to continue to recruit and train new testers throughout the duration of the study, which required considerable staff time.

The Availability of Reliable Transportation

Before the start of testing, the project team assessed the availability of transportation options for testers in wheelchairs in each of the study's 30 sites and found tremendous differences in the availability and cost of public and private transportation options. In many sites, public paratransit services were deemed too unreliable to be used. Even in cities where wheelchair accessible transportation options were more readily available, many services required testers to allocate additional transit time to ensure they would arrive on time for appointments with housing providers; if testers arrived too early or too late, indications of differential treatment might be attributed to the timeliness of the tester rather than the housing provider's behavior. In some sites, testers had lengthy wait times for drivers taking them to site visits, which were exacerbated during inclement weather. During HDS-Disabilities testing, extreme weather events (for example, heavy snow, ice storms, extreme cold, and flooding) slowed or halted testing temporarily in 18 of the 30 project sites. Some testers used their own vehicles, which they or personal aides drove, but encountered problems finding places to park when they reached their appointments. At apartment complexes with parking lots, some testers described that golf carts (used by agents to tour the facilities with prospective tenants) were parked in accessible spaces. Even when they did not block the spaces entirely, the presence of the carts affected the ability of testers to exit their vehicles with their wheelchairs. Although previous HDSs required testers to travel the same relative distances—testers could receive assignments sending them anywhere within the metropolitan area¹³—the transportation challenges that testers with mobility disabilities faced ultimately were tackled, site by site, with strategic, advance planning and with the extraordinary patience and perseverance of the testers.

The Risk of Detection by Housing Providers

One of the most significant challenges the project team faced was avoiding detection by landlords while completing the required number of paired tests in each study site.¹⁴ Because a sudden influx

¹³ Metropolitan statistical area definitions were used to determine geographic boundaries for study sites.

¹⁴ Households that include a person with a mobility disability—defined as having serious difficulty walking or climbing stairs—represent less than 5 percent of the total metropolitan renter population—about 1 in 21 renter households (2010 American Community Survey [ACS] Public Use Microdata analysis; Levy et al., 2014).

of renters in wheelchairs within a small metropolitan area could be particularly conspicuous, the project team decided that the sampling frame for study sites be restricted to areas with a minimum population threshold of 450,000 (Levy et al., 2014). The Urban Institute field team implemented specific procedures to minimize the inherent disclosure risk. Local project staff used detailed tracking logs to ensure no single property was tested more than once for the study. In metropolitan areas with limited availability of eligible properties (that is, housing that could be entered by people using wheelchairs), local organizations could assign a subsequent test to a previously tested property if a sufficient period of time (typically no less than a period of several months) had elapsed from the date of the first site visit. Project staff based at the Urban Institute authorized such assignments on a case-by-case basis. In addition, local testing organizations worked to maintain a steady but moderate testing pace from week to week, which enabled them to keep careful watch on where testers had been previously while also avoiding large surges of site visits in a concentrated period.

As with previous HDSs, the Urban Institute field team used careful oversight and regular communication to anticipate operational challenges and correct problems as they developed in the study sites. By maintaining daily contact with test coordinators and monitoring incoming data (submitted via an online data collection system), regional coordinators at the Urban Institute provided timely feedback to sites and helped ensure tester adherence to reporting requirements. The structure of the project team enabled the team at the Urban Institute to facilitate the sharing of best practices from the field in tester recruitment and retention, test coordination, and project management among the many local testing organizations, all of which helped the team overcome the study's toughest hurdles.

Key Study Findings

The key findings of HDS-Disabilities, which are briefly summarized in this section, highlight the incidence and forms of differential treatment experienced by homeseekers who are deaf or hard and hearing and by homeseekers who use wheelchairs.

Discrimination Against Homeseekers Who Are Deaf or Hard of Hearing

The deaf and hard of hearing study of HDS-Disabilities included 1,665 tests, which were conducted remotely via telephone or the Internet in 168 metropolitan areas that account for more than four-fifths (82 percent) of the renter population that is deaf or hard of hearing (Levy et al., 2014).¹⁵ Three local testing organizations were responsible for coordinating tests, and the tester pool was composed of people who are deaf and those who are hard of hearing. Testers who are deaf or hard of hearing contacted housing providers by using one of three commonly used TRSs, and hearing testers used the telephone to make contact with housing providers. Testers inquired about available rental housing, including rent costs and terms, and requested an appointment to meet in person but then cancelled more than an hour before the meeting time. Deaf and hard of hearing testers did not request any reasonable accommodations or modifications.

HDS-Disabilities found that housing providers are less likely to communicate with homeseekers who are deaf or hard of hearing and also tell those homeseekers about fewer available units (Levy

¹⁵ The data were drawn from the ACS, which asks if a person is deaf or has serious difficulty hearing.

et al., 2014). Hearing testers successfully reach an agent in 95.8 percent of tests compared with 90.7 percent for deaf and hard of hearing testers, a statistically significant difference of 5.1 percentage points. In one specific example, a deaf tester contacted the housing provider via VRS. After the CA explained the call was from a deaf person, the provider said that she did not take those kinds of calls, apologized, and hung up. On another test, a deaf tester used the IP CTS to make contact with the housing provider. He informed the housing provider that he was deaf before asking about the advertised apartment, but the provider said she was too busy but could e-mail the tester later.

When both testers of a pair do reach a housing provider, testers who are deaf or hard of hearing are 2.3 percentage points less likely to be told about any available units. When the housing providers' willingness to communicate with a homeseeker was combined with the availability of units, the study found housing providers tell deaf and hard of hearing testers about 0.14 fewer housing units per inquiry than they do hearing testers. In other words, for every seven attempts to find out about available rental housing, a homeseeker who is deaf or hard of hearing learns about one fewer available units than a comparable hearing homeseeker (Levy et al., 2014). In addition, housing providers who are willing to communicate with testers who are deaf or hard of hearing were found to be equally likely to schedule an appointment with both testers in a pair. Overall, deaf and hard of hearing testers were provided with the same average yearly net cost of units by housing providers, regardless of their disability status.

Discrimination Against Homeseekers Who Use Wheelchairs

The wheelchair component of HDS-Disabilities included 1,265 tests conducted in 30 metropolitan areas that represented nearly three-fourths (73 percent) of the population that has a mobility disability and resides in rental housing (Levy et al., 2014).¹⁶ After testers were given assignments, they contacted housing providers by telephone and made requests for appointments; those who were able to secure an appointment conducted in-person visits. Testers inquired about available housing for rent, including rent costs and terms, and testers who use wheelchairs made requests for up to three reasonable modifications but did not request any reasonable accommodations.

When renters who use wheelchairs inquire about advertised housing that appears to be accessible, they are treated less favorably on several key indicators than equally qualified renters who are ambulatory (Levy et al., 2014). Housing providers are 1.7 percentage points less likely to make an appointment with homeseekers who use wheelchairs than with control testers. When both testers of a pair are able to meet with a provider and a suitable unit is available, users of wheelchairs are 2.4 percentage points less likely to be told about any available units. When housing providers tell both testers about available units and where units can be inspected by a person who uses a wheelchair, providers are 3.1 percentage points less likely to show any units to those who use wheelchairs. For example, in one test, the housing provider showed apartment floor plans to the tester who uses a wheelchair. When the tester asked to view an available apartment, the housing provider said it could not be shown because it was occupied. The housing provider did show an apartment to the control tester, however, and the control tester noted that the apartment was

¹⁶ These data are drawn from the ACS, which asks if a person has serious difficulty walking or climbing stairs.

accessible. The findings indicated no statistically significant difference between average yearly net costs for people who use wheelchairs and for people who do not. Further, no factor consistently contributes to variations in treatment of testers using wheelchairs.

Housing providers were 33.7 percentage points more likely to make comments about housing accessibility, 6.1 percentage points more likely to make comments about people with disabilities, and 4.8 percentage points more likely to make comments about fair housing to a homeseeker who uses a wheelchair than they were to a homeseeker who does not use a wheelchair (Levy et al., 2014). Most housing providers' comments documented by testers tended to be neutral, informative, or helpful. For example, agents told testers that other renters in the building used a wheelchair, mentioned that a loved one used a wheelchair so they understood accessibility needs, or pointed out a unit with a wheelchair ramp at the entrance. Other housing providers' comments and actions, however, were negative. One agent told a tester, "Oh, you're disabled ... you don't work." Another agent said that she thought the tester would be able to get out of the wheelchair to enter the building. One tester wrote in the test narrative that a housing provider said he had "never seen any disabled people for the building so this is new for him, especially someone who is severely disabled as I appear to be" (Levy et al., 2014: 50).

HDS-Disabilities also found that, on average, one-half of all advertisements for privately owned rental housing in the 30 study sites appeared to lead to units accessible by people who use wheelchairs (Levy et al., 2014). Overall, only 44 percent of advertisements for rental units randomly selected for paired testing led to units identified as accessible. As previously noted in this article, however, because HDS-Disabilities used an operational standard of accessibility and not one equivalent to specific laws or design and construction standards, the report urges caution in the interpretation of this measure. The study did not conduct a formal survey of the rental housing stock, and the rates of accessibility do vary considerably across the 30 metropolitan areas included in the study, from a low of 11 percent to a high of 87 percent. Sites with a higher proportion of rental units in multifamily buildings rather than single-family housing and sites with a higher proportion of rental units in buildings constructed after 1990 have higher rates of accessible units.

Limitations

Although the power of the paired-testing methodology is undeniable because of the direct observations it provides of differences in treatment, it does have limitations. As in the previous HUD-funded studies, HDS-Disabilities provides information on the inquiry and information-gathering stages of the homeseeking process but does not capture all the differential treatment a tester might experience as a renter (Levy et al., 2014). The study does not reveal the outcomes a homeseeker might have later in the process, such as during the application phase or lease signing. Paired-testing also would be difficult to use in the post-occupancy phase to learn more about the experience of residents; because a housing provider is already aware of tenants' financial and household characteristics, the challenge of forming appropriately matched pairs could prove insurmountable. Further, because HDS-Disabilities testers were assigned incomes that made them well qualified for the housing about which they inquired, the results may not reflect the incidence of discrimination that more marginally qualified disabled homeseekers with higher rent burdens might experience.

Further Directions

The findings of HDS-Disabilities strongly suggest that important next steps should be taken along three categories: (1) future research, including future testing studies; (2) housing provider education and outreach; and (3) action, including the improvement of TRS technologies and the increase of the accessible rental housing stock.

Future Research

Future paired-testing studies are needed to provide additional evidence on the experience of people across the spectrum of disability and to produce estimates of housing structures that are in compliance with the design and construction requirements. Future testing studies could highlight the following—

- **Treatment of more marginally qualified homeseekers with disabilities.** As with previous HDSs, testers participating in HDS-Disabilities were assigned household incomes that made them well qualified for the housing about which they inquired. Although testers were rarely asked about their income by housing providers, they were commonly asked about the maximum rent they were willing to pay, which always was higher than the rent of the unit about which they inquired. For example, even if a tester began a test inquiring about a one-bedroom apartment, they were provided sufficient income to consider larger available units. As a result, the maximum rent a tester provided signaled to the housing provider how much more they could afford to spend overall, which gave some indication of a tester's financial capacity. In addition, if the agent raised the subject of credit standing, testers were able to volunteer they had excellent credit. The topic frequently arose during discussions about the security deposit when an agent might indicate that the cost was dependent on one's credit score. Taken together, these indicators may have helped convey the strength of the testers' financial qualifications. Without subsequent testing, it cannot be known definitely if (and to what extent) people with disabilities who are less well qualified (for example, lower income or blemished or no credit) or who receive public assistance (for example, Supplemental Security Income or Social Security Disability Insurance) would experience treatment comparable to more highly qualified homeseekers.
- **Treatment of homeseekers who are deaf or hard of hearing during in-person visits.** Because the tests that were conducted for the deaf and hard of hearing component of the study were done remotely, and because most comments that housing providers made expressed concern about the ability to communicate with the homeseekers, a future testing study could include in-person site visits to examine how the treatment of testers might vary during face-to-face meetings with housing providers. In addition, the study could include requests for modifications, such as the installation of flashing lights for the doorbell, which were not included as part of the HDS-Disabilities remote testing effort.
- **Treatment of homeseekers who are blind or visually impaired.** As the exploratory component of the 2005 disabilities pilot study demonstrated, the paired-testing methodology can be used to examine the level of differential treatment against people who are blind or visually impaired. A future study also could focus on housing providers' willingness to make

reasonable accommodations for people using assistance animals. Given the extent to which testers in the pilot study had difficulty finding front doors, using intercoms or buzzer systems, or accessing offices or properties, the testing protocol would need to outline whether aides would be needed to help testers with transportation and with gaining entry to the housing.

- **The compliance of housing with the Fair Housing Act design and construction standards.** Although HDS-Disabilities did not produce national estimates of accessibility among rental units, the number of wheelchair-accessible units identified in the study sites may indicate a problem requiring additional attention. A future study could use a carefully defined population of housing structures from which to draw a representative sample of units covered by the Fair Housing Act's design and construction requirements for housing first occupied after March 13, 1991. The study could be carried out by single testers (instead of pairs) trained to conduct design and construction assessments and could produce regional or national estimates.

Additional research methods may shed light on the experience of in-place tenants with disabilities. For example, tenant surveys could provide information on modification requests made before lease signing or subsequent to occupancy. Although HDS-Disabilities found that more than 70 percent of modification requests were approved, without conducting subsequent tenant surveys, the amount of time that tenants waited for official approval or implementation of the requested modifications and the ultimate cost of making specific modifications cannot be known. In addition, because HDS-Disabilities found differences in the willingness of housing providers to communicate with deaf or hard of hearing testers based on the type of TRS used, more details should be known about the users of each TRS type, such as demographic and socioeconomic characteristics and whether those who are deaf and hard of hearing face barriers in accessing particular TRS types.

Housing Provider Education and Outreach

The findings of HDS-Disabilities also strongly point to the need for ongoing housing-provider education and outreach. First, housing providers must be trained on the Fair Housing Act and other laws, such as the Americans with Disabilities Act, prohibiting discrimination against people with disabilities. The comments that housing providers made about housing accessibility and modification requests suggest that some landlords and property managers do not understand their legal obligations. Second, housing providers should increase their awareness of the accessibility of their properties and the internal procedures for requesting and approving reasonable modification requests. When a tester asked over the telephone or during e-mail contact whether a property was wheelchair accessible, a number of housing providers did not know. Providers also could not (or did not) always respond to testers' requests for permission to make reasonable modifications to lobby areas or apartment interiors.

Action

Finally, the findings of the study suggest that action is needed on three important fronts. First, public and private fair housing organizations should continue to aggressively pursue complaint-based and systemic testing investigations. Without ongoing monitoring and enforcement, discriminatory housing provider patterns and practices may persist in communities across the country and new forms of differential treatment may be allowed to extend their reach. The resources allocated to

testing programs also should reflect the diverse needs and demographics (for example, ethnic and racial minorities, immigrants, and seniors) of the disability community in any jurisdiction. Second, improvements to TRS technologies are needed. Study findings show that housing providers contacted by people using VRS technology were more likely to communicate with the caller than with those using the other TRSs. Compared with the two other technologies used in HDS-Disabilities, VRS supports a smoother and quicker pace of communication between a person who is deaf and a person who is hearing. Use of VRS relies on a person's ability to communicate in sign language, however, so this technology is not an option for everyone who needs to use a TRS. Improvements in communication technologies could improve the housing search, and possibly the outcomes, for people who begin their housing search remotely. Finally, the stock of accessible rental housing must increase. Discrimination-based impediments to housing access, along with inaccessible housing stock and population trends, likely will increase the need for accessible housing in cities across the United States. Findings from this study show that people who use wheelchairs face reduced housing options compared with people who are ambulatory. As the U.S. population trends older and rates of disability increase, competition could increase among renters for accessible apartments and homes. Increasing pressures on housing stock could be particularly strong in markets where a predominance of housing was built for first occupancy before March 13, 1991, when the Fair Housing Act's design and construction requirements went into effect. Housing policy and industry professionals in cities with an older housing stock and aging resident populations should consider how to meet an increased demand for accessible units.

Conclusion

Because HDS-Disabilities is the first paired-testing study to produce national estimates of rental housing discrimination against people who are deaf or hard of hearing and people who use wheelchairs, the evidence it provides can serve as a useful benchmark for policymakers and practitioners alike as they address the discriminatory treatment of people with disabilities. These findings also may help practitioners assess the magnitude of barriers to accessible housing as an aging population further increases the demand, particularly in communities with an older housing stock. In addition, the lessons learned by the study's field implementation team may also be helpful to local testing organizations in communities across the country as they conduct disability-based testing. Future paired-testing studies are needed to help measure change in the forms and incidence of discrimination over time and to provide estimates on differential treatment against varying subgroups of people with disabilities.

Acknowledgments

HDS-Disabilities could not have been achieved without the extraordinary dedication and talent of field staff based at local organizations across the country, including disability advocacy groups, centers for independent living, and fair housing organizations. The perseverance of testers, including those who traveled significant distances (sometimes to other cities and states) to help conduct tests, provided an inspiring example for the entire project team based at the Urban Institute and led by Diane K. Levy, Margery A. Turner, Rob Santos, Doug Wissoker, Rob Pitingolo, Helen Ho, and Dave

D’Orio. Tremendous credit is also due to the regional coordinators who ensured each test met the highest standards: Jim Crawford, Lainie Dulaney, Laura Loessner, Thu Nguyen, Max Pollock, and Sarale Sewell.

Author

Claudia L. Aranda is a senior research associate in the Metropolitan Housing and Communities Policy Center at the Urban Institute.

References

- Hoffman, Denise W., and Gina A. Livermore. 2012. “The House Next Door: A Comparison of Residences by Disability Status Using New Measures in the American Housing Survey,” *Cityscape* 14 (1): 5–33.
- Jacobsen, Linda A., Mary Kent, Marlene Lee, and Mark Mather. 2011. “America’s Aging Population,” *Population Reference Bureau* 66 (1): 1–16.
- Joint Center for Housing Studies (JCHS) of Harvard University. 2014. *Housing America’s Older Adults—Meeting the Needs of an Aging Population*. Cambridge, MA: Joint Center for Housing Studies of Harvard University.
- Levy, Diane K., Margery Austin Turner, Rob Santos, Doug Wissoker, Claudia L. Aranda, Rob Pitingolo, and Helen Ho. 2014. *Discrimination in the Rental Housing Market Against People Who Are Deaf and People Who Use Wheelchairs: National Study Findings*. Washington, DC: U.S. Department of Housing and Urban Development.
- National Council on Disability. 2010. *State of Housing in America in the 21st Century: A Disability Perspective*. Washington, DC: National Council on Disability.
- National Fair Housing Alliance (NFHA). 2014. *2014 Fair Housing Trends Report*. Washington, DC: National Fair Housing Alliance.
- Nelson, Kathryn. 2008. *The Hidden Housing Crisis: Worst Case Housing Needs Among Adults With Disabilities*. Washington, DC: Consortium for Citizens With Disabilities Task Force.
- Rolka Loube Saltzer Associates. 2014. “Interstate TRS Fund Performance Status Report: August 2014.” http://media.wix.com/ugd/455e4d_88b979484cb24c14a934ba343bb812a2.pdf.
- . 2012. “Interstate TRS Fund Performance Status Report: Funding Year July 2012–June 2013.” http://media.wix.com/ugd/455e4d_ba61f9b0649f4357945bd3b7679a1318.pdf.
- She, Peiyun, and Gina A. Livermore. 2009. “Long-Term Poverty and Disability Among Working-Age Adults,” *Journal of Disability Policy Studies* 19 (4): 244–256.
- Smith, Stanley K., Stefan Rayer, and Eleanor A. Smith. 2008. “Aging and Disability: Implications for the Housing Industry and Public Policy in the United States,” *Journal of American Planning Association* 74 (93): 289–306.

Smith, Stanley K., Stefan Rayer, Eleanor A. Smith, Zhenglian Wang, and Yi Zeng. 2012. "Population Aging, Disability and Housing Accessibility: Implications for Sub-National Areas in the United States," *Housing Studies* 27 (2): 252–266.

Turner, Margery Austin, Carla Herbig, Deborah Kaye, Julie Fenderson, and Diane K. Levy. 2005. *Discrimination Against Persons With Disabilities: Barriers at Every Step*. Washington, DC: U.S. Department of Housing and Urban Development.

Turner, Margery Austin, Rob Santos, Diane K. Levy, Doug Wissoker, Claudia L. Aranda, and Rob Pitingolo. 2013. *Housing Discrimination Against Racial and Ethnic Minorities 2012*. Washington, DC: U.S. Department of Housing and Urban Development.

Additional Reading

Turner, Margery Austin, and Stephen Ross. 2003. *Discrimination Against African Americans and Hispanics: Supplemental Results From Phase II of HDS2000*. Washington, DC: U.S. Department of Housing and Urban Development.

Turner, Margery Austin, Ray J. Struyk, and John Yinger. 1991. *Housing Discrimination Study Synthesis*. Washington, DC: U.S. Department of Housing and Urban Development.

Wienk, Ronald E., Clifford E. Reid, John C. Simonson, and Frederick J. Eggers. 1979. *Measuring Discrimination in American Housing Markets: The Housing Market Practices Survey*. Washington, DC: U.S. Department of Housing and Urban Development.

Other Protected Classes: Extending Estimates of Housing Discrimination

Margery Austin Turner
Urban Institute

Abstract

This article discusses the challenges involved in extending the paired-testing methodology from its original purpose of measuring discrimination based on race and ethnicity to rigorously measuring the incidence and forms of discrimination against other protected classes of homeseekers. It highlights three critical design challenges that any such study must resolve and draws on three recent pilot studies to illustrate these challenges and how they can be resolved. The power of paired testing is greatest when three key conditions are met: (1) the standard for comparison is unambiguous, (2) the relevant segments of the housing market can be identified and tested, and (3) testers' status is apparent to housing providers.

Introduction

During the past four decades, paired testing has proven to be an effective tool for measuring the incidence and forms of discrimination against minority homeseekers. The matched pairs directly control for other factors (like income, wealth, household composition, and housing preferences) that contribute to differences in housing outcomes to reveal differences in treatment based solely on race or ethnicity. Results that document unequal treatment of equally qualified homeseekers are easily understandable and offend most people's idea of basic fairness. As Oh and Yinger's (2015) article in this issue of *Cityscape* discusses, findings from paired-testing studies have influenced federal fair housing legislation, funding for fair housing enforcement, and public understanding of the persistence of discrimination in housing market.

Given the effectiveness of paired testing, the idea of extending the methodology to measure the incidence and forms of discrimination against other groups is appealing to policymakers and practitioners. In recent years, the U.S. Department of Housing and Urban Development (HUD) has funded pilot or exploratory testing studies of discrimination against families with children; lesbian,

gay, and transgender homeseekers; and renters using housing vouchers.¹ For all these classes of homeseekers, other research provides evidence of poor housing outcomes, and it seems straightforward—at least at first blush—to use paired-testing to find out whether and how discriminatory treatment by housing providers may be contributing to those outcomes.

Extending the paired-testing methodology highlights three key conditions that undergird its power to compellingly document discriminatory treatment of minority homeseekers. First, ***the standard for comparison is unambiguous***. Ample evidence demonstrates that minority homeseekers experience inferior housing outcomes compared with White homeseekers, and it is straightforward to match minority and White testers so that their race or ethnicity is the only difference observed by a housing provider. Second, ***both minority and White households can reasonably be expected to occupy all sizes and types of housing available in the marketplace***. Testers can easily be assigned financial and other characteristics to correspond to the characteristics of any advertised house or apartment. Any available housing unit should, in principle, be offered to both on the same terms. Third, ***a person's race or ethnicity is (in most cases) immediately apparent to housing providers***, based on appearance, name, and possibly even speech patterns.

These three conditions do not necessarily apply for other classes of homeseekers. As a consequence, researchers seeking to extend the paired-testing methodology face thorny design and implementation challenges to produce credible and compelling estimates of differential treatment discrimination. To be specific, efforts to extend paired testing to new classes of homeseekers must effectively address one or more of the following three questions.

1. What is the appropriate comparison group?
2. Should some segments of the housing stock be either excluded or oversampled?
3. When and how should a tester's status be disclosed to housing providers?

This article focuses in turn on three pilot testing studies sponsored by HUD to assess the feasibility of effectively testing for discrimination against families with children; lesbian, gay, and transgender people; and housing voucher recipients.² As of this writing, results of these pilot studies have not been released, so this article focuses not on findings but on the key design challenges, discussing in depth how each study resolved the three questions in the preceding list. The article then concludes with implications for future paired-testing studies aimed at measuring the incidence of discrimination for other groups of homeseekers.

Families With Children

In 1988, the Federal Fair Housing Act of 1968 was amended to protect families with children and pregnant women from discrimination in the housing market. Before these protections were

¹ Note that the federal Fair Housing Act prohibits discrimination based on the presence of children and on gender, but not based on sexual orientation or identity or on housing subsidy receipt. Some states and localities provide legal protections against discrimination based on sexual orientation and some prohibit housing discrimination based on source of income.

² Aranda's (2015) article in this issue of *Cityscape* addresses the challenges of effectively measuring discrimination against people with disabilities.

in place, it was not uncommon for rental properties to exclude children under 18 or restrict the number of children allowed (Colten and Marans, 1982). Today, families with children face significant challenges in the rental housing market compared with childless renters (Aratani et al., 2011; JCHS, 2013). Too little is known, however, about the extent and forms of discrimination against families with children. Analysis of fair housing complaints suggests that families with children experience discrimination from rental housing providers, but no recent research has investigated the problem systematically (NFHA, 2012, 2011).

Therefore, HUD commissioned the Urban Institute to adapt the paired-testing methodology to measure discrimination against families with children seeking rental housing and apply it in three metropolitan areas. The primary goals of this pilot study were to develop a preliminary estimate of the incidence of discrimination against families with children in the rental housing market and to assess the feasibility of conducting a larger national paired-testing study (Aron et al., forthcoming). All three challenges introduced previously had to be addressed in the design of this pilot effort.

What Is the Appropriate Comparison Group?

At first, the answer to this question may appear to be straightforward: the comparison should be a childless renter household with the same financial qualifications and housing preferences. But a childless household with three or four members is fundamentally different from a family with children with the same number of members. Consider, for example, an approach in which the two renter households of each matched pair are the same size but, in one case, some of the household members are children and, in the other, all are adults. Under this design, the control household paired to a married-couple family with two children would be a married couple living with two other adults. This would be problematic for several reasons. First, it represents a household structure that is rarely seen in the United States—only 2 percent of all rental households have four or more adults and no children (2010 American Community Survey). Second, to assign this four-adult household the same combined income as a married couple with two children would require that all four worked at lower paying jobs or that one or two are unemployed. Third, having four adults in a household resembles a group quarter's situation (for example, five college students sharing a house), an extended family situation, or a household of related adults sharing housing perhaps for reasons of economic hardship. In some jurisdictions, renting to a large group of unrelated adults violates local occupancy regulations.

Similar incongruences emerge with most other family sizes and scenarios. Even matching a single parent and one child to a childless couple raises comparability concerns, because housing providers might react more to the parent's nonmarried status than to the presence of a child. Thus, a paired-testing design for families with children that holds household size constant seems destined to yield inappropriate, apples-to-oranges comparisons.

Analysis of national data comparing renter households with and without children suggests that most childless renters are singles or couples as opposed to larger groups of childless adults. So for this pilot study, the counterfactuals for families with children were childless singles and childless couples. To be more specific, for single-parent female-headed households, the counterfactual was a childless single female, and for married couples with one or two children, the counterfactual was a childless couple. This design reflects the fact that renters with children compete in today's housing market with smaller, childless households.

However, the difference in household size between the two members of each tester pair raises concerns about how to interpret some differences in treatment that might be observed. The two testers in each matched pair of this pilot study inquired about the same size of housing unit, even though their assigned household sizes differed. In every test, the family with children was larger than the childless household. If housing providers recommend or show larger units—and possibly, higher cost units—to families with children than to their childless counterparts, should this difference in treatment be classified as discrimination? This issue was resolved in the pilot study by first ensuring that both testers in each pair asked about the same size unit and reported whether it was available to them. They also documented the size of other units they were recommended and shown and the asking rent for each of these units. Measures of differential treatment were then constructed for the following indicators.

- Was the tester told that any units were available?
- How many units of the requested size were available? How many units of any size?
- What was the average size of the available units?
- What was the average asking rent for available units, controlling for size?

If families with children were told about fewer available units than their childless counterparts, denied information about units of the requested size, or steered toward larger and more expensive units, this treatment could be considered discriminatory. If, on the other hand, families with children were told about the same units as their childless counterparts *plus* some larger units, this treatment probably could be considered nondiscriminatory, even though it might result in a higher average asking rent across all available units. In fact, some might view this outcome as favoring the family with children relative to the childless individual or couple, by offering more—or more desirable—housing options.

Should Some Segments of the Housing Stock Be Either Excluded or Oversampled?

Most renter families with children (79 percent) live in two- or three-bedroom units, with only 8 percent living in one-bedroom units. Even among families with one child, only 11 percent occupy one-bedroom units. This evidence suggests the possibility of excluding one-bedroom units from the sample of advertised housing selected for testing. However, one reason so few families occupy these units may be that they are in high demand (given their affordability relative to larger apartments) and landlords may know they can easily be rented to childless renters. In other words, data on where families currently live may reflect ongoing discriminatory practices rather than families' needs or preferences. Therefore, the pilot study included one-bedroom units in its testing sample.

The inclusion of one-bedroom units raises a concern, however, about whether occupancy standards might complicate the study findings, because some tests would involve inquiries about units that would involve more than two people per bedroom. Local occupancy laws vary widely, but no more than two people per bedroom is widely considered a standard rule of thumb. Some landlords may think that local laws limit occupancy to two people per bedroom even though this may not be the case, and others may use the two-people-per-bedroom limit as an excuse for excluding families with children. Actual occupancy standards are considerably more nuanced and typically consider both the number of bedrooms and the square footage of the unit as a whole.

Because the central purpose of the pilot study was to measure differential treatment of renter households based on the presence of children—not to assess landlords’ adherence to varying and potentially complex occupancy standards—all tests of one-bedroom units were assigned either single or married parents with only one child. Even so, a married couple with one child might stretch a landlord’s perception of occupancy limits for a one-bedroom unit. If large shares of landlords were using occupancy standards as an excuse to exclude families with children or believe such families violate the standards, this pattern would be reflected in the results of the one-bedroom tests.³

When and How Should a Tester’s Status Be Disclosed to Housing Providers?

In most paired-testing studies, one member of the fictional household calls or visits the sampled housing provider to inquire about the availability and terms of a house or an apartment. The presence or absence of children from the household is not discernable unless the tester discloses it. Therefore, all testers were directed to disclose their assigned household composition at the beginning of a telephone conversation or during a telephone message or e-mail. For example, if a tester was assigned a married profile, he told the housing provider that he and his wife were looking for housing. If a tester was assigned children for a given test, she indicated the age and sex of the children. Testers also disclosed their assigned household composition at the beginning of their in-person visits to housing providers. This disclosure protocol did not raise significant doubts or concerns, because it seems natural for a homeseeker to describe the composition of his or her household early in any inquiry about available housing.

Exhibit 1 sums up the pilot study’s response to the challenges of extending the paired-testing methodology to measure rental housing discrimination against families with children. This design differs from conventional paired testing in that the households in the protected group are larger than those in the comparison group, making it plausible for housing providers to offer them larger and more expensive housing units. To some modest degree, this undermines the power of the paired testing to measure differences in treatment all other things being equal.

Exhibit 1

Paired-Testing Methodology Extended to Families With Children

	Comparison Group	Housing Types Excluded or Oversampled	Disclosure of Status
Household composition	Childless singles and couples	Efficiencies excluded; one-bedroom units oversampled	Mention children early in both phone and in-person contact

³ This issue suggests the value of a rigorous study of local occupancy standards, how they are enforced, and how housing providers understand and apply them.

Lesbian, Gay, and Transgender People

Until very recently, much of what we knew about housing discrimination against lesbian, gay and transgender people came from surveys that asked respondents to report whether they have experienced discrimination while searching for housing (Colvin, 2004; Grant, Mottet, and Tanis, 2011; Kaiser Family Foundation, 2001). These surveys found that many lesbian women, gay men, and transgender people feel discriminated against, but the surveys do not provide rigorous estimates of the incidence or forms this discrimination can take. To the extent that self-reports of discrimination gathered by the surveys might capture only the most blatant forms of discrimination, the findings likely underestimate the actual occurrence of discrimination.

A handful of other studies have used e-mail contact to measure discrimination. The most comprehensive of these is HUD's recent study on housing discrimination against same-sex couples, which found that male same-sex couples were slightly less likely to receive e-mail responses from housing providers relative to comparable heterosexual couples. Other differences in treatment were not statistically significant (Friedman et al., 2013). In addition, a number of fair housing organizations in the United States have conducted in-person, paired tests of housing discrimination based on sexual orientation and gender identity for both research and enforcement purposes. For example, the Fair Housing Centers of Michigan studied housing discrimination based on sexual orientation with testers posing as same-sex couples matched with testers posing as heterosexual couples (Fair Housing Centers of Michigan, 2007). The tests, which targeted differences in treatment in rental housing, homes for sale, and home financing, found widespread discrimination against same-sex couples. In-person testing captured a much broader range of treatment than the very limited number of variables observed in e-mail-testing studies, reporting differential treatment related to the rent amount, the number of units discussed, offers of rental applications, and offers of move-in incentives. Testers also reported on subtler forms of treatment, such as whether an agent made any comments about gay people or homosexuality.

This experience led HUD to commission a pilot paired-testing study to measure the incidence and forms of discrimination against lesbian and gay renters and against transgender individuals seeking rental housing. This study will record and compare treatment both during telephone inquiries and during in-person visits with rental housing providers. The toughest design challenge for this pilot study revolves around how to disclose to housing providers the fact that a homeseeker is lesbian, gay, or transgender. In fact, the pilot study resolves this challenge differently for lesbian and gay people than for transgender people.

What Is the Appropriate Comparison Group?

For lesbian women and gay men, it seems quite clear that the appropriate comparison group consists of heterosexual women and men, respectively, with the same household composition. In other words, a single lesbian woman would be matched with a single heterosexual woman, and a gay man with a husband (or partner) and a child would be matched with a heterosexual man with a wife (or partner) and child. For purposes of the pilot study, all testers were assigned a partner; none posed as single renters, whether gay or straight, in part because of the challenge of

establishing a credible protocol for disclosing a single person's sexual orientation (discussed further below). Future studies might explore strategies for measuring discrimination against single lesbian women and gay men.

The pilot study matched transgender people with cisgender⁴ individuals whose gender matched the gender identity of the transgender person. Transgender people who did not identify as either female or male (who identified as gender queer) were randomly assigned a male or female cisgender match. In the pilot study, all transgender and cisgender testers posed as single individuals.

When and How Should a Tester's Status Be Disclosed to Housing Providers?

When searching for a house or an apartment to rent, lesbian women, gay men, and transgender people might choose to explicitly convey their sexual orientation or gender identity early in the interaction with housing providers in an effort to weed out disrespectful providers. On the other hand, they might choose *not* to convey identity, on the grounds that their sexual orientation or gender identity should not be a concern of the housing provider. Three options were considered for the pilot study, each offering different advantages and disadvantages.

1. Disclose sexual orientation or gender identity during the initial telephone inquiry and again at the start of the in-person visit. This option would ensure that housing providers are aware of each tester's status from the earliest point in the transaction and will presumably treat lesbian, gay, and transgender people unfavorably if they are so inclined. The disadvantage is that disclosure—especially during an initial telephone inquiry—may seem awkward and unnatural and may therefore lead to detection. Do most lesbian, gay, or transgender people actually make a point of disclosing their status when making an appointment to see a house or an apartment?
2. Wait until the in-person visit to disclose sexual orientation or gender identity. This option limits the analytic value of information gathered at the telephone inquiry stage, but it may reduce the risk of detection. In addition, it still raises the question of whether lesbian, gay, and transgender people proactively disclose their status to housing providers.
3. Do not explicitly disclose sexual orientation or gender identity at any stage, but rely instead on the housing provider's perceptions. This option minimizes the risk of detection but may also reduce the study's ability to observe differential treatment, especially if the identifiability of testers as lesbian, gay, or transgender varies or is unknown. In the most recent testing study of racial and ethnic discrimination, testers' identifiability was assessed by independent coders and used to analyze the extent to which testers who are more identifiably minority experienced more discrimination. A similar analysis of identifiability could potentially be conducted for lesbian, gay, and transgender testers.

Using feedback from expert advisors and from focus groups with transgender people, the pilot study implemented the first option for tests of discrimination against lesbian women and gay men and experimented with *both* the second and third options for tests of discrimination against transgender people. To be more specific, in the tests involving lesbian and gay couples, testers will refer to their partners or spouses by name during the initial telephone inquiry and again at the

⁴ A *cisgender* person is one whose gender identity corresponds with the biological sex assigned at birth.

start of the in-person visits with housing providers. In one-half of the tests involving transgender people, the transgender testers will refer to their identity as a transgender person early in the in-person visit. In the other one-half of the tests, they will not explicitly disclose their gender identity. Possible scripts for disclosure include explaining that the person's legal identification does not yet correspond to his or her identity and signing the guest register with both a chosen name and a legal name or referencing a different legal name on a credit report if the tester were to submit an application.

Exhibit 2 sums up the pilot study's response to the challenges of extending the paired-testing methodology to measure rental housing discrimination against lesbian and gay couples and against transgender individuals. The main issue here is when and how to disclose testers' status as members of the potentially disadvantaged class, and, for transgender homeseekers, the researchers have chosen to test the implications of two possible strategies. Future studies could potentially assess the incidence and forms of adverse treatment against lesbian and gay homeseekers who do not proactively disclose their identity.

Exhibit 2

Paired-Testing Methodology Extended to Lesbian and Gay Couples and Transgender Individuals

	Comparison Group	Housing Types Excluded or Oversampled	Disclosure of Status
Sexual preference	Heterosexual couples	None	Mention partner (by name) early in both phone and in-person contact
Gender identity	Cisgender individuals, matched to the transgender tester's gender identity	None	Two solutions: 1. Refer to transition in context of form of identification, legal name, or credit check 2. Do not explicitly disclose status

Housing Voucher Recipients

The federal Housing Choice Voucher program supplements the rent low-income households pay for homes and apartments in the private market, enabling them to obtain decent, affordable housing in neighborhoods of their choice. Three-and-a-half decades of experience have proven vouchers to be an extremely effective tool for addressing the housing needs of low-income families. Most households that get a voucher succeed in finding a house or an apartment where they can receive assistance, and recipients generally live in better quality housing and pay more affordable rents than similar unassisted households.

Despite its overall success, the voucher program falls short of its potential to provide access to rental housing in safe, opportunity-rich neighborhoods. Evidence suggests that housing providers in these neighborhoods may refuse to accept vouchers, leaving voucher recipients with limited options about where to live, but there is a dearth of knowledge about discrimination against voucher

holders in the private rental market. Fair housing groups have conducted local studies, mostly for enforcement purposes, to investigate landlord behavior. Methods for these studies vary widely and include some mix of passive screening of advertisements, screening landlords over the phone to see if they accept vouchers, and sending testers to screen landlords during on-site visits, sometimes with a matched pair. No matter the methodology, high rates of voucher refusal were common across all studies. These tests, however, were not designed for research purposes, and, thus, have significant limitations in terms of generalizability and replicability.⁵

Therefore, HUD commissioned a pilot study to explore the feasibility of using paired testing to more rigorously estimate the incidence and forms of discrimination against housing voucher recipients. This study has important implications for federal housing assistance policy and program implementation, but it poses daunting design challenges, involving the appropriate comparison group, when and how testers should disclose that they have housing vouchers, and how to sample available rental housing units.

What Is the Appropriate Comparison Group?

To capture adverse treatment based on voucher receipt, control testers should pose as comparably qualified renters without housing vouchers. What does it mean to be comparably qualified? A housing voucher effectively increases a household's available income for housing, so it would not be entirely satisfactory to match the incomes of the two testers at prevoucher levels, with one tester in possession of a voucher. The purpose of the voucher is to allow low-income households to obtain housing that would otherwise be unaffordable to them. Comparing the treatment of voucher holders with that of households with the same prevoucher income would mean that control testers were inquiring about housing units with unaffordably high rent levels, placing them at a clear disadvantage in the eyes of landlords.

Setting control testers' income too high, however, could potentially bias results in the opposite direction, creating a scenario in which landlords are presented with the choice between a low-income household with a voucher and a moderate- to middle-income household without one. Preferential treatment of the control tester could be attributed to its relative affluence (or perceived class) and not to the landlord's aversion to the voucher.

The pilot study addresses this quandary by assigning each control tester an income equal to the voucher holder's income plus the approximate value of the voucher itself. This gives the control tester sufficient income to rent the same unit for which the voucher household is qualified, but not so much income as to place the control household in a higher socioeconomic class. In effect, the two testers in each pair will have the same income, but from different sources.

It is still possible, however, that the extra income assigned to the control tester will—in the eyes of a housing provider—give her greater socioeconomic status (or perceived class status) than the voucher holder. If this perception is the case, differences in treatment could be interpreted as

⁵ On state and local levels, 12 states, the District of Columbia, and more than 30 cities and counties have enacted statutes that prohibit discrimination in the housing market based on source of income (PRRAC, 2014). Some evidence suggests that these protections have a positive effect on voucher lease rates and neighborhood choice, but no conclusive evidence supports this claim (Finkel and Buron, 2001; Galvez, 2011, 2010).

reflecting a preference for higher status tenants (possibly because of expectations about reliability, behavioral issues, or crime), rather than discrimination against voucher recipients per se. One potential strategy for exploring this issue further (after the testing is complete) would be to compare outcomes for tests in which the housing provider asked testers about their incomes with tests in which testers' assigned incomes were not disclosed to the housing provider.

Should Some Segments of the Housing Stock Be Either Excluded or Oversampled?

Existing evidence suggests that voucher recipients face higher rates of refusal or discrimination when they inquire about housing in high-quality, opportunity-rich neighborhoods than when they inquire about housing in distressed neighborhoods. To explore this issue fully, the sample of available rental homes and apartments for which tests are initiated must include units from these more opportunity-rich neighborhoods. In fact, depending on the geographic distribution of voucher-affordable rental housing in a given metropolitan area, it might make sense to oversample from these types of neighborhoods.

The pilot study will select a geographically representative sample of all voucher-affordable rental housing from a central city and a suburban jurisdiction in each of five metropolitan areas, without oversampling specific low-poverty areas or excluding areas with high concentrations of assisted residents. Analysis of the distribution of voucher-affordable housing in these target jurisdictions suggests that this approach is likely to yield a sufficient number of units for testing within low-poverty neighborhoods to generate reasonably precise estimates of discrimination—without the need for oversampling. It will not, however, yield metropolitanwide estimates of discrimination against voucher holders.

It is also possible that units in high-opportunity neighborhoods may be effectively eliminated from the in-person testing component of the study if landlords in these areas are more likely to (1) routinely refuse vouchers or (2) refuse voucher tenants during the phone tests. High rates of refusal at either of these steps would result in few in-person tests in these neighborhoods. If this proves to be the case, the pilot methodology would effectively capture the most important considerations about voucher access to high-opportunity neighborhoods: the overall rate of refusal of voucher holders as a matter of landlord policy and refusal to make appointments with voucher holders who inquire by telephone. It would not, however, be able to effectively capture any differences in treatment during in-person visits to housing in high-opportunity neighborhoods.

When and How Should a Tester's Status Be Disclosed to Housing Providers?

Housing agencies and advocates for housing voucher recipients express differing opinions about when households should disclose the fact that they have a voucher and about what most voucher recipients actually do. Some suggest that voucher holders should disclose their status in the first inquiry so they do not waste precious time visiting properties that do not accept vouchers. Others argue that voucher holders should wait until they have met a housing provider and established a positive relationship, in hopes that the provider will agree to accept the voucher because they feel comfortable with the potential tenant. Given the available evidence about the frequency with which housing providers refuse vouchers, the decision about when testers should disclose their status has important implications for the study's potential to capture specific forms of adverse treatment.

The pilot study arrived at a three-step testing process that is intended to capture both outright refusal to accept vouchers *and* any possible differential treatment of voucher holders by housing providers who accept (or at least say they accept) vouchers.

1. Advertised rental units that fall within voucher rent limits will receive a single screening call, asking if the provider accepts housing vouchers. This first step (which does not involve a matched pair) will yield an estimate of the incidence of outright refusal across the rental market.
2. For each unit that accepts vouchers, two testers will call to inquire about the unit's availability and terms and to make appointments for in-person visits. The tester who is assigned a voucher will disclose during the call that she is a voucher holder. This step should yield estimates of major differences in treatment between voucher recipients and nonvoucher households by housing providers who reportedly accept vouchers.
3. For each case in which both testers are able to obtain an appointment, they will make in-person visits, during which they again inquire about the advertised unit and try to inspect it and other available homes or apartments. The tester who is assigned a voucher discloses this fact at the start of the in-person visit. This step should yield estimates of any additional treatment differences by housing providers who agree to meet with voucher holders.

If the incidence of outright refusal (in step 1) is high, this approach requires that large numbers of advertised units are screened to produce sufficiently large samples at steps 2 and 3 to reliably capture any differences in treatment among housing providers that accept vouchers. It is possible that this three-step approach will yield minimal (or unmeasurable) differences in treatment between voucher and nonvoucher households, after providers who refuse to accept vouchers are screened out in the first step.

Exhibit 3 sums up the pilot study's response to the challenges of extending the paired-testing methodology to measure discrimination against households that receive vouchers. This study will yield estimates of the share of available housing units whose owners refuse outright to accept vouchers *and* the incidence of differential treatment voucher recipients experience when they inquire in person about available units where owners are willing to accept vouchers. The experience of voucher recipients will be compared with that of low-income households with the same effective purchasing power, so their incomes will be higher by the amount of the housing voucher subsidy amount. The sample of advertised rental housing for which tests are initiated will be limited to units with rents that fall within local payment standards so that they are indeed appropriate for voucher recipients. Units in low-poverty census tracts will be oversampled, if necessary, to ensure that the study can provide information about access to affordable rental housing outside poor neighborhoods.

Exhibit 3

Paired-Testing Methodology Extended to Housing Choice Voucher Holders

	Comparison Group	Housing Types Excluded or Oversampled	Disclosure of Status
Source of income (housing subsidy)	Unsubsidized households with the same income plus the value of the voucher subsidy	Units with rents more than payment standard excluded; units in low-poverty neighborhoods oversampled	Ask about voucher acceptance in advance phone contact and disclose early during in-person contact

Future Paired-Testing Research

Looking ahead, evidence about barriers to housing choice and unequal housing outcomes for other groups may lead to interest in further extensions of paired-testing research. In particular, given the effectiveness of paired-testing research to date, the methodology might be considered for measuring the incidence of housing discrimination on the basis of national origin, English fluency, other disabilities, or religious affiliation. Not every extension of paired testing will be straightforward and, in fact, some might prove infeasible. As policymakers and researchers consider these and other possible applications of paired-testing research, they should explicitly address each of the following three design questions.

1. What is the appropriate comparison group? Paired testing works as a methodology only when the treatment experienced by the group of interest should—in principle—be exactly the same as another, more privileged group.
2. Should some segments of the housing stock be either excluded or oversampled? Obtaining a meaningful understanding of a group's treatment may require a sampling design that screens for available housing units with characteristics that align with the group's needs or capacities, complicating the logistics of sampling and potentially undermining the generalizability of the results.
3. When and how should a tester's status be disclosed to housing providers? If housing providers cannot predictably discern the difference in status between two matched testers, the paired-testing methodology cannot reliably detect systematic differences in treatment.

For some classes of homeseekers, one or more of these questions may be easy to resolve, but others will be much more difficult—potentially so difficult that the power of the paired-testing methodology is undermined. For example, in a study designed to test for discrimination based on national origin, each tester should be matched to a native-born American of the same race or ethnicity. How would his or her national origin be disclosed during an initial telephone inquiry or in-person visit with a housing provider? In a study exploring potential housing discrimination against people of the Muslim faith, what religion should be assigned to the control testers and how would testers disclose their religious affiliation? Finally, in a study designed to test for discrimination against people whose English is not fluent, the testers' status will be disclosed as soon as they begin speaking to housing providers. If the housing provider has difficulty understanding or communicating with the tester, it might be implausible to interpret differences in treatment (such as information about fees or incentives, or even the number of units shown) to discrimination.

When the pilot studies discussed in this article are complete, we can assess how effectively each study tackled these design challenges and how useful paired testing proves to be for the populations on which they focus. These findings will also help inform future discussions about whether and how to extend paired-testing research. Paired testing may not prove to be the best tool for rigorously measuring the extent and forms of discrimination against all potentially vulnerable classes of homeseekers. In some circumstances, researchers may need to consider other methods—potentially including surveys of housing providers or households—either in combination with or as alternatives to paired testing.

It may not be possible to determine in advance whether and how the design challenges highlighted here can be effectively addressed for a new class of homeseekers. Recent experience with extensions of the methodology suggests that these challenges are fully appreciated—and resolved—only when they are addressed in practice. It is not always clear in theory whether adaptations of the method will work. Therefore, before launching a costly nationwide study, it is essential to invest in a smaller, pilot effort that provides the opportunity to fully explore options, assess the feasibility of modified testing protocols, and determine whether a full-scale, national study is likely to yield worthwhile information to guide policy and practice.

Acknowledgments

The author thanks Laudan Aron, Diane Levy, and Mary Cunningham of the Urban Institute, who are leading the three pilot studies discussed in this article, and the anonymous reviewers.

Author

Margery Austin Turner is Senior Vice President for Program Planning and Management at the Urban Institute.

References

- Aranda, Claudia L. 2015. "Targeting Disability Discrimination: Findings and Reflections From the National Study on Housing Discrimination Against People Who Are Deaf and People Who Use Wheelchairs," *Cityscape* 17 (3): 103–121.
- Aratani, Yumiko, Michelle Chau, Vanessa R. Wight, and Sophia Addy. 2011. *Rent Burden, Housing Subsidies and the Well-being of Children and Youth*. New York: National Center for Children in Poverty. <http://www.nccp.org>.
- Aron, Laudan, Claudia Aranda, Douglas Wissoker, Molly Scott, Brent Howell, Robert Santos, and Margery Austin Turner. Forthcoming. *Discrimination Against Families in Rental Housing Markets: Findings of the Pilot Study*. Washington, DC: U.S. Department of Housing and Urban Development.
- Colten, Mary, and Robert Marans. 1982. "Restrictive Rental Practices and Their Impact on Families," *Population Research and Policy Review* 1 (1): 43–58.
- Colvin, R. 2004. *The Extent of Sexual Orientation Discrimination in Topeka, Kansas*. Topeka, Kansas: National Gay and Lesbian Task Force Policy Institute.
- Fair Housing Centers of Michigan. 2007. *Sexual Orientation and Housing Discrimination in Michigan: A Report of Michigan's Fair Housing Centers*. Ann Arbor, MI: Fair Housing Centers of Michigan.
- Finkel, Meryl, and Larry Buron. 2001. *Study on Section 8 Voucher Success Rates: Volume I Quantitative Study of Success Rates in Metropolitan Areas*. Cambridge, MA: Abt Associates. huduser.gov/publications/pdf/sec8success.pdf.

Friedman, S., A. Reynolds, S. Scovill, F. R. Brassier, R. Campbell, and M. Ballou. 2013. *An Estimate of Housing Discrimination Against Same-Sex Couples*. Philadelphia: M. Davis and Company.

Galvez, Martha. 2011. "Defining 'Choice' in the Housing Choice Voucher Program: The Role of Market Constraints and Household Preferences in Location Outcomes." Ph.D. diss., New York University.

———. 2010. "What Do We Know About Housing Choice Voucher Program Location Outcomes? A Review of Recent Literature." <http://www.urban.org/uploadedpdf/412218-housing-choice-voucher.pdf>.

Grant, J. M., L. A. Mottet, and J. Tanis, 2011. *Injustice at Every Turn: A Report of the National Transgender Discrimination Survey*. Washington, DC: National Center for Transgender Equality; National Gay and Lesbian Task Force.

Joint Center for Housing Studies (JCHS) of Harvard University. 2013. *America's Rental Housing, Evolving Markets and Needs*. Cambridge, MA: Joint Center for Housing Studies of Harvard University.

Kaiser Family Foundation. 2001. *Inside-Out: A Report on the Experiences of Lesbians, Gays, and Bisexuals in America and the Public's View on Issues and Politics Related to Sexual Orientation*. Menlo Park, CA: Kaiser Family Foundation.

National Fair Housing Alliance (NFHA). 2012. *Fair Housing in a Changing Nation: 2012 Fair Housing Trends Report*. Washington, DC: National Fair Housing Alliance.

———. 2011. *The Big Picture: How Fair Housing Organizations Challenge Systemic and Institutionalized Discrimination: 2011 Fair Housing Trends Report*. Washington, DC: National Fair Housing Alliance.

Oh, Sun Jung, and John Yinger. 2015. "What Have We Learned From Paired Testing in Housing Markets?" *Cityscape* 17 (3): 15–59.

Poverty and Race Research Action Council (PRRAC). 2014. *Appendix B: State, Local, and Federal Laws Barring Source-of-Income Discrimination*. Washington, DC: Poverty and Race Research Action Council. <http://www.prrac.org/pdf/AppendixB.pdf>.

Commentary: Some Thoughts on Field Experiments on Housing Discrimination From a European View

Ali M. Ahmed

Linköping University

Discrimination violates the basic human right of equal treatment, adopted, for example, in the *Universal Declaration of Human Rights* by the United Nations in 1948. Discrimination is without doubt a threat against the cohesion in a society and leads to the exclusion of people. The international human rights law also recognizes everyone's right to adequate housing and standard of living. Furthermore, having a home is a fundamental need, which is necessary to be an integrated part of society on equal terms. Housing discrimination, therefore, concurrently violates two fundamental human rights and is a severe problem for any society.

One question that naturally arises when discussing housing discrimination is, "Who has access to what housing and why?" In any country, if city maps are marked on the basis of this question, clear ethnic and racial dividing lines will soon emerge. That racial and ethnic majority and minority populations, figuratively speaking, live in different physical worlds is particularly evident in most larger cities throughout the world. Socioeconomic models and individuals' own residential choices and preferences have often been put forward as possible explanations for the ethnic and racial residential segregation. Yet, these explanations have their deficiencies and do not tell the whole story. Scholars, therefore, recognize the role of housing discrimination in generating and maintaining ethnic and racial residential segregation as central.

To combat housing discrimination, we need to have knowledge of its extent, how it manifests itself, and how it can best be countered. Researchers and various organizations in the United States have systematically accumulated such knowledge by using field experiments for several decades. The most appealing feature of these field experiments is that they provide clear and indisputable evidence of housing discrimination. The current symposium in *Cityscape* contains pertinent and updated contributions concerning field experiments on housing discrimination in the United States. The articles in the symposium provide a complete background and overview of this research, a thorough discussion of the methodological challenges and problems, and a detailed discussion of future developments and directions. In this brief commentary, I share a few thoughts related to field experiments on housing discrimination from a European perspective, inspired by the articles in this symposium.

Research involving field experiments on housing discrimination in European countries was first conducted during the 1960s and 1970s (Bovenkerk et al., 1979; Daniel, 1968; McIntosh and Smith, 1974). Since then, however, this powerful methodology has not been used to its full extent. Only recently have researchers in European countries picked up where the original studies left off. Why is it that field experiments have not been used to the same degree in European countries as they have been in the United States? Let me start by addressing this question. As Margery Austin Turner and Judson James mention in their introduction to the symposium, the two main reasons for conducting field experiments on housing discrimination in the United States are (1) to expose and document discrimination in the housing market and (2) to enforce housing discrimination laws (Turner and James, 2015). The lack of field experiments in European countries is related to these two purposes and the associated methodology.

For the purpose of exposing and documenting discrimination in the housing market, field experiments have long been considered to be unethical by many European scholars, because the methodology involves the use of deception. At most, two types of deception may occur, depending on the experimental design chosen by the researcher (Riach and Rich, 2004). The methodological differences between the various types of field experiments are discussed in the excellent review article by Sun Jung Oh and John Yinger in this symposium and, therefore, will not be repeated here.

First, deception of research subjects will always occur, regardless of the experimental design. Field experiments on housing discrimination involve introducing fictitious applicants to the housing market. Real estate agents are deceived in the sense that they are engaged with individuals who do not actually want to buy or rent. Hence, real estate agents are never given the opportunity to give their consent in these experiments on discrimination. Most researchers find this type of deception acceptable nowadays, even in European countries. As Riach and Rich (2004: 463) put it, "In summary, the justification which we offer for the deception of subjects in field experiments of labour, housing and product markets is that a lack of veracity is endemic in these markets; that great harm is done to the social fabric by discriminatory practices in such markets; that minimal inconvenience is imposed on the entrepreneurs in the experiment, and that the technique provides evidence with a degree of accuracy and transparency which is not available from any other procedure." Still, not long ago, I recall that when a colleague and I conducted the first field experiment on discrimination in the Swedish housing market and we published our preliminary results, we were bombarded with criticism, claiming we had violated the ethical code of practice for research in Sweden by using the field experimental methodology to prove discrimination (Ahmed and Hammarstedt, 2008).

Second, deception of testers occurs if field experiments on housing discrimination involve personal attendance at real estate offices, and testers recruited for the experiments are unaware of the hypotheses of the studies. The idea of keeping testers ignorant of the study purpose was proposed by Heckman and Siegelman (1993) to avoid experimenter effects. Although deception of real estate agents is regarded as acceptable nowadays, deception of testers is considered unethical. For instance, Riach and Rich (2004) argued that deception of testers cannot be justified because it may do harm when testers ultimately become aware of the true purpose of their involvement in the study and because researchers have viable alternatives to this type of experimental design, such as correspondence testing experiments. This argument explains why nearly all field experiments on

housing discrimination in European countries have been correspondence tests (see, for example, Ahmed, Andersson, and Hammarstedt, 2010, 2008; Ahmed and Hammarstedt, 2009, 2008; Andersson, Jakobsson, and Kotsadam, 2012; Baldini and Federici, 2011; Bosch, Carnero, and Farré, 2010).

Oh and Yinger (2015) discuss the practice of using field experiments with bogus applicants to enforce antidiscrimination laws and prepare court cases against alleged discriminators in the United States. This type of application of field experiments is controversial in many European countries (De Schutter, 2003). Court cases based on field experimental-like evidence are, however, becoming more common; for example, in Belgium, the Czech Republic, Denmark, Finland, France, Hungary, Latvia, the Netherlands, Slovakia, and the United Kingdom, where evidence based on experimental tests has been used (Rorive, 2009). Many of these examples, however, involve contexts other than the housing market. According to Rorive (2009: 47), the main criticism against using field experimental-like evidence for law enforcement is that “it does not correspond to the principle of fairness of evidence; it could amount to provocation to commit a crime and it threatens the right to respect for private life.” Hence, the critics view experimental tests on discrimination as a form of entrapment.

From my point of view, several European countries should adopt the practice of using field experimental evidence in court and take advantage of the experiences of the United States. In Sweden, for example, the Discrimination Act protects people from unequal treatment, and an authority—the Equality Ombudsman—ensures the law is followed. If someone has been discriminated against, he or she can report it to the Equality Ombudsman. The problem, however, is that the Equality Ombudsman faces difficulties in proving that discrimination has occurred. According to a report from the European Commission against Racism and Intolerance (2012), only 1 percent of all complaints to the Equality Ombudsman regarding discrimination based on ethnicity or religion results in a lawsuit in Sweden. That figure demonstrates that the Discrimination Act and the Equality Ombudsman currently are merely symbolic and that European countries need powerful tools to enforce the Discrimination Act.

The article by Fred Freiberg and Gregory D. Squires in this symposium highlights the fact that housing discrimination is changing its nature and has become subtler in recent years. In line with these changes, the authors explore how field experiments can be refined to more effectively identify housing discrimination (Freiberg and Squires, 2015). I am in total agreement with the authors in this respect, and I find the authors’ ideas for expanding the use of field experiments refreshing. In line with these ideas, I will share a Swedish example of how field experiments on discrimination can be extended to get a deeper understanding of discrimination. Although Rooth (2010) studied labor market discrimination, I believe a similar methodological approach could be applied to the housing market. Rooth (2010) examined the relationship between implicit attitudes against minorities and discriminatory behavior in a real hiring situation. He first conducted a traditional field experiment on discrimination in the Swedish labor market and then approached the responsible recruiters at each firm, asking them to participate in an implicit association test (Greenwald, McGhee, and Schwartz, 1998). In a nutshell, this test offers the researcher a way to probe unconscious biases. Rooth (2010) found the stronger the unconscious biases recruiters had, as measured by the implicit association test, the lower the probability was that the recruiters would invite the minority

applicant for a job interview. This kind of experimental refinement may both capture subtle forms of discrimination and give a better understanding of various underlying mechanisms of discrimination.

Two articles in this issue of *Cityscape* discuss how field experiments on discrimination can be extended to include protected classes other than ethnic and racial minorities. Claudia L. Aranda gives a concrete example of how exactly field experiments can be used to document discrimination against people who are deaf or hard of hearing and people who use wheelchairs. She elaborates on the difficulties involving designing an experiment for testing disability discrimination and then presents the results of a recent study. She found that people who are deaf or hard of hearing and people who use wheelchairs all face barriers in the housing market (Aranda, 2015). It is interesting to compare these findings with a very recent European study on disability discrimination in the housing market. Fumarco (2015) focused on whether blind tenants assisted by guide dogs are discriminated against in the Italian rental housing market. By contrast with the American study, Fumarco (2015) relied solely on correspondence tests and on a randomized design; that is, only one inquiry was sent to each real estate agent. One out of three possible potential tenants—a married couple, a married couple in which the wife was blind and had a guide dog, and a married couple with a dog—was randomly assigned to each real estate agent. His results showed that some real estate agents discriminated against the blind tenants because of the presence of the guide dog, not because of the disability.

Margery Austin Turner suggests in her article ways of extending the estimates on housing discrimination to an additional three protected classes: families with children; lesbian, gay, and transgender homeseekers; and renters using housing vouchers. Along this line, I would like to add another protected class: elderly people. A field experiment in the Swedish labor market, for instance, provided clear evidence of ageism (Ahmed, Andersson, and Hammarstedt, 2012). It is, therefore, reasonable to hypothesize that age discrimination might exist in the housing market as well. Of course, complexities are involved in designing an experiment for testing age discrimination, and we may not be able to address all three questions Turner (2015) poses in her article as a guideline for designing an experiment for other protected classes. Nonetheless, attention to the circumstances of the elderly population is of utmost importance. The growing elderly population is one of the biggest social and economic challenges for our societies in the future. This development will affect most policy areas, housing being one of them.

Author

Ali M. Ahmed is a professor of economics at Linköping University in Sweden.

References

Ahmed, Ali M., Lina Andersson, and Mats Hammarstedt. 2012. "Does Age Matter for Employability? A Field Experiment on Ageism in the Swedish Labour Market," *Applied Economics Letters* 19 (4): 403–406.

———. 2010. “Can Discrimination in the Housing Market Be Reduced by Increasing the Information About the Applicants?” *Land Economics* 86 (1): 79–90.

———. 2008. “Are Lesbians Discriminated Against in the Rental Housing Market? Evidence From a Correspondence Testing Experiment,” *Journal of Housing Economics* 17 (3): 234–238.

Ahmed, Ali M., and Mats Hammarstedt. 2009. “Detecting Discrimination Against Homosexuals: Evidence From a Field Experiment on the Internet,” *Economica* 76 (303): 588–597.

———. 2008. “Discrimination in the Rental Housing Market: A Field Experiment on the Internet,” *Journal of Urban Economics* 64 (2): 362–372.

Andersson, Lisa, Niklas Jakobsson, and Andreas Kotsadam. 2012. “A Field Experiment of Discrimination in the Norwegian Housing Market: Gender, Class, and Ethnicity,” *Land Economics* 88 (2): 233–240.

Aranda, Claudia L. 2015. “Targeting Disability Discrimination: Findings and Reflections From the National Study on Housing Discrimination Against People Who Are Deaf and People Who Use Wheelchairs,” *Cityscape* 17 (3): 103–121.

Baldini, Massimo, and Marta Federici. 2011. “Ethnic Discrimination in the Italian Rental Housing Market,” *Journal of Housing Economics* 20 (1): 1–14.

Bosch, Mariano, M. Angeles Carnero, and Lidia Farré. 2010. “Information and Discrimination in the Rental Housing Market: Evidence From a Field Experiment,” *Regional Science and Urban Economics* 40 (1): 11–19.

Bovenkerk, Frank, Benjamin Kilborne, Francois Raveau, and David Smith. 1979. “Comparative Aspects of Research on Discrimination Against Non-White Citizens in Great Britain, France and the Netherlands.” In *Problems in International Comparative Research in the Social Sciences*, edited by Jan Berting, Felix Geyer, and Ray Jurkovich. Oxford, United Kingdom: Pergamon Press: 105–124.

Daniel, William W. 1968. *Racial Discrimination in England*. Middlesex, United Kingdom: Penguin.

De Schutter, Olivier. 2003. “Methods of Proof in the Context of Combating Discrimination.” In *Proving Discrimination—The Dynamic Implementation of EU Anti-Discrimination Law: The Role of Specialised Bodies*, edited by Janet Cormack. Brussels, Belgium: Migration Policy Group: 22–37.

European Commission Against Racism and Intolerance. 2012. *ECRI Report on Sweden: Fourth Monitoring Cycle*. Strasbourg, France: Council of Europe.

Freiberg, Fred, and Gregory D. Squires. 2015. “Changing Contexts and New Directions for the Use of Testing,” *Cityscape* 17 (3): 87–101.

Fumarco, Luca. 2015. *Disability Discrimination in the Rental Housing Market*. Linköping, Sweden: Linnaeus University Centre for Labour Market and Discrimination Studies.

Greenwald, Anthony G., Debbie E. McGhee, and Jordan L.K. Schwartz. 1998. “Measuring Individual Differences in Implicit Cognition: The Implicit Association Test,” *Journal of Personality and Social Psychology* 74 (6): 1464–1480.

Heckman, James, and Peter Siegelman. 1993. "The Urban Institute Audit Studies: Their Methods and Findings." In *Clear and Convincing Evidence: Measurement of Discrimination in America*, edited by Michael Fix and Raymond Struyk. Washington, DC: Urban Institute Press: 187–258.

McIntosh, Neil, and David J. Smith. 1974. *The Extent of Racial Discrimination*. London, United Kingdom: Political and Economic Planning.

Oh, Sun Jung, and John Yinger. 2015. "What Have We Learned From Paired Testing in Housing Markets?" *Cityscape* 17 (3): 15–59.

Riach, Peter A., and Judith Rich. 2004. "Deceptive Field Experiments of Discrimination: Are They Ethical?" *Kyklos* 57 (3): 457–470.

Rooth, Dan-Olof. 2010. "Automatic Associations and Discrimination in Hiring: Real World Evidence," *Labour Economics* 17 (3): 523–534.

Rorive, Isabelle. 2009. *Proving Discrimination Cases—The Role of Situation Testing*. Brussels, Belgium: Kaligram.

Turner, Margery Austin, and Judson James. 2015. "Discrimination as an Object of Measurement," *Cityscape* 17 (3): 3–13.

Turner, Margery Austin. 2015. "Other Protected Classes: Extending Estimates of Housing Discrimination," *Cityscape* 17 (3): 123–136.

Commentary: Housing Discrimination Research in the 21st Century

Samantha Friedman

University at Albany, State University of New York

Housing discrimination continues to be a significant problem in America nearly a half century after the passage of the Fair Housing Act. It is estimated that, annually, 4 million people experience discrimination in the rental housing market (NFHA, 2015). A very small number of those experiencing discrimination, however, actually report it. In 2014, for example, only about 27,000 housing discrimination complaints in both the rental and sales markets were filed with the U.S. Department of Housing and Urban Development (HUD), U.S. Department of Justice, and other substantially equivalent fair housing agencies (NFHA, 2015).

Given that data on complaints are quite limited, the paired-testing methodology has been an important tool that researchers and policymakers have used to study housing discrimination. The articles included in this symposium—especially the article by Sun Jung Oh and John Yinger—offer reviews of the literature that uses paired-testing methodology, present critiques of the current methodology, and provide insights on how to broaden the scope of housing discrimination research beyond racial and ethnic discrimination. I consider the latter two dimensions of this discussion and offer my own insights on current housing discrimination research and the future of that research.

One key issue discussed in this symposium is whether the current use of paired testing adequately captures the actual levels of racial and ethnic discrimination that exist in the housing market. The articles by Rob Pitingolo and Stephen L. Ross and by Fred Freiberg and Gregory D. Squires comment on this important issue, and Pitingolo and Ross (2015) offer new analyses that speak to some of the key criticisms. Another important issue is the extent to which all forms of housing discrimination, not just those based on race and ethnicity, are being measured by the current methodology. The limited complaint data that I mentioned in the first paragraph reveal that the bases of housing discrimination complaints have changed. In fiscal year (FY) 2013, disability status was the protected class forming the largest basis of the complaints, or 53 percent of complaints, up from 44 percent in FY 2008 (HUD 2013, 2011). Race and Hispanic origin together comprised 37 percent of the complaints in FY 2013, down from 43 percent in FY 2008. Claudia L. Aranda and Margery Austin Turner's articles offer insights into the changing nature of housing discrimination and the associated challenges in conducting research on other protected classes, such as those involving people with disabilities or families with children. In this commentary, I discuss and expand on the critiques and insights raised in all these articles.

Racial and Ethnic Discrimination in the Housing Market

Although residential segregation between Black and White households has declined over time, 53 percent of metropolitan Black residents continue to live in metropolitan areas that are either hypersegregated or in the high segregation range (Massey and Tannen, 2015). Results from the 2012 Housing Discrimination Study (HDS2012), which are based on the paired-testing methodology, reveal that, since HDS1989, discrimination against Black homeseekers, particularly in the rental market, has declined (Turner et al., 2013). For example, in HDS1989, White homeseekers were favored over Black homeseekers by 7 percentage points in being told about the availability of the advertised rental unit, but, in HDS2012, the net discrimination rate had reduced to 0, or no difference between the two groups (see Oh and Yinger, 2015). With respect to inspecting more rental units, White homeseekers were favored over Black homeseekers by about 19 percentage points in HDS1989; by HDS2012, the net discrimination rate was reduced to 3 percentage points (see Oh and Yinger, 2015). Among homebuyers, steering increased between HDS1989 and HDS2012, but the relative share of White- to Black-favored audits in HDS2012—or the *net measure of discrimination*—was rather small. White homeseekers were not favored over Black homeseekers in inspecting homes in tracts with majority White populations in HDS1989, but in HDS2012, they were favored over Black homeseekers in 5 percent of the audits (see Oh and Yinger, 2015). Moreover, the percent White in the average neighborhoods to which White and Black homeseekers were steered differed by only about 2 percentage points (Turner et al., 2013). Given that a majority of metropolitan Black households continues to be highly segregated from White households, it is surprising that HDS has not found more discrimination against Black households.

The articles by Pitingolo and Ross (2015) and Freiberg and Squires (2015) in this symposium consider various reasons why HDS may underestimate the level of housing discrimination against Black homeseekers that could be contributing to high levels of Black-White residential segregation in many metropolitan areas. I consider two important methodological issues highlighted in these critiques that relate to HDS and to housing discrimination research more generally. The first relates to the sampling design.

Both articles identify an important critique of HDS sampling design, namely that the housing stock in the sampling frame may not be representative of the housing available in the market. Landlords who want to discriminate do not publicly advertise their units, and the units that are advertised may not represent all the units available for rent or for sale. Using the 2011 American Community Survey (ACS) data, Pitingolo and Ross (2015) determined the number of available rental and owner-occupied units, defined as those units where the household changed between 2010 and 2011. Then they reweighted the analyses of housing discrimination based on the ACS data. Their new analyses are similar to those tabulated in HDS, suggesting that the units being advertised, which are captured in the HDS sample, are similar to the distribution of units in the ACS that were identified as available. As they point out, however, “if variation in discrimination across neighborhoods exists within metropolitan areas, the area estimates on which the national estimates are based may have considerable measurement error because the number of tests per site is too small to accurately cover the many distinct regions or neighborhoods in each site” (Pitingolo and Ross, 2015: 68).

To maintain comparability across time, the sampling strategy of HDS was designed to be nationally representative in scope and locally representative in a handful of metropolitan areas. As a result, the number of audits within each of these metropolitan areas is small, with most having around 44 tests, except for the eight areas where rental tests were oversampled. Whether this design affects the overall results is a question that should guide future housing discrimination research. Hanson and Hawley (2011) conducted more than 4,000 correspondence tests in 10 metropolitan rental housing markets. In 8 of the markets, about 500 tests were conducted; in Washington, D.C., and Houston, 160 and 296 tests, respectively, were conducted. The results reveal that White homeseekers were favored over Black homeseekers in getting a response from the landlord in 6.3 percent of the tests overall and, in Boston and Los Angeles, the net measures were 12.1 and 11.37 percent, respectively. HDS2012 found no significant discrimination at the stage of whether auditors could make appointments to see the advertised units, although their contact was made primarily by telephone rather than e-mail.

In addition to potentially revealing higher levels of discrimination, collecting data from a larger number of audits, either through in-person audits or correspondence tests, for a substantial number of metropolitan areas would enable researchers to create metropolitan-specific estimates of housing discrimination. One of the biggest challenges that researchers studying residential segregation face is the lack of data on housing discrimination. If future research could generate such estimates, it would greatly enhance the research on residential segregation by enabling researchers to assess the direct effect of housing discrimination on residential segregation. In the past, very little research has made this direct link and, as a result, the extent to which housing discrimination directly affects residential segregation remains unknown.

Aside from increasing the sample sizes of data collected within metropolitan areas, future research should consider implementing a more complex sampling design that stratifies neighborhoods by their racial composition and selects advertised housing units based on this stratification. Hanson and Hawley's (2011) results reveal that, in terms of receiving a response, White homeseekers are favored over Black homeseekers in 10.62 percent of the tests in neighborhoods where White households comprise 80 to 95 percent of the population in the tract, or what they call *tipping point* neighborhoods. Moreover, controlling for other factors, Black homeseekers inquiring about rental housing are significantly less likely to receive responses in tipping point neighborhoods than in neighborhoods with other levels of White racial composition. These results suggest that employing a stratified sampling design in future housing discrimination research could potentially reveal greater levels of discrimination than those found in existing studies that sample housing units irrespective of neighborhood racial composition.

The second methodological issue that is important to consider in this research relates to the timing of when the housing discrimination could happen. Pitingolo and Ross (2015) state that one limitation of HDS is that housing discrimination may occur early in the process before the in-person audit occurs. As mentioned previously, however, using HDS2012 data, Turner et al. (2013) found no evidence of differing experiences between White and Black homeseekers when they made appointments by telephone to visit sampled units. The results from the Hanson and Hawley (2011) study and from other recent research (for example, Ewens, Tomlin, and Wang, 2014; Hogan and Berry, 2011), however, contradict these findings and suggest that additional research is needed to fully uncover the extent to which discrimination exists in the earlier part of the process of acquiring a home.

Freiberg and Squires (2015) comment on the later part of the discrimination process, which has received much less attention in the housing discrimination literature. According to their article, testing practitioners report that landlords often provide equal treatment to testers in the initial contact phase but, in subsequent connections with landlords, discrimination is more likely to occur. To echo the recommendations of Freiberg and Squires (2015), future research should build on the experience of housing practitioners and incorporate multiple interactions with providers into the paired-testing methodology. Housing discrimination is a “moving target” (Massey, 2005) and, until housing discrimination research is designed to capture all the time points at which housing discrimination could occur, it will fail to capture the true level of racial and ethnic discrimination that exists.

Broadening the Scope of Housing Discrimination Research

One of the biggest contributions of this symposium on paired testing is that it highlights research on housing discrimination against other protected groups. Such research has received very little attention in the existing literature, despite the changing nature of the bases of complaints mentioned previously, the fact that a growing share of the population is disabled, and the fact that the treatment of the lesbian, gay, bisexual, transgender (LGBT) community is under more scrutiny because of the recent liberalization of marriage laws. Aranda (2015) summarizes new research on housing discrimination in the rental market against people with disabilities. People who are deaf or hard of hearing and those who use wheelchairs face significant barriers in obtaining rental housing. Turner (2015) offers insights on studies in progress aimed at documenting discrimination against families with children, the LGBT community, and people receiving housing vouchers. Significant challenges in studying different protected groups clearly have not arisen in the research on housing discrimination against racial and ethnic minorities. The research discussed in these articles should serve as a point of departure for future studies on these groups to deepen our understanding of the nature of discrimination that these protected classes face.

In addition, this new research can offer methodological insights for research that needs to be conducted on other protected groups that have received very little attention, such as discrimination on the basis of religion, which is also protected by the Fair Housing Act. Oh and Yinger (2015) show in their table in appendix A that significant discrimination exists against Arabic and Moroccan homeseekers—who are likely to be of the Muslim faith—in Italy, Norway, Spain, and Sweden. Only one such study has been conducted in the United States—in Los Angeles, and it also reveals unfavorable treatment against Arabic homeseekers relative to White homeseekers, although the study was not based on paired testing (Carpusor and Loges, 2006). More research should be focused on housing discrimination against Muslim homeseekers in the United States, especially given that Muslim civil rights complaints rose from slightly more than 1,000 in 2003 to more than 2,700 in 2008 (CAIR, 2009).

The research on people with disabilities and other protected classes highlighted in this symposium should be used to guide research that might combine two or more protected classes. Massey and Lundy (2001) found that poor, Black females received the worst treatment by landlords in a telephone audit study of rental housing in Philadelphia. No followup studies, however, have examined

gender, race, and class, or other combinations of protected classes, such as family status, despite the important findings that this study yielded. Desmond (2012) showed that eviction is a more significant problem for Black women in inner-city Black neighborhoods than it is for Black men, with a majority of such women parenting children on their own. Eviction leads to more mobility and potentially more housing discrimination. Given that 30.1 percent of Black family households are single female-headed households (Lofquist et al., 2012) and that 50.5 percent of Black children live in such families (U.S. Census Bureau, 2013), future research should consider how combinations of protected classes like race, sex, and family status may detrimentally affect the treatment of groups in their quest for better housing. Perhaps different segments of the Black population are experiencing different levels of discrimination and that could be another reason why a broad focus on Black-White discrimination underestimates the true level of housing discrimination experienced by Black homeseekers.

Conclusion

Paired-testing methodology has been instrumental in the study of housing discrimination. The articles in this symposium should be commended for broadening our thinking about how to conduct future research on housing discrimination against racial and ethnic minorities and against other protected groups. A large agenda of research lies ahead for the next generation of housing discrimination researchers to capture the moving target of housing discrimination. Funding is extremely limited, however. In recent years, linkages have been made between racial and ethnic disparities in perceived discrimination and various health outcomes (see, for example, Williams and Mohammed, 2009). This work and future research on housing discrimination against people with disabilities could be attractive to funding agencies focusing on health. Researchers clearly will have to be much more innovative and entrepreneurial to acquire sufficient funds to fulfill the challenging objectives of this 21st century housing discrimination research agenda.

Author

Samantha Friedman is an associate professor of sociology at the University at Albany, State University of New York.

References

- Aranda, Claudia L. 2015. "Targeting Disability Discrimination: Findings and Reflections From the National Study on Housing Discrimination Against People Who Are Deaf and People Who Use Wheelchairs," *Cityscape* 17 (3): 103–121.
- Carpusor, Adrian G., and William E. Loges. 2006. "Rental Discrimination and Ethnicity in Names," *Journal of Applied Social Psychology* 36 (4): 934–952.
- Council on American-Islamic Relations (CAIR). 2009. *The Status of Muslim Civil Rights in the United States 2009: Seeking Full Inclusion*. Washington, DC: Council on American-Islamic Relations.

- Desmond, Matthew. 2012. "Eviction and the Reproduction of Urban Poverty," *American Journal of Sociology* 118 (1): 88–133.
- Ewens, Michael, Bryan Tomlin, and Liang Choon Wang. 2014. "Statistical Discrimination or Prejudice? A Large Sample Field Experiment," *Review of Economics and Statistics* 96 (1): 119–134.
- Freiberg, Fred, and Gregory D. Squires. 2015. "Changing Contexts and New Directions for the Use of Testing," *Cityscape* 17 (3): 87–101.
- Hanson, Andrew, and Zackary Hawley. 2011. "Do Landlords Discriminate in the Rental Housing Market? Evidence From an Internet Field Experiment in U.S. Cities," *Journal of Urban Economics* 70 (2): 99–114.
- Hogan, Bernie, and Brent Berry. 2011. "Racial and Ethnic Biases in Rental Housing: An Audit Study of Online Apartment Listings," *City & Community* 10 (4): 351–372.
- Lofquist, Daphne, Terry Lugalia, Martin O'Connell, and Sarah Feliz. 2012. *Households and Families: 2010*. C2010BR-14. Washington, DC: U.S. Census Bureau.
- Massey, Douglas S. 2005. "Racial Discrimination in Housing: A Moving Target," *Social Problems* 52 (2): 148–151.
- Massey, Douglas S., and Garvey Lundy. 2001. "Use of Black English and Racial Discrimination in Urban Housing Markets: New Methods and Findings," *Urban Affairs Review* 36 (4): 452–469.
- Massey, Douglas S., and Jonathan Tannen. 2015. "A Research Note on Trends in Black Hypersegregation," *Demography* 52 (3): 1025–1034.
- National Fair Housing Alliance (NFHA). 2015. *Where You Live Matters: 2015 Fair Housing Trends Report*. Washington, DC: National Fair Housing Alliance. <http://www.nationalfairhousing.org>.
- Oh, Sun Jung, and John Yinger. 2015. "What Have We Learned From Paired Testing in Housing Markets?" *Cityscape* 17 (3): 15–59.
- Pitingolo, Rob, and Stephen L. Ross. 2015. "Housing Discrimination Among Available Housing Units in 2012: Do Paired-Testing Studies Understate Housing Discrimination?" *Cityscape* 17 (3): 61–85.
- Turner, Margery A., Rob Santos, Diane K. Levy, Doug Wissoker, Claudia Aranda, and Rob Pitingolo. 2013. *Housing Discrimination Against Racial and Ethnic Minorities 2012*. Washington, DC: U.S. Department of Housing and Urban Development.
- Turner, Margery Austin. 2015. "Other Protected Classes: Extending Estimates of Housing Discrimination," *Cityscape* 17 (3): 123–136.
- U.S. Census Bureau. 2013. "Table C9—Children by Presence and Type of Parent(s), Race, and Hispanic Origin: 2013." Current Population Survey, 2013 Annual Social and Economic Supplement. <http://www.census.gov/hhes/families/data/cps2013C.html>.

U.S. Department of Housing and Urban Development (HUD). 2013. *Annual Report on Fair Housing: Fiscal Year 2012–2013*. Washington, DC: Government Printing Office. <http://portal.hud.gov/hudportal/HUD?src=/annualreport>.

———. 2011. *Annual Report on Fair Housing: Fiscal Year 2011*. Washington, DC: Government Printing Office.

Williams, David R., and Selina A. Mohammed. 2009. “Discrimination and Racial Disparities in Health: Evidence and Needed Research,” *Journal of Behavioral Medicine* 32 (1): 20–47.

Commentary: Expanding the Fair Housing Testing Landscape

James Perry

In my 12 years as a fair housing nonprofit organization CEO, I have never read more thorough and in-depth analyses of fair housing testing than those presented in these several articles. Although I have occasional disagreements, the authors succeed in presenting compelling analyses of testing history, methodology, and results.

When read as a group, the articles are thorough in their analyses of testing specifics, while occasionally referencing important macro considerations. The authors presumably prioritized sufficiency of technique, assuming that an important path to resolution of greater challenges was testing efficacy.

I, however, think it important to consider macro questions that the articles do not raise directly. What point is there to perfecting testing and the collection of its data if policymakers fail to use it when promulgating policy? The real estate market crash of the late 2000s was predicted by testing data evidencing reverse redlining in minority neighborhoods. Black and Latino neighborhoods became foreclosure ghost towns that, years in advance, foretold the fate of the American real estate market. The testing data that unearthed this phenomenon should have informed a course correction to stem discriminatory redlining and prevent the crash. The failure of policymakers to take action based on testing data, however, resulted in historic losses of equity, first in minority neighborhoods and then across all neighborhoods.

In the end, the quality of testing matters little if policymakers refuse to see it as the all-important tool it is. It can be argued that there is no more sophisticated and advanced tool for determining trends in the real estate market. Policymakers would do well to prioritize the role of testing in setting housing policy for the nation.

Turning more directly to the science of testing, authors Fred Freiberg and Gregory D. Squires raise important concerns and poignantly think through the future of testing. The authors use experiences and lessons from enforcement investigations in New York, America's most competitive housing market, to argue that the U.S. Department of Housing and Urban Development's (HUD's) decennial paired-testing studies must evolve should HUD intend the studies to remain relevant. The authors correctly argue that the "predominantly complaint-responsive approach to enforcing fair housing laws is inadequate" and HUD should "place a greater emphasis on proactive testing to uncover systemic discrimination" (Freiberg and Squires, 2015: 92).

Freiberg and Squires (2015) prescribe multiple-contact testing rather than simple initial-contact testing to provide fuller insight into discriminatory practices. I agree. Testing in academic spheres often chases data purity so aggressively that organizers lose sight of the goal. Rather than achieving perfect data collection, testing should seek to determine whether or not differential treatment has occurred. The methods of deterring differential treatment unfortunately do not always fit neatly in data collection schematics. Like Freiberg and Squires (2015), I urge academic and audit-based test organizers to evolve their thinking and testing in a manner that always prioritizes uncovering discrimination.

This concept comports perfectly with an overall, but understated, finding that weaves through each article. Testing is evolutionary. The articles document the transition from print advertising to online advertising in determining which housing providers to investigate. Testing organizers have adopted digital audio and video recording in jurisdictions that allow it. Testing has emerged as a method for investigating not only race-based discrimination but also discrimination based on national origin, disability, religion, familial status, income, sex, and sexual orientation. In recent years, organizers have been able to conduct some tests wholly via digital communication without need for in-person contact or voice conversations. As housing markets have evolved, so has testing. These articles are testament.

A final issue that is evident but, again, not directly confronted in the articles is the lack of funding for additional audit and enforcement testing. The housing market is huge. The market produces hundreds of billions of dollars in transactions each year. In the third quarter of 2014, housing accounted for 15.2 percent of U.S. Gross Domestic Product.¹ President Obama's 2016 budget proposal, however, recommends only \$71 million for fair housing programs.² A serious commitment to eradicating housing discrimination requires funding levels that allow for testing to become commonplace in American housing transactions. Government spending on testing is less than 1 percent of the value of American housing transactions. Most people in the real estate business can practice their entire careers without ever being audited for fair housing practices. Unused tools are of no value. These articles demonstrate the sophistication and broad ability of the testing tool. So let us put it to work. Government should give testing value and purpose by expanding its use in the American housing market. Absent that increased use, it is unlikely that we will ever fully eradicate discrimination in American housing transactions.

Acknowledgments

The author thanks Sarah Rosen Wartell for ushering Urban Institute's policy brilliance into social and broadcast media, where it is accessible to everyday Americans, and Margery Turner for her relentless optimism in pursuit of inclusive housing policy.

¹ <http://www.probuilder.com/housing-policy-update-gdp-and-housing-industry's-impact>.

² <http://portal.hud.gov/hudportal/documents/huddoc?id=36-FY16CJ-FHPrograms.pdf>.

Author

James Perry is a fair housing advocate who previously served as Executive Director of the Greater New Orleans Fair Housing Action Center and the Mississippi Gulf Coast Fair Housing Center.

References

Freiberg, Fred, and Gregory D. Squires. 2015. "Changing Contexts and New Directions for the Use of Testing," *Cityscape* 17 (3): 87–101.

Commentary: Testing Benefits Housing Providers and the Industry

Fred Underwood

National Association of Realtors®

Paired testing is useful for detecting discriminatory differences in treatment and can be used to address cases of alleged discrimination, to assess market levels of discrimination, or to educate the housing industry so it can monitor itself and make corrections to possible discriminatory practices. Changes in how discrimination is manifested in the marketplace do not remove the value of effective paired tests, but they may make such testing more difficult or expensive. Working with the housing industry, particularly practitioners with a deep commitment to fair housing and an intimate understanding of the market, can be beneficial in refining and crafting measures to test today's markets. Participation and cooperation with the industry becomes critical as testing seeks to dive deeper into a transaction.

It is important to note how significantly demographic change is affecting the housing market. According to the Joint Center for Housing Studies (JCHS) at Harvard University (2015), the millennial generation is 45 percent multicultural (JCHS uses the term *minority*) and 22 percent of the millennial generation is Hispanic. In the next 10 years, 76 percent of housing growth will be multicultural. The *2014 State of Hispanic Homeownership Report*, issued by the National Association of Hispanic Real Estate Professionals (2015), found that Hispanic homeowners accounted for 50 percent of the net growth in owner households between 2000 and 2014.

As is often said at multicultural real estate conferences, "We are the market." Real estate in the coming decades will increasingly be multicultural, meaning that real estate agents' clients will come from multicultural populations and the profession itself will increasingly reflect the market. That demographic change cannot predict a change in the level of housing discrimination, but it does mean that the real estate profession has a stronger commitment to and identification with the goals of fair housing. At the same time, the declines in homeownership rates and household wealth need to be examined to determine causes of this decline, including the effect of discrimination at the transaction level.

Rooting out housing discrimination is understood by a greater number of people in the housing industry than in the past. The housing industry is more diverse as it reflects today's multicultural housing market. Support for fair housing has moved from passive acceptance of the law to active support. For example, recent policy positions at the National Association of Realtors® support the concept of disparate impact and call for fair housing protections based on sexual orientation and gender identity.

The analysis presented by Fred Freiberg and Gregory D. Squires may be most helpful in employing testing as a continued and effective tool for identifying discriminatory behavior. Differences in the traditional measures of discrimination, such as making an appointment or housing availability, have declined. Freiberg and Squires (2015) discuss the need to do more followup testing and to use more creative testing structures. Changes in the way housing is marketed, including on the Internet, and the actions of some who may be intent on finding new ways to discriminate also will shift the focus of testing to actions later in the transaction. Many housing providers, sharing a deep commitment to fair housing, can become strong allies to the fair housing community and help develop new and focused investigations. Although the housing industry has employed self-testing in relatively few instances, further examination of these self-testing methods and results could inform this discussion.

Collaboration between experienced fair housing organizations and state or local Realtor® associations can be of great help in the field of testing. A housing provider seeking to learn whether his or her agents are discriminating will not be satisfied simply knowing that every prospect gets a cup of coffee and four listings to look at. The provider will be looking to see if the agents steer, provide assistance equally later in the home search process, follow up equally, and so on. Brokers may also be interested in determining if other actors in the housing transaction are treating their clients fairly and may provide access to the state of the housing search to help test those others in the housing transaction. Michigan Realtors® contracted with experienced fair housing centers in Michigan to make self-testing available to brokers. Although the results of self-testing are generally not available for enforcement actions, results were used to improve fair housing education for Realtors®, and the knowledge of housing practices and the positive relationships that develop can be invaluable for designing effective tests.

Freiberg and Squires (2015) also discuss the cost of testing on housing providers and state that the value of testing also accrues to the housing provider by dispelling false claims of discrimination. More can be done to provide value back to the housing industry in this regard. Case studies are one of the best tools used in fair housing training sessions, office meetings, or discussions at conferences to educate real estate professionals. A contemporary and local example of a housing transaction enables discussion and learning about housing discrimination and helps agents and brokers identify ways to change their practices to reduce or eliminate discrimination. Whether the tests are conducted for enforcement, research, or self-correction, the act of being shown how certain practices or actions can be discriminatory helps practitioners improve their skills and their service to all customers.

Testing to identify general levels of discrimination in a market is required for both enforcement and education of the housing industry. As the articles of both Freiberg and Squires (2015) and of Rob Pitingolo and Stephen L. Ross (2015) point out, changes in how real estate is marketed and shown present challenges to traditional testing methods, but these challenges can be met. Pitingolo and Ross (2015) conclude that these challenges have not resulted in an underestimate of housing discrimination. Marjorie Austin Turner focuses on how to adjust testing to address bases of discrimination in which the tester's status is not apparent to housing providers. Addressing these changes in the market may require a renewed commitment to testing resources, however,

the general reaction to being tested and its cost in terms of lost time and business, is negative. Documenting the benefits of testing to housing providers and the industry may help to develop the support needed for continued funding for testing.

Author

Fred Underwood is the Director of Diversity for the National Association of Realtors®.

References

Freiberg, Fred, and Gregory D. Squires. 2015. "Changing Contexts and New Directions for the Use of Testing," *Cityscape* 17 (3): 87–101.

Joint Center for Housing Studies (JCHS) at Harvard University. 2015. *The State of the Nation's Housing*. Cambridge, MA: Joint Center for Housing Studies at Harvard University.

National Association of Hispanic Real Estate Professionals. 2015. *2014 State of Hispanic Homeownership Report*. San Diego: National Association of Hispanic Real Estate Professionals.

Pitingolo, Rob, and Stephen L. Ross. 2015. "Housing Discrimination Among Available Housing Units in 2012: Do Paired-Testing Studies Understate Housing Discrimination?" *Cityscape* 17 (3): 61–85.

Turner, Margery Austin. 2015. "Other Protected Classes: Extending Estimates of Housing Discrimination," *Cityscape* 17 (3): 123–136.

Refereed Papers

Refereed papers that appear in Cityscape have undergone a thorough and timely double-blind review by highly qualified referees. The managing editor reviews submitted manuscripts or outlines of proposed papers to determine their suitability for inclusion in this section. To submit a manuscript or outline, send an e-mail to cityscape@hud.gov.

What Happens to Housing Assistance Leavers?

Robin E. Smith

DeBruce Foundation

Susan J. Popkin

Urban Institute

Taz George

Federal Reserve Bank of Chicago

Jennifer Comey

Washington, D.C. Office of the Deputy Mayor for Education

The views expressed in this article do not necessarily represent the views of the Federal Reserve Bank of Chicago.

Abstract

To assess whether federal housing assistance can encourage asset building and self-sufficiency, we need to know why households leave housing assistance and how they fare on their own. As a group, housing assistance leavers appear to be doing better than those still in public housing or receiving rent subsidies; they have higher incomes, are more likely to be married, and live in lower poverty, safer communities. Dividing unassisted households into those who left housing assistance for negative reasons and those who left for positive reasons highlights how those leaving for negative reasons are worse off and how those leaving for positive reasons are struggling. Such findings suggest the need for targeted approaches to support both groups.

Introduction

Housing assistance¹ in the United States is unusual: unlike many other forms of public assistance, it is not an entitlement and serves only about one-fourth of eligible households (Turner and Kingsley, 2008). In many jurisdictions, waiting lists for public housing and vouchers are closed or very long; applicants can wait years before they reach the top of the list. Those households lucky enough to successfully navigate the system receive deep subsidies that require them to pay one-third of their income for housing; they generally are able to keep their housing assistance as long as they remain income eligible and a tenant in good standing.² Despite the open-ended nature of the subsidy—and the fact that after having left housing assistance, getting back on is extremely difficult—people nationwide remain on assistance for only a few years (Turner and Kingsley, 2008). Evidence suggests, however, that households in distressed urban public housing typically remain on assistance much longer than households receiving other forms of housing assistance, such as vouchers. Most HOPE VI Panel Study respondents who were residents of distressed public housing developments slated for demolition had lived in their developments for 10 years or longer (Popkin et al., 2002). Likewise, a long-term study of Chicago public housing residents found an average tenure of 28 years (Popkin et al., 2013).

Housing assistance meets its basic goals for those households lucky enough to receive it. Having a voucher or living in public housing improves stability, dramatically reduces homelessness,³ and may lead to better outcomes for children (Mills et al., 2006; Newman and Harkness, 1999). On the contrary, individuals lacking stable housing or at risk of losing their permanent housing face serious mental and physical health consequences (Burgard, Seefeldt, and Zelner, 2012). The public and assisted housing programs, however, have significant flaws that leave them open to criticism from advocates, policymakers, and researchers, particularly the fundamental problem that the programs serve only a fraction of those in need. Observers are also concerned about the racial and economic segregation of public and assisted housing (Popkin et al., 2012) and whether the programs should be administered locally or regionally (Katz and Turner, 2013). In addition, like other safety-net programs, housing assistance has been the target of congressional budget cuts because of its relatively high cost.

Despite the large body of research on housing assistance programs—especially on housing choice vouchers (formerly Section 8 vouchers)—few researchers have attempted to study what happens to recipients when they leave assisted housing. Given the research and policy attention to “welfare

¹ Two major types of federal housing subsidies are public housing and housing choice vouchers (formerly Section 8 vouchers). For residents, public housing is very low cost and sometimes provides a place-based concentration of services and programs for children and adults. Such public housing neighborhoods concentrate poverty, however, and can have higher crime, poor-performing neighborhood schools, low-quality housing, and the stigma of living in “the projects.” Housing choice vouchers are portable, with greater neighborhood and housing choice and less stigma. Residents often face higher utility bills and must navigate the private market and landlords (who may or may not accept the subsidy as payment), however, while complying with program rules about using the voucher within a set timeframe.

² Eligibility for housing assistance is determined by household income. Although targeting requirements may influence the income of a household offered assistance, in general, a household with less than 80 percent of the Area Median Income is eligible.

³ In this article, we use “literal homelessness” in reference to the HUD definition of homeless: residing in places not meant for human habitation, residing in a homeless shelter or supportive housing, or facing imminent loss of their permanent housing. “Doubled up” refers to people living in accommodations designed for a smaller number of occupants. In our analysis, we examine both of these living conditions as separate categories and together under the heading of “homeless.”

leavers” when welfare reform was first implemented, it is surprising that stakeholders know so little about why households stop receiving federal housing assistance and even less about how households fare afterward.⁴ We know that welfare leavers continue to experience economic hardship, with the average monthly household income for leavers near the poverty level, and that the most challenging barriers to self-sufficiency often involve unemployment and the inability to maintain or find work owing to poor health (Acs and Loprest, 2004, 2001). Still, most households who leave welfare are at least slightly better off than those who remain on assistance. Most welfare leavers are employed in the months after leaving assistance, and nearly two-thirds of all exits are associated with work (Hofferth, Stanhope, and Harris, 2002). In fact, Acs and Loprest found that “hourly wage rates of working leavers in NSAF and SIPP are consistently higher than those of current recipients, suggesting that those who can earn higher wages are more likely to exit or less likely to continue to be eligible for TANF [Temporary Assistance for Needy Families]” (Acs and Loprest, 2001: 78). Literature has also suggested, however, that employment rates of welfare leavers vary by year of exit (Acs and Loprest, 2004, 2001).

Housing is the biggest expense for most households; in many cities, minimum-wage earners cannot afford even basic two-bedroom apartments (DeCrappeo et al., 2010). Further, evidence shows that welfare recipients who also receive housing assistance have lower incomes and less social support than other TANF recipients; they also surprisingly report high levels of material hardship (Zedlewski, 2002). Although income cutoffs for housing assistance are much higher than those for TANF benefits—households are eligible for vouchers or public housing as long as their household income does not exceed 80 percent of the Area Median Income (AMI)⁵—housing assistance leavers are likely at risk for hardship and instability because they still have low (and often extremely low) incomes.⁶

During the past 20 years, successive U.S. Department of Housing and Urban Development (HUD) administrations have promoted the potential for housing assistance to help recipients build assets and improve their circumstances. Federal programs have targeted housing assistance recipients for help toward homeownership as a way for low-income households to increase housing stability and build wealth. The HOPE I, II, and III programs experimented with allowing public housing residents to purchase their units and provided resources to prepare them for homeownership. HUD has also aimed to use tenant-based assistance as a steppingstone to homeownership and, thus, economic stability. The Section 8 homeownership program allows eligible participants to use their vouchers toward the purchase of a home; the program is relatively small and operates at the discretion of individual housing authorities. In addition, HUD’s Family Self-Sufficiency (FSS) program enables housing authorities to support participants’ moving toward self-sufficiency in the form of educational and employment opportunities, and, in some cases, homeownership. The key benefit for participants is that as their income increases, their rent increases as well, but the

⁴ Some evidence suggests that elderly household heads and disabled household heads are less likely to stop receiving housing choice vouchers (Olsen, Davis, and Carrillo, 2005).

⁵ Although households may be *eligible* for admission if their income is less than 80 percent of AMI, targeting requirements are often much lower.

⁶ According to Turner and Kingsley (2008), HUD classifies a household’s income in relation to the median income for the local housing market area, known as Area Median Income, or AMI (an approach considered more equitable than the federal poverty level because it roughly takes differences in cost of living into account). According to HUD definitions, low income is less than 80 percent of AMI, very low income is less than 50 percent of AMI, and extremely low income is less than 30 percent of AMI.

housing authority directs the difference in rent payments into an escrow account. Participants can claim the escrow after completing the program and use it for a downpayment on a home, education expenses, or a car to help them maintain employment.

To test whether housing assistance can encourage asset building and self-sufficiency, it would help if policymakers knew more about how households fared after they made their transition off assistance, to judge the success of the program as a springboard to better outcomes.⁷ To date, little systematic research has been conducted on the effects of these efforts on households after they leave housing assistance, particularly whether these programs help recipients successfully make the transition off housing assistance and build long-term assets, such as a home or car.⁸

The Urban Institute's HOPE VI Panel Study is one of the few studies that has attempted to look at what happens to housing assistance leavers. The researchers took advantage of the study's longitudinal panel design to explore what happened to participants who had left or lost their assistance (McInnis, Buron, and Popkin, 2007). The study tracked a sample of 887 residents from five housing developments targeted for HOPE VI redevelopment from 2001 to 2005; during that period, 103 households left housing assistance. About one-half of households left for positive reasons, such as marriage or a wage increase that made them ineligible for assistance; some of these residents became homeowners. The rest left for negative reasons, such as breaking program rules, being evicted, or being relocated and unable to move back. Among the Panel Study's findings were that the unassisted households seemed to be highly mobile and that, although many were apparently doing better economically than their counterparts still on assistance, they still experienced substantial material hardship. These preliminary findings from the HOPE VI Panel Study suggest that despite efforts to turn housing assistance into a steppingstone for economic stability, the trajectory for those who leave is likely similar to that of welfare leavers: ongoing struggles and insecurity. Given the small size of the sample, however, more research is required before definitively concluding how housing assistance leavers fare.

HUD's Moving to Opportunity (MTO) demonstration provides a unique opportunity to explore what happens to housing assistance leavers in greater depth. The MTO final evaluation survey tracked a sample of nearly 5,000 public housing households in five cities from 1994 (the baseline) through the final evaluation surveys approximately 10 to 15 years after their initial moves (Sanbonmatsu et al., 2011). The MTO research tested the effect of offering very low-income public housing residents the opportunity to move to low-poverty communities; the hope was that moving would improve adults' access to jobs, children's access to better schools, and economic outcomes overall.⁹ The study provides a rich dataset.

In this article, we take advantage of the tracking of participants over time (including after they leave housing assistance) to study the factors that cause households to leave assistance and how the experiences of leavers compare with households that remain on assistance. We supplement the data from the MTO final evaluation survey with new, qualitative indepth interviews with a small number of housing assistance leavers from two MTO sites. We use the MTO survey data

⁷ Our analysis describes the economic, social, and physical well-being of formerly assisted households and those remaining on assistance. We do not attempt to directly detect an effect of assistance receipt on these outcomes.

⁸ HUD is currently funding a major MDRC study of the FSS program.

⁹ The findings from the MTO experiment have been reported extensively elsewhere (see Ludwig et al., 2008; Sanbonmatsu et al., 2011).

to describe the characteristics of households who leave assistance and to describe how they were faring at the time of the final MTO survey (in quality of life, housing, finances, family stability, employment, and mental health) compared with their counterparts still on housing assistance. We also explore how the Great Recession (December 2007 to June 2009) may have influenced the lives of housing assistance leavers, especially those who had attempted to become homeowners.

The picture for housing assistance leavers is complex. New policies could help support households when they make the transition off assistance to ensure they do not experience severe hardship or fall back into poverty.

Research Questions and Methodology

This article explores the following research questions.

- How many MTO participants left housing assistance during the demonstration?
- Why did households leave housing assistance? Can leavers be classified into those leaving for positive and negative reasons?
- How did households describe leaving assistance?
- How do the characteristics and experiences of households leaving for positive reasons compare with those leaving for negative reasons?
- How do households no longer receiving federal housing assistance compare with households still receiving it?
- How did households describe their lives after leaving housing assistance?
- How did households describe their experiences with homeownership, and how were these experiences affected by the recession?

We used two different data sources on MTO participants for this analysis: (1) the MTO final evaluation survey (Sanbonmatsu et al., 2011) and (2) new indepth interviews with MTO households no longer receiving federal housing assistance. Although all MTO households received some form of housing assistance at random assignment (1994 to 1998), many were no longer on assistance at the time of the final evaluation interview (2008 to 2010). These two data sources provide a unique opportunity to study what happens to households when they leave assisted housing, why they leave, and how they compare with their still-assisted peers.

MTO Survey Data

HUD launched the MTO demonstration in 1994 in Baltimore, Boston, Chicago, Los Angeles, and New York City. MTO was a voluntary relocation program, targeted at very low-income residents of distressed public housing in high-poverty neighborhoods (Orr et al., 2003; Sanbonmatsu et al., 2011). About 4,600 households, mostly African-American and Latino, were randomly assigned to one of three treatment groups: (1) a control group, in which families retained their public housing unit and received no new assistance related to MTO; (2) a Section 8 comparison group, in which

families received the standard counseling and a voucher subsidy for use in the private housing market; or (3) an experimental group, in which households received special relocation counseling and search assistance, along with a voucher designed to encourage relocating to a low-poverty neighborhood for at least 1 year. Slightly less than one-half of households in this group took advantage of the special voucher. Households participated in extensive surveys at three points during the length of the 15-year study: (1) at baseline, when randomization occurred; (2) in 2002 for an interim evaluation; and (3) between 2008 and 2010.

The University of Michigan's Institute for Social Research collected the MTO final evaluation survey data between June 2008 and April 2010 under its contract with the National Bureau of Economic Research. The database includes 3,273 adult household heads and 5,105 youths who were between the ages of 10 and 20 at the end of 2007 (Sanbonmatsu et al., 2011). The response rate was approximately 90 percent for adults and youths. The survey covered a wide variety of outcomes and mediators in six domains: (1) housing mobility; (2) adult education, employment, and earnings; (3) household income and public assistance; (4) adult, youth, and child mental and physical health; (5) youth and child social well-being; and (6) child and youth educational performance.

Analytic Approach

For the MTO final impacts experimental analysis, the Urban Institute developed a unique multi-source process to more accurately identify whether each MTO head of household was receiving *any* federal rental assistance and to determine the *specific type* of assistance received.¹⁰ Although housing assistance status is a key outcome for the MTO demonstration, it is difficult to determine whether a household is still receiving a subsidy and, if so, what type. Recipients often misidentify the type of housing assistance they receive or erroneously report not receiving any assistance (see the appendix of Shroder, 2002). Relying on administrative housing assistance data can also be unreliable, because resident annual recertification records are not always entered into the appropriate databases (Olsen, Davis, and Carrillo, 2005).

To reduce participant misreporting, the MTO final evaluation survey included a new series of housing assistance status questions.¹¹ We compared these responses with two administrative sources—Multifamily Tenant Characteristics System (MTCS)/Public and Indian Housing Information Center (PIC) and Tenant Rental Assistance Certification System (TRACS)/Multifamily data—to identify each MTO participant's type of housing assistance.¹² HUD staff from the Office of Policy Development and Research (PD&R) successfully matched approximately 90 percent of MTO heads of household to at least one administrative source.

¹⁰ The information in this section is excerpted from Comey, Popkin, and Franks (2012). The assistance types are public housing, tenant-based federal rental assistance, project-based nonpublic housing federal rental assistance, and no federal rental assistance (including owners, unassisted renters, the homeless, and those with other statuses).

¹¹ The researchers based the new questions on the MTO interim survey (Orr et al., 2003) and the HOPE VI Panel Study, a five-site study that tracked outcomes for 887 residents of public housing developments targeted for redevelopment. See Popkin et al. (2002) for a full description of the study.

¹² MTCS/PIC data contain longitudinal information on households living in public housing or receiving tenant-based housing vouchers (Form 50058), whereas TRACS/Multifamily data contain longitudinal information on households living in project-based Section 8 housing (Form 50059). See Comey, Popkin, and Franks (2012) for a complete description of the methodology for identifying households' housing assistance status.

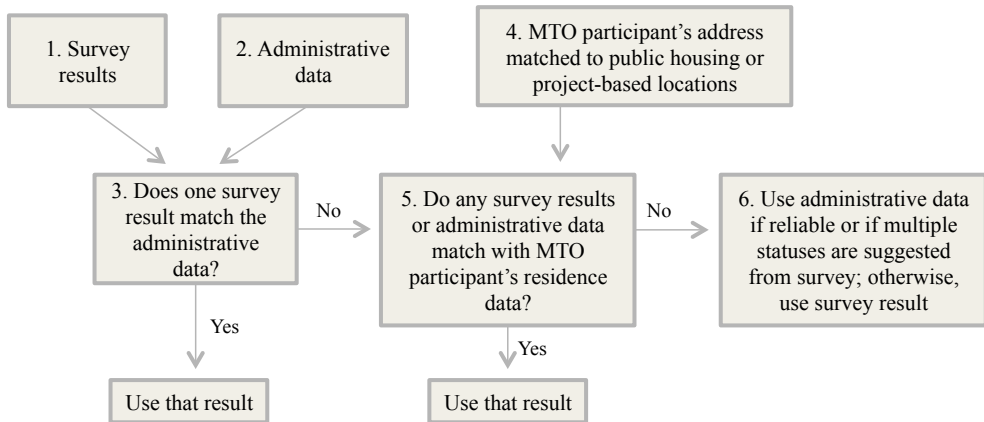
In the first step of this new process, housing assistance survey responses were coded as either eliminating or not eliminating each of eight possible housing assistance statuses (exhibit 1).¹³ For MTO heads of household who were linked to the MTCS/PIC and TRACS/Multifamily data, the administrative housing status was determined (step 2). Researchers then compared this status with its corresponding survey information, and, if the survey analysis matched a status from the administrative data, they used the resulting assistance status (step 3).

For the 14 percent of participants whose survey responses did not agree with administrative sources, analysts compared participants' residences at the time of the final survey with the known addresses of the housing authorities' housing developments and project-based assistance buildings (step 4). They also compared MTO participants' ZIP Codes at the time of the final survey with the survey responses and administrative data (step 5). For the 7 percent of MTO participants who still had conflicting housing assistance statuses after this step, analysts selected the housing assistance status from the administrative data if the participant's administrative records matched residents' characteristics from the survey file and they found no duplicate records (step 6). Otherwise, analysts assigned participants a status based on the survey result.

After all households that completed the final survey were classified as assisted or unassisted, we compared key outcomes of these groups to assess how unassisted households fared relative to assisted households. We identified 40 items from the MTO final evaluation survey related to quality-of-life issues that housing assistance policies are designed to improve for participating households,

Exhibit 1

Multistep Triangulation Process To Identify Housing Assistance Status



MTO = Moving to Opportunity.

¹³ The eight possible housing categories are (1) renter with tenant-based assistance, (2) renter in public housing, (3) renter with project-based assistance, (4) renter without housing assistance, (5) homeowner, (6) homeless individual, (7) individual who lives with family or friends and does not pay rent, and (8) individual with another housing arrangement. The researchers could not determine assistance status for owners, because most owners were not asked about housing assistance. For this reason, the final categories include information only on rental assistance, not on homeownership assistance.

such as housing stability, neighborhood quality, income and benefits, and material hardship. We also examined demographic characteristics to explore what kinds of households left housing assistance and for what reasons.

Finally, we created a classification scheme to sort assistance leavers into two groups—those that left housing assistance for positive reasons and those that left for negative reasons—and analyzed differences among unassisted households related to their motivation for leaving assistance. We then calculated mean values of these outcome and demographic characteristics and performed *t*-tests to assess whether any differences between assisted and unassisted households and between positive-reason and negative-reason leavers were statistically significant.

Our primary method of classifying leavers as positive or negative involved a range of indicators, the most reliable being the household's survey response to the question of why they left assistance. This item has a low response rate, however, and could be used to classify only approximately one-half of the unassisted households. The other indicators used in the classification process, including income and homeownership status at time of the final interview, are noncontemporaneous to the time each household actually left assistance. To address concerns that imputing the positive and negative classifications with these additional indicators may have significantly altered the comparison of positive and negative leavers, we conducted a separate analysis that relied on only the direct survey item for classifying leavers as positive, negative, or unknown. These secondary results are presented and discussed in the appendix.

Qualitative Indepth Interviews

The second data source was new, indepth interviews conducted with MTO households that no longer received housing assistance, including those that had left for positive and negative reasons. Indepth interviews with household heads no longer receiving assistance highlighted both the reasons households discontinued housing assistance and how unassisted households adjusted to make ends meet. In the fall of 2011, we conducted in-person interviews with 24 households in Boston and Los Angeles.

We included in the eligible pool households whose current address was within a 30-minute drive from the metropolitan area center or within a cluster of households 2 hours or less from the metropolitan area center, given limited time for research staff to complete interviews. To create the eligible pool, we identified households in these areas whose final survey was conducted in English: 126 families in the Boston area and 165 families in the Los Angeles area. In total, we interviewed representatives from 24 households, which included 24 adult household heads (11 from the Boston area and 13 from the Los Angeles area) and 13 youths (5 from the Boston area and 8 from the Los Angeles area).

Trained Urban Institute research staff conducted separate but concurrent interviews with the head of household (parent or guardian) and youth, when present. Interviews were held in respondents' homes and lasted approximately 60 minutes. Respondents (adult and youth) were given \$40 to compensate them for their time. The semistructured interview guides covered topics that included reasons for leaving assistance, housing history after assistance (including homelessness and homeownership), and finances and family life after housing assistance. Interviews were recorded,

transcribed, and reviewed for themes. Summary memos were prepared for each household, reviewing household composition and housing history, with an emphasis on reasons for leaving assistance and life after assistance.

We sent recruitment letters to all eligible households, providing a toll-free number to call if they wanted to participate. We followed the letters with telephone calls, soliciting participation after respondents answered a series of screening questions to ensure they no longer received federal housing assistance. We monitored the categorization of positive leavers and negative leavers during recruitment; when we found more positive leavers were being recruited for interviews, we placed additional emphasis (primarily telephone calls) on identifying negative leavers to increase their participation rates. The characteristics of those we recruited or those who had valid addresses up to 2 years after the final contact for the final evaluation were similar in many ways to the whole MTO population we identified as unassisted at the time of the final evaluation using our triangulation method, as shown in exhibit 2. Households that agreed to participate in the research project had

Exhibit 2

Interviewed Households Compared With All MTO Households and the Recruitment Pool

	All MTO Housing Assistance Leavers	Recruitment Pool		Interviewed	
		Boston	Los Angeles	Boston	Los Angeles
		N (%)	N (%)	N (%)	N (%)
Total number of households	1,149	126	165	11	13
Type of leaver					
Positive	603 (52)	84 (67)	91 (55)	9 (82)	11 (85)
Negative	546 (48)	42 (33)	74 (45)	2 (18)	2 (15)
Average age of head of household (years)	43.9	42.5	42.7	43.1	43
Gender of head of household					
Female	1,115 (97)	125 (99)	161 (98)	11 (100)	12 (92)
Male	34 (3)	1 (0.8)	4 (2.5)	0 (0)	1 (8)
Marital status					
Single	766 (69)	86 (71)	117 (72)	9 (82)	5 (39)
Married	352 (32)	35 (29)	45 (28)	2 (18)	8 (62)
Average income (final)	\$23,915	\$35,035	\$25,608	\$47,205	\$39,052
Employment					
Employed in past 2 weeks	733 (64)	87 (69)	104 (63)	10 (91)	11 (85)
Not employed in past 2 weeks	414 (36)	39 (31)	61 (37)	1 (9)	2 (15)
Education					
Not a high school graduate	375 (33)	23 (19)	62 (38)	0	4 (31)
High school graduate/GED®	611 (54)	71 (57)	90 (55)	6 (55)	9 (69)
College graduate	151 (13)	30 (24)	11 (7)	5 (46)	0 (0)
Tenure					
Renter	739 (64)	71 (56)	129 (78)	7 (64)	10 (77)
Owner	317 (28)	46 (37)	28 (17)	4 (36)	3 (23)
Homeless (doubled up or literally)	79 (7)	6 (5)	7 (4)	0	0
Other	14 (1)	3 (2)	1 (0.6)	0 (0)	0 (0)
Average years at current address	4.8	5	5	4.2	6.2

GED® = general educational development. MTO = Moving to Opportunity.

Note: When categories do not total 100 percent (or to the column totals), the rest are missing.

higher incomes, education levels, and employment, however, than the greater unassisted pool. Participants were also much more likely to be classified as leaving for positive reasons. Therefore, the qualitative findings may not represent the most challenged households.

Results

In this section, we first present our findings on the number of people who left housing assistance and our analysis of why they left assistance. Then we describe how and why we categorize unassisted households into those leaving for positive versus negative reasons and relate how households in both situations describe leaving assistance. We detail how unassisted households compare with those still on assistance, with particular attention to how positive and negative reasons for leaving influence results. Finally, we present how households describe their lives after leaving housing assistance and examine how leaver households—particularly those who made the transition to homeownership—may have been affected by the Great Recession.

How Many MTO Participants Left Housing Assistance During the Demonstration?

Using the triangulation methodology for determining assistance status, we find that 35 percent of all MTO households (1,149 heads of household) were no longer receiving housing assistance at the time of the final outcomes survey.¹⁴ This figure is smaller than the proportion we would expect if we looked only at national averages; according to HUD data, the median length of time households use housing assistance is 4.7 years for those living in public housing and 3.1 years for voucher holders (Turner and Kingsley, 2008). The national data include *all* types of recipients (senior citizens, people with a disability, and families), however, from *all* housing authorities (small rural authorities to large, urban agencies). A better benchmark for the MTO sample is *families living in distressed public housing*, such as the HOPE VI Panel Study, which shows residents having much longer tenures than national averages.¹⁵

Can Housing Assistance Leavers Be Classified Into Those Leaving for Positive and Negative Reasons?

Determining why households leave housing assistance is even more challenging than determining their housing assistance status. Limited information on reasons households leave assistance is available in the MTO final evaluation survey. Only households that reported during the survey that they no longer received housing assistance were asked why they left assistance.¹⁶ Of the 1,149

¹⁴ The number and the percentage reflect unweighted respondents because we matched actual survey respondents to administrative data. This share differs from other published reports and articles about the unassisted MTO population, such as Sanbonmatsu et al. (2011) and Comey, Popkin, and Franks (2012), because those studies report the share of the control group only. This article does not differentiate between the MTO control and treatment groups.

¹⁵ See, for example, Popkin et al. (2013, 2010, 2002).

¹⁶ As noted previously, self-reports can be unreliable and inaccurate. For example, 13 percent of respondents who said they no longer received assistance said they left assistance because they had relocated from public housing and later could not move back. Given the frequency of redevelopment in MTO communities (primarily from HOPE VI), some families probably relocated and could not later move back. What is unclear is if all these families left assistance entirely or if some had tenant-based vouchers (and thus were still assisted).

households we determined were no longer receiving assistance, only 630 self-identified as such, meaning that only a little more than one-half responded to the survey item asking why they left assistance.¹⁷ Exhibit 3 shows that 26 percent of these households were unclear or not sure in their response. Of reported housing assistance leavers, 25 percent said they left because they were income ineligible. We do not know the financial circumstances that made these households ineligible but could expect some saw wage gains after promotions and/or steady employment, others completed education or training that prompted higher earnings, and still others added another wage earner (spouse, adult child, or other household member) to the household. “Incoming out,” or seeing household income increases high enough that the resulting subsidy given to the household is very low (or zero), can be a relatively positive reason for leaving assistance.¹⁸ On the opposite end of the spectrum, households that leave because they are evicted or violate program rules (22 percent) left for negative reasons. Family turmoil and economic instability that may contribute to program departure may also increase potential challenges for the household after it no longer receives a housing subsidy.

We created a strategy to enable us to classify all 1,149 households in the MTO final evaluation survey that were no longer receiving housing subsidies into two categories: (1) participants who left assistance for mainly positive reasons (such as homeownership or incoming out) and (2) participants who lost their housing assistance for primarily negative reasons (such as lease violations, evictions, or inability to lease up during the period). We used a combination of information from the MTO final evaluation survey, including stated reason for leaving housing assistance when available, reported homeownership, and income, to separate households into the two categories.¹⁹

Exhibit 3

Reasons for Leaving Assistance Reported in MTO Final Evaluation Survey

Reason	Percent
Other reasons or not sure	26
Incomed out	25
Evicted or terminated from program for violating rules	22
Relocated from public housing and could not move back	13
Moved in with a partner, friend, or relative	9
Purchased a home	5

MTO = Moving to Opportunity.

Note: Unweighted sample is 630 households of 1,149 that Urban Institute determined were assistance leavers. The remaining 519 households did not self-identify as assistance leavers in the survey. Of the responses, 35 percent were written in by participants and back coded.

¹⁷ The weighted share of households classified as unassisted who self-identified as such was 54.1 percent.

¹⁸ “Incoming out” can be associated with tenant-based assistance as a family’s rising income reduces the subsidy it receives of its private-market rent. Households in public housing face a similar phenomenon: income increases can push their public housing rent up to fair-market or ceiling rent.

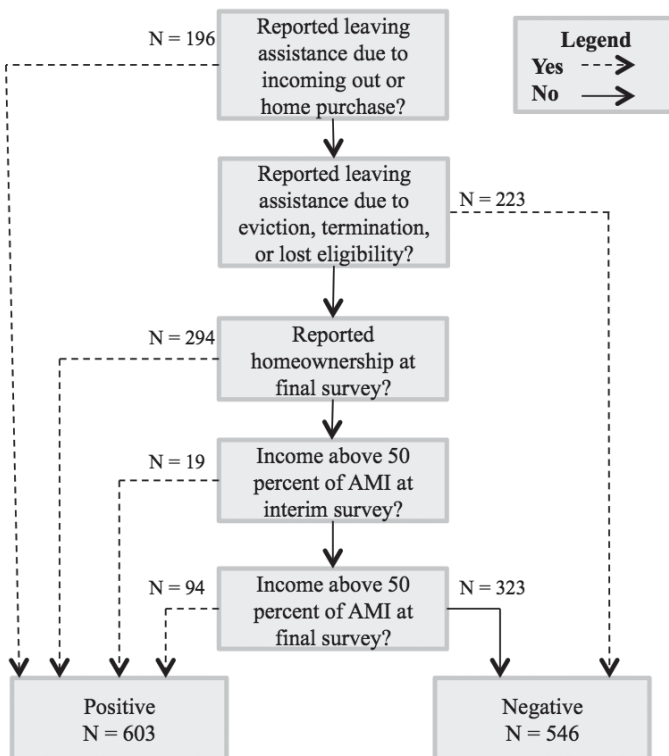
¹⁹ Heads of household identified as a positive leaver because they incomed out if (1) they answered the survey question about why they no longer received housing assistance because they incomed out or purchased a home, (2) they responded that they were a homeowner during the final survey, or (3) the reported household total income was above the HUD Section 8 limit reported in either the interim (2002) or final (2009) surveys. Heads of household identified as a negative leaver (or forced out) if (1) they answered the survey question about why they no longer received housing assistance because they were evicted, terminated, or forced out of the program or (2) the reported household total income at interim or final was below the HUD Section 8 limit. Income limits were applied using the household sizes at interim and final and addresses at interim and final.

A key dimension of the classification process is that preference was given to the most reliable indicators of reasons for leaving assistance whenever possible. Exhibit 4 illustrates the order in which these factors were considered, along with the number of households categorized at each stage. The available indicators were ordered so less reliable measures were invoked only when the stronger ones did not provide a clear answer. For example, the lowest measure on the figure—income at final survey—was used only to categorize households that could not be categorized with the early measures because of inconclusive responses.

Of the available indicators that suggest why a household left assistance, we give preference to the survey question directly asking households’ reasons for leaving. Inferences based on homeownership and income are the next-best available means for classifying households because they correspond to the time of the interim and final survey, not to the exact time of a household’s transition off assistance. Therefore, the first step was to categorize unassisted households, whenever possible, as either positive or negative leavers based on their recorded reason for leaving assistance. Of

Exhibit 4

Classifying an Unassisted Household as a Positive or Negative Leaver



AMI = Area median income. MTO = Moving to Opportunity.
 Note: Each number is unweighted total households classified as either positive or negative assistance leavers at each particular step in the sorting process.
 Source: MTO interim and final survey evaluation data

the households responding to this question, 25 percent reported being income out and were classified as positive leavers, and 22 percent were recorded as evicted or losing eligibility and were classified as negative leavers. The remaining households either did not respond to this question or provided a response that we did not count as positive or negative, such as those who were not sure of their reason for leaving or those who moved in with a partner, friend, or relative.

The remaining households were then assessed based on reported homeownership status at the time of the final survey. Homeownership is a probable indicator of a household leaving housing assistance for positive reasons because the means necessary to become a homeowner potentially reflect a stronger household financial position and intent to leave assistance.²⁰ The 294 households that reported owning a home were recorded as positive leavers, and the remaining households were then categorized based on interim income. If interim income was more than 50 percent of the AMI, households were considered positive leavers. If not, the final assessment was made based on final income relative to 50 percent of area median income. We recognize that even for this group, all leavers at or more than 50 percent of AMI will not have left the program for positive reasons. We use this threshold to approximate those households for whom the declining level of subsidy could likely be a reason for departure.²¹

Via this process, we classified 603 households, a weighted share of 53 percent of all unassisted households,²² as leaving for predominantly positive reasons, and 546 households (47 percent weighted) as leaving for negative reasons.

Survey and administrative data do not reliably indicate the date at which households left assistance, so classifying households as positive or negative leavers using noncontemporaneous variables (income and homeownership status at time of final survey) is an imperfect process. For example, the classification process assumes that any unassisted household that did not report eviction, termination of eligibility, or incoming out, but does report homeownership, left assistance for positive reasons. For such a household, we do not know the circumstances of its departure from assistance but infer from its reported homeownership that the family left assistance to pursue better opportunities of their own will or because its income then exceeded eligibility limits. The appendix includes results achieved by classifying leavers based only on this survey item, albeit with a large group of leavers classified as leaving assistance for unknown reasons.

Although the reliance on noncontemporaneous variables in the primary methodology is imperfect, it alleviates concerns of a potentially large selection bias into the “unknown” category due to survey nonresponse or unclear response. Households in the unknown category of the secondary

²⁰ Given the recession and housing crash that began soon after the end of the MTO experiment, homeownership was not necessarily a long-term positive outcome for many households. We consider it a positive reason for leaving assistance, however, because it suggests the household had the means and intent to leave assistance.

²¹ We employed a 50-percent threshold to infer households whose departure decisions may be influenced by declining subsidy because this level is (1) used by HUD PD&R as an element in the definition of worst case need for housing assistance, (2) a criterion for initial eligibility in the MTO demonstration, and (3) part of HUD’s explanation of eligibility for housing choice vouchers, suggesting that households above this threshold receive little or no tenant-based assistance (see http://portal.hud.gov/hudportal/HUD?src=/program_offices/public_indian_housing/programs/hcv/about/fact_sheet).

²² This percentage is similar to the portion of households who leave welfare because of increased employment. More than one-half of all welfare leavers cite increased earnings, finding a job, or working more on the same job as the primary reason for leaving welfare (Acs and Loprest, 2001).

methodology in the appendix are exclusively those that no longer received assistance according to administrative records but that nonetheless reported either never receiving assistance, continuing to receive assistance, or being unsure of why they no longer received assistance. A range of conditions could influence whether a household's awareness of its assistance status (for example, the type of assistance received in the first place, the actual reason for departing, and the financial literacy of the householder). It would be unreasonable to potentially introduce this bias to our sample of positive and negative leavers based on the households' lack of awareness of assistance status.

How Do Families Describe Leaving Assistance?

To better understand the nuances within the positive and negative reasons families no longer receive housing assistance, we spoke at length with unassisted MTO households regarding the situations surrounding their departure from assistance. We know from the MTO final evaluation survey that some families are unclear about why they lost their housing assistance. For some of our interview respondents, the incident that resulted in termination from housing (such as eviction) was often clear.

The court sent me a letter, saying it was an eviction. (LA159)

Other households found program rules difficult. Some families missed mandatory appointments (such as recertification) and did not fulfill obligations required to maintain assistance. Others struggled with program rules, such as using their voucher within a set timeframe or identifying a unit that passes inspection. In particular, respondents discussed difficulties navigating the private market with poor credit and insufficient security deposits. One woman described her situation this way:

I had to find another place. And it was kind of hard because every place wanted the first month and first and last ... my voucher had a certain amount of days, months, to move into another place, otherwise the voucher was going to expire, and I couldn't find...the places that I did find weren't approved ... my coupon just expired. (LA147)

As expected, some families who left for more positive reasons describe steady employment and increasing paychecks, which helped them move off the program. One woman said she got off because of "a better job, making more money, you know" (LA156). Other families described how, as their income increased, their housing assistance decreased.

I think I kind of weaned, got, kind of got weaned off [assistance] because as I grew on my job, and financially, you get more money, your rent increases slowly, slowly, slowly. So for me, it was like a gradual thing. It just wasn't one day a low amount, and then the next month, you know, a high amount. So I was able to grow gradually and get off of it. (BOS57)

Yeah, to take it [my job] more seriously, you know, the career, so I did that. And that's when they [the housing authority] started paying less and less and less as my income increased, increased, increased. (LA129)

For some, the amount of assistance became small enough that they left the program before being formally terminated by the housing authority.

The amount [of rent] that Section 8 was paying was less than what I was paying. It was like not even \$200 of the rent they were covering. And I felt secure enough in my own income and stability with my job that I didn't need the third party. You know, renewing and verification and this, that, and inspections, and just every 6 months, and just I didn't feel that is was ... it wasn't worth it. (BOS57)

Other respondents said changes in family status, particularly getting married, added income to their household, making them ineligible. One woman described how she married her long-time boyfriend (and father of her children) when he got out of prison. Even those who earn their way off assistance or leave because of a marriage or an additional income earner moving in may perceive themselves as having been forced off assistance rather than voluntarily deciding to leave. It is not surprising that many people are reluctant to give up assistance, given that the subsidy is deep and difficult to get. One respondent described the difficulty of willingly giving up housing assistance:

It's hard because it takes, you have to go through hell and high water to get housing. And I thought, what if I can't afford full rent? Where will my kids be, in a shelter? So you get scared because it takes so long to get housing. You know what I mean? It's like a trap. It's hard to get in, and because of that, you're scared to get out. (BOS106)

Several respondents who left assistance because they had too much income told interviewers that they did not consider themselves as exceeding the income limit because the housing authority included the earnings from adult children or extended family in the income calculation when the household head did not believe he or she had access to the additional funds. When this happened to one woman during recertification, she said—

I was surprised, and I regret it today, because housing is so expensive. How I wish I had that Section 8. (BOS26)

Other respondents noted that family members (particularly husbands or boyfriends) pressured them to get off housing assistance.

Well, once I started working, and I got in, I got back with their dad, I think the income was one of the issues. And he was the type of person that didn't want to be involved in anything like that [housing assistance]. (LA147)

Even families who noted significant income changes, such as through marriage, used language that suggested they felt pushed out of the program rather than voluntarily left. People commonly described when they “lost” Section 8 or had it “taken away.”

I had to come off Section 8 because I got married a year later, and my income was over the income limit. And that's when I lost my Section 8. (BOS2)

I was working at the apprenticeship program, it, I was going to school like, they had us go to school every 6 months for 2 weeks ... every time I left there, I got a raise. So it was just constantly rising, so that's why, you know, they took the Section 8 from me. (LA156)

How Do the Characteristics and Experiences of Households Leaving for Positive Reasons Compare With Those Leaving for Negative Reasons?

Our interviews indicate that even those who left for positive reasons often think their exit was not fully voluntary. Our analysis of the MTO survey data, however, shows clear differences in the trajectories for positive leavers and negative leavers, even with few demographic differences between the groups (exhibit 5). Positive leavers were more likely to be married at the study end.²³ African-American leavers who were unassisted were more likely to leave for negative reasons. This finding requires deeper analysis; it could reflect that Hispanic households were more likely to be two-parent households or that housing authority policies on eviction vary across the five MTO sites, which themselves vary considerably in demographic composition.²⁴

The striking difference between the two groups is income—not surprisingly, given its role in positive departures from assistance. Households that left for positive reasons reported a median income of \$37,865, while families leaving for negative reasons reported a median income of \$13,950 (exhibit 6). This difference in income has dramatic implications for the ability of these households to function in the private housing market. Indeed, those who left for positive reasons were also more likely to have better outcomes, presumably linked to their higher incomes. For instance, positive-leaver heads of households were more likely to be married, less likely to have experienced homelessness or overcrowding in their household, and less likely to experience housing cost burdens than households leaving for negative reasons.

Those people who moved for positive reasons were also more likely to have moved because they wanted a better or bigger housing unit, although negative leavers characterized their most recent move as prompted by problems with their landlords (exhibit 7). Positive leavers were also more

Exhibit 5

Demographics of All MTO, Assisted, and Unassisted Households

Characteristic	All MTO Households	Assisted Households	Unassisted Households		
			All	Positive Leavers	Negative Leavers
Age of household head (years)	45	45	44***	44***	44*
Female household head (%)	98	99	97	97	97
Household size (people)	3.7	3.7	3.8*	3.7	3.8*
African-American, non-Hispanic (%)	63	62	64	59**	71***
Hispanic (%)	31	32	30	34	25**
Married at end of study (%)	20	13	32***	43***	21***

MTO = Moving to Opportunity.

* $p < .10$. ** $p < .01$. *** $p < .001$.

Notes: To test statistical significance in difference between the assisted households and each group of unassisted households (all, positive leavers, and negative leavers), chi-squared tests were performed on categorical variables and t-tests were performed on continuous variables. In each test, the assisted households are the reference group.

²³ As noted previously, adding an additional wage earner to a household could push income high enough that the corresponding housing subsidy is very low or zero.

²⁴ Two of the five sites—Baltimore and Chicago—had entirely African-American populations. Both those sites had large numbers of HOPE VI grants and were relocating many of their residents. Because housing authorities have great discretion in setting lease requirements and enforcing one-strike rules, these sites could have used different standards for eviction than the other sites.

Exhibit 6

Housing Cost Burden of All MTO, Assisted, and Unassisted Households (percent)

Characteristic	All MTO Households	Assisted Households	Unassisted Households		
			All	Positive Leavers	Negative Leavers
High housing cost burden	68	67	70	66	74**
Severely high housing cost burden	42	41	43	36**	52***
Ever homeless	23	17	34***	24***	46***
Ever doubled up	20	14	31***	23***	40***
Ever literally homeless	5	4	7***	3*	12***
Overcrowded housing	21	20	23*	20	27***
Severely overcrowded housing	8	7	11***	9*	14***

MTO = Moving to Opportunity.

*p < .10. **p < .01. ***p < .001.

Notes: To test statistical significance in difference between the assisted households and each group of unassisted households (all, positive leavers, and negative leavers), chi-squared tests were performed on categorical variables and t-tests were performed on continuous variables. In each test, the assisted households are the reference group.

Exhibit 7

Housing, Neighborhood, and Health Characteristics of Positive and Negative Leavers (percent, except where noted)

Characteristic	Positive Leavers	Negative Leavers
Moved to attain improved housing	34	23
Moved because of landlord problems	13	22
Rated housing as excellent or good	70	56
Neighborhood satisfaction rating ^a	1.92	2.4
Neighborhood feels safe at night	79	64
Adult rates health as good or better	74	53
Takes medicine for blood pressure	32	38
Problems with depression	14	19

^a Collected on a 5-point scale, with 1 representing the highest level of satisfaction.

likely to be satisfied with their housing quality and neighborhood and to feel safe than negative leavers. Finally, those who left for positive reasons felt better physically, took fewer medications, and were less likely to face mental health problems than those who left for negative reasons.

How Do Families No Longer Receiving Federal Housing Assistance Compare With Households Still Receiving It?

Outcomes for households that leave federal housing assistance for positive reasons differ significantly from outcomes for households that leave for negative reasons. As detailed in the following section, substantial differences are evident in outcomes for the positive-leaver households and those for households still on assistance. By contrast, outcomes for negative leavers look remarkably similar to—or sometimes worse than—those for the still-assisted. Still, the only notable difference in demographic characteristics among the three groups is the proportion of families who report being married at the final evaluation survey (exhibit 5). Positive leavers are about twice as likely to report being married at the end of the study than negative leavers; likewise, negative leavers are nearly twice as likely to be married than those still on assistance.

Unassisted Households That Left for Positive Reasons Are Better Off Financially

The average MTO head of household who was unassisted at the time of the MTO final evaluation survey was better off financially than his or her assisted peers.²⁵ The income difference between unassisted and assisted households is almost entirely driven by the relatively higher incomes of the positive leavers; negative leavers' income is only slightly higher than those still on assistance. One reason for this difference might be the larger number of two-parent households among the positive leavers: one-third of unassisted heads of household (43 percent of positive leavers) were married at the time of the final survey compared with 13 percent of those who were assisted.

Similarities between welfare leavers and housing leavers suggest that income stability, or lack thereof, is central to why families decide to leave assistance and, over time, is how families continue to face the social and economic hardships of living in poverty without the housing safety-net assistance previously provided. Although unassisted families have higher incomes on average than assisted families, families leaving for negative reasons fall well below the federal poverty level. This difference across housing assistance leavers is consistent with research on welfare leavers that finds a growing inequality across TANF recipients, with "some families ... moving up and out of poverty, but some families ... moving down into extreme poverty" (Acs and Loprest, 2001: 83).

Unassisted Households See More Income Growth

Although the two groups had similar incomes at baseline, the unassisted group experienced a significantly larger increase in household income during the demonstration period than the assisted group (an increase of \$13,216 from baseline to interim for the unassisted compared with \$1,441 for assisted households). Both positive (\$23,627) and negative (\$4,498) leavers experienced greater increases in income than still-assisted families. Still, the difference is mainly driven by the positive leavers; as exhibit 8 shows, the median incomes of still-assisted households and

Exhibit 8

Income and Benefit Receipt of All MTO, Assisted, and Unassisted Households

Characteristic	All MTO Households	Assisted Households	Unassisted Households		
			All	Positive Leavers	Negative Leavers
Median income at study end	\$15,521	\$13,153	\$23,915***	\$37,865***	\$13,950
Median income change over study period	\$3,466	\$1,441	\$13,216***	\$23,627***	\$4,498
Receives food stamps (%)	48	57	33***	17***	51**
Has Medicaid coverage (%)	37	46	22***	12***	33***
Receives Social Security benefits (%)	30	36	18***	11***	26***
Receives TANF benefits (%)	16	19	12***	6***	19
Adult has health insurance (%)	84	87	79***	84*	73***

MTO = Moving to Opportunity. TANF = Temporary Assistance for Needy Families.

*p < .10. **p < .01. ***p < .001.

Notes: To test statistical significance in difference between the assisted households and each group of unassisted households (all, positive leavers, and negative leavers), chi-squared tests were performed on categorical variables and t-tests were performed on continuous variables. In each test, the assisted households are the reference group.

²⁵ This finding is consistent with the HOPE VI Panel Study research (McInnis, Buron, and Popkin, 2007) and with the fact that many households leave assistance because their income rises to a level where they receive little or no benefit.

negative leavers are very similar. In addition, like the trends in income overall, this difference is almost certainly attributable to the higher proportion of two-parent households among the unassisted.

Unassisted Households Receive Less Other Public Assistance

Given the higher incomes of the unassisted group, it is not surprising that, overall, the group was less likely to report receiving benefits such as TANF, Social Security, Medicaid, and food stamps throughout the demonstration period and especially by the time of the MTO final evaluation survey. Although positive leavers may require less financial help, the findings for negative leavers are worrisome: they are actually less likely to report receiving food stamps or Medicaid than still-assisted households. Without the buffer of housing assistance, this finding suggests that these households are at risk of significant hardship and instability. Indeed, negative leavers are more likely than both positive leavers and the still-assisted to report food insecurity, which is the prevalence of food hardship defined by the Community Population Survey-Food Security Scale (exhibit 9). Further, more than one-third of both the still-assisted and negative leavers report difficulty in making utility payments.

Exhibit 9

Hardship and Debts of All MTO, Assisted, and Unassisted Households (percent)

Characteristic	All MTO Households	Assisted Households	Unassisted Households		
			All	Positive Leavers	Negative Leavers
Household is food insufficient	30	32	28*	20***	37*
Any credit card or medical bills debt	45	41	52***	64***	40
More than \$5,000 of credit card debt	8	7	10***	17***	3**
More than \$5,000 of medical bills debt	6	5	9***	10***	8***

MTO = Moving to Opportunity.

* $p < .10$. ** $p < .01$. *** $p < .001$.

Notes: To test statistical significance in difference between the assisted households and each group of unassisted households (all, positive leavers, and negative leavers), chi-squared tests were performed on categorical variables and t-tests were performed on continuous variables. In each test, the assisted households are the reference group.

Positive Leavers Still Report Hardship and Have High Debts

Still, one in five positive leavers reports food insecurity. Further, nearly two-thirds report having medical or credit card debt: 17 percent report having more than \$5,000 in credit card debt, and 10 percent report the same level of medical debt (see exhibit 9). Negative leavers and those still on assistance also report having some debt, but relatively few report carrying credit card or medical balances greater than \$5,000.

Positive Leavers Have Better Housing and Neighborhood Outcomes

Heads of household who left for positive reasons report higher levels of satisfaction with their housing and neighborhood (exhibit 10). Positive leavers also report feeling safer and having fewer problems with crime and disorder. The differences could be particularly striking for families leaving public housing. For example, after graphically describing the environment she and her

sons experienced in public housing, one respondent said she “cried tears of joy” when she moved to a better neighborhood (BOS124). By contrast, negative leavers reported only marginally better neighborhood conditions than those still on assistance.

Looking at national census data, unassisted households, especially positive leavers, lived in neighborhoods with slightly higher but still statistically different socioeconomic indicators (exhibit 11). For instance, the median income in census tracts where the unassisted lived at the time of the final survey was \$37,436 compared with \$29,346 for tracts where assisted households lived; in addition, 23 percent of households were below the federal poverty level in tracts where unassisted

Exhibit 10

Housing and Neighborhood Satisfaction of All MTO, Assisted, and Unassisted Households (percent, except where noted)

Characteristic	All MTO Households	Assisted Households	Unassisted Households		
			All	Positive Leavers	Negative Leavers
Rated housing as excellent or good	60	58	63**	70	56
Neighborhood satisfaction rating (1–5) ^a	2.39	2.53	2.15***	1.92***	2.4*
Moved to a better neighborhood	23	22	26**	27**	24
Neighborhood feels safe at night	63	58	72***	79***	64
Neighborhood has drug problems	26	30	20***	13***	27
Neighborhood has alcohol problems	48	54	38***	30***	47**
Neighborhood has loitering problems	53	59	42***	34***	51***
Neighborhood has trash, graffiti, and abandoned buildings	66	71	58***	52***	65**
No transportation access problems	94	93	96***	98***	94

MTO = Moving to Opportunity.

*p < .10. **p < .01. ***p < .001.

^a Collected on a 5-point scale, with 1 representing the highest level of satisfaction.

Notes: To test statistical significance in difference between the assisted households and each group of unassisted households (all, positive leavers, and negative leavers), chi-squared tests were performed on categorical variables and t-tests were performed on continuous variables. In each test, the assisted households are the reference group.

Exhibit 11

Neighborhood Quality for All MTO, Assisted, and Unassisted Households

Characteristic	All MTO Households	Assisted Households	Unassisted Households		
			All	Positive Leavers	Negative Leavers
Median neighborhood income	\$32,372	\$29,346	\$37,436***	\$41,996***	\$33,449***
Neighborhood poverty rate (%)	29	32	23***	20***	26***
Neighborhood single-parent rate (%)	46	50	40***	37***	43***
Share of employed residents (%)	86	85	87***	88***	86**
Adult has no friends in neighborhood (%)	58	57	60*	62*	59

MTO = Moving to Opportunity.

*p < .10. **p < .01. ***p < .001.

Notes: To test statistical significance in difference between the assisted households and each group of unassisted households (all, positive leavers, and negative leavers), chi-squared tests were performed on categorical variables and t-tests were performed on continuous variables. In each test, the assisted households are the reference group.

households lived at final survey compared with 32 percent for tracts where assisted households lived. This comparison is somewhat misleading, however: a substantial proportion of assisted households still live in public housing communities, which often have extreme levels of concentrated poverty. As one interview respondent noted, “I guess I should be grateful in a sense because if it [public housing] wasn’t horrendous, I’d probably still be there, afraid to leave” (BOS106). The figures in exhibits 10 and 11 make clear that even the former residents who were best off—the positive leavers—mostly still lived in neighborhoods with high rates of poverty, crime, and disorder.

Positive Leavers Have Better Physical and Mental Health

Exhibit 12 shows health outcomes for all MTO participants at the final evaluation survey. Positive leavers report the best health status overall and have substantially lower levels of hypertension and depression. By contrast, negative leavers report the worst health status and are about as likely as those still on assistance to report depression, hypertension, or disability. Given that the negative leavers are also the least likely to have Medicaid coverage, these findings are particularly worrisome and highlight this group’s high vulnerability.

Exhibit 12

Health of All MTO, Assisted, and Unassisted Households (percent)

Characteristic	All MTO Households	Assisted Households	Unassisted Households		
			All	Positive Leavers	Negative Leavers
Adult rates health as good or better	57	52	64***	74***	53
Takes medicine for blood pressure	39	41	35***	32***	38*
Has trouble lifting or climbing stairs	48	54	38***	30***	46***
Problems with depression	18	18	16	14**	19

MTO = Moving to Opportunity.

*p < .10. **p < .01. ***p < .001.

Notes: To test statistical significance in difference between the assisted households and each group of unassisted households (all, positive leavers, and negative leavers), chi-squared tests were performed on categorical variables and t-tests were performed on continuous variables. In each test, the assisted households are the reference group.

Negative Leavers Experience Worrisome Levels of Housing Hardship and Instability

Assisted and unassisted households generally did not differ statistically in levels of housing cost burden and severe housing cost burden, in part because assistance programs often require participants to pay for utilities and a varying portion of rent, and because assisted households have a lower median income (Comey, Popkin, and Franks, 2012). Housing cost burden is defined as paying more than 30 percent of a family’s income on housing-related expenses; severe housing cost burden is defined as paying more than 50 percent of a family’s income for housing-related expenses. More than one-half of negative leavers report severe housing cost burdens (exhibit 6). Even more worrisome, these households face high levels of housing instability and hardship.

Approximately one-third of the unassisted household heads reported a time when they were without their own place to stay compared with only 17 percent of assisted households; the figure for the negative leavers was 46 percent. One in five positive leavers and more than one in three negative leavers reported having to double up with friends and family. Of the unassisted, 7 percent

had been literally homeless; the figure for negative leavers was a shocking 12 percent. Negative leavers were also the most likely to report overcrowded—and severely overcrowded—housing.

For example, one interview respondent said that after she lost her housing assistance, she and her four children had to move into one bedroom in a family member’s apartment:

I couldn’t find a place, so I didn’t have nowhere else to go, so I moved in with them [family members] because they had a spare bedroom.... everybody moved in.... by that time, I only had four [children] because the older one had already [moved out]. (LA147)

Unassisted Households Struggle With High Utility and Housing Costs

Both assisted and unassisted households struggle with utility payments (exhibit 13). Other research has documented that utilities are often a problem for residents who leave public housing for vouchers (see Popkin et al., 2013). A respondent from LA told us—

And the utilities were more over here than they were there [in public housing]... we had a water bill that we weren’t used to having, and trash. So those was things we didn’t prepare for, and repairs. (LA127)

Still, even though all respondents report problems with utility payments, unassisted households face higher utility costs and nearly double total housing costs compared with assisted families. These higher costs put them at risk for food insecurity and housing instability. This LA respondent spoke of having to cut her grocery expenses in half to pay her other bills:

We cut back on the groceries now, because... everything, like I said, is more expensive. So we cut back on that. Instead of spending like \$600 [on food] where we were before, we only spend \$300. (LA127)

Exhibit 13

Hardship and Housing Costs for All MTO, Assisted, and Unassisted Households

Characteristic	All MTO Households	Assisted Households	Unassisted Households		
			All	Positive Leavers	Negative Leavers
Ever late on utilities payment (%)	36	39	31***	22***	41
Ever threatened to shut off utilities (%)	27	28	24*	18***	32*
Ever had utilities shut off (%)	6	6	6	3***	10***
Household electricity cost	\$122	\$109	\$144***	\$157***	\$129**
Household gas cost	\$77	\$70	\$90***	\$99***	\$81*
Household rent/mortgage cost	\$493	\$335	\$785***	\$997***	\$555***
Total monthly housing costs	\$691	\$514	\$1,017***	\$1,248***	\$765***
Housing payments ever 15 days late (%)	22	21	25**	27***	22

MTO = Moving to Opportunity.

*p < .10. **p < .01. ***p < .001.

Notes: To test statistical significance in difference between the assisted households and each group of unassisted households (all, positive leavers, and negative leavers), chi-squared tests were performed on categorical variables and t-tests were performed on continuous variables. In each test, the assisted households are the reference group.

Likewise, this Boston homeowner reported falling behind on her mortgage and “shorting” her food budget to make payments:

Yeah. Yeah. Yeah, I had, at one point it was scary... I never really got far behind, but I had, and that generate a paper but I've never really fallen behind because I would pay it and not have the food. The food is, is always shorted of food. (BOS124)

Others discussed purchasing only necessities and then only enough to meet immediate needs, such as buying one roll of toilet paper at a time.

Even when we take a shower, we cannot buy shampoo... we use soap. (LA209)

These intense financial pressures take a toll on families' well-being. Interview respondents discussed being stressed and worried about paying the rent, the mortgage, and other bills. Some also commented on how financial pressures affected their relationship with children and spouses.

[Financial pressure] wasn't the reason why I, we got a divorce. There was other things going on, but on top of the house putting more pressure on us, it was a lot of pressure. So it, I would have rather loved to continue to rent, you know. (BOS2)

How Do Families Describe Their Lives After Leaving Housing Assistance?

Analysis of the MTO final evaluation survey shows that in many areas households that leave assistance for positive reasons are better off than either negative leavers or those still receiving housing assistance. To put “better off” in perspective, we spoke with unassisted families about their lives since leaving assistance. Our respondents described their battles with unpredictable income, unsteady employment, and unstable housing, as well as the resulting financial pressures. With little savings and a weak safety net, they related how a health problem, divorce, or job loss quickly negated their previous financial gains. These stories are particularly troubling because, as noted previously, our sample of interview respondents may not represent the most challenged families. Our respondents are much more likely to include families who left assistance for positive reasons and have incomes higher than the average MTO positive leaver.

Unpredictable Income

Foreclosure, economic uncertainty, and employment instability fueled by the Great Recession devastated families across the income scale. Low-income families were particularly hard hit, however (Kingsley, Smith, and Price 2009),²⁶ and our MTO respondents were no exception. These respondents from LA described precarious situations that developed when their husbands' employment situations became unstable:

It was scary because, you know, you're not always sure from month to month with the financial. I mean, he [husband] ended up losing one job, went to another job, you know. So that's part of life. It's just like any other depending checks, the welfare, depending on that. (LA159)

²⁶ See also Eckholm (2010).

Actually, it [his income] went down, because he went from being a supervisor to working for himself. So from steady pay to whatever he really makes on his service calls and things like that. So it went down.... It's up and down. (LA127)

Another respondent talked about losing her good-paying job when her company began to lose money because of the recession:

So when all the work got slow, and, you know, they lost a little money, had to wait, that's how I ended up at [name of company], which I make [dollar amount] an hour there, so that's a big cut. (LA156)

Health Problems That Derail Previous Gains

Health problems were another major factor that destabilized families, causing them to lose jobs or income. Even families who left assistance for positive reasons and had been moving toward self-sufficiency could easily be derailed by a health crisis. Some respondents described losing steady jobs after they had surgery or major illness:

So I had to stop working to have the surgery. But when I went back there, they had no openings, so I never really went back. (LA136)

Things changed for me too.... I'm not working now, but in 2009 I almost died. I ended up losing my job because I couldn't go back to work. So I was unemployed for a little bit. And then I got a temporary position that I just finished off, so now I'm looking again. (BOS97)

Other family members' (spouses', children's) health problems also created challenges, reducing income and increasing household costs:

My daughter, she got diagnosed... with [serious medical condition] and she became very ill.... I didn't have enough for her copayments, for the health piece, because it can run from \$500 to \$1,500 to \$3,000. I didn't know what I was going to do. (BOS106)

This mother ultimately turned to credit cards to finance her daughter's medical care, house repairs, and household bills. This decision contributed to high balances and tremendous stress as she worried about how to pay her bills.

Unstable Housing

As noted previously, unassisted families (both positive and negative leavers) are more likely than their assisted peers to deal with housing instability and overcrowding. Although some interview respondents had been at their current address for many years, it was common for them to have had multiple moves after leaving assistance. In some cases, major life changes (divorce, job loss, or illness) prompt housing instability. One woman described what happened a few years after her family had left housing assistance that started them moving from place to place.

Back then I was making a little money, you know, and I was working pretty steady, you know what I'm saying, to pay my rent. What took a big toll on me was when my husband left [went to jail] because that was a lot of help, you know, two incomes are better than one.... as a matter of fact, he went to jail 2004 November and we moved out in 2005. (LA156)

It is interesting to note that some respondents who were relatively housing stable attributed the affordability of their current situations to the assistance of family and friends rather than their ability to function on the private market.

My rent was extremely high [area of Boston]. I just couldn't afford it, it was extremely high. And I tried, again to get some type of assistance when I was living there because I love that area, and I couldn't. I'm living here now. My girlfriend owns this place here.... she said, "I'll rent it to you and give it to you for less... of what you're paying now." And that is everything included. So I moved here. (BOS2)

How Do Families Describe Their Homeownership Experiences, Particularly in Light of the Great Recession?

Homeownership was a dream for some of these MTO families, who hoped that it would help them gain a measure of control over their housing situations and insulate them from capricious landlords and unexpected rent increases. One respondent spoke movingly of her concerns, "constantly worrying" about being unable to afford her apartment and saying she "would stay with relatives or a shelter." She followed up these comments on financial instability and concerns about being able to afford an apartment by explaining that she wanted to buy a house.

You know, I mean, I'm tired of just giving people my money, working hard and giving up my money. Where I feel like I can take that money, and put it toward something that I know someday is going to be mine. (BOS26)

In reality, however, it is often difficult for low-income homeowners to sustain homeownership (Van Zandt and Rohe, 2011). Although some respondents thought their move to homeownership made positive contributions to their lives, others described homeownership as a millstone rather than a steppingstone. Our interviews with MTO families suggest that many were struggling during the recession. Particularly troubling is that some of the families who seemed most financially stable when they left the program—those that left assistance for positive reasons and became homeowners—were now under severe financial pressure.²⁷ This finding suggests that federal policy encouraging voucher holders to move directly to homeownership may not always serve families well in the long run. The Section 8 Homeownership program allows participants to use their vouchers toward a mortgage, and the Family Self-Sufficiency program intends to help participants build escrow that they can use for a downpayment. Our interviews indicate, however, that this strategy may have backfired for some MTO families, especially those who got caught up in the housing bubble that preceded the recession. Some had taken out unmanageable interest-only loans and could afford housing only in poor, minority areas—housing that rapidly lost value after the crash and left them "under water." One respondent said, "Now I owe so much money on my credit cards from trying to keep up with this house because I bought a piece of sh—" (BOS106).

When asked when they bought their home, one respondent replies, "I think it was just before the fall." She went on to explain that she "wasn't even ready to buy a house" but it was her husband's

²⁷ Some recipients of federal housing assistance participate in programs that support and encourage homeownership through escrow accounts, mortgage support, counseling, or other inducements. It is unclear how many interview respondents, if any, participated in such programs.

“dream.” She said, “We really didn’t have the, you know, funds to do it. You know, we didn’t really save to do it. So I think that made a burden, you know” and explained that “we didn’t put our, no substantially amount down to get the house. So that’s why my mortgage was so high” (BOS2).

The lure of homeownership is strong. One family struggling to make mortgage payments described how it felt when they left Section 8 and purchased a home.

I felt important, because... when you [are able] to do something for yourself and provide for your family the good things and feel help, you know, their lives are important, it was very, very happy at that time. (LA209)

This family went on to struggle with its payments and was worried about losing its home. Other respondents had lost a home purchased after leaving housing assistance.

We bought a house. That’s why we moved, we bought a house. But then, 18 months later we got a divorce [lost house in bankruptcy]. (BOS2)

Discussion and Policy Implications

This article explores in depth the experiences of federal housing assistance leavers, both the factors that cause families to leave assistance and how their experiences compare with their counterparts still in public housing or using a voucher. Our analysis takes advantage of the MTO research platform, using the rich data in the MTO final evaluation survey and new qualitative interviews with families in Boston and Los Angeles. Building on the methodology developed for the experimental analysis of housing outcomes in the MTO final evaluation survey reported in Comey, Popkin, and Franks (2012), we determined that 35 percent of MTO participants (1,149 households) were no longer receiving federal housing assistance at the time of the final survey. Our analysis of the survey data and administrative data showed that, of those households, 52 percent left for positive reasons (for example, increased earnings or homeownership) and the remaining 48 percent left for negative reasons (for example, lease violations, eviction, or inability to find a unit before their voucher expired).

Because a mix of factors drives families to leave assistance, it is difficult to draw conclusions about how leavers compare with those still on assistance. If we look at the unassisted as a group, they appear to be doing better than those still on assistance in many ways: higher income, more likely to be married, and living in lower poverty, safer communities. When we divide the unassisted into positive and negative leavers, however, we see a much more complex picture, one that highlights the ways that negative leavers are struggling and the challenges that still put positive leavers at risk of instability. Those who left for negative reasons look much like those still on assistance: they are single-parent households that live in poor-quality housing in high-poverty, high-crime neighborhoods—although not in neighborhoods with the poverty and violence levels that characterize public housing. Most worrisome, their household incomes are barely higher than those still on assistance, which means that, in reality, they are *worse off* because they lack the economic buffer that housing assistance provides. The consequences are apparent in the high levels of instability and hardship negative leavers report: slightly less than 50 percent of households reported experiencing spells of homelessness, with 12 percent reporting being literally homeless at some point.

By contrast, positive leavers have average household incomes above the poverty level, are more likely to be two-parent households, and are in better health than both negative leavers and those still on assistance. They live in safer, lower poverty neighborhoods and are generally satisfied with their housing. Although positive leavers are better off on average, they are still at significant risk for instability. Positive leavers are the most likely to report high credit card debt, and one in five report difficulty affording food or paying their utility bills. Further, our qualitative interviews suggest that this group may have been particularly vulnerable to the effects of the Great Recession: some had become homeowners and were struggling to make payments on large interest-only loans. The housing they had been able to afford was often in the least desirable neighborhoods and likely to rapidly lose value, leaving them under water. Finally, they faced the possibility of losing the employment that had helped them leave assistance in the first place and being unable to make mortgage or rent payments. For some of these households, the dream of building enough assets to move out of poverty permanently was becoming increasingly elusive.

These findings suggest the need for new, targeted approaches to support both households at risk of losing their assistance and those moving toward leaving for positive reasons. According to our analysis, households that leave assistance for negative reasons are at risk of falling into the homeless system. Given the high costs and negative outcomes (especially for children) associated with homelessness, Congress should require housing authorities to target households that appear to be at risk of lease violations or eviction with intensive supports. We believe it is possible to develop a targeting strategy based on the Urban Institute's work on vulnerable public housing families (Popkin and McDaniel, 2013; Theodos et al., 2012). That work shows that the highest risk households are those with no steady income earner and the household is struggling with rent arrears; housing authorities should be able to readily identify some of those households in public housing, although not voucher recipients, through their administrative and property management data systems. Although intensive case management can be costly (Popkin et al., 2010), it is almost certainly less expensive than the multisystem costs associated with being homeless or unstably housed.

Households making the transition off assistance also require assistance, especially given the difficulties of getting back on federal housing assistance. Again, housing authorities should be able to use their annual recertification data to identify some of the households with income increases or new members. Congress could require (or housing authorities could voluntarily target) these households for such services as financial counseling, budgeting assistance, and links to community services. FSS programs sometimes offer these kinds of supports, but the services offered vary considerably across housing authorities.

Our qualitative interviews with MTO families also suggest a need to carefully evaluate the true potential of HUD's homeownership promotion strategies to help assisted households build assets and achieve economic stability. The families we interviewed faced all the worst problems of the housing bubble: high-interest or interest-only loans and housing that lost rather than gained value, leaving them under water. Even those who had managed to hold onto their housing faced serious challenges; they were often able to afford only older units in relatively poor neighborhoods and were challenged by the substantial costs to keep up their units. Before continuing to move forward with homeownership promotion as a strategy for families on housing assistance—some of the poorest families in the United States—we recommend HUD commission a thorough long-term evaluation of its Section 8 homeownership. An evaluation of FSS escrow programs is currently under way.

Finally, we are in an era of shrinking safety nets that is likely to leave these families ever more vulnerable. Congress has gradually squeezed funds for housing assistance, reducing the already-low odds that families who leave and then fall into homelessness will be able to get back on assistance. Signs are emerging of a renewed policy interest in poverty and inequality, however. As part of any new debate, we need a real policy conversation about both the costs of *not* providing support to families who leave assistance and also about not serving the far larger number of families who never manage to receive assistance in the first place.

Appendix

Our classification of leavers into the positive and negative groups was based on a range of indicators. Household income and homeownership status at final survey are only proxy measures of whether a household left assistance for positive or negative reasons, and we rely on these measures because slightly less than one-half of unassisted households could be classified via their survey response to why they left assistance. To assess whether the use of proxy measures impacted our findings of differences between the positive and negative leavers, we employed an alternative classification process using only households’ direct survey response when available, classifying those with an uncertain or missing response as a third “unknown” group. A comparison of the results of the two classification methods is shown in exhibit A-1 and exhibit A-2.

In this secondary method, while slightly more than one-half of leavers fell into the unknown group, the ratio of positive to negative leavers was nearly identical to the primary method, as shown in exhibit A-1. Several differences, shown in exhibit A-2, were evident in the comparison of socioeconomic characteristics between positive and negative leavers, although, as we expected, negative leavers remained more distressed than positive leavers by nearly every measure. In the secondary methodology, the differences in median income and housing-related expenses between positive and negative leavers were much smaller. Positive leavers in the secondary methodology had lower levels of indebtedness than in the primary methodology. Both positive and negative leavers had higher rates of housing instability in the secondary methodology, although the unknown category of leavers had much lower instability.

As previously explained, although the reliance on noncontemporaneous variables in the primary methodology is imperfect, it alleviates concerns of a potentially large selection bias into the “unknown” category due to survey nonresponse or unclear response.

Exhibit A-1

Comparing Methods of Designating Positive and Negative Leavers

	Primary Methodology: Survey Responses Plus Additional Indicators	Secondary Methodology: Survey Responses Only
Remain assisted	2,124	2,124
Total leavers	1,149	1,149
Positive leavers (% of leavers)	603 (52.5%)	250 (21.8%)
Negative leavers (% of leavers)	546 (47.5%)	223 (19.4%)
Unknown reason leavers (% of leavers)	0 (0%)	676 (58.8%)
Ratio of positive to negative leavers	1.10	1.12

Exhibit A-2

Comparing Methods of Designating Positive and Negative Leavers (1 of 2)

	All	Remain Assisted	Left Assistance	Primary Methodology: Survey Responses Plus Additional Indicators			Secondary Methodology: Survey Responses Only		
				Positive Leavers	Negative Leavers	Unknown Leavers	Positive Leavers	Negative Leavers	Unknown Leavers
Adult age	45	45	44	44	44	43	43	44	
African-American, non-Hispanic (%)	63	62	64	59	71	66	73	61	
Hispanic (%)	31	32	30	34	25	28	24	32	
Female head of household (%)	98	99	97	97	97	97	97	97	
Ever married (%)	32	27	42	52	31	45	33	44	
Married at end of study (%)	20	13	32	43	21	35	22	35	
Household size	3.71	3.67	3.77	3.74	3.82	3.66	4.02	3.74	
Ever homeless (%)	23	17	34	24	46	31	54	29	
Ever doubled up (%)	20	14	31	23	40	29	48	27	
Ever literally homeless (%)	5	4	7	3	12	3	13	7	
High housing cost burden (%)	68	67	70	66	74	67	74	70	
Severely high housing cost burden (%)	42	41	43	36	52	36	51	44	
Housing payments ever 15 days late (%)	22	21	25	27	22	26	26	24	
Ever late on utilities payment (%)	36	39	31	22	41	48	42	22	
Ever had utilities shut off (%)	6	6	6	3	10	8	11	4	
Formally threatened to shut off utilities (%)	27	28	24	18	32	38	34	16	
Household rent/mortgage costs (\$)	493	335	785	997	555	802	655	822	
Household electricity cost (\$)	122	109	144	157	129	145	145	143	
Household gas cost (\$)	77	70	90	99	81	104	83	88	
Total monthly housing cost (\$)	691	514	1,017	1,248	765	1,053	884	1,047	
Any credit card or medical bills debt (%)	45	41	52	64	40	56	45	54	
More than \$5,000 of credit card debt (%)	8	7	10	17	3	8	3	13	
More than \$5,000 of medical bills debt (%)	6	5	9	10	8	6	10	9	
Overcrowded housing (%)	21	20	23	20	27	27	34	19	
Severely overcrowded housing (%)	8	7	11	9	14	12	20	9	
Rated housing as excellent or good (%)	60	58	63	70	56	61	51	68	
Moved to attain improved housing (%)	31	32	29	34	23	24	21	33	

Exhibit A-2

Comparing Methods of Designating Positive and Negative Leavers (2 of 2)

	All	Remain Assisted	Left Assistance	Primary Methodology: Survey Responses Plus Additional Indicators		Secondary Methodology: Survey Responses Only		
				Positive Leavers	Negative Leavers	Positive Leavers	Negative Leavers	
				Unknown Leavers	Unknown Leavers	Unknown Leavers	Unknown Leavers	
Moved because of landlord problems (%)	19	19	17	13	22	21	24	14
Moved to a better neighborhood (%)	23	22	26	27	24	25	30	25
Neighborhood employment rate (%)	86	85	87	88	86	88	87	87
Neighborhood poverty rate (%)	29	32	23	20	26	21	27	22
Median neighborhood income (\$)	32,372	29,346	37,436	41,996	33,449	40,647	33,601	38,703
Neighborhood single-parent household rate (%)	46	50	40	37	43	37	42	40
Neighborhood is safe at night (%)	63	58	72	79	64	74	62	74
No transportation access problems (%)	94	93	96	98	94	98	93	97
Neighborhood satisfaction rating (1-5)	2.39	2.53	2.15	1.92	2.4	2.05	2.42	2.1
Neighborhood drug problems (%)	26	30	20	13	27	14	30	18
Neighborhood alcohol problems (%)	48	54	38	30	47	36	52	34
Neighborhood loitering problems (%)	53	59	42	34	51	36	52	40
Neighborhood has trash, graffiti, or abandoned buildings (%)	66	71	58	52	65	55	69	56
Problems with depression (%)	18	18	16	14	19	15	17	17
Adult has good or better health (%)	57	52	64	74	53	69	53	66
Takes medicine for blood pressure (%)	39	41	35	32	38	32	39	34
Has trouble lifting or climbing stairs (%)	48	54	38	30	46	33	43	38
Receives food stamps (%)	48	57	33	17	51	19	49	32
Has Medicaid coverage (%)	37	46	22	12	33	14	28	22
Receives Social Security benefits (%)	30	36	18	11	26	9	23	20
Receives TANF benefits (%)	16	19	12	6	19	6	18	12
Household is food insufficient (%)	30	32	28	20	37	28	37	25
Adult has health insurance (%)	84	87	79	84	73	81	71	81
Median income change over study period (\$)	3,466	1,441	13,216	23,627	4,498	20,386	9,778	12,651
Median income at study period end (\$)	15,521	13,153	23,915	37,865	13,950	31,041	18,932	23,591
Adult has no friends in neighborhood (%)	58	57	60	62	59	62	63	59

TANF = Temporary Assistance for Needy Families.

Authors

Robin E. Smith is a senior research director at the DeBruce Foundation.

Susan J. Popkin is Director of the Urban Institute's Program on Neighborhoods and Youth Development and a senior fellow in the Metropolitan Housing and Communities Policy Center.

Taz George is a research analyst in the Division of Community Development and Policy Studies at the Federal Reserve Bank of Chicago. When this article was written, he was a research associate with the Urban Institute's Housing Policy Finance Center.

Jennifer Comey is a senior policy advisor at the Washington, D.C. Office of the Deputy Mayor for Education.

References

Acs, Gregory, and Pamela Loprest. 2004. *Leaving Welfare: Employment and Well-Being of Families That Left Welfare in the Post-Entitlement Era*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

———. 2001. *Final Synthesis Report of Findings from ASPE's 'Leavers' Grants*. Washington, DC: Urban Institute.

Burgard, Sarah A., Kristin S. Seefeldt, and Sarah Zelner. 2012. "Housing Instability and Health: Findings From the Michigan Recession and Recovery Study," *Social Science and Medicine* 75 (12): 2215–2224.

Comey, Jennifer, Susan J. Popkin, and Kaitlin Franks. 2012. "MTO: A Successful Housing Intervention," *Cityscape* 14 (2): 87–106.

DeCrappeo, Megan, Danilo Pelletiere, Sheila Crowley, and Elisabeth Teater. 2010. *Out of Reach*. Washington, DC: National Low Income Housing Coalition.

Eckholm, Erik. 2010. "Recession Raises Poverty Rate to a 15-Year High," *New York Times*, September 16.

Hofferth, Sandra, Stephen Stanhope, and Kathleen Mullan Harris. 2002. "Exiting Welfare in the 1990s: Did Public Policy Influence Recipients' Behavior?" *Population Research and Policy Review* 21 (5): 433–472.

Katz, Bruce, and Margery Austin Turner. 2013. "Invest but Reform: Streamline Administration of Housing Choice Voucher Program." Washington, DC: The Brookings Institution.

Kingsley, G. Thomas, Robin Smith, and David Price. 2009. *The Impacts of Foreclosure on Families and Communities*. Washington, DC: Urban Institute.

Ludwig, Jens, Jeffrey B. Liebman, Jeffrey R. Kling, Greg J. Duncan, Lawrence F. Katz, Ronald C. Kessler, and Lisa Sanbonmatsu. 2008. "What Can We Learn about Neighborhood Effects from the Moving to Opportunity Experiment?" *American Journal of Sociology* 114 (1): 144–188.

McInnis, Debi, Larry Buron, and Susan J. Popkin. 2007. *Are HOPE VI Families at Greater Risk for Homelessness?* Washington, DC: Urban Institute. <http://www.urban.org/sites/default/files/alfresco/publication-pdfs/311490-Are-HOPE-VI-Families-at-Greater-Risk-for-Homelessness-.PDF>

- Mills, Gregory, Daniel Gubits, Larry Orr, David Long, Judie Feins, Bulbul Kaul, Michelle Wood, Amy Jones & Associates, Cloudburst Consulting, and The QED Group. 2006. *Effects of Housing Vouchers on Welfare Families*. Cambridge, MA: Abt Associates.
- Newman, Sandra J., and Joseph M. Harkness. 1999. "The Long-Term Effects of Public Housing on Self-Sufficiency," *Journal of Policy Analysis and Management* 21 (1): 21–43.
- Olsen, Edgar O., Scott E. Davis, and Paul E. Carrillo. 2005. "Explaining Attrition in the Housing Voucher Program," *Cityscape* 8 (2): 95–113.
- Orr, Larry, Judith D. Feins, Robin Jacob, Erik Beecroft, Lisa Sanbonmatsu, Lawrence F. Katz, Jeffrey B. Liebman, and Jeffrey R. Kling. 2003. *Moving to Opportunity Interim Impacts Evaluation*. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research.
- Popkin, Susan, Megan Gallagher, Chantal Hailey, Elizabeth Davies, Larry Buron, and Christopher Hayes. 2013. *CHA Residents and the Plan for Transformation*. Washington, DC: Urban Institute.
- Popkin, Susan, Diane K. Levy, Laura E. Harris, Jennifer Comey, Mary K. Cunningham, and Larry Buron. 2002. *HOPE VI Panel Study: Baseline Report*. Washington, DC: Urban Institute.
- Popkin, Susan, Michael Rich, Leah Hendey, Christopher Hayes, and Joe Parilla. 2012. *Public Housing Transformation and Crime: Making the Case for Responsible Relocation*. Washington, DC: Urban Institute.
- Popkin, Susan, Brett Theodos, Liza Getsinger, and Joe Parilla. 2010. *An Overview of the Chicago Family Case Management Demonstration*. Washington, DC: Urban Institute.
- Popkin, Susan J., and Marla McDaniel. 2013. *HOST: Can Public Housing Be a Platform for Change?* Washington, DC: Urban Institute.
- Sanbonmatsu, Lisa, Jens Ludwig, Lawrence F. Katz, Lisa A. Gennetian, Greg J. Duncan, Ronald C. Kessler, Emma Adam, Thomas W. McDade, and Stacy Tessler Lindau. 2011. *Moving to Opportunity for Fair Housing Demonstration Program: Final Impacts Evaluation*. Washington, DC: U.S. Department of Housing and Urban Development.
- Shroder, Mark. 2002. "Does Housing Assistance Perversely Affect Self-Sufficiency? A Review Essay," *Journal of Housing Economics* 11 (4): 381–417.
- Theodos, Brett, Susan J. Popkin, Joe Parilla, and Liza Getsinger. 2012. "The Challenge of Targeting Services: A Typology of Public Housing Residents," *Social Service Review* 86 (3): 517–544.
- Turner, Margery, and G. Thomas Kingsley. 2008. *Federal Programs for Addressing Low-Income Housing Needs*. Washington, DC: Urban Institute.
- Van Zandt, Shannon, and William M. Rohe. 2011. "The Sustainability of Low-Income Homeownership: The Incidence of Unexpected Costs and Needed Repairs among Low-Income Home Buyers," *Housing Policy Debate* 21 (2): 317–341.
- Zedlewski, Sheila Rafferty. 2002. *The Importance of Housing Benefits to Welfare Success*. Washington, DC: The Brookings Institution; Urban Institute.

Do the GSEs Meet the Credit Needs of Underserved Communities of Color?

Michela Zonta

Center for American Progress

Abstract

The government-sponsored enterprises (GSEs) are required by Congress to promote access to mortgage credit in underserved markets by meeting explicit affordable housing goals. Although the GSEs have met these goals in the aggregate, previous research suggests that the GSEs' targeted purchases have not encouraged sufficient lending to the most underserved homebuyers. By comparing primary-market lending and GSE secondary-market purchases in the periods before and after the Housing and Economic Recovery Act of 2008, this study revisits the questions of whether the GSEs lead the market and serve all members of underserved markets equally or serve primarily the least underserved of the underserved, especially when it relates to communities of color, who tend to be concentrated in many of the geographically targeted areas. Results from a series of logit models of the determinants of GSE purchases suggest that, although the new designations of underserved markets seem to do a better job in pinpointing low-income and minority communities compared with the early broader definitions, they do not guarantee that the GSEs serve the most underserved of the underserved, especially when it relates to communities of color. Policymakers should revisit the criteria currently adopted for the designation of underserved markets and consider incorporating race and ethnicity in the formulation of affordable housing goals, revising the designation of geographically targeted areas, and establishing subgoals that are specific to geography.

Introduction

In light of the severe distress recently experienced by the government-sponsored enterprises (GSEs; that is, Fannie Mae and Freddie Mac) and in a time of great uncertainty in the direction and extent of housing finance reform, this article revisits the question of whether and how the GSEs can ensure and promote lending among underserved communities of color. The 2013 Home Mortgage

Disclosure Act (HMDA) data indicate that people of color continue to lose ground in the home-ownership market. In particular, African-American and Hispanic households represent an ever-shrinking fraction of homeowners and continue to receive higher cost mortgage loans compared with White borrowers. These patterns give cause for concern, particularly if we consider that people of color will account for three-fourths of household growth during the coming decade (Joint Center for Housing Studies, 2014).

This study calls into question the efficacy of the GSEs in promoting lending in communities of color. Significant government benefits are granted to Fannie Mae and Freddie Mac in exchange for their commitment to meet numerical targets specified by the affordable housing goals for purchasing mortgages made to borrowers from underserved markets. Past research provides little evidence that the affordable housing goals have spurred sufficient lending to the needs of the most underserved homebuyers, especially among people of color. Because this research largely predates the financial crisis and the revisions to the goals made by the Housing and Economic Recovery Act of 2008 (HERA), it is timely to examine whether the revisions have helped the GSEs improve their performance in serving the most underserved markets. In particular, this article brings attention to the fact that the affordable housing goals were never established in terms of race or ethnicity of the borrower and suggests that subgoals targeted at particular race and ethnicity groups might be needed.

By comparing primary-market lending and GSE secondary-market purchases in the pre- and post-HERA periods, this analysis addresses the question of whether the GSEs lead the market. Through a series of logit models, we test the alternative hypotheses either that the GSEs serve all members of underserved markets equally or that they serve primarily the least underserved of the underserved especially when it comes to communities of color, who tend to be concentrated in many of the geographically targeted areas. Further, we examine whether the criteria currently adopted for the designation of underserved markets, which are predominantly based on economic factors, should be revisited to better target communities of color that are still underserved and are in chronic need of mortgage credit, especially in the wake of the foreclosure crisis. Previous research has shown that the influence of the GSEs goes beyond those loans they actually purchase and their procedures and actions may affect the entire mortgage market (Williams, McConnell, and Nesiba et al., 2001).

After providing a background on the GSEs and the affordable housing goals, this article presents a literature review of research on the performance of the GSEs in underserved areas to set the empirical study in its larger theoretical framework. It then describes the methods and data used for the logistic regression analysis and presents results on the lending trends in income-based and race-based underserved markets and the determinants of GSE purchases in these markets. It concludes with a recommendation that the Federal Housing Finance Agency (FHFA), which regulates the GSEs, consider housing goals that include explicitly race-based criteria to encourage primary-market lending to African-American and Hispanic borrowers.

Background

This section provides a brief background on the GSEs, their history, and the issues that led to HERA. In addition, this section briefly discusses the evolution of the affordable housing goals and changes to their formulation, especially those related to the geographically targeted underserved areas.

The Government-Sponsored Enterprises

Fannie Mae and Freddie Mac, the GSEs on which this study focuses, are privately owned, federally chartered entities that purchase residential home mortgages from primary-market lenders and package most of the purchased loans into securities to be sold to private investors with a guarantee of full payment of principal and interest.¹ They are required by Congress to provide stability in the secondary market for residential mortgages and to promote access to mortgage credit in underserved markets. By enabling mortgage lenders to offer housing finance at lower mortgage interest rates, Fannie Mae and Freddie Mac are expected to make homeownership affordable to a wider range of households. In the early 1990s, the Federal Housing Enterprises Financial Safety and Soundness Act (the GSE Act) of 1992 expanded the housing mission for both Fannie Mae and Freddie Mac and called for the U.S. Department of Housing and Urban Development (HUD) to establish annual affordable and geographic goals for the GSEs' loan purchases. Significant government benefits were granted to the GSEs in exchange for their commitment to serve members of the markets specified by the housing goals.²

The regulatory structure, operations, and financial conditions of the GSEs have been subject to intense controversy, most notably in the early 1990s and during the subprime mortgage crisis that led to the collapse of the GSEs in 2008.³ The housing crisis, in particular, has raised large questions about the future of the two entities and whether they should be nationalized, privatized, or extinguished or if they should maintain their current structure. Accounting scandals, a weak regulatory structure—especially regarding capital standards—and credit risk are among the issues related to the GSEs' solvency problems that have been intensely debated over time. With the substantial deterioration of the housing markets that materialized in 2007, Fannie Mae's and Freddie Mac's financial conditions were severely damaged and left them unable to fund mortgages and fulfill their mission without government intervention.

Important controversies emerged also in relation to the affordable housing goals. The goals have been revised several times since their early inception (appendix A). During the period from 1993 to 2009, the numeric targets were periodically increased based on the premise that if the affordable

¹ The federal government established Fannie Mae in 1938 to support the stabilization of the financial conditions of mortgage lenders. Its initial purpose was to increase liquidity for investment by purchasing mortgages insured by the Federal Housing Administration through funds raised by the sale of government-backed securities. The government created Freddie Mac in 1970 to provide a secondary market for conventional mortgage loans. Formerly a government agency, Fannie Mae was converted into a private-public entity in 1968 and became recognized as a GSE responsible for serving low-income and minority borrowers. In 1989, Freddie Mac also acquired the status of GSE.

² For instance, Fannie Mae and Freddie Mac are exempt from Securities and Exchange Commission regulations and state securities laws and pay no state or local income tax (Williams, 2006). The performance of the GSEs would be assessed in terms of their share of the conventional conforming market in each goal category and their ability to lead or meet the industry in making loans in each target category.

³ In the early 1990s, concerns about the GSEs' capital adequacy encouraged HUD and Congress to perform a series of stress tests to calculate the amount of capital that each GSE would need to survive a serious economic downturn. According to HUD's capitalization study (HUD, 1991), neither GSE could survive 3 years of a severe recession. Despite this warning, the GSEs decided not to hold any more capital than they were forced to hold and, instead, to rely heavily on their government guarantees to borrow cheaply (Weicher, 2010).

housing goals were set at less than the primary market they would not be very effective in achieving the GSEs' public purpose of promoting homeownership. The GSEs stated that the affordable housing goals established in 2005 and later years were too high. Critics have also claimed that the affordable housing goals were substantially responsible for the two entities' collapse in 2008 when the purchases of single-family mortgages that the GSEs made to meet the goals targeted to low-income individuals and the potential for moral hazard induced by implicit government backing drove lending to high-risk borrowers (Roberts, 2010; Wallison, 2011). Counterarguments, however, have pointed out that other factors, such as misjudgments about capital requirements and risks associated with sub-prime activity rather than the housing goals, were responsible for the GSEs' collapse (Bolotnyy, 2012; Weicher, 2010).

In an effort to restore confidence in Fannie Mae and Freddie Mac by providing stronger regulation of the GSEs and injecting capital into the two entities, HERA created a new regulator, FHFA, which placed Fannie Mae and Freddie Mac into conservatorship, a legal status similar to Chapter 11 bankruptcy. In addition, HERA transferred the authority to establish, monitor, and enforce the GSEs' annual affordable housing goals from HUD to FHFA.

The Affordable Housing Goals and Underserved Areas

The affordable housing goals were originally formulated to foster one of the public purposes of the GSEs: to provide ongoing assistance to the secondary market for conforming home mortgage loans, in particular those for low- and moderate-income families (Weicher, 2010).⁴ The goals, which were specified in terms of total units financed by GSE purchases, addressed three segments of the mortgage market: (1) low- and moderate-income families; (2) borrowers in geographically targeted underserved areas; and (3) very low-income families and low-income families in low-income areas.⁵ Defining the goals and establishing numerical targets for each goal involved the complex task of (1) determining who the target population was, both in terms of income and location; and (2) setting the targets with reference to the performance and effort of the GSEs toward achieving the targets in previous years.⁶

⁴ Fannie Mae and Freddie Mac are restricted by law to purchasing single-family mortgages with origination balances that are less than a specific amount, known as the conforming loan limit. Loans that are more than this limit are known as jumbo loans. The national conforming loan limit for mortgages for the purchase of single-family, one-unit properties increased from \$203,150 in 1993 to \$417,000 in the 2006-to-2014 period. Since 2008, legislation increased the loan limits in certain high-cost areas in the United States (<http://www.fhfa.gov/DataTools/Downloads/Pages/Conforming-Loan-Limits.aspx>).

⁵ The goals were statutorily specified as follows: (1) The Low- and Moderate-Income Housing Goal: loans to borrowers with incomes at or below the median income for the market area in which they live; (2) The Special Affordable Goal: loans to very low-income borrowers (those with incomes at or below 60 percent of the Area Median Income [AMI]), or to low-income borrowers living in low-income areas (borrowers with incomes at or below 80 percent of the AMI, living in census tracts in which the Median Family Income is at or below 80 percent of the AMI); and (3) The Underserved Areas Goal: loans to borrowers living in low-income census tracts (tracts in which the median income of residents is at or below 90 percent of the AMI) or high-minority tracts (tracts in which minorities comprise at least 30 percent of residents, and the median income of residents in the tract does not exceed 120 percent of the AMI).

⁶ Weicher (2010) discussed this process at length and pointed out that the low-income category is not uniform but varies based on the median income in different metropolitan areas and nonmetropolitan counties.

The designation of affordable housing goals changed periodically based on the performance and efforts of the GSEs toward achieving the targets in previous years.⁷

The affordable housing goals were never established in terms of race or ethnicity of the borrower. With the exception of Goal 3, for which a minority presence in geographically targeted areas is mentioned, the affordable housing goals have continued to be based predominantly on economic factors. Further, where the goals explicitly address minority status, they do not distinguish among the racial and ethnic groups that make up minority neighborhoods.

The definition of geographically targeted areas has also changed considerably since the GSE Act of 1992⁸ because of changes in the criteria and data used for the various definitions.⁹ Until HERA, the geographic areas targeted by the GSEs covered a very large portion of the country,¹⁰ leading to the question of whether such a broad definition of geographically targeted areas could actually be effective in addressing the needs of specific neighborhoods characterized by a consistently limited access to credit. With the adoption of different criteria by FHFA, the newly designated underserved areas occupy a much narrower portion of the nation's territory.¹¹

⁷ HERA modified the housing goals as follows. Goal 1: A low-income home purchase goal for home purchase mortgages to families with incomes of no greater than 80 percent of the Area Median Income (AMI). Goal 2: A very low-income home purchase goal for home purchase mortgages to families with incomes of no greater than 50 percent of AMI. Goal 3: A low-income area home purchase subgoal for mortgages to families living in census tracts with tract incomes of no greater than 80 percent of AMI or to families with incomes of no greater than 100 percent of AMI who live in census tracts with a minority population of 30 percent or more and a tract median income of less than 100 percent of AMI.

⁸ The legislation provided operational definitions for a 3-year transition period (1993 through 1995) during which underserved areas consisted of central cities, as defined by the Office of Management and Budget. HUD replaced the interim definition in 1996 based on research that demonstrated that low-income and high-minority census tracts have high mortgage denial rates and low mortgage origination rates. The new designation defined underserved areas as follows. (1) Within metropolitan areas: census tracts with a median income of less than or equal to 90 percent of Area Median Family Income (AMFI) or a minority population of more than or equal to 30 percent and a median income of less than or equal to 120 percent of AMFI. (2) Within rural areas: counties with a median income of less than or equal to 95 percent of the greater of statewide nonmetropolitan or national nonmetropolitan median income, or counties with a minority population of more than or equal to 30 percent and a median income of less than or equal to 120 percent of the statewide nonmetropolitan median income (appendix A).

⁹ The criteria used for the various definitions have largely been based on data provided by the U.S. Bureau of the Census. These criteria include information on race, ethnicity, and income from the decennial censuses and, more recently, from the American Community Survey (ACS). Census 1990 data provided the basis for the designation of underserved tracts from 1996 to 2004 and census 2000 data were used for the same purpose in subsequent years until 2011. With the releases of the 2010 census and the annual ACS, the definition and geographic distribution of underserved tracts have varied since 2012. In particular, the base data for the identification of low-income census tracts have been based on the annual releases of the 5-year ACS, yielding a more current designation of low-income areas. Because census tract boundaries have changed over time, the various definitions of targeted areas reflect these changes as well. In particular, census 1990 boundaries were used for the definitions provided from 1996 to 2002. Starting in 2003, census 2000 boundaries were used until 2011, whereas census 2010 boundaries have been reported since 2012 (appendix A).

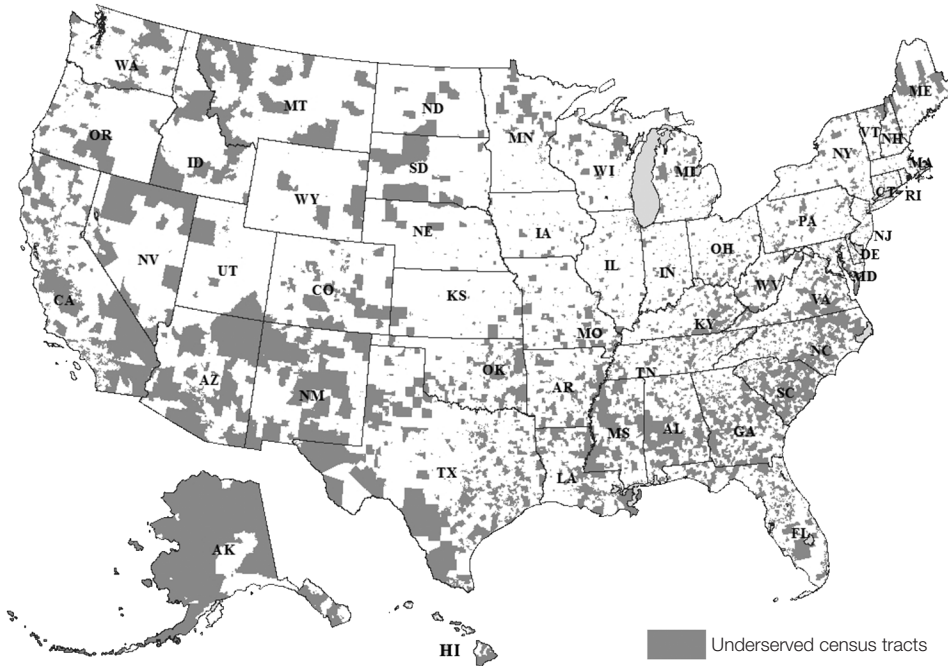
¹⁰ From 1996 to 2004, nearly one-half of all census tracts were designated as underserved (48 percent). The percentage increased to 52 percent from 2005 to 2009.

¹¹ The percentage of targeted tracts dropped in 2010 (36 percent). HERA redefined "underserved areas" and the new geographic targets for the purchase of single-family owner-occupied homes, which came into effect in 2010, now include the following. (1) Census tracts or block numbering areas in which the median income does not exceed 80 percent of the Area Median Income (AMI). (2) Census tracts with a minority population of at least 30 percent and a median income of less than 100 percent of the AMI. (3) Designated disaster areas. See <https://www.fhfa.gov/SupervisionRegulation/Rules/Pages/2009-Enterprise-Transition-Affordable-Housing-Goals-Mortgage-Market-Assessment-Final-Rule.aspx>.

As exhibit 1 illustrates, geographically targeted areas are unevenly distributed throughout the country and most seem to be clustered predominantly in the South and West.

Exhibit 1

2014 Underserved Areas



Source: Federal Housing Finance Agency 2014 low-income areas file

Literature Review

Numerous studies have examined the performance of Fannie Mae and Freddie Mac in relation to the affordable housing goals.¹² A few have specifically focused on the geographically targeted goal. Early research found that the GSEs tended to purchase loans on homes located in low- and moderate-income tracts at a rate that was, in general, less than that of the industry as a whole (Bunce and Scheessele, 1996; Lind, 1996). Gyourko and Hu (1999) found that GSE-purchased loans tended to be overrepresented in higher income census tracts characterized by higher ownership rates.¹³

More recent research has similarly found that, although Fannie Mae and Freddie Mac are meeting their housing goals, they tend to purchase loans in underserved areas with higher median incomes compared with other geographically targeted areas (Case, Gillen, and Wachter, 2002; Williams,

¹² See, for instance, Bunce (2002); Canner, Passmore, and Surrence (1996); Case, Gillen, and Wachter (2002); Weicher (2010); Williams (2006); Williams and Bond (2002); and Williams, McConnell, and Nesiba (2001).

¹³ Gyourko and Hu (1999) noted that expected default costs probably are less in neighborhoods with more, rather than fewer, homeowners.

McConnell, and Nesiba, 2001). GSE purchase shares tend to be lower in central cities, lower in tracts with the highest minority concentrations, and lower in tracts with high vacancy rates (An and Bostic, 2008). An et al. (2007) showed that loan-purchase activity declined with tract median income and purchase activity decreased as minority share increased. Williams similarly found that the GSEs tended to serve underserved markets where borrowers have higher incomes, are less likely to be minorities, and are more likely to live in higher income neighborhoods and metropolitan statistical areas (MSAs; Williams, 2006).

Bhutta (2009) suggested that the GSEs' lending activity is not reaching the lower income neighborhoods within the designated underserved areas because it is more costly to do so and so is expected to have lower returns. Ambrose, Thibodeau, and Temkin (2002) similarly suggested that the GSEs were seeking to mitigate risk in underserved areas by purchasing loans from higher income borrowers located in underserved areas. They also found that GSE minority purchases are concentrated outside underserved areas.

In general, these pre-HERA studies show that the GSEs' lending activity led to limited improvements in housing market conditions in targeted neighborhoods along such indicators as homeownership rates, housing stock appreciation, and vacancy rates. Gabriel and Rosenthal (2007), for instance, examined the degree to which the GSEs focused more intensively on underserved markets and the extent to which their purchase activity served to crowd out private-sector loan purchases.¹⁴ They found that gains in liquidity resulting from the GSEs' activities in these areas were partly offset by losses in the nonconforming sector and that GSE loan purchases were crowding out purchases by unsubsidized, private secondary-market intermediaries. They suggested that the increased activity in the conforming sector may come at the expense of lending activity in the nonconforming sector or may crowd out other unsubsidized, private secondary-market intermediaries. Bostic and Gabriel (2006) similarly found that GSE-targeted tracts on average did not show statistically significant improvements in housing market conditions, suggesting that the affordable housing goals were doing little to improve local homeownership rates or improve local housing conditions.

Research also stresses the importance of geographic variations: although patterns, in general, are consistent across metropolitan areas overall, variation in behavior exists across areas.¹⁵ Pearce's study of GSE purchases in 10 high- and low-cost metropolitan areas, for instance, indicated that mortgage lending activity in low-income tracts is less than in middle-income tracts more often in low-cost MSAs than in high-cost MSAs (Pearce, 2001). Williams (2006) observed that the affordable housing goals set standards for the GSEs' nationwide performance and failed to recognize important variations across regions based on housing market characteristics and costs.

A few studies have questioned the effectiveness of the official designations of underserved areas and have suggested that refining the approach to designating which census tracts are to be considered underserved may have some benefit (Pearce, 2001). In general, these studies argue that, although low-income and high-minority representation, on average, is highly correlated with low levels of

¹⁴ See also Gabriel and Rosenthal (2010, 2008).

¹⁵ See, for instance, Boxall and Silver (2001); Case, Gillen, and Wachter (2002); Gyourko and Hu (1999); Lind (1996); MacDonald (2001). Gyourko and Hu, in particular, found that in the Boston, New York, and Los Angeles markets, a tract is more likely to be underrepresented the higher its income is relative to the area median.

mortgage activity, the low levels may not be sufficient to identify underserved neighborhoods, especially across different markets. Such proxies may not account for factors responsible for variations in mortgage availability in low-income neighborhoods across states or metropolitan areas. The GSEs may be meeting the housing goals by purchasing loans from low-income areas that have mortgage origination rates comparable with those of middle-income neighborhoods such as, for example, low-income areas characterized by a large presence of non-Hispanic Whites. Moreover, if a neighborhood does not have mortgage credit, it does not necessarily mean that it is underserved, and a neighborhood can be considered underserved only if both mortgage credit is absent *and* demand exists for that credit. McClure (2001) argues that directing credit to underserved areas is helpful only insofar as it helps to direct credit to neighborhoods that are marginally less desirable than the neighborhoods deemed to be well served. To the author's knowledge, however, no studies have specifically suggested incorporating race and ethnicity as a critical criterion in the designation of underserved areas.

Methods and Data

This study covers the periods immediately preceding and following HERA legislation to understand the trajectory of lending in underserved markets. The analysis focuses both on officially designated underserved markets (Final Rule underserved markets)—whose definition is predominantly based on income—and on markets that we define in terms of race and ethnicity. Given the continuing importance of race in mortgage lending, we specifically employ a definition of race-based underserved markets by making a distinction among the major minority groups that make up the underserved market. The analysis makes a distinction among the segments of these markets that overlap with each other to highlight trends specific to geography and market overlaps. Indeed, significant overlap exists among the various underserved markets. For example, any very low-income family is also a member of the low-income market. Low-income families may reside in geographically targeted areas. Minority borrowers may not necessarily reside in geographically targeted areas. In summary, the study analyzes the following underserved markets.

We first analyze the following four Final Rule underserved markets.

1. Low-income market alone (Goal 1).
2. Very low-income market alone (Goal 2).
3. Geographically targeted low-income market alone (Goal 3).
4. Overlapping Final Rule underserved markets: any of the previous markets listed simultaneously.

We then consider the following three underserved markets, which, when contrasted with the official designation, are based on disaggregated racial and ethnic groups. This consideration is important to identify any underserved markets that are not currently captured by income-based official designations.

1. African-American borrowers.
2. Asian and Pacific Islander (API) borrowers.
3. Hispanic borrowers.

In addition, part of the analysis includes the following two served areas for comparison purposes.

1. Balance of geographically targeted areas: moderate- and high-income borrowers purchasing in geographically targeted areas.
2. Market outside geographically targeted areas: moderate- and high-income borrowers purchasing outside geographically targeted areas.

The analysis is divided into three parts. First, the study compares GSE purchases of underserved market loans with the proportion of those loans held in the primary market; that is, the lenders who make the loans in the first place. Comparisons are made throughout 8 years, including those immediately preceding HERA (2004 through 2007) and the 4 years of post-HERA economic recovery (2010 through 2013). Looking at data over time helps gauge whether changes in the designations of underserved markets and the affordable housing goals brought about by HERA may be associated with any improvements of GSE performance in these markets. In addition, by looking at changes in the composition of GSE purchases, it is possible to understand whether the GSEs are leading the market or simply reflecting the primary market. For instance, if the changes observed in GSE purchases follow similar changes in primary-market lending, then the GSEs are likely just reflecting the market. If increases in GSE purchases from underserved markets are followed by increased primary-market lending to those groups, then the GSEs are likely leading the market. Second, the study evaluates GSE performance by examining whether they serve all members of underserved markets equally or benefit primarily the least underserved of the underserved, consistent with the literature discussed previously. Third, to examine the determinants of GSE purchases in underserved areas, a set of logistic regressions—described in more detail in the results section of this article—is performed for the study years.

Data used for the analyses come from several sources.

1. The annual lists of geographically targeted areas provided by HUD and FHFA. These lists include the geographic identifiers of census tracts that can be merged with boundary files provided by the Census Bureau (2000 and 2010) to represent the targeted areas on maps with the use of Geographic Information Systems, or GISs.
2. HUD- and FHFA-provided information on conforming loan limits and annual housing goals.
3. National HMDA loan application registers from 2004 to 2007 and from 2010 to 2013. The series has consistent variable definitions (for example, racial and ethnic borrower characteristics and high-cost loans).¹⁶ Coded to the census tract level, the data allow more detailed exploration

¹⁶ HMDA data, made available by the Federal Financial Institutions Examination Council, provide information on mortgage loan applications received annually by depository and nondepository institutions. Institutions report information about each application or loan—loan type, purpose, occupancy, amount, and action taken—and about the census tract of the dwelling to which it relates. Only applicants for whom racial/ethnic background is reported and applications that are complete are included in the analysis. HMDA records with edit failures are omitted from the analysis.

of the factors influencing GSE performance compared with aggregated data. The analysis relies on HMDA data and not on the Public Use GSE datasets, because the construction of the latter greatly limits their usefulness for the sort of analysis undertaken here.¹⁷

4. The American Community Survey (ACS) 2008/2012 (5-year average) summary file by census tract.

The study focuses on the applications for conventional conforming loans for the purchase of one- to four-family owner-occupied units that resulted in either originations or denials. Federal Housing Administration (FHA) and U.S. Department of Veterans Affairs, or VA, loans were excluded. Jumbo loans, high-cost loans, and records with high loan-to-income ratios (6 or more) are excluded. Also excluded are cases with missing data on either applicant income or loan amount.

Lending Trends in Underserved Conventional Markets

Exhibit 2 compares the underserved market performance of the GSEs with that of the primary market in the years before and following the HERA Act. For any given year, the numbers in the table indicate the percentage of loans made to or purchased from a particular underserved market. Notwithstanding some short-term fluctuations occurring from year to year, private-market lending in underserved markets increased throughout the study period while purchases by the GSEs tended to decrease. Consistent with previous studies (Williams, 2006), GSE-purchased loans in underserved markets continued to be underrepresented compared with the percentage of such loans made in the primary market (31 versus 37 percent in the 2010-to-2013 period).¹⁸ The performance of the GSEs has distinct variations, however, both regarding the primary market and in trends over time, as the figures disaggregated by segments of the underserved market show. The share of the GSEs' purchases in low-income markets and geographically targeted areas decreased from the pre-HERA period to post-HERA years, whereas the share of purchases in very low-income markets increased. The share in overlapping markets remained the same. By contrast, primary-market lending increased in all but the geographically targeted areas.

Data related to the race-based markets show that the percentage of all loans going to borrowers of color decreased from 20.7 percent in the pre-HERA period to 19.4 percent in post-HERA years. The GSEs seem to lead the market by a small margin in the 2010-to-2013 period. Based on the race and national origin of borrowers, however, important variations are present. Despite a decrease in their share of loans in markets of color, primary-market lenders led the African-American and Hispanic markets in both periods, whereas the GSEs led the API markets throughout the years, with a substantial increase in their share of loans in the post-HERA period.

¹⁷ For proprietary reasons, the GSE datasets are divided into three unlinkable datasets. The census tract file does not distinguish between home-purchase and refinance loans. Also, race and national origin of applicant and coapplicant have no distinction. Previous research, however, finds that the HMDA data provide very good estimates of GSE activity in underserved markets (Williams, McConnell, and Nesiba, 2001).

¹⁸ Williams suggests that the GSEs have never been "leading the market," explaining that "underserved market loans that others were willing to buy or hold in portfolio were loans that the GSEs were either unwilling or unable to purchase" (Williams, 2006: vii).

Exhibit 2

Trends in Underserved Markets

Underserved Market	Type of Lender/ Purchaser	Percent per Year								Percent per Period	
		Pre-HERA				Post-HERA				2004–2007	2010–2013
		2004	2005	2006	2007	2010	2011	2012	2013		
Final Rule Underserved Markets											
Goal 1—Low-income borrowers (alone)	Primary Market	13.2	16.2	15.7	13.5	15.8	16.0	15.2	13.6	14.6	15.2
	GSE	16.0	15.2	13.9	14.0	13.4	14.4	13.8	12.6	14.8	13.6
Goal 2—Very low-income borrowers (alone)	Primary Market	3.8	4.5	4.1	3.3	6.0	6.9	5.8	4.6	3.9	5.8
	GSE	4.5	4.2	3.6	3.4	4.7	5.4	4.7	3.6	3.9	4.6
Goal 3—Geographically targeted low-income market (alone)	Primary Market	8.6	8.5	9.4	9.9	7.2	6.5	7.5	8.1	9.1	7.3
	GSE	7.2	7.3	7.7	8.3	7.1	5.5	6.2	7.1	7.6	6.5
Goals 1–3 (overlap)	Primary Market	7.4	8.4	8.9	8.2	9.4	7.5	9.8	8.9	8.2	8.9
	GSE	7.0	6.1	5.8	6.5	7.2	5.1	6.5	6.3	6.3	6.3
All	Primary Market	33.0	37.5	38.1	34.9	38.4	36.8	38.3	35.2	35.9	37.2
	GSE	34.7	32.8	31.0	32.2	32.3	30.5	31.2	29.7	32.7	30.9
Race-Based Markets											
African-American borrowers	Primary Market	5.8	5.9	7.4	7.4	3.9	4.0	3.9	3.9	6.6	3.9
	GSE	4.7	4.1	4.6	5.7	1.5	2.0	1.9	2.2	4.8	1.9
Asian/Pacific-Islander borrowers	Primary Market	7.6	5.1	5.3	5.9	12.8	7.9	7.8	8.2	6.0	9.1
	GSE	7.1	6.3	6.2	6.2	18.3	10.8	9.9	9.8	6.4	12.2
Hispanic borrowers	Primary Market	10.5	9.2	10.6	11.2	5.7	5.9	5.9	5.9	10.4	5.8
	GSE	9.1	7.4	7.6	8.8	6.1	5.1	5.0	5.8	8.2	5.5
All	Primary Market	23.9	20.3	23.2	24.5	22.4	17.8	17.6	17.9	23.0	18.9
	GSE	20.9	17.8	18.4	20.6	25.9	17.9	16.8	17.9	19.4	19.6

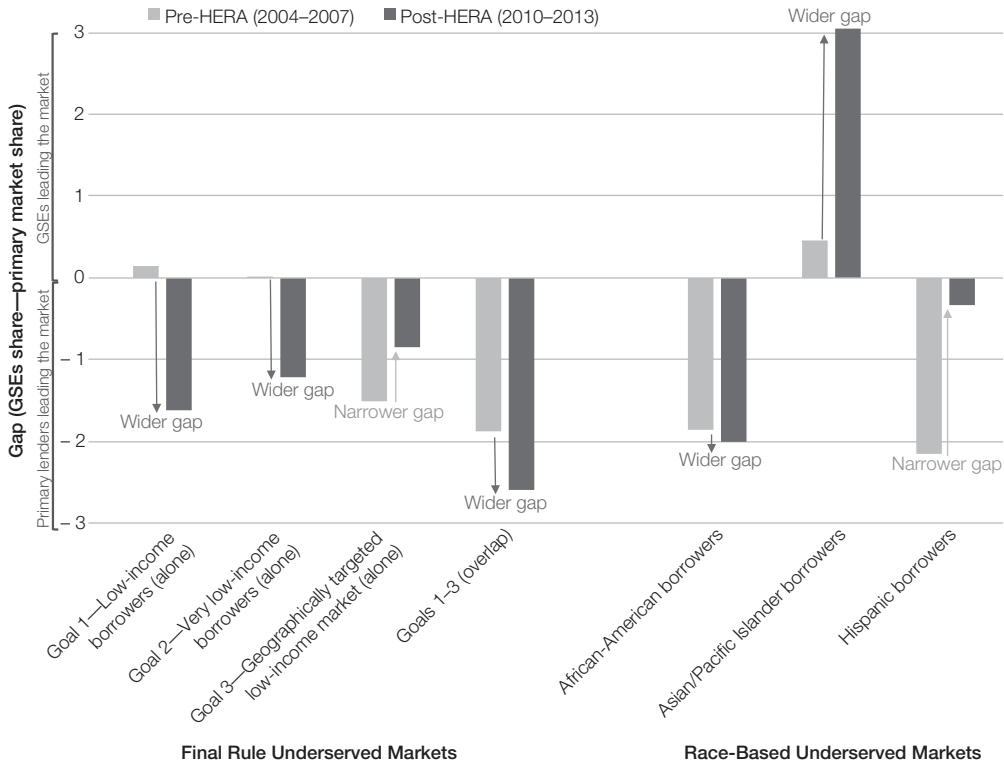
GSE = government-sponsored enterprise. HERA = Housing and Economic Recovery Act of 2008.

Exhibit 3 provides a graphic summary of the findings discussed previously and shows that the GSEs have trailed the primary market as a whole in both periods, with the exception of the pre-HERA period when the GSEs led the low-income market. For every Final Rule underserved market, except geographically targeted areas alone, the gap between the GSEs' performance and that of the primary market has widened over time. The gap increased especially in the low-income and very low-income markets. The panel on the right side of exhibit 3 illustrates that the gap between the GSEs and the primary market has increased by a large margin in API markets, whereas it has narrowed substantially in Hispanic markets, although, in the latter, primary lenders still lead the market, as in the case of African-American borrowers.

Existing studies of GSE performance suggest that, even in underserved markets, the GSEs tend to serve the least underserved. To explore whether this trend is still the case, exhibit 4 illustrates mean incomes of borrowers by underserved market segment for the pre- and post-HERA periods. Data for the 4 years comprising each period are adjusted for inflation and pooled to facilitate interpretation and minimize the noise represented by yearly fluctuations. Mean incomes associated with

Exhibit 3

Gaps Between GSEs and Primary Market in Underserved Market Shares Pre- and Post-HERA Periods



GSE = government-sponsored enterprise. HERA = Housing and Economic Recovery Act of 2008.

primary-market loans and GSEs’ purchases are specifically compared with each other and with those of all applicants eligible for a loan in each underserved market segment. For example, the top rows of the table show that in both periods the mean income of all applicants for conventional conforming loans in the low-income bracket of the market (Goal 1) was \$47,400. The mean income of borrowers of loans purchased by the GSEs was more than that of both all applicants and those with loans held in the primary market both in the pre- and post-HERA periods, with a slight increase in the latter years. Further, while in the first period the primary market seemed to favor those with higher average incomes, in the second period it clearly tended to hold loans made to borrowers with an average income that is less than that of all applicants in the low-income segment of the market.

Overall, the figures presented in exhibit 4 suggest that the GSEs have tended to mirror the primary market by favoring loans associated with borrowers with higher average incomes in all segments of the Final Rule underserved market, except the geographically targeted low-income market alone (Goal 3), where average incomes associated with GSE purchases are slightly less than those of all applicants in the same market. It is important to note that average incomes in the geographically

Exhibit 4

Borrower Income by Underserved Market and Type of Lender/Purchaser, Pre- and Post-HERA Periods

Underserved Market	Type of Lender/ Purchaser	Mean Income (2013 thousands of dollars)	
		Pre-HERA (2004–2007)	Post-HERA (2010–2013)
Final Rule Underserved Markets			
Goal 1—Low-income borrowers (alone)	Primary Market	48.2	47.1
	GSEs	48.3	48.4
Goal 2—Very low-income borrowers (alone)	Primary Market	29.8	29.1
	GSEs	30.0	29.7
Goal 3—Geographically targeted low-income market (alone)	Primary Market	107.4	110.2
	GSEs	100.0	103.9
Goals 1–3 (overlap)	Primary Market	43.6	41.2
	GSEs	43.6	43.1
Race-Based Markets			
African-American borrowers	Primary Market	83.6	75.8
	GSEs	83.0	96.3
Asian/Pacific Islander borrowers	Primary Market	120.4	114.9
	GSEs	105.8	110.0
Hispanic borrowers	Primary Market	88.5	80.1
	GSEs	84.6	87.1

GSE = government-sponsored enterprise. HERA = Housing and Economic Recovery Act of 2008.

targeted segment of the market have experienced substantial variations throughout the study period and especially in the most recent years, when they, in general, have decreased, most likely as a result of the economic downturn.¹⁹

Data related to the race-based underserved markets show that, in the post-HERA period, the GSEs have clearly favored loans associated with borrowers of color with higher average incomes. It is interesting to note that, although average incomes associated with loans held in the primary market have tended to decrease across all markets of color between the two periods, average incomes associated with GSE purchases, in general, have climbed across the board. The gaps between the GSEs and the primary market regarding average incomes of borrowers have substantially increased in the case of African-Americans (from \$500 in the pre-HERA period to \$20,500 in the second period) and Hispanic borrowers (from \$3,900 to \$7,000), whereas the gap has narrowed in the case of API borrowers (from \$14,600 to \$4,900).

To understand whether the intersection between borrower race and ethnicity and each segment of the Final Rule underserved market contributes to the variations in market shares described previously, exhibit 5 illustrates mean incomes and percentage distributions of borrowers of color by

¹⁹ Average incomes for all applicants in the geographically targeted low-income market alone decreased from \$111,000 in 2010 to \$98,000 in 2013.

Exhibit 5

Income and Percentage Distribution of Borrowers by Underserved Market Type of Lender/Purchaser and Applicant/Borrower Race and Ethnicity, 2013

Underserved Market	Type of Lender/ Purchaser	Mean Income (thousands of dollars)					Percent of Total					N
		Non- Hispanic White	African- American	API	Hispanic	Total	Non- Hispanic White	African- American	API	Hispanic	Total	
Goal 1—Low-income borrowers (alone)	Primary Market	44.1	45.9	50.8	43.6	44.8	86	3	5	5	33,004	
	GSEs	45.5	47.6	50.6	45.0	46.1	85	2	7	5	69,249	
Goal 2—Very low-income borrowers (alone)	Primary Market	27.5	28.3	30.2	27.9	27.8	85	3	5	6	11,293	
	GSEs	28.4	29.4	31.2	28.2	28.7	85	2	7	5	20,303	
Goal 3—Geographically targeted low-income market (alone)	Primary Market	105.4	84.7	104.8	84.1	103.9	72	7	11	8	19,020	
	GSEs	96.7	90.4	101.6	81.5	96.8	72	4	13	10	38,099	
Goals 1–3 (overlap)	Primary Market	38.6	39.0	43.3	36.1	39.1	60	13	10	16	21,453	
	GSEs	40.6	42.0	45.2	38.3	41.2	65	5	14	15	34,595	
All underserved markets	Primary Market	50.6	48.7	60.7	45.9	51.7	76	6	8	9	84,770	
	GSEs	53.4	57.5	62.4	51.4	55.1	78	3	10	8	162,246	

API = Asian/Pacific Islander. GSE = government-sponsored enterprise.

Final Rule underserved market segment and type of lender/purchaser. The table, which shows data for 2013, also presents income and racial distributions of the sample of applicants selected for this study.²⁰ As the last row of data indicates, the racial and ethnic breakdown of borrowers served in all underserved markets combined reveals that most borrowers with GSE loans (88 percent) are either non-Hispanic Whites or APIs, whereas GSE loans are underrepresented in the African-American and Hispanic markets. Only 3 and 8 percent of GSE-purchased loans serve African-American and Hispanic borrowers, respectively, compared with 6 and 9 percent of loans held in the primary market. Moreover, the table shows that the average income of borrowers whose loans were purchased by the GSEs tends to be higher than that associated with loans held in the primary market and that of the whole applicant pool. The differences between the incomes of borrowers with GSE loans and those with loans held in the primary market are especially high among African-American and Hispanic borrowers—nearly \$9,000 and \$6,000, respectively.

In general, the patterns described previously hold for each segment of the underserved market except the geographically targeted areas (Goal 3). Overall, borrowers in the latter feature much higher incomes than borrowers in the other segments of the market. It is interesting to note that average incomes of non-Hispanic White and API borrowers tend to be less than those of the respective pools of applicants, although these borrowers are clearly overrepresented. This income trend is in sharp contrast with other borrowers of color, especially African-Americans, who tend to be underrepresented among borrowers with GSE loans but feature much higher average incomes—\$90,400 compared with \$85,500 for the African-American pool of this segment of the market.

In summary, the descriptive statistics presented previously suggest that, consistent with earlier research, the GSEs reflect primary-market activity and tend to serve the least underserved; that is, those with higher average incomes. It is most important to note that the GSEs seem to underserve African-American and Hispanic markets by focusing predominantly on non-Hispanic White and API borrowers and the most affluent segments of the African-American and Hispanic markets.

It is possible, however, to estimate the likelihood that mortgage loans in underserved communities of color will be purchased by the GSEs after controlling for some key variables.

Determinants of GSE Purchases in Underserved Markets of Color

Exhibit 6 illustrates a set of logistic regressions modeling the determinants of GSE purchases and compares them with loans held in the primary market in both pre- and post-HERA periods (2004 through 2007 and 2010 through 2013), after controlling for the characteristics of the borrowers, neighborhoods, lenders, and geographic areas. To identify which markets are most in need of the GSEs' attention, the models estimate simultaneously the effects of each of the following underserved market: (1) those defined by the affordable housing goals for single-family home purchase loans; (2) the segment of these markets that overlap with each other; (3) the geographically targeted areas not affected by the affordable housing goals rule; (4) the served markets outside targeted census

²⁰ Equivalent tables for the previous years are available from the author.

Exhibit 6

Logistic Regressions of Determinants of GSE Purchases, Odds Ratios, 2004–2007 and 2010–2013 (1 of 2)

Predictors	Average							Average 2010– 2013		
	2004	2005	2006	2007	2004– 2007	2010	2011		2012	2013
	Odds Ratios									
Goal 1	0.64***	0.85***	0.75***	0.75	0.75	0.83***	0.86***	0.84***	0.85***	0.84
Goal 2	0.54***	0.94***	0.76***	1.35***	0.90	0.75***	0.77***	0.77***	0.75***	0.76
Goal 3	0.84***	0.93***	0.94***	0.98*	0.92	0.95**	0.91**	0.93***	0.92***	0.93
Goals 1–3 (overlap)	0.52***	0.77***	0.73***	1.19***	0.80	0.69***	0.71***	0.71***	0.70***	0.70
Targeted areas—balance	0.95	0.99	1.08	1.02	1.01	0.99	0.99	1.04	0.96	1.00
African-American borrower	0.69***	0.63***	0.63***	0.71***	0.67	0.43**	0.50***	0.52***	0.60**	0.51
Asian/Pacific Islander borrower	1.11**	1.05***	1.09***	0.93***	1.05	1.05***	0.97*	0.93***	0.87***	0.96
Hispanic borrower	0.78***	0.70***	0.70***	0.71***	0.72	0.91***	0.75***	0.76***	0.81***	0.81
Borrower of other race/ethnicity	0.95**	0.91***	0.91***	0.84***	0.90	0.89***	0.75***	0.77***	0.75***	0.79
Loan amount	0.99	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00
Borrower income	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AMI	1.01	1.00	1.00	0.98	1.00	1.00	1.00	1.01	1.00	1.00
Census tract minority population (%)	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00
Census tract median income/AMI ratio	1.47***	1.38***	1.48***	0.89***	1.30	1.00	1.00	1.33***	1.30***	1.16
Owner-occupied units in census tract	1.03	0.98	1.00	1.02	1.01	1.02***	1.04**	1.06***	1.05***	1.04
CRA lender	0.22***	0.25***	0.47***	0.41***	0.34	0.92***	0.43***	0.29***	0.23***	0.47
Lender size (assets)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Midwest	1.04***	1.35***	1.40***	1.45***	1.31	1.85***	1.83***	1.93***	1.84***	1.86
South	0.95***	1.04***	1.11***	1.16***	1.06	1.02*	1.23***	1.20***	1.27***	1.18
West	0.92***	1.13***	1.27***	0.83***	1.04	2.08***	2.03***	2.03***	2.15***	2.07
N	1,160,338	1,019,886	1,009,773	1,039,269	4,229,266	404,538	419,973	580,628	773,596	2,178,735

Exhibit 6

Logistic Regressions of Determinants of GSE Purchases, Odds Ratios, 2004–2007 and 2010–2013 (2 of 2)

Predictors	2004	2005	2006	2007	Average 2004– 2007	2010	2011	2012	2013	Average 2010– 2013
	Odds Ratios									
Pseudo <i>r</i> squared	0.136	0.0763	0.0314	0.1027	0.0479	0.0731	0.0815	0.098		
Log likelihood	-657,943***	-594,725***	-638,055***	-636,346***	-239,749***	-258,003***	-334,279***	-429,047***		

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

AMI = Area Median Income. CRA = Community Reinvestment Act. GSE = government-sponsored enterprise.

tracts; and, most importantly, (5) the markets consisting of borrowers of color. The models include regional indicators to account for variations across different housing markets. The national sample used for the regressions includes all originated conforming loans for the purchase of one- to four-family owner-occupied dwellings. Records for which a loan-to-income ratio is equal to or exceeds 6 are omitted, following Williams, Mcconnell, and Nesiba (2001). Excluded are also second lien loans and records for which the race and income of the applicant is not known. Binomial logistic regression is appropriate for this analysis because it involves estimating equations for a dependent variable with two categories. For each year examined, the logistic regression equation takes the following form—

$$\ln \left[\frac{P(Y = 1|X)}{P(Y = 0|X)} \right] = \sum^k b_k X_k \quad (1)$$

where Y is coded 1 for GSE-purchased loans. Loans held in the primary market represent the reference category ($Y = 0$). Each equation predicts the log odds of a loan being purchased by the GSEs rather than being held in the primary market, after controlling for a vector of k borrower, lender, neighborhood, and geographic characteristics, X_k , each associated with a specific coefficient b . The odds are estimated for each underserved market. Applicant characteristics include annual family income, loan amount, and race and ethnicity. Income and loan amount are used as proxies of risk of default in the absence of any credit history information in HMDA datasets. Non-Hispanic White applicants represent the reference racial group. Neighborhood characteristics include the percentage of minority residents, the number of owner-occupied units in the census tract, and the ratio of the census tract median family income to the Area Median Family Income, or AMFI. The independent variables include lender size in terms of assets and an indicator coded 1 for lenders subject to Community Reinvestment Act (CRA) regulations. Variations in local housing markets are captured through the inclusion of the Area Median Income, or AMI, associated with each loan and the geographic region in which the loan was made.²¹ Loans made in the Northeast represent the reference category.

Regression results indicate that the likelihood of loans being purchased by the GSEs is significantly less that of a loan being held in primary-market portfolios in the underserved markets defined by Goal 1 and Goal 2 and in the markets characterized by an overlap of the affordable housing goals in both periods, after controlling for all other variables.²² The likelihood of loans being purchased by the GSEs is also significantly less in the underserved markets defined by borrower race, especially in African-American markets and, to a lesser extent, in Hispanic markets. Further, the odds ratios associated with contextual variables and lender characteristics indicate that, in general, the likelihood of loans being purchased by the GSEs is significantly less in the case of lenders subject to CRA regulations, whereas it tends to be more in more affluent census tracts, in the Midwest, and in the Western region.

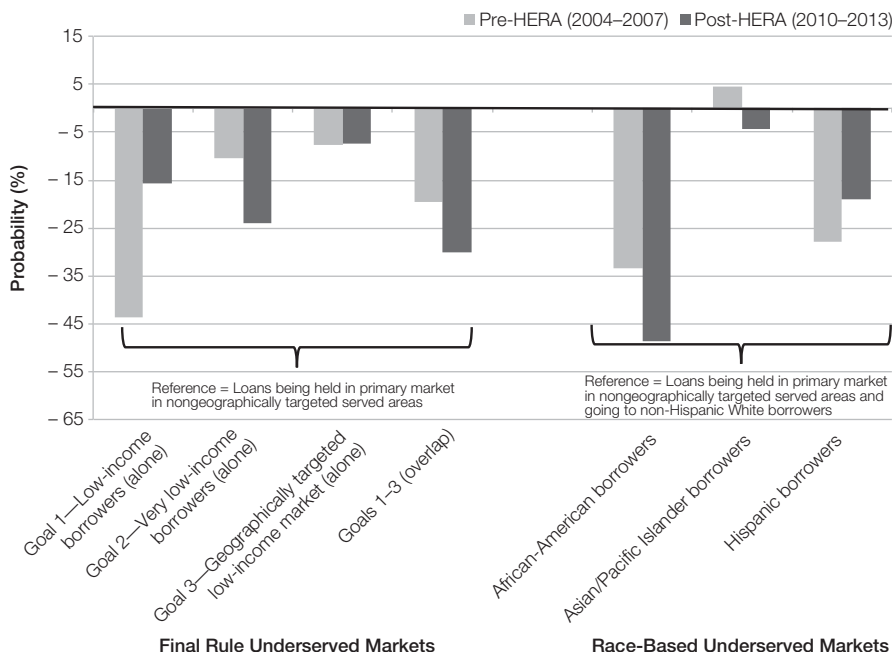
To facilitate the interpretation of the logit model results, exhibit 7 summarizes the regression results by focusing on the net probabilities of GSE purchases in both the Final Rule underserved

²¹ Note that geographically targeted areas tend to be concentrated in the South and the West.

²² Note that the p-values are very low for most predictors, possibly due to the large sample size of the data used for this analysis. The p-values have been omitted for variables presenting odds ratios equal to 1.

Exhibit 7

Conditional Probabilities of GSE Purchases Versus Probability of Loans Being Held in Primary Market



GSE = government-sponsored enterprise. HERA = Housing and Economic Recovery Act of 2008.

market segments and the race-based underserved markets in the pre- and post-HERA years. The graph specifically compares average probabilities for the two periods. After controlling for all other variables, the loans purchased by the GSEs clearly are less likely than the loans held in the primary market to serve the underserved markets designated in the affordable housing goals rule in both periods, despite some variations across the different segments of the market. In the pre-HERA era, for instance, the average likelihood of a loan in a low-income market being purchased by the GSEs was 44 percent less than a loan in primary-market portfolios. Since 2010, however, the GSEs have performed better in this market, with an average probability of serving this market just 16 percent less than that of primary-market lenders. The GSEs similarly have performed slightly better in geographically targeted markets alone. By contrast, in very low-income markets and in overlapping markets, the likelihood of the GSEs purchasing loans has decreased—from 10 to 24 percent and from 20 to 30 percent, respectively, less than the likelihood of loans being held in the primary market.

In race-specific markets, the performance of the GSEs has been generally higher in API markets compared with other markets of color, although the probability of the GSEs purchasing loans in this market has decreased over time. The likelihood of the GSEs purchasing loans issued to African-American and Hispanic borrowers, in general, has been much less than in the case of API borrowers. It is interesting to note, however, that while the GSEs' performance has improved over time in Hispanic markets, it has worsened further in African-American markets, after controlling

for all other variables, including income levels. While the likelihood of GSE purchases in African-American markets was 33 percent less than the probability of loans being held in the primary market in the pre-HERA period, the GSEs were, on average, 49 percent less likely than the primary market to hold loans issued to African-American borrowers in the post-HERA period.

Discussion

This study shows that the GSEs' performance in underserved markets has continued to trail behind that of the primary market in the post-HERA period. Further, it shows that the GSEs extend credit largely to non-Hispanic White borrowers, API borrowers, and other minority borrowers with higher income levels. For every Final Rule underserved market except the geographically targeted market alone, the gap between the GSEs' performance and that of the primary market has widened over time. This gap increased especially in the low-income and very low-income segments of the market. Further, primary-market lenders still lead the African-American and Hispanic markets, although, in the latter, the gap between the GSEs' performance and that of the primary market has narrowed. The GSEs have continued to lead the API market, especially after HERA. Moreover, the GSEs have tended to mirror the primary market by favoring loans associated with more affluent borrowers of any race and ethnicity. This behavior is particularly clear in low-income and very low-income markets and in overlapping markets. Borrowers of color tend to be underrepresented in the underserved markets targeted by the GSEs, except in the case of APIs, who are particularly overrepresented in geographically targeted areas. Although in geographically targeted areas average incomes are generally higher than in other segments of the underserved market—most likely driven by designation criteria—it is important to note that in these areas the GSEs clearly favor African-American borrowers with average incomes that are much higher than those served by the primary market. Even after controlling for several borrower, neighborhood, and lender characteristics, the GSEs' performance in Hispanic and, especially, in African-American markets lags behind that of the primary market.

In summary, although the new designations of underserved markets seem to do a better job in pinpointing low-income and minority communities compared with the early broader definitions, especially from a geographic perspective, they do not guarantee that the GSEs serve the most underserved of the underserved, especially when it comes to communities of color. This finding may have several explanations that unfortunately cannot be fully spelled out with currently available HMDA data. HMDA data do not report any information on credit history, net worth, and total indebtedness, which differ systematically among racial groups and may play a role in the GSEs' decision to purchase loans from groups that are perceived as risky, especially in light of the recent controversial issues related to the GSEs' solvency. The economic recession and stricter underwriting criteria have made loans more difficult to access by communities of color that have been disproportionately hit by the subprime lending and foreclosure crisis. The primary market and the GSEs may refrain from taking risks in these communities. Further, FHA loans may be perceived as the channel of choice by borrowers of color and seem to compete with the conventional market in these communities. Some have also argued that lenders subject to CRA—commercial banks, savings and loans—deliberately hold in portfolio loans to low-income and minority communities that are likely to make them look good from a CRA point of view, thus reducing the likelihood

that these loans are sold to the GSEs (Williams, 2006). Additional explanations may attribute the observed disparities to differences among the lending institutions involved based on their size, scale, legal obligations, the federal agencies to which they report, and the geographic areas in which they operate. It is also important not to discount racial discrimination in mortgage lending. What may be perceived as the GSEs' neglect of communities of color may well be a reflection of the disparate treatment that these borrowers may receive in the primary market. Notwithstanding the frequent overlap between low-income markets and race-based underserved markets, the GSEs are held accountable only formally on the basis of their performance in markets defined predominantly by their income level and only tangentially by their racial and ethnic composition. Yet, the GSEs' performance seems to fare worse in the very markets that are characterized by high concentrations of people of color.

Limitations in available data do not allow for a thorough testing of all the previously mentioned hypotheses, especially when it comes to the incorporation of credit history data in any analysis.²³ The findings presented here, however, support the argument that basing the designation of underserved markets solely on economic factors may lead the GSEs to miss an important segment of the underserved market that has been historically excluded from broad access to mortgage credit. The most direct way that the GSEs can affect home mortgage lending is through the loans they purchase. They can also have indirect effects on lending, because GSE activity in an area may encourage more lenders to be active there (Williams, 2006). Therefore, a shift of focus to race and ethnicity in the GSEs' practices may have the potential to influence the mortgage market at large. This focus is particularly crucial in the wake of the foreclosure crisis. The evidence in this paper suggests that policymakers should revisit the criteria currently adopted for the designation of underserved markets, in particular by incorporating race and ethnicity in the formulation of affordable housing goals, revising the designation of geographically targeted areas, and establishing subgoals that are specific to geography to promote the GSEs' outreach in communities of color that still lag behind the mainstream market in terms of their access to mortgage capital.

In addition, it is important to fine-tune race-based designations of underserved markets based on a disaggregation of borrowers of color by race and ethnicity. This redesignation is particularly critical in API markets that are characterized by a significantly heterogeneous population in terms of national origins, immigrant status, and socioeconomic characteristics. The findings of this study, as they apply to API markets, may be misleading because of the lack of disaggregated data for these markets. It is worth noting that aggregated data generally have shown that API borrowers have experiences that are similar to those of non-Hispanic White borrowers. Yet, HMDA data have failed to reveal disparities in access to mortgage lending experienced by different groups of Asian descent. It would be very useful for HMDA data to break down information related to APIs into subcategories that represent the largest ethnic groups in the nation to better understand the mortgage lending experience of these groups, both in the primary and secondary markets.

By incorporating race and ethnicity in the formulation of affordable housing goals and fine-tuning the designations of underserved markets, the benefits the GSEs receive in exchange for promoting underserved market lending might be put to some more effective use.

²³ Limitations are notably attributed to HMDA's absence of data on borrowers' credit history and default risk.

Appendix

Exhibit A-1

Definitions of Underserved Areas and Affordable Housing Goals (1993–2014)

Year	Underserved Area Definition	Tract Boundary (Census Year)	Data on Which the Definition Is Based	Affordable Housing Goals		
				Geographically Targeted (Under-served) Area Goal	Low- and Moderate-Income Goal	Special Affordable Goal
1993	Central cities			30	30	NA
1994				30	30	NA
1995				30	30	NA
1996	Metropolitan census tracts with (1) tract MFI of less than or equal to 90 percent of AMI or (2) minority concentration of at least 30 percent and tract MFI of less than or equal to 120 percent of AMI.	1990	Census 1990	21	40	12
1997		1990	Census 1990	24	42	14
1998		1990	Census 1990	24	42	14
1999		1990	Census 1990	24	42	14
2000		1990	Census 1990	24	42	14
2001		1990	Census 1990	31	50	20
2002	Nonmetropolitan counties with (1) MFI of less than or equal to 95 percent of the greater of state or national nonmetropolitan median income or (2) minority concentration of at least 30 percent and county MFI of less than or equal to 120 percent of the greater of state or national nonmetropolitan median income (HUD 1996–2008; FHFA 2009).	1990	Census 1990	31	50	20
2003		2000	Census 1990	31	50	20
2004		2000	Census 1990	31	50	20
2005		2000	Census 2000	37	52	22
2006		2000	Census 2000	38	53	23
2007		2000	Census 2000	38	55	25
2008		2000	Census 2000	39	56	27
2009		2000	Census 2000	32	43	18
2010	Census tracts or block numbering areas in which the median income does not exceed 80 percent of the AMI; census tracts with a minority population of at least 30 percent and a median income of less than 100 percent of the AMI; designated disaster areas (FHFA 2010–2014).	2000	Census 2000	13	27	8
2011		2000	Census 2000	13	27	8
2012		2010	Census 2010	11	23	7
2013		2010	Census 2010, ACS 2007–2011	11	23	7
2014		2010	Census 2010, ACS 2008–2012	11	23	7

ACS = American Community Survey. AMI = Area Median Income. FHFA = Federal Housing Finance Agency. HUD = U.S. Department of Housing and Urban Development. MFI = Median Family Income. NA = Data not available.

Acknowledgments

The author thanks Mark Shroder and three anonymous referees for their helpful comments. The author also thanks Jim Carr, Julia Gordon, Kevin Park, and David Sanchez for their feedback on early versions of this article.

Author

Michela Zonta is a senior policy analyst at the Center for American Progress.

References

- Ambrose, Brent W., Thomas G. Thibodeau, and Kenneth Temkin. 2002. *An Analysis of the Effects of the GSE Affordable Goals on Low- and Moderate-Income Families*. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research.
- An, Xudong, and Raphael Bostic. 2008. "GSE Activity, FHA Feedback, and Implications for the Efficacy of the Affordable Housing Goals," *The Journal of Real Estate Finance and Economics* 36 (2): 207–231.
- An, Xudong, Raphael Bostic, Yongheng Deng, and Stuart A. Gabriel. 2007. "GSE Loan Purchases, the FHA, and Housing Outcomes in Targeted, Low-Income Neighborhoods." Paper presented at the Brookings-Wharton Conference on Urban Affairs, October 2006. Washington, DC: The Brookings Institution.
- Bhutta, Neil. 2009. GSE Activity and Mortgage Supply in Lower-Income and Minority Neighborhoods: The Effect of the Affordable Housing Goals. Working paper. Finance and Economics Discussion Series, 2009-3. Washington, DC: Federal Reserve Board, Divisions of Research and Statistics and Monetary Affairs.
- Bolotnyy, Valentin. 2012. The Government-Sponsored Enterprises and the Mortgage Crisis: The Role of the Affordable Housing Goals. Working paper. Finance and Economics Discussion Series, 2012-25. Washington, DC: Federal Reserve Board, Divisions of Research and Statistics and Monetary Affairs.
- Bostic, Raphael, and Stuart A. Gabriel. 2006. "Do the GSEs Matter to Low-Income Housing Markets? An Assessment of the Effects of GSE Loan Purchase Activity on California Housing Outcomes," *Journal of Urban Economics* 59 (3): 458–475.
- Boxall, Patrick, and Joshua B. Silver. 2001. "Performance of the GSEs at the Metropolitan Level," *Cityscape* 5 (3): 145–218.
- Bunce, Harold L. 2002. The GSEs' Funding of Affordable Loans: A 2000 Update. Housing Finance Working Paper Series, Working Paper No. HF-013. Washington, DC: U.S. Department of Housing and Urban Development.

Bunce, Harold L., and Randall M. Scheessele. 1996. *The GSEs' Funding of Affordable Loans*. Housing Finance Working Paper Series, Working Paper No. HF-001. Washington, DC: U.S. Department of Housing and Urban Development.

Canner, Glen B., Wayne Passmore, and Brian J. Surrette. 1996. "Distribution of Credit Risk Among Providers of Mortgages to Lower-Income and Minority Homebuyers," *Federal Reserve Bulletin* 82 (December): 1077–1102.

Case, Bradford, Kevin Gillen, and Susan M. Wachter. 2002. "Spatial Variation in GSE Mortgage Purchase Activity," *Cityscape* 6 (1): 9–84.

Gabriel, Stuart A., and Stuart S. Rosenthal. 2010. "Do the GSEs Expand the Supply of Mortgage Credit? New Evidence of Crowd Out in the Secondary Mortgage Market," *Journal of Public Economics* 94 (11–12): 975–986.

———. 2008. "The GSEs, CRA, and Homeownership in Targeted Underserved Neighborhoods." Paper presented at the conference "Built Environment: Access, Finance, and Policy," in honor of Karl Case, December 7–8, 2007. Cambridge, MA: Lincoln Institute of Land Policy.

———. 2007. *HUD Purchase Goals and Crowd Out: Do the GSEs Expand the Supply of Mortgage Credit?* UCLA Anderson School Finance Series Working Papers. Los Angeles: University of California, Los Angeles, Anderson School of Finance.

Gyourko, Joseph, and Dapeng Hu. 1999. "The Intra-Metropolitan Area Distribution of GSE Mortgage Purchases Made in Support of Low-Income Related Goals." <http://EconPapers.repec.org/RePEc:wop:pennzl:317>.

Joint Center for Housing Studies at Harvard University. 2014. *The State of the Nation's Housing*. Cambridge, MA: Harvard University, Joint Center for Housing Studies at Harvard University.

Lind, John E. 1996. *A Comparison of the Community Reinvestment and Equal Credit Opportunity Performance of Fannie Mae and Freddie Mac Portfolios by Suppliers from the 1994 HMDA Data*. San Francisco, CA: Cannicor.

MacDonald, Heather I. 2001. "Fannie Mae and Freddie Mac in Nonmetropolitan Housing Markets: Does Space Matter?" *Cityscape* 5 (3): 219–264.

McClure, Kirk. 2001. "The Twin Mandates Given to the GSEs: Which Works Best, Helping Low-Income Homebuyers or Helping Underserved Areas?" *Cityscape* 5 (3): 107–144.

Pearce, James E. 2001. "Fannie Mae and Freddie Mac Mortgage Purchases in Minority Neighborhoods: 1994–96," *Cityscape* 5 (3): 265–316.

Roberts, Russell. 2010. *Gambling with Other People's Money: How Perverted Incentives Caused the Financial Crisis*. Arlington, VA: George Mason University, Mercatus Center.

Ross, Stephen L., and John Yinger. 2002. *The Color of Credit: Mortgage Discrimination, Research Methodology, and Fair-Lending Enforcement*. Cambridge, MA: The MIT Press.

U.S. Department of Housing and Urban Development (HUD). 1991. *Capitalization Study of the Federal National Mortgage Association and the Federal Home Loan Mortgage Corporation*. Washington, DC: U.S. Department of Housing and Urban Development.

Wallison, Peter J. 2011. *Dissent from the Majority Report of the Financial Crisis Inquiry Commission*. Washington, DC: AEI Press.

Weicher, John C. 2010. "The Affordable Housing Goals, Homeownership and Risk: Some Lessons from Past Efforts to Regulate the GSEs." Paper presented at the conference "The Past, Present and Future of the Government-Sponsored Enterprises," November 17. St. Louis: Federal Reserve Bank of St. Louis.

Williams, Richard. 2006. *Alternative Assessments of GSE Performance, Influence, and Impact 1993–2003*. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research.

Williams, Richard, and Carolyn Bond. 2002. *Are the GSEs Leading, and if so Do They Have Any Followers? An Analysis of the GSEs' Impact on Home Purchase Lending to Underserved Markets During the 1990s*. Report prepared for the U.S. Department of Housing and Urban Development. Working paper and technical report Series No. 2003-2. Notre Dame, IN: University of Notre Dame, Department of Sociology.

Williams, Richard, Eileen Diaz McConnell, and Reynold Nesiba. 2001. "The Effects of the GSEs, CRA, and Institutional Characteristics on Home Mortgage Lending to Underserved Markets," *Cityscape* 5 (3): 9–106.

Additional Reading

Ambrose, Brent W., and Thomas G. Thibodeau. 2004. "Have the GSE Affordable Housing Goals Increased the Supply of Mortgage Credit?" *Regional Science and Urban Economics* 34 (3): 263–273.

Bradford, Calvin. 2002. "The Patterns of GSE Participation in Minority and Racially Changing Markets Reviewed From the Context of the Levels of Distress Associated With High Levels of FHA Lending," *Cityscape* 6 (1): 145–311.

Briggs, Xavier de Souza, ed. 2005. *The Geography of Opportunity: Race and Housing Choice in Metropolitan America*. Washington, DC: Brookings Institution Press.

Frame, W. Scott., and Lawrence J. White. 2004. *Fussing and Fuming Over Fannie and Freddie: How Much Smoke, How Much Fire?* Working Paper 2004-26. Atlanta: Federal Reserve Bank of Atlanta

Freeman, Lance, and George Galster. 2006. "The Impact of Secondary Mortgage Market and GSE Purchases on Underserved Neighborhood Housing Markets: A Cleveland Case Study," *Urban Affairs Review* 42 (2): 193–223.

Goetzmann, William N., and Matthew Spiegel. 2000. "The Policy Implications of Portfolio Choice in Underserved Mortgage Markets" Working Paper No. 00-18. New Haven, CT: Yale University, Yale International Center for Finance.

Hu, Dapeng. 2000. "A Semiparametric Investigation of Lower-Income Home Mortgage Purchases in the Secondary Mortgage Market." Working paper No. 350. Philadelphia, PA: Samuel Zell and Robert Lurie Real Estate Center, The Wharton School, University of Pennsylvania.

Immergluck, Dan. 2004. *Credit to the Community: Community Reinvestment and Fair Lending Policy in the United States*. Armonk, NY: M.E. Sharpe.

Jaffee, Dwight M., and John M. Quigley. 2007. Housing Subsidies and Homeowners: What Role for Government-Sponsored Enterprises? Working Paper Series, Working Paper No. W06-006. Berkeley, CA: University of California, Berkeley, Program on Housing and Urban Policy.

Manchester, Paul B. 1998. Characteristics of Mortgages Purchased by Fannie Mae and Freddie Mac: 1996–97 Update. Housing Finance Working Paper Series, Working Paper No. HF-006. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research.

Manchester, Paul. B., Sue George Neal, and Harold L. Bunce. 1998. Characteristics of Mortgages Purchased by Fannie Mae and Freddie Mac, 1993–95. Housing Finance Working Paper Series, Working Paper No. HF-003. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research.

Myers, Samuel L. 2002. "Government-Sponsored Enterprise Secondary Market Decisions: Effects on Racial Disparities in Home Mortgage Loan Rejection Rates," *Cityscape* 6 (1): 85–113.

Olinger, Jillian. 2010. *Fannie, Freddie, and the Future of Fair Housing*. Columbus, OH: The Ohio State University, Kirwan Institute for the Study of Race and Ethnicity.

Passmore, Wayne, Shane M. Sherlund, and Gillian Burgess. 2005. The Effect of Housing Government-Sponsored Enterprises on Mortgage Rates. Finance and Economics Discussion Series, Working Paper 2005-06. Washington, DC: Federal Reserve Board, Divisions of Research and Statistics and Monetary Affairs.

Schwartz, Alex F. 2015. *Housing Policy in the United States*. New York: Routledge.

Shear, William B., James Berkovec, Ann Dougherty, and Frank E. Nothaft. 1995. "Unmet Housing Needs: The Role of Mortgage Markets," *Journal of Housing Economics* 4: 291–306.

Departments

In this issue—

- *Data Shop*
- *Graphic Detail*
- *Industrial Revolution*
- *Foreign Exchange*
- *SpAM*
- *Evaluation Tradecraft*

Data Shop

Data Shop, a department of Cityscape, presents short articles or notes on the uses of data in housing and urban research. Through this department, the Office of Policy Development and Research introduces readers to new and overlooked data sources and to improved techniques in using well-known data. The emphasis is on sources and methods that analysts can use in their own work. Researchers often run into knotty data problems involving data interpretation or manipulation that must be solved before a project can proceed, but they seldom get to focus in detail on the solutions to such problems. If you have an idea for an applied, data-centric note of no more than 3,000 words, please send a one-paragraph abstract to david.a.vandenbroucke@hud.gov for consideration.

Measuring Neighborhood Opportunity With AFFH Data

Brent D. Mast

U.S. Department of Housing and Urban Development

The views expressed in this article are those of the author and do not represent the official positions or policies of the Office of Policy Development and Research, the U.S. Department of Housing and Urban Development, or the U.S. government.

Abstract

HUD's new Affirmatively Furthering Fair Housing (AFFH) database is designed to help U.S. Department of Housing and Urban Development (HUD) program participants affirmatively further the purposes of the Fair Housing Act. Along with the AFFH database, HUD is providing a geospatial tool to generate a series of maps of tables with the AFFH data. Both the tool and database provide a new means for HUD program participants, researchers, and the public to assess neighborhood opportunity on a national basis.

This article introduces readers to the new AFFH database and compares it with other sources of data on neighborhood opportunity.

As an example of the type of data analysis possible with AFFH data, I analyze the relationship among school proficiency, the minority population, and poverty for 23 census tracts in Roanoke, Virginia. Results indicate that school proficiency is negatively related with both the percent of the population that is non-White and the poverty rate. Eight geographically contiguous tracts with the highest percent non-White and highest poverty also tend to have the lowest school proficiency.

Introduction

The Fair Housing Act (Title VIII of the Civil Rights Act of 1968) prohibits housing discrimination based on race, color, national origin, religion, or sex.¹ Amendments to the Act in 1988 further banned discrimination against families with children and people with disabilities, and they greatly increased the U.S. Department of Housing and Urban Development's (HUD's) enforcement role.

Local governments and states receiving Community Development Block Grants (CDBGs); HOME Investment Partnerships, or HOME; Emergency Solutions Grants, or ESGs; and Housing Opportunities for Persons with AIDS, or HOPWA, are obligated to affirmatively further the purposes of the Fair Housing Act, as are public housing agencies (PHAs). To help program participants meet this obligation, HUD's Affirmatively Furthering Fair Housing (AFFH) initiative² provides guidance, data, and an assessment template from which the participants will complete an assessment of fair housing (the AFH).

The AFH focuses program participants' analysis on four primary goals—

1. Reduce segregation and build on the nation's increasing racial, geographic, and economic diversity.
2. Eliminate racially and ethnically concentrated areas of poverty.
3. Reduce disparities in access to opportunities such as high-quality schools, job centers, and transit through both mobility and neighborhood reinvestment.
4. Narrow gaps that leave families with children; people with disabilities; and people of different races, colors, and national origins with disproportionate housing needs.

HUD's AFFH database provides nationally available data on these four areas. After analyzing the HUD data and any supplemental information they choose to add, program participants identify the primary determinants influencing fair housing conditions, prioritize addressing these conditions, and set one or more goals to further fair housing.

This article introduces readers to the data HUD is providing to grantees and PHAs to help complete their AFHs. The following sections (1) describe the AFFH database in greater detail; (2) compare the AFFH data with alternative sources of data on neighborhood opportunity; (3) present a data analysis example, analyzing the relationship among school proficiency, the minority population, and poverty for census tracts in Roanoke, Virginia; and (4) present concluding remarks.

¹ For more information on the Fair Housing Act, see HUD (2015a).

² See HUD (2015b) for more information about the AFFH final rule.

AFFH Database

The AFFH database contains property-level, block group-level, and census tract-level information from numerous sources. HUD provides a geospatial tool³ that enables program participants and the public to generate a series of tables and maps with the data required for an AFH.⁴ The tool also enables users to download the data used to populate the tables and maps.

Socioeconomic and Demographic Data

The AFFH database includes demographic data from the 2010 decennial census (for example, block-group data on race and ethnicity). Demographic and socioeconomic data (for example, data on people with disabilities, people in poverty, and unemployment) are also taken from the 5-year American Community Survey (ACS) for various timeframes. To keep margins of error within reasonable bounds, ACS estimates are not reported below the census tract level. Longitudinal socioeconomic and demographic tract data for 1990 and 2000 are from Brown University's Longitudinal Tract Database (Brown University, 2015), based on decennial census and ACS data.

Housing Data

The AFFH database includes property-level and census tract-level data on households receiving public housing and HUD multifamily rental assistance and tract-level data on households in the Housing Choice Voucher (HCV) program. Public housing and HCV data are from HUD's Inventory Management System, or IMS/Public and Indian Housing, or PIH, Information Center, or PIC (HUD, 2015f); multifamily data are from HUD's Tenant Rental Assistance Certification System, or TRACS (HUD, 2015g). Data on Low-Income Housing Tax Credit (LIHTC) Program properties are from HUD's LIHTC database (HUD, 2015h). Tract-level data on households with disproportionate housing needs are from HUD's Comprehensive Housing Affordability Strategy, or CHAS, database (HUD, 2015i).

Opportunity Indices

The AFFH database contains seven percentile indices to measure neighborhood opportunity.⁵ Described in more detail in the data analysis section, the block-group school proficiency index is based on the percent of fourth grade students proficient on state math and reading exams. The low-poverty index is based on the census tract family poverty rate.

Also computed at the tract level, the labor market index is based on the unemployment rate, the labor force participation rate, and the percent of the older-than-25 population with at least a bachelor's degree.

³ See HUD (2015c) for release 1 of the AFFH Tool.

⁴ See HUD (2015d) for proposed tables and HUD (2015e) for proposed maps for the local jurisdictions, such as CDBG grantees. States, PHAs, and regional consortia will have separate templates with possibly different tables and maps.

⁵ More information about the indices is available in the AFFH data documentation (HUD, 2015j).

The tract-level environmental health index is a linear combination of standardized estimates of air-quality carcinogenic, respiratory, and neurological hazards. Environmental hazard data are from the Environmental Protection Agency's (EPA's) National-Scale Air Toxics Assessment, or NATA, program (EPA, 2015a).

The jobs accessibility index for a given residential block group is measured as a function of its distance to all job locations within a Core Based Statistical Area (CBSA), with distances to larger employment centers weighted more heavily. Data on jobs and employment are from the Census Bureau's Longitudinal Employer-Household Dynamics, or LEHD, program (U.S. Census Bureau, 2015).

Two indices measure transportation opportunity for a household profile consisting of a single parent family of three, renting, with income equal to 50 percent of Area Median Income. The low transportation cost index is based on modeled transportation costs as a percent of household income. The transit trips index is based on modeled annual household transit trips. Tract data for both indices are from HUD's Location Affordability Index database (HUD, 2015k).

Other Sources of Neighborhood Opportunity Data

In this section, I discuss other sources of neighborhood opportunity data and compare them with the AFFH database.

Smart Location Database

The Smart Location Database (SLD) is EPA's geographic database for measuring location efficiency (EPA, 2015b). It includes more than 90 attributes summarizing characteristics such as housing density, diversity of land use, neighborhood design, destination accessibility, transit service, employment, and demographics. Most variables are available for all U.S. block groups.

The SLD contains measures of job accessibility via cars and mass transit compared with the AFFH jobs accessibility index, which is based on geodesic distance. The SLD transit measures are available only for participating General Transit Feed Specification, or GTFS, transit agencies compared with the AFFH transit index, which is available for all U.S. states and Washington, D.C.

EJSCREEN and C-FERST

EJSCREEN (EPA, 2015c), the EPA's environmental justice (EJ) screening and mapping tool, provides a nationally consistent dataset and methodology for calculating EJ indices at the block-group level. Each of the 12 EJ indices combines an environmental indicator (for example, a lead paint indicator) with demographic indicators (predictors of health status and of potential vulnerability to environment).

C-FERST, EPA's Community-Focused Exposure and Risk Screening Tool (EPA, 2015d), is being developed as a community mapping, information access, and assessment tool designed to help assess risk and assist in decision-making with communities. It will incorporate research estimating human exposures to toxic substances in the environment.

Compared with the AFFH environmental health index, EJSCREEN and C-FERST will contain a much richer set of data on environmental health risks.

The Kirwan Institute

The Kirwan Institute’s opportunity mapping initiative (Kirwan Institute, 2015) includes projects for numerous metropolitan areas. For instance, they partnered with the Puget Sound Regional Council (PSRC, 2015) to develop a series of 20 indicators (for example, percent of an area that is within a food desert) that represent five major categories of opportunity: education, economic health, housing and neighborhood quality, transportation/mobility, and health and environment.

The Brandeis University site <http://www.diversitydatakids.org> includes information about a research project designed to provide national, integrated information about demographics, outcomes, and factors driving outcomes for children. Its child opportunity index (developed in conjunction with the Kirwan Institute) is calculated using 19 indicators (for example, proximity to parks and open spaces) in three defined opportunity domains: (1) educational opportunity, (2) health and environmental opportunity, and (3) social and economic opportunity. Its child opportunity maps visualize the geographic distribution of the index in the 100 largest U.S. metropolitan areas.

Although the Kirwan Institute and their partners have developed many more indicators of neighborhood opportunity than are contained in the AFFH database, the indicators are available for only select metropolitan areas.

National Neighborhood Indicators Partnership

The National Neighborhood Indicators Partnership (NNIP) is a collaboration of the Urban Institute and more than 30 city local partners to further the development and use of neighborhood-level data (NNIP, 2015). The NNIP data inventory contains a wealth of information on neighborhood characteristics such as demographics, education, health, public assistance, and business/economy data. The NNIP data inventory also includes data on crime for participating partners. The AFFH database does not include crime indicators, because neighborhood-level crime data are not nationally available.

Data Analysis

In this section, I analyze the relationship among school proficiency, the minority population, and poverty for 23 census tracts in Roanoke, Virginia.

The school proficiency index is based on the percent of fourth grade students proficient on state math and reading exams in up to three schools closest to the block-group centroid.⁶

$$Proficiency_i = \sum_{j=1}^3 \frac{e_j}{E} * \left[\frac{r_{j+} + m_j}{2} \right], \quad (1)$$

where i denotes a block group; e denotes fourth grade enrollment in the j th school; E denotes total fourth grade enrollment in the j schools; and r and m are percentages of fourth grade students proficient in reading and math, respectively, standardized by state. Proficiency data are from Great Schools for school year 2011–2012, and school location and enrollment data are from the U.S. Department of Education’s Common Core of Data (ED/NCES, 2015).

⁶ Elementary schools are linked with block groups based on a geographic mapping of attendance area zones from the School Attendance Boundary Information System, where available, or within-district proximity matches within 1.5 miles.

The school proficiency index is a within-state percentile of the variable defined above. While the index is measured at the block-group level, for this analysis I created a tract index by computing a tract mean of the block-group indices, weighting by fourth grade enrollment.

Exhibit 1 reports a linked micromap⁷ of Roanoke tracts with data on the school index, percent of the population that is non-White (hereafter referred to as percent non-White), and the poverty rate. Data on percent non-White are from the 2010 decennial census, and poverty data are from the 2006–2010 ACS. Data in exhibit 1 are reported by ascending values of the school index; the data indicate that school proficiency is negatively related with both percent non-White and the poverty rate.

For further analysis, it might be helpful to classify tracts according their percent non-White and poverty rate. The AFFH data include an indicator for tracts classified as racially/ethnically concentrated areas of poverty (R/ECAPs). In CBSAs, R/ECAPs are defined as having percent non-White of at least 50 percent and a poverty rate that is at least 40 percent or three times the average tract poverty rate for the CBSA.

Because Roanoke has only three R/ECAP tracts, I employ an alternate approach categorizing tracts into two percent non-White categories (less than 37.9 percent and greater than or equal to 37.9 percent) and two poverty rate categories (less than 10.9 percent and greater than or equal 10.9 percent). I will refer to the lower categories for both variables as “low,” and the higher categories as “high.” I chose cut points with a regression tree.⁸ In a least squares regression, the two categorical variables explain 65.7 percent of the variation in the school index.

Exhibit 2 reports school proficiency index summary statistics grouped by the two categorical variables; exhibit 3 reports a conditioned choropleth map⁹ with the school proficiency index mapped conditioned on the two categorical variables. No tracts have a high percent non-White category and low-poverty category, seven tracts have low percent non-White and low-poverty categories, eight tracts have low percent non-White and high poverty categories, and eight geographically contiguous tracts have high percent non-White and high poverty categories.

Tracts with the low percent non-White and low-poverty categories tend to have the highest school proficiency (mean of 74.7), while the tracts with high percent non-White and high poverty categories tend to have the lowest school proficiency (mean of 45.8). Tracts with low percent non-White and high poverty categories have a mean school index of 66.3.

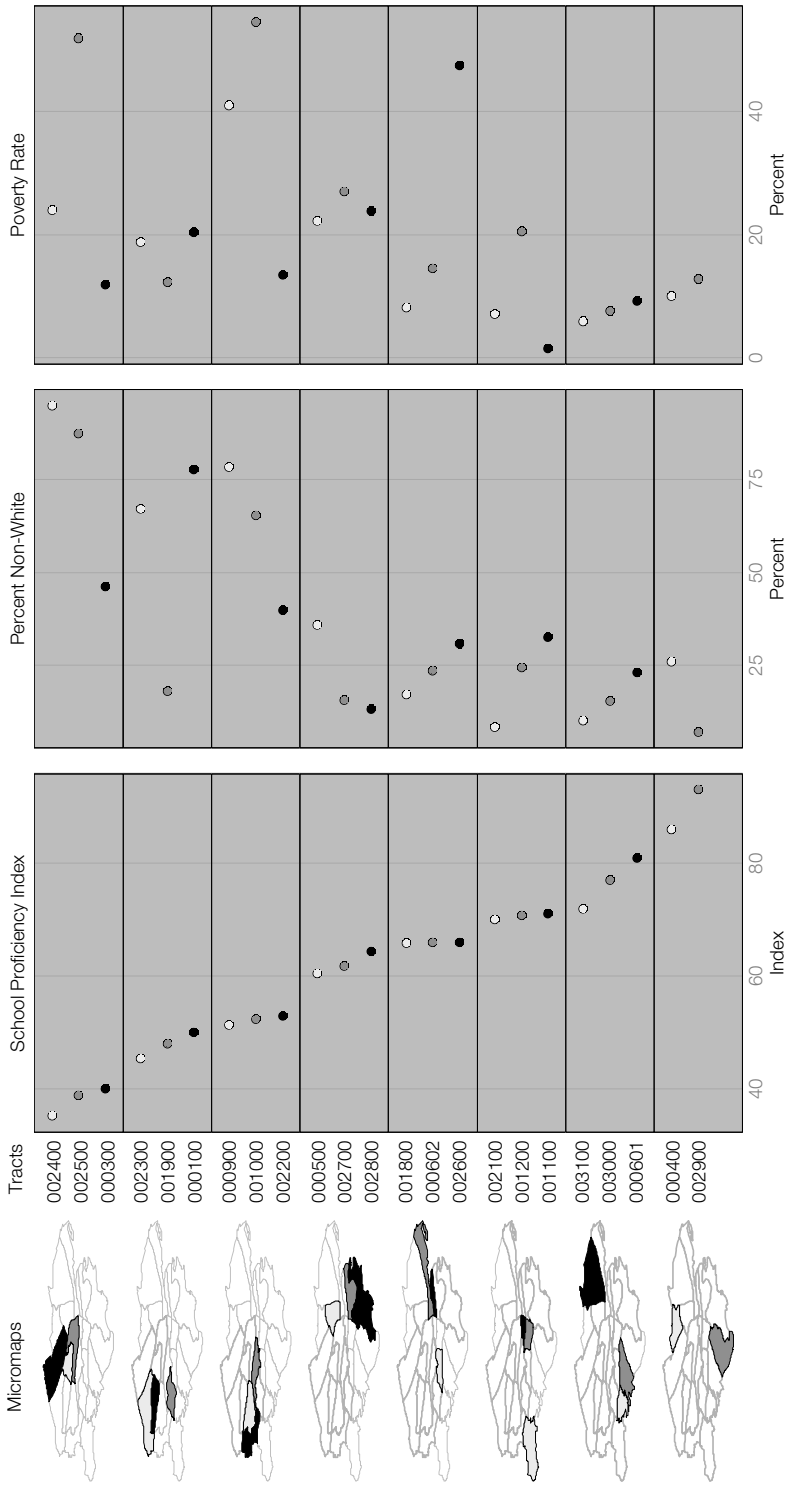
⁷ The linked micromap in exhibit 1 was generated with the R “micromap” package (program available upon request).

⁸ The regression tree was estimated with the R “rpart” package (estimates and program available upon request).

⁹ The conditioned choropleth map in exhibit 3 was generated with the R “mapproj” package (program available upon request).

Exhibit 1

Linked Micromap With School Proficiency Index, Percent Non-White, and Poverty Rate



Sources: Great Schools 2011–2012 (school proficiency index); 2010 decennial census (percent non-White); 2006–2010 American Community Survey (poverty rate)

Exhibit 2

School Proficiency Index Summary Statistics

Percent Non-White Category	Poverty Rate Category	School Proficiency Index					Standard Deviation
		n	Minimum	Median	Mean	Maximum	
Low: < 37.9%	Low: < 10.9%	8	65.9	71.9	74.7	86.0	7.0
Low: < 37.9%	High: ≥ 10.9%	7	48.0	65.2	66.3	93.0	12.7
High: ≥ 37.9%	High: ≥ 10.9%	8	35.3	47.7	45.8	53.0	6.9

Sources: Great Schools 2011–2012 (school proficiency index); 2010 decennial census (percent non-White); 2006–2010 American Community Survey (poverty rate)

Exhibit 3

Map of School Proficiency Index Conditional on Percent Non-White and Poverty Rate



Sources: Great Schools 2011–2012 (school proficiency index); 2010 decennial census (percent non-White); 2006–2010 American Community Survey (poverty rate)

Conclusion

To help HUD program participants affirmatively further the purposes of the Fair Housing Act, HUD launched a new AFFH database. HUD is also providing a geospatial tool to generate a series of maps of tables with the AFFH data. Both the tool and database provide a new means for HUD program participants, researchers, and the public to assess neighborhood opportunity on a national basis.

The AFFH database contains a large amount of demographic, socioeconomic, and housing data at the census-tract and block-group levels. The database also includes seven indices to measure neighborhood school proficiency, jobs accessibility, environmental health, poverty, labor market conditions, transit use, and transportation costs.

This article introduces readers to the new AFFH database and compares it with other sources of data on neighborhood opportunity.

As an example of the type of data analysis possible with the AFFH database, I analyzed the relationship among school proficiency, the minority population, and poverty for 23 census tracts in Roanoke, Virginia. Results indicate that school proficiency is negatively related with both the percent of the population that is non-White and the poverty rate. Eight geographically contiguous tracts with the highest percent non-White and highest poverty also tend to have the lowest school proficiency.

Author

Brent D. Mast is a social science analyst at the U.S. Department of Housing and Urban Development, Office of Policy Development and Research, Program Monitoring and Research Division.

References

- Brown University. 2015. "Census Geography: Bridging Data From Prior Years to the 2010 Tract Boundaries." <http://www.s4.brown.edu/us2010/Researcher/Bridging.htm>.
- Kirwan Institute. 2015. "Opportunity Mapping Initiative." <http://kirwaninstitute.osu.edu/opportunity-communities/mapping/>.
- National Neighborhood Indicators Partnership (NNIP). 2015. "National Neighborhood Indicators Partnership." <http://www.neighborhoodindicators.org/>.
- Puget Sound Regional Council (PSRC). 2015. "Opportunity Mapping." <http://www.psrc.org/growth/tod/growing-transit-communities-strategy/equity/opportunity-mapping/>.
- U.S. Census Bureau. 2015. "Longitudinal Employer-Household Dynamics." <http://www.lehd.ces.census.gov/>.
- U.S. Department of Education, National Center for Education Statistics (ED/NCES). 2015. "Common Core of Data." <http://www.nces.ed.gov/ccd/>.
- U.S. Department of Housing and Urban Development (HUD). 2015a. "Fair Housing—It's Your Right." http://www.portal.hud.gov/hudportal/HUD?src=/program_offices/fair_housing_equal_opp/FHLaws/yourrights.
- . 2015b. "Affirmatively Furthering Fair Housing (AFFH) Final Rule." <https://www.hudexchange.info/programs/affh/>.

- . 2015c. “Affirmatively Furthering Fair Housing Tool.” <http://www.affht.vsolvit.com/>.
- . 2015d. “AFFH Assessment Tool Data Tables.” huduser.gov/portal/publications/pdf/AFFH-Assessment-Tool-data-tables.pdf.
- . 2015e. “AFFH Template Maps.” huduser.gov/portal/publications/pdf/AFFH-Template-Maps-2014.pdf.
- . 2015f. “Inventory Management System (IMS)/PIH Information Center (PIC).” http://www.portal.hud.gov/hudportal/HUD?src=/program_offices/public_indian_housing/systems/pic.
- . 2015g. “Tenant Rental Assistance Certification System (TRACS).” http://portal.hud.gov/hudportal/HUD?src=/program_offices/housing/mfh/trx/trxsum.
- . 2015h. “Low-Income Housing Tax Credits.” huduser.gov/portal/datasets/lihtc.html.
- . 2015i. “Consolidated Planning/CHAS Data.” huduser.gov/portal/datasets/cp.html.
- . 2015j. “Affirmatively Furthering Fair Housing (AFFH) Data Documentation.” huduser.gov/portal/sites/default/files/docs/AFFH_Documentation_forPublic_07172015.docx.
- . 2015k. “Location Affordability Index.” http://www.locationaffordability.info/about_lai.aspx.
- U.S. Environmental Protection Agency (EPA). 2015a. “National-Scale Air Toxics Assessment.” <http://www.epa.gov/airtoxics/natamain/>.
- . 2015b. “Smart Location Database.” <http://www2.epa.gov/smart-growth/smart-location-mapping>.
- . 2015c. “EJSCREEN: Environmental Justice Screening and Mapping Tool.” <http://www2.epa.gov/ejscreen/>.
- . 2015d. “Community-Focused Exposure and Risk Screening Tool.” <http://www.epa.gov/heasd/c-ferst/>.

Graphic Detail

Geographic Information Systems (GIS) organize and clarify the patterns of human activities on the Earth's surface and their interaction with each other. GIS data, in the form of maps, can quickly and powerfully convey relationships to policymakers and the public. This department of Cityscape includes maps that convey important housing or community development policy issues or solutions. If you have made such a map and are willing to share it in a future issue of Cityscape, please contact john.c.huggins@hud.gov.

Civil Unrest and Marginalization in Baltimore

John C. Huggins

U.S. Department of Housing and Urban Development

The views expressed in this article are those of the author and do not represent the official positions or policies of the Office of Policy Development and Research, the U.S. Department of Housing and Urban Development, or the U.S. government.

The map in exhibit 1 illustrates incidents of civil unrest that occurred in Baltimore, Maryland, on April 27, 2015. Incident locations were mapped relative to the 2010 U.S. census tracts designated by the U.S. Department of Housing and Urban Development as Racially and Ethnically Concentrated Areas of Poverty (R/ECAP), and regional Labor Market Engagement Scores also were denoted by 2010 U.S. census tracts.¹

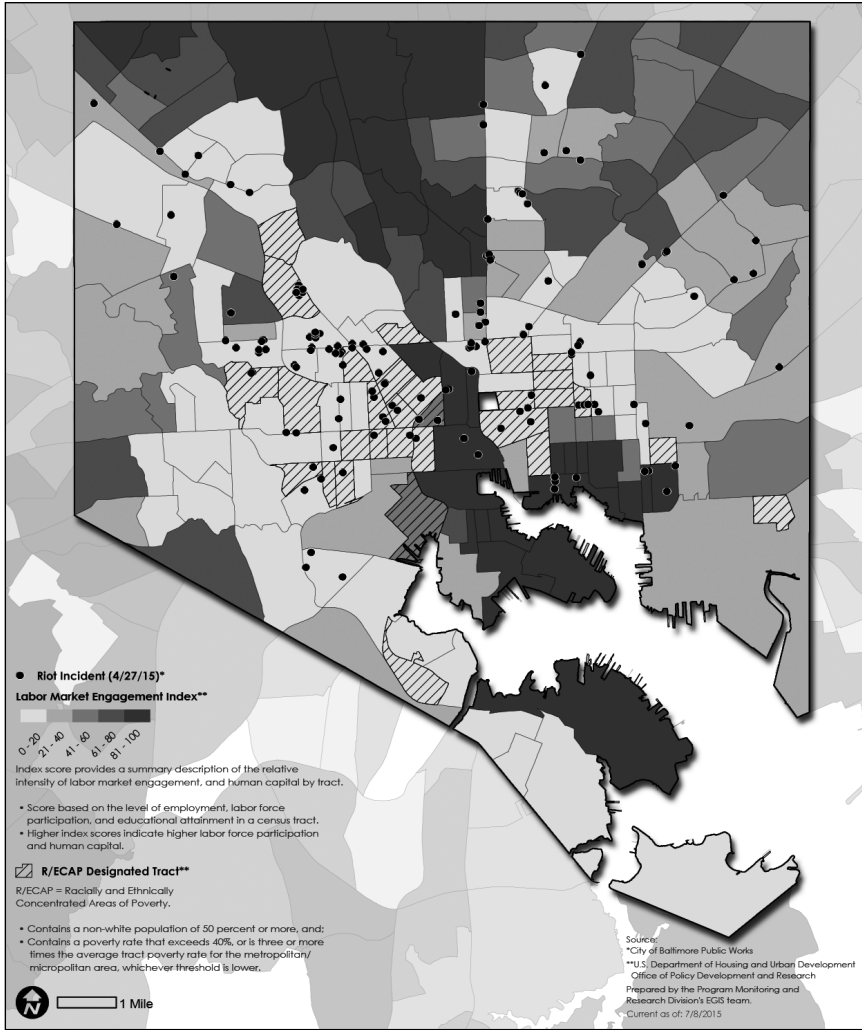
Although incidents of civil unrest occurred citywide, the map demonstrates that many incidents took place within R/ECAP-designated census tracts. Moreover, most events occurred in areas with low labor-market engagement scores. Therefore, the map suggests a strong relationship among riot events, R/ECAP tracts, and areas of low labor-market engagement.

The selection of points and tracts relative to one another reveals that approximately 26 percent of all rioting events occurred within R/ECAP-designated tracts. Furthermore, 86 percent of the tracts where civil unrest occurred score 60 percent or less on the Labor Market Engagement Index, and, conversely, 88 percent of all incident points fall within those low scoring tracts.

¹ R/ECAPs include tracts that contain a non-White population of 50 percent or more and a poverty rate that exceeds 40 percent or that is three or more times the average tract poverty rate for the metropolitan/micropolitan area, whichever threshold is lower. The Labor Market Engagement Score provides a summary description of the relative intensity of labor-market engagement and human capital by census tract. Scores are based on the level of employment, labor force participation, and educational attainment in a census tract. Higher index scores indicate higher labor force participation and human capital.

Exhibit 1

City of Baltimore: Civil Unrest, Labor Market Engagement, and Concentrated Poverty



Maps provide outstanding tools that can be used to identify spatial correlation. In this case, the map clearly reveals a strong correlation among the locations where incidents of civil unrest occurred, R/ECAP-designated census tracts, and areas of low labor-market engagement. Of course, the old adage *correlation does not imply causation* applies here as elsewhere, and additional analysis remains necessary to better understand the correlation alluded to in this particular map.

Author

John C. Huggins is a social science analyst and research geographer at the U.S. Department of Housing and Urban Development, Office of Policy Development and Research.

Industrial Revolution

Every home makes compromises among different and often competing goals: comfort, convenience, durability, energy consumption, maintenance, construction costs, appearance, strength, community acceptance, and resale value. Often consumers and developers making the tradeoffs among these goals do so with incomplete information, increasing the risks and slowing the adoption of innovative products and processes. This slow diffusion negatively affects productivity, quality, performance, and value. This department of Cityscape presents, in graphic form, a few promising technological improvements to the U.S. housing stock. If you have an idea for a future department feature, please send your diagram or photograph, along with a few, well-chosen words, to elizabeth.a.cocke@hud.gov.

Rural America: Perceptions of Residential Energy Retrofits

Nathan Barry
University of Nebraska at Kearney

Abstract

Residential energy consumption accounts for most of the overall energy use in the United States (Pew Center Global Climate Change, 2008). Although many larger municipalities are successfully implementing residential energy retrofit programs, these benefits are currently not being seen in most rural townships. Seeking an opportunity to understand and reduce energy consumption in these areas, this study examines how the aging housing stock and owners' perceptions of energy retrofits affect the overall performance of the home. Using the small agricultural community of Woodbine, Iowa, the study compares 3 years of actual historical energy data with the homeowners' perceptions of energy retrofits. The statistical analysis of the perceptions survey not only shows that homeowners changing their energy use habits has a positive effect on lowering energy consumption, but also that homeowners perceiving that they have changed their energy use habits also plays a critical role in reducing their use of energy. The results indicate that the homeowners' simple awareness of their improved energy use practices can have a positive effect on lowering their monthly utility bills without having to invest large amounts of money into energy reduction.

Status Quo

Multiple surveys since the 1990s demonstrate national trends of consumers' preference for renewable energy. Further U.S. Department of Energy research at the regional and state levels confirmed that residential customers favor renewable sources and are willing to pay more to power their home with those resources (Stein and Meier, 2000). Using Woodbine, Iowa, as a representative example of Midwestern rural America, this case study investigates if awareness programs and basic energy conservation information do in fact reduce energy bills.

Located in western Iowa, Woodbine is an agriculturally based community approximately 50 miles northeast of the Omaha, Nebraska metropolitan area. The town has 647 households consisting of a representative sample of the U.S. housing stock. With a medium household income of only \$30,083, the residents are a mix of blue collar, lower to middle-income families. Most of the homes were built before 1979 and many have had additions or renovations made throughout the life of the home. With property values stagnant, investing in energy retrofit projects has been financially difficult.

Importance of the Study

Residential energy savings remain the largest untapped opportunity to reduce our dependence on fossil fuels whether domestically or abroad. The public and private communities continue to invest in new energy sources. Today, the average cost for private, large-scale wind production is \$0.08 per kilowatt hour and solar production can reach upwards of \$0.20 per kilowatt hour. By simply retrofitting residential properties, an energy company conserving kilowatt hours enables the power companies to reuse the energy at a cost of \$0.03 per kilowatt hour (McGraw Hill Construction, 2009). The combination of this economic benefit with the pride a homeowner gains when improving the performance of his or her home highlights the importance of residential energy retrofit and the sense of empowerment owners gain. A challenge in the aging housing stock is the necessary structural work that is needed to increase the energy performance of the home. In an effort not to exclude older homes, the necessary question arises: How much, if any, can an energy savings education or awareness program affect the overall energy savings of these homes?

Homeowner Perceptions Survey

To help understand the effects that the homeowners' perception of overall energy efficiency and habitual habits have on consumption, the research team developed a customized homeowner survey. The 20-question survey asked a range of questions regarding the residents' feelings toward their personal utility bills, thoughts on energy conservation nationally, and perceived condition of their home. The surveys were delivered, along with monthly utility bills, to all 647 residences of Woodbine. The following month, 92 residences returned completed surveys to the local utility provider. Although the survey data gave a broad range of very valuable information about each household, two specific questions were designed to be compared against actual utility consumption data of the same household. Focusing on these two questions, to what extent do actual changes to living habits have and whether the basic perception of the owners' living habits affected utility costs in their home.

In collaboration with the local utility provider, the researcher was granted access to 3 years of monthly energy use data for each home. Participants in the study also concurrently were provided a homeowner perception survey in their monthly utility bill. The research team analyzed the actual changes in each participant’s energy use habits and his or her awareness of habit changes to determine the effect of each action; for example, the effect of change of habit versus awareness of their action.

To examine the importance of habit changes versus perceptions of habit changes, we examined the responses to questions 10 and 14. For question 10, “How much do you feel your living habits contribute to your overall utility costs?” participants were given four response choices: (1) a lot, (2) some, (3) a little, and (4) none. Responses *a lot* and *some* were coded as positive, and responses *a little* and *none* were coded as negative. For question 14, “Have you changed your habits in the last 12 months to help lower your utility bills?” respondents were given a response choice of either *yes* or *no*. Again, *yes* was coded as positive and *no* was coded as negative. A multiple regression correlation analysis was run on the two questions using their actual utility consumption data to determine the effects of homeowners’ habits and actual energy usage.

Of the 92 completed surveys, 40 surveys indicated positive responses for both questions 10 (perception of impact of changes in habits) and question 14 (real changes of habits). Of the 40 completed surveys, residents in homes with lower energy usage completed 28. With 70 percent (28 of 40 respondents) of the homes showing actual lower energy consumption in comparison with their neighbors, all indicators would confirm that a homeowner’s perception has an influence on energy consumption during a 3-year period. Exhibit 1 highlights the findings, dividing the observed residences into two types—the lower 50 percent of energy consumers and the upper 50 percent of energy users, when normalized by BTU/SQ/DD¹ calculations.

Exhibit 1

Positive Responses to Questions 10 and 14

Observed	Positive Response to Questions 10 and 14	Negative Response to at Least One Question	Total
Lower 50%	28	29	57
Upper 50%	12	21	33
Total	40	50	90

Lessons Learned

The homeowners’ perception survey provided valuable information for this study and could assist researchers conducting future community perspective studies. This study upholds the argument that both positive homeowner habits and homeowners’ perceptions of being energy efficient in the home are effective when working to reduce household energy consumption.

¹ BTU/SQ/DD—Standardized energy equation used to normalize energy usage. British Thermal Unit divided by the Square Foot of the residence divided by the Degree Days.

The survey data collected by this research were consistent with the findings of the Green Trends and perception studies that McGraw Hill Construction and the National Association of Home Builders conducted (McGraw Hill Construction, 2008; NAHB, 2010). Unique to this study was the ability to use a single community to statistically validate the qualitative data of the survey with the quantitative data of historical energy consumption. The statistical analysis of the homeowners' perceptions survey showed that changing habits has a positive effect on lowering energy consumption, but that recognizing improved ways to reduce energy consumption, regardless of actual changes in day-to-day practice to reduce energy consumption, plays a critical role in reducing energy consumption. The results indicate the key role of an educational or awareness program with a community-based retrofit program.

The research suggests that the first approach in working with a smaller, rural community is a simple awareness and educational program. Such programs can be facilitated by the local volunteer community action boards or by energy providers looking for ways to cut down on peak load stresses. Making homeowners aware of visual inspection items, such as inappropriately installed insulation, turning down or using a programmable thermostat, or easily using the storm windows that are sitting in the corner of the shed, can have positive effects on reducing energy consumption with no out-of-pocket cost to the owner. Exhibit 2 shows a crawl space under a conditioned living space

Exhibit 2

Crawl Space Under a Conditioned Living Area With Missing Floor Insulation



where the floor insulation has fallen away, a condition that can require a home's mechanical system to work overtime, costing the owner more in utility cost each month. Exhibit 3 serves as reminder that by simply turning down the thermostat or installing a programmable thermostat that lowers while the homeowner is away, the homeowner can realize utility savings. Exhibit 4 illustrates that installing storm windows over older windows can improve the thermal performance of a home at a much lower cost than installing new windows.

Being able to analyze and understand consumption patterns in the aging housing stock is best accomplished through a partnership with local utility providers. Access to actual historical home consumption is critical to the accuracy of preauditing retrofit candidates. With the variables of the housing stock, any computer simulation simply cannot predict savings calculations. Homeowners can always provide personal consumption data, but the utility provider's database enables the research team to analyze the homes as a community and then comparatively on an individual basis. Homeowners' habits and lifestyles must also play a role in the data collection. Habitual information enables the preauditor to provide and assess what influence, if any, these habits and lifestyles are having on consumption patterns. Researchers can collect all of the above-mentioned preaudit information before visually inspecting the property. This preaudit collection can enable a community retrofit strategist to prioritize candidates and significantly reduce the audit investment time and to increase the potential of audits being performed on homes that will lead to retrofits with significant energy savings.

Exhibit 3

A Programmable Thermostat Can Help Reduce Utility Costs



Exhibit 4

Storm Windows Can Improve the Thermal Performance of a Home



Research needs to continue to implement pilot projects in multiple communities across the country. Increased partnerships with private industries will enable researchers and professionals to develop retrofit practices that are not only accurate and sustainable but also profitable for privatized energy retrofit business models.

Author

Nathan Barry is an assistant professor of construction management at the University of Nebraska at Kearney.

References

- McGraw Hill Construction. 2009. *Green Outlook 2009: Trends Driving Change*. Columbus, OH: McGraw Hill Publishing.
- . 2008. *Key Trends in the European and U.S. Construction Marketplace: SmartMarket Report*. Columbus, OH: McGraw Hill Publishing.
- National Association of Home Builders (NAHB). 2010. *New and Existing Home Sales, U.S.* Washington, DC: National Association of Home Builders, Economics Department.
- Pew Center for Global Climate Change. 2008. *Energy Use by Building Sector*. Arlington, VA: Center for Climate and Energy Solutions.
- Stein, Jeff Ross, and Alan Meier. 2000. "Accuracy of Home Energy Rating Systems," *Energy* 25 (11): 339–354.

Additional Reading

- Friedrich, K., M. Eldridge, D. York, P. Witte, M. Kushler. 2009. *Saving Energy Cost Effectively: A National Review of the Cost of Energy Saved Through Utility-Sector Energy Efficiency Programs*. Washington, DC: American Council for an Energy-Efficient Economy.
- Hunt, R. 2009 (September 9). Personal communication (interview). C.E.M. Energy Efficiency Team.
- Neels, K. 1982. *Reducing Energy Consumption in Housing: An Assessment of Alternatives*, Vol. 7. Santa Monica, CA: International Regional Science Review.
- U.S. Census Bureau. 2007. *Current Housing Reports, Series H1 50/07, American Housing Survey for the United States*. Washington, DC: U.S. Government Printing Office.

Foreign Exchange

Foreign Exchange, a department of Cityscape, reports on what the U.S. Department of Housing and Urban Development's Office for International and Philanthropic Innovation has learned about new departures in housing and development policy in cities and suburbs throughout the world that might have value if applied in U.S. communities. If you have a recent research report or article of fewer than 2,000 words to share in a forthcoming issue of Cityscape, please send a one-paragraph abstract to matthew.l.hennessy@hud.gov.

Preparing Our Housing for the Transition to a Post-Baby Boom World: Reflections on Japan's May 26, 2015 Vacant Housing Law

Peter Manda

Boston University School of Law

Abstract

For more than a decade, policymakers and planners around the United States have increasingly been making a concerted effort to address the needs of the retiring baby boomers, particularly in funding for health care, long-term social services, elder justice, and retirement security. The 2015 Japanese Special Measures Law to Further a Response to Vacant Housing provides local governments and municipalities in Japan with expansive powers to identify vacant homes and compel owners to repair or remediate them. The Japanese Vacant Housing Law asks us to consider more carefully (1) what will happen to housing when the baby boomers die, and (2) what measures can be taken to prevent an abandoned housing crisis of equal scale in the United States. This article reviews general demographic trends in the United States and Japan, provides a summary of the 2015 Japanese Vacant Housing Law, and recommends development of data to assist with policies in the United States that can better address, and possibly prevent, a potential exacerbation of vacant and abandoned housing related blight over the course of the next 45 years.

Introduction

Although Japanese welfare and social planners developed policies during the 1980s and 1990s to address health and welfare issues related to aging, they only recently began to address the significant increase in vacant and abandoned housing stock at unexpected rates because of the country's aging demographic. The 2015 Japanese Special Measures Law to Further a Response to Vacant Housing¹ (Vacant Housing Law, or VHL) provides local governments and municipalities in Japan with expansive powers to identify vacant homes and compel owners to repair or remediate them. For cases in which an owner refuses to comply with an order to repair, cannot be identified, or has died intestate, the law grants the local municipality expansive taking powers. Properties thus acquired may be used only for cultural, social, or governmental purposes that are predefined in an approved smart growth-oriented response plan. Municipal and prefectural governments are granted the authority to raise taxes to fund implementation of the law.

For more than a decade, policymakers and planners around the United States have increasingly been making a concerted effort to address the needs of the retiring baby boomers, particularly in funding for health care, long-term social services, elder justice, and retirement security. On the housing front, investments in assisted living and nursing and retirement homes have rapidly expanded. State and local governments have focused on supporting “aging-in-place” to enable individuals to remain in their homes past retirement and still receive the services and access to amenities that they need (2005 WHCoA, 2005). Little policy research, however, has been published to investigate what may happen to housing as the baby boomer generation passes away. In many ways, this failure to consider the housing effects of the baby boomers dying reflects the Japanese policy approach toward aging during the 1980s and 1990s. The Japanese Vacant Housing Law asks us to consider more carefully (1) what will happen to housing when the baby boomers die and (2) what measures can be taken to prevent an abandoned housing crisis of equal scale in the United States.

Demographic Background

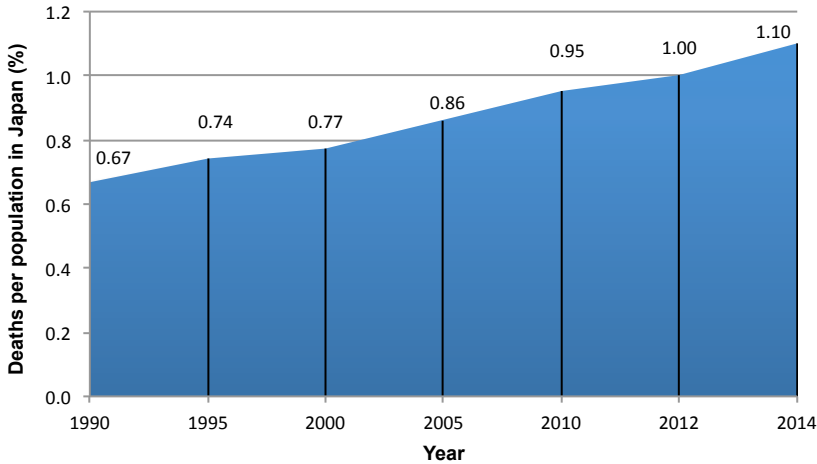
Japan's population is aging and declining. Between today and 2060, Japan's population is projected to decline from 125 million inhabitants to roughly 81 million inhabitants, an overall 35-percent projected decline (Nikkei, 2015). In 2012, the annual death-to-population ratio exceeded 1 percent of the population per year, and that ratio is projected to continue rising at a steady rate in the foreseeable future. (See exhibit 1 and Japanese Ministry of Internal Affairs and Communications, Statistics Bureau, 2015).

American baby boomers, those born between 1945 and 1965, comprise 80 million individuals. In 2010, the first of this generation began to retire (Colby and Ortman, 2014). By 2030, 60 million Americans are projected to be between the ages of 66 and 84. Their numbers will dwindle to 2.4 million by 2060. Based on these assumptions, by 2040, the U.S. death-to-population ratio is expected to reach and surpass the 1-percent death-to-population ratio currently being experienced in Japan. The U.S. death-to-population ratio is projected to remain above 1 percent through 2060 (exhibit 2; U.S. Census Bureau, 2014). In numbers, annual deaths are projected to exceed 4 million individuals beginning in 2041 and will remain at or between 4 and 4.1 million individuals through 2057.

¹ Akiya Nado Taisaku no Suishinn ni Kansuru Tokubetsu Sochihou, *Law Heisei 26-127*, 2015 (<http://www.mlit.go.jp/common/001080536.pdf>).

Exhibit 1

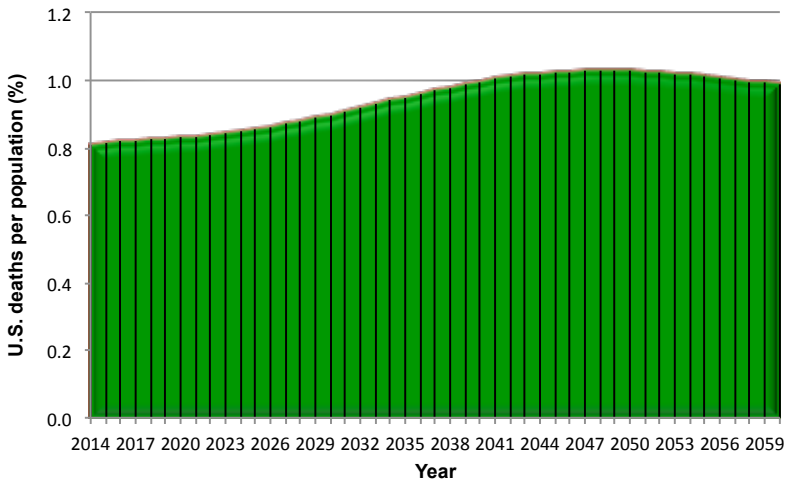
Percentage of Deaths per Population in Japan Passed the 1-Percent Threshold in 2012



Source: Japan Statistical Yearbook 2015

Exhibit 2

Percentage of U.S. Deaths per Population Is Projected To Pass the 1-Percent Threshold in 2041



Source: U.S. Census Bureau (2014)

As the death-to-population ratio remains constant during the 2040-through-2060 period at 1 percent, natural population growth in the United States is estimated to decline from 1 percent to 0.8 percent per year (Ortman, 2013). The gap in population (to thus secure a constant population growth) is assumed to be controlled through immigration growth (Colby and Ortman 2014; Ortman, 2013). Myers and Pitkin predicted in 2009 that “the eventual housing sell-off among the

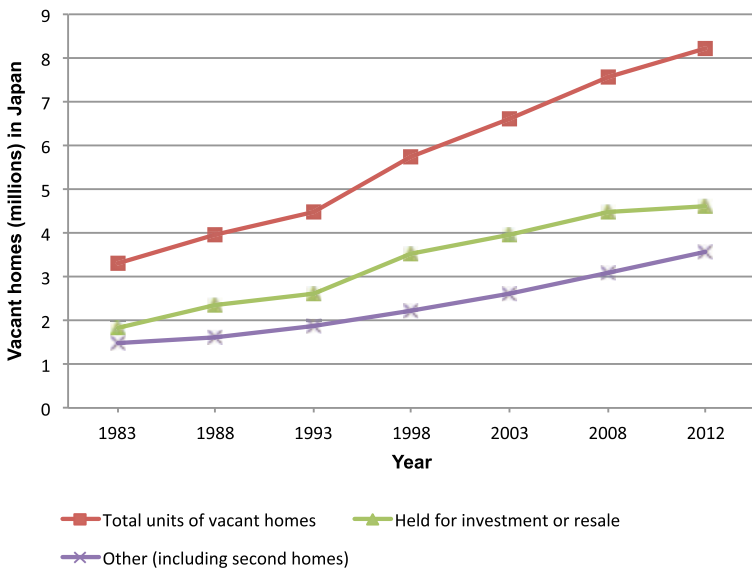
boomers will create a substantial imbalance of supply relative to demand, particularly in states where there is not a substantial, growing, younger generation or large immigrant inflows to absorb the homes for sale” (Myers and Pitkin, 2009: 6; see also Khimm, 2012). Others believe, however, that the “[d]eclining household growth (in the United States) because of increased household dissolutions among the elderly will be spread out over many decades” and may not be as precipitous as is taking place in Japan (Masnick, 2015: 3; Lee, 2014). With these conflicting predictions, Japan’s need to enact the May 26, 2015 Japanese Vacant Housing Law warns of a potentially significant vacant housing problem between 2040 and 2060 when the death-to-population ratio reaches (and depending on the number of immigrants, potentially exceeds) Japan’s current ratio. Thus, the Vacant Housing Law calls us to think about proactive steps we could take today to prevent exacerbating housing issues as the baby boomer generation leaves us.

The Japanese Vacant Housing Law

The decline in Japan’s population described previously has been visibly reflected in Japan’s housing issues. In 2008, there were 7.6 million vacant and abandoned homes in Japan (Mallach, 2014). As of October 2013, there were 8.2 million vacant and abandoned homes in Japan—double the amount of vacant housing in the mid-1990s, reflecting 13.5 percent of Japan’s total housing stock (MLIT, 2015a). Exhibit 3 is an annotated copy of a chart based on Japan’s Ministry of Internal Affairs data showing the increase over time in vacant housing, measured as the percentage of vacant and abandoned housing over other housing, in Japan from 8.6 percent (3.3 million homes) in 1983 to 13.5 percent (8.2 million homes) in 2013 (MLIT, 2015a).

Exhibit 3

General Data on Vacant Housing in Japan Between 1983 and 2012



Source: MLIT (2015b)

In response to this vacant and abandoned housing crisis, on May 26, 2015, the Japanese Diet adopted the Vacant Housing Law with the purpose of addressing the burgeoning abandoned housing problem in the country (see exhibit 4). The 2015 Japanese Vacant Housing Law provides local governments and municipalities in Japan with expansive powers to identify, remediate and repair, or tear down and develop alternative uses for homes abandoned or neglected by senior citizens who have either passed away intestate or who are physically or mentally incapable of taking care of their properties. The core policy reason for adopting the law was to “prevent fires and emergencies, health and sanitation issues, and aesthetic and landscape deterioration that have a severe effect on the living environment and to protect the lives, health, and assets of area residents” (VHL Article 1).

The law thus authorizes local municipalities and wards to develop local response plans based on a survey of vacant housing that identifies and classifies the housing. The plan is then used to implement regulations developed by the Ministry of Land, Infrastructure, Transport, and Tourism and the Ministry of Internal Affairs and Communications (VHL Articles 4–6).

To assist with implementation efforts, local municipalities are authorized to create Consultation Councils constituted of local residents, members of the municipal council, lawyers, real estate agents, and representatives of construction companies and cultural associations (VHL Article 7). Prefectural and national tax authorities are instructed to provide local municipalities with metadata relating to ownership of identified vacant properties (VHL Article 10). The information thus gathered is entered into a database that helps keep track of the housing and of any progress in remediation (VHL Article 11). At the same time, municipal and ward employees are granted authority to enter and search vacant homes to provide notice to identified owners of the remedial actions needed to return properties to habitable use (VHL Articles 9, 12–14). Owners who are unable or unwilling to take remedial measures are summoned to (or may petition for) a public hearing at which they can present evidence and witnesses in their defense (VHL Article 14). Failure to comply will result in fines (VHL Article 16).

Exhibit 4

Overgrown Vegetation, Trees, and General Dilapidation Are Clearly Visible at an Abandoned Home in the Middle of Nakano Ward in Tokyo



Source: Sato (2015)

To raise funds for the attendant costs, the national government is authorized to implement a national tax, and municipalities and wards are authorized to charge attendant fees and levy local taxes (VHL Articles 15–16). Properties that revert to the local municipality or ward because owners cannot be identified or refuse to take action can be demolished or refurbished, but they may be used only for cultural, social, or governmental purposes that are predefined in an approved smart growth-oriented response plan (VHL Article 1).²

Proactive Responses for America’s Elderly Citizens

Proactive measures for America’s elderly citizens will allow policymakers and planners to formulate policies for predicted shrinking demographics and shifting housing needs beginning over the course of the next 25 years and lasting through 2060. Designing policies around death, in both Japan and the United States, touches on deep cultural taboos. Thus, it is much easier to design policies for elderly citizens that focus on facilitating graceful aging, including the preference of aging-in-place. A response to the shifting demographics of a rapidly expanding elderly population and contracting younger population that includes planned, focused, and proactive policies must nevertheless include a “what then?” calculus. By including the “and after death?” question into housing policy, policymakers can make the tough optimization choices necessary to spare the next generation the costs of dealing with a national blight problem similar to the one currently experienced in Japan. (Compare Faiola, 2006; Johnston, 2015; Mallach, 2014; Myers and Ryu 2008.)

Database

While the demographic changes in Japan led to the adoption of the Vacant Housing Law, policymakers in the United States need appropriate tools to understand and predict the likelihood of a similar housing crisis here in the United States. Census data and U.S. Department of Housing and Urban Development’s (HUD’s) United States Postal Service (USPS) Vacant Address Data allow for a strong understanding of the number of vacant housing units. Interactive tools, such as the Mapping America’s Futures tool developed by the Urban Institute, have been used by municipalities to provide equally valuable information on the demographic shifts that municipalities and local governments face (Mayor’s Housing Task Force, 2014; Pendall, 2015; see also Misra, 2015).

Similar data in Japan have proven insufficient, however, to predict and prevent the current vacant and abandoned housing problem. Thus, the Vacant Housing Law requires localities to build and gather databases that track vacant and abandoned housing and include information on the status of the homes and their ownership and relevant metadata from tax rolls and similar governmental sources (VHL Article 11). The information gathered enables Japanese municipalities to sort out which homes are truly abandoned and which are neglected (VHASS, 2015). Local governments

² By October 2015, the Ministry of Land, Infrastructure, Transport, and Tourism had promulgated regulations and guidelines as required under VHL Article 10 (providing municipalities with metadata from the tax rolls), Article 14 (hearing rules and concomitant forms), and Article 15 (regarding the appropriate use of fees and tax levies raised). At the same time, the Ministry (in association with a nonprofit organization created for the purpose) was administering a publicly accessible Internet page that details implementation progress in municipalities throughout the country. The website also identifies abandoned homes and provides information as to whether they are available for purchase by private parties before reverting to a local municipality (*Akiya Soudann Madoguchi Kennsaku Shisutemu*, “Vacant Housing Assistance Service System [VHASS],” 2015).

then can (1) require identified owners to remove nuisances by repairing or remediating the properties or (2) acquire properties that have been abandoned and convert them for public uses. A similar database in the United States could inform local governments' development of specific responses that target local and regional responses to addressing blight. (See Lee, Terranova, and Immergluck, 2013, discussing the Vacant Property Registration Ordinance Database that keeps track of local vacant property registration ordinances in the United States. See also Popper and Popper, 1987, who proposed converting abandoned areas of the Great Plains into a "Buffalo Commons" and Mallach, 2015, who proposed converting vacant and abandoned properties in urban environments to agricultural uses).

Immigration

It is not altogether clear how immigration patterns will change in the United States and how these potential changes might affect the housing market. Even if immigration increases, immigrants will not necessarily settle in areas currently inhabited with large aging adult populations. Given the current political climate, it could well be that decisionmakers will prefer a more restrictive immigration policy. In either case, immigration policy outcomes could increase the death-to-population ratio and (without appropriate and preventative land use planning) intensify a potential abandoned housing problem as profound as the one that Japan faces today (See, in general, AILA, 2008; FAIR, 2008; Hipsman and Meissner, 2013). A public discussion on the economic effects of immigration from this vantage point is therefore in order. (For discussions of the economic effects of immigration, see, in general, AILA, n.d.; National League of Cities, 2013; Rogers, 2013.)

Conclusion

With a more robust database that builds on census data and HUD's USPS Vacant Address Data, local governments will be in a position to better determine which solutions would be most appropriate in preventing a Japanese-style blight problem. Understanding changing demographics may also inform immigration reform policies. Better information would permit a robust national debate on how to implement policies that could allow for managing properties abandoned by elderly citizens.

Acknowledgments

The author thanks his colleagues and mentors at Bloustein, Fels, and Penn and the Boston University School of Law for their support and encouragement. He also thanks the editors at the U.S. Department of Housing and Urban Development, Alan Mallach, and Erin Nowak, for their support and for reviewing and editing previous versions of this article.

Author

Peter Manda is a Master of Laws degree candidate at the Boston University School of Law.

References

- 2005 White House Conference on Aging (2005 WHCoA). 2005. "Report to the President and the Congress. The Booming Dynamics of Aging: From Awareness to Action." <http://nicoa.org/wp-content/uploads/2014/05/2005-WHCOA-Final-Report.pdf>.
- Akiya Soudann Madoguchi Kennsaku Shisutemu ("Vacant Housing Assistance Service System"). 2015. *Ippan Shadan Houjinn Sumaizukuri Machizukuri Senta-Rennogokai* ("General Incorporated Housing and Community Support Association"). <http://www.sumikae-nichiikikyoju.net/madoguchi/akiya/>.
- American Immigration Lawyers Association (AILA). 2008. "AILA Issue Paper: Comprehensive Immigration Reform." <https://shusterman.com/pdf/aila-issues304.pdf>.
- . n.d. "Economic Outlook Shows Vital Need for Immigrants in U.S. Economy." <https://www.google.com/url?sa=t&rc=j&q=&resrc=s&source=web&cd=1&ved=0CB8QFjAA&url=http%3A%2F%2Fwww.aila.org%2Ffile%2FDownloadEmbeddedFile%2F49643&ei=y8qhVdueIYSgNo2-hugK&usg=AFQjCNGcV4UO9JIF5AKXpdxowStpfrsM7Q&bvm=bv.97653015,d.eXY>.
- Colby, Sandra L., and Jennifer M. Ortman. 2014. *The Baby Boom Cohort in the United States: 2012 to 2060*. Washington, DC: U.S. Census Bureau. <https://www.census.gov/prod/2014pubs/p25-1141.pdf>.
- Faiola, Anthony. 2006. "For Elderly in Japan, a Very Long Winter: Social Changes Leave Many To Fend for Themselves," *The Washington Post*, February 19.
- Federation for American Immigration Reform (FAIR). 2008. "Population Growth (2008)." <http://www.fairus.org/issue/population-growth>.
- Hipsman, Faye, and Doris Meissner. 2013. *Immigration in the United States: New Economic, Social, Political Landscapes With Legislative Reform on the Horizon*. Washington, D.C.: Migration Policy Institute. <http://www.migrationpolicy.org/article/immigration-united-states-new-economic-social-political-landscapes-legislative-reform>.
- Japanese Ministry of Internal Affairs and Communications, Statistics Bureau. 2015. "Population and Households: Deaths and Death Rates by Age." *Japan Statistical Yearbook 2015*. <http://www.stat.go.jp/english/data/nenkan/1431-02.htm>.
- Johnston, Eric. 2015. "Is Japan Becoming Extinct?" *Japan Times*, May 17. <http://www.japantimes.co.jp/news/2015/05/16/national/social-issues/japan-becoming-extinct/>.
- Khimm, Suzy. 2012. "Baby Boomers Will Be Selling Their Homes, but Will Their Children Want To Buy Them?" *The Washington Post*, March 7. http://www.washingtonpost.com/blogs/ezra-klein/post/baby-boomers-will-be-selling-their-homes-but-will-their-children-want-to-buy-them/2012/03/07/g1QAQ05ywr_blog.html?wprss=linkset.
- Lee, Ronald D. 2014. "Macroeconomic Consequences of Population Aging in the United States: Overview of a National Academy Report," *American Economic Review: Papers & Proceedings* 104 (5): 234–239. <https://www.aeaweb.org/articles.php?doi=10.1257/aer.104.5.234>.

- Lee, Yun Sang, Patrick Terranova, and Dan Immergluck. 2013. "New Data on Local Vacant Property Registration Ordinances," *Cityscape* 15 (2): 259–66. huduser.gov/portal/periodicals/cityscape/vol15num2/ch22.pdf.
- Mallach, Alan. 2015. "Vacant Properties: An Urban Epidemic." American Bar Association Section of State and Local Government Law, Spring Meeting 2015: Land Use Planning and Zoning Section panel. "Turning Ugly Ducklings Into Swans: Converting Vacant Municipal Properties Into Urban Gardens and Farms." Slides available at http://www.americanbar.org/content/dam/aba/administrative/state_local_government/TurningUglyDucklings42015.authcheckdam.pdf.
- . 2014. "Japan's Unintentional Social Experiment." *Rooflines: The Shelterforce* (blog), August 5. National Housing Institute. http://www.rooflines.org/3807/japans_unintentional_social_experiment/.
- . 2011. "Building Sustainable Ownership: Rethinking Public Policy Toward Lower-Income Homeownership." *Discussion Papers: Community Development Studies and Education*. Philadelphia: Federal Reserve Bank of Philadelphia.
- Masnack, George. 2015. "What Will Happen to Housing When the Baby Boomers Are Gone?" *Housing Perspectives* (blog), February 17. <http://housingperspectives.blogspot.com/2015/02/what-will-happen-to-housing-when-baby.html>.
- Mayor's Housing Task Force. 2014. "Mayor Martin J. Walsh: Housing a Changing City—Boston 2030." http://www.cityofboston.gov/dnd/pdfs/boston2030/Housing_A_Changing_City- Boston_2030_full_plan.pdf.
- Ministry of Land, Infrastructure, Transportation, and Tourism; Ministry of Internal Affairs and Communications (MLIT). 2015a. *Akiya nado ni kannsuru taisaku wo sogouteki katsu keikakutekini jisshisuru tame no kihonntekina shishinn* ("Basic Policy To Implement a Planned and General Response to Vacant Housing"). <http://www.mlit.go.jp/common/001080537.pdf>.
- . 2015b. *Akiya no jokyaku wo sokushinnsuru tame no tochi ni kakaru koteishisannzei nado ni kannsuru shoyou no sochi* ("Required Measurements Related to Real Estate Taxes on Land in Order To Track Progress on the Elimination of Vacant Homes"). <http://www.mlit.go.jp/common/001091835.pdf>.
- Misra, Tanvi. 2015. "See What Your City Will Be Like in 15 Years," *The Atlantic Citylab*. <http://www.citylab.com/work/2015/01/see-what-your-city-will-be-like-in-15-years/384648/>.
- Myers, Dowell, and John Pitkin. 2009. "Demographic Forces and Turning Points in the American City, 1950–2040," *Annals of the American Academy of Political and Social Science* 626: 91–111.
- Myers, Dowell, and SungHo Ryu. 2008. "Aging Baby Boomers and the Generational Housing Bubble," *Journal of the American Planning Association* 74 (1): 17–33.
- National League of Cities. 2013. "Fix the Nation's Broken Immigration System." *Legislative Issue Brief*. http://www.nlc.org/Documents/Influence_Federal_Policy/Advocacy/Legislative/issue-brief-immigration-Jan2013.pdf.

Nikkei Keizai Shimbun, Topics. 2015. *Anata no machi ha? 'Jinnko Gennshou Chizu' de shiru gennjou to shourai* ("And What About Your Town? Seeing the Present and Future Through the 'Population Reduction Map'"). <http://www.nikkei.com/article/DGXZZO77905320T01C14A0000000/>.

Ortman, Jennifer M. 2013. "U.S. Population Projections: 2012 to 2060." Presentation for the FFC/GW Brown Bag Seminar Series on Forecasting, George Washington University, Washington, DC.

Pendall, Rolf. 2015. *Mapping America's Futures*. Washington, DC: Urban Institute. <http://datatools.urban.org/features/mapping-americas-futures/>.

Popper, Deborah Epstein, and Frank J. Popper. 1987. "The Great Plains: From Dust to Dust," *Planning* 53: 12–18.

Rogers, Marie Lopez. 2013. "Immigration Reform Tied to Future of Cities, Towns." <http://archive.azcentral.com/opinions/articles/20130207immigration-reform-cities-future.html>.

Sato, Asato. 2015. Personal communication and photos. Google coordinates: <https://www.google.com/maps/place/2+Chome-16-5+Nakano,+Nakano-ku,+Tōkyō-10+164-0001,+Japan/@35.703313,139.6691825,19z/data=!3m1!4b1!4m2!3m1!1s0x6018f29515f3f761:0xcc7d819685a15829>.

U.S. Census Bureau. 2014. "National Population Projections: Projected Deaths by Single Year of Age, Sex, Race, Hispanic Origin, and Nativity for the United States: 2014 to 2060." <http://www.census.gov/population/projections/data/national/2014/downloadablefiles.html>

SpAM

SpAM (Spatial Analysis and Methods) presents short articles on the use of spatial statistical techniques for housing or urban development research. Through this department of Cityscape, the Office of Policy Development and Research introduces readers to the use of emerging spatial data analysis methods or techniques for measuring geographic relationships in research data. Researchers increasingly use these new techniques to enhance their understanding of urban patterns but often do not have access to short demonstration articles for applied guidance. If you have an idea for an article of no more than 3,000 words presenting an applied spatial data analysis method or technique, please send a one-paragraph abstract to rwilson@umbc.edu for review.

Predicting Local Crime Clusters Using (Multinomial) Logistic Regression

Martin A. Andresen
Simon Fraser University

Abstract

Understanding hot spots of crime has been a concern of spatial criminology for nearly 200 years. A number of methods are used to identify and calculate hot spots, such as dot maps, surface interpolation (kernel density estimation), and statistically identified cluster analysis. Relating to the latter set of methods, local Moran's I is one of the more commonly used methods for identifying local crime clusters. One important aspect for this method for subsequent analysis is that it uses areas, such as census boundary units, to identify local clusters of crime. Consequently, census data may be used to predict and better understand these local crime clusters. In this article, I use multinomial logistic regression and census variables to predict the local crime clusters identified by local Moran's I. This analysis shows a number of nuances regarding local crime clusters, and the spatial patterns of crime, more generally, can be identified using this two-stage approach. Such an approach provides a better understanding of spatial crime patterns than the more common regression methods.

Introduction

The use of spatial statistical methods in criminology is becoming increasingly popular with the recognition that spatial relationships exist in crime data that complicate traditional statistical estimation. In addition, spatial statistical techniques are becoming applied more because of their ability to ask new questions to better understand the criminal event. One of these statistical techniques is explicitly spatial and considers local relationships: local crime clusters using local Moran's *I*. This technique and others like it are "local" because, rather than calculating a statistic for the entire study area (a "global" statistic), it calculates a statistic for each unit under analysis (census tracts, for example).

Although instructive on their own, simply with their identification, a better understanding of the determinants of these local crime clusters is of interest. This article discusses a two-stage research methodology that incorporates local crime clusters and their prediction. The two-stage approach first identifies the local crime clusters in the study area and then a multinomial logistic regression to identify the predictors for each local crime cluster type. This two-stage approach shows that the output from a local Moran's *I* and multinomial logistic regression is able to identify factors specific to each local crime cluster type beyond what would be found in more common regression analyses.

Data and Methods

The data for the following analyses come from Vancouver, British Columbia, Canada. Vancouver, a city on the west coast of Canada, just north of the U.S. border with Washington State, is Canada's third largest census metropolitan area (CMA), with a population that is slightly more than 2.3 million. Although Vancouver is the third largest CMA in Canada, it historically has had the highest crime rate for CMAs in the country. In 2013, for example, Vancouver's total crime rate (excluding traffic) was 6,897 per 100,000 residents, more than double the crime rate in the largest Canadian CMA, Toronto, at 2,941 per 100,000 (Boyce, Cotter, and Perreault, 2014).

Data and Spatial Units of Analysis

The spatial unit of analysis for Vancouver is the dissemination area (DA). The DA, approximately the size of a census block group in the U.S. census, contains approximately 400–700 people and is composed of one or more blocks. In Vancouver, 990 DAs are used for our analyses.

Crime data for Vancouver consist of calls for service made to the Vancouver Police Department for the year 2001. The calls-for-service database is the set of calls made to the Vancouver Police Department directly, calls allocated to them through the 911 emergency system, and calls made by police while on duty. These data consist of automotive theft (theft of vehicle and theft from vehicle), burglary (commercial and residential), and violent crime (assault, fighting, holdups, homicide, robbery, sexual assault, and stabbing). Each call for service includes the complaint code/description, listed previously, and the location in the form of an address of the call for subsequent mapping and spatial analysis. These data were geocoded to the street network with a 94-percent success rate, exceeding the minimum acceptable hit rate of 85 percent set by Ratcliffe (2004).

The census data used in the inferential analyses represent the appropriate years of crime data for Vancouver (2001) at the dissemination area level. These explanatory variables were chosen, in both cases, to represent social disorganization theory and routine activity theory (Cohen and Felson, 1979; Shaw and McKay, 1942). For the Vancouver analyses, the following variables were employed: population change, percent; males ages 15 to 24, percent; single-parent families, percent; recent immigrants, percent; ethnic diversity, measured using the Blau Index; unemployment, percent; post-secondary completion, percent; average income in thousands of dollars; population density; average dwelling value in thousands of dollars; rentals, percent; and housing in major repair, percent.

Local Indicator of Spatial Association, Local Moran's *I*

As mentioned previously, local Moran's *I* has been used in a number of criminological contexts to represent local clusters of crime. Local Moran's *I* is classified as a local indicator of spatial association (LISA) because it is mathematically related to the global spatial statistic, Moran's *I* (Anselin, 1995). This local spatial statistic identifies spatial clustering at the local level, and it does so for each spatial unit of analysis, indicating if each spatial unit of analysis is surrounded by similar or dissimilar values.

The local Moran's *I* statistic is calculated in the following manner—

$$I_i = \frac{(x_i - x^*) \sum_j w_{ij} (x_j - x^*)}{\sum_i (x_i - x^*)^2 / n}, \quad (1)$$

where x_i is the value of variable x in spatial unit i , x^* is the mean of x , n is the number of spatial units, and w_{ij} is the spatial weights matrix that measures the strength of the relationship between two spatial units. In the analyses that follow, spatial weights are defined using first order Queen's contiguity, such that all spatial units that are contiguous are considered neighbors, even if they touch at only a corner. The local Moran's *I* statistic ranges from -1 (perfect negative spatial autocorrelation) to +1 (perfect positive spatial autocorrelation). These values are then used for the following classifications of local clusters: high-high, low-low, low-high, and high-low. High-high and low-low represent local positive spatial autocorrelation, high crime-rate areas surrounded by other high crime-rate areas (hot spots of crime) and low crime-rate areas surrounded by other low crime-rate areas (cool spots of crime), respectively. Low-high and high-low represent local negative spatial autocorrelation, low crime-rate areas surrounded by high crime-rate areas and high crime-rate areas surrounded by low crime-rate areas, respectively. A final and fifth classification also represents no statistically significant spatial clustering.

Maps of these local crime clusters are shown in the Presence and Prediction of Local Crime Clusters section, but these local crime cluster classifications can also be used as the dependent variable in a regression context. As such, we can use a regression technique—(multinomial) logistic regression—to predict the respective categories.

(Multinomial) Logistic Regression

Because of the categorical nature of the local crime cluster classifications, a logistic regression model is appropriate (McFadden, 1974, 1981). It is important to note, however, that just because it is possible to have a total of five local crime cluster classifications, it does not mean that all will be present in any given analysis. In the research on Vancouver, as shown in the following section, all five classifications are present.

Multinomial logistic regression is a statistical technique that specifies the dependent variable as a category, rather than as a continuous or count-based variable. The local crime clusters that emerge from local Moran's I is an example of such data. The output from the multinomial logistic regression is a set of parameters for each local crime cluster (no statistically significant clustering is used as the base category in the regression model) that can then be used to calculate the probability that each local crime cluster will occur, given the values of the estimated parameters and the explanatory variables.

The formula for the multinomial logistic model is as follows—

$$\text{Prob}(Y = J) = \frac{e^{X\beta_j}}{1 + \sum_{i=1}^J e^{X\beta_i}}, \quad (2)$$

where Y is the local cluster type, $J + 1$ is the number of alternatives, e is the natural exponential function, X is the matrix of independent variables, and β_i are estimated parameters. In the analyses that follow, all variables listed previously are included in the specification of equation 2. All statistically insignificant variables subsequently were removed using t -tests for individual variables and likelihood-ratio tests for joint significance tests. This step is done to avoid removing relevant variables because of multicollinearity.

One final consideration for the multinomial logistic regression model is how to assess the individual impact of explanatory variables on the probability that each local crime cluster will occur. This assessment can be done by calculating marginal effects or using odds-ratios. Because of the nonlinearity of the multinomial logistic functional form, the marginal effect of the explanatory variable on the outcome variable is not constant. Rather, the marginal effect is a function of all the estimated coefficients in the model (Greene, 2000; Kennedy, 2003; Wooldridge, 2002). For example, the change in probability for choice J is the partial derivative of the likelihood function with respect to x_i —

$$\text{Marginal Effect} = \beta_j e^{X\beta_j} (1 + e^{X\beta})^{-2}. \quad (3)$$

Clearly evident from equation 3 is that the partial derivative varies with the values of *all* the explanatory variables in the estimated model. As such, a choice needs to be made regarding which values for the explanatory variables should be used to calculate the marginal effects. The most common method of calculating the marginal effects is to use the average values for each explanatory variable (Kennedy, 2003). The marginal effects shown in the tables that follow are calculated in this manner.

The odds-ratio is the exponential function of the estimated parameter, e^β . This odds-ratio represents the relative change in the probability of a local crime cluster when one unit increases the value of

the explanatory variable. For example, if the odds-ratio is 1.20, then a one-unit increase in an explanatory variable leads to a 20-percent increase in the probability of a local crime cluster; similarly, if the odds-ratio is 0.80, then a one-unit increase in an explanatory variable leads to a 20-percent decrease in the probability of a local crime cluster. An example of each method is provided in the following section for the Vancouver analyses. All estimation was undertaken using R: A Language and Environment for Statistical Computing (<http://www.R-project.org>).

Presence and Prediction of Local Crime Clusters

The results for the local crime clusters in Vancouver are shown in exhibits 1 through 4. Exhibit 1 shows the counts and percentages of local crime cluster types for the dissemination areas. Immediately evident is the insignificant clustering that is, by far, the most common result for Vancouver dissemination areas, 77.0 to 83.0 percent. This exhibit clearly shows that hot and cool spots of crime are far from a dichotomy. The high-high clusters account for 4.0 to 8.5 percent of the dissemination areas, whereas low-low clusters account for 8.0 to 14.0 percent of the dissemination areas. As such, in Vancouver, statistically defined cool spots of crime are uncommon but more present than hot spots of crime. The remaining two local crime cluster types, low-high and high-low, represent negative spatial autocorrelation and are rather rare, 1.0 to 2.5 percent of dissemination areas.

Turning to the maps, exhibit 2 shows the local crime clusters for automotive theft in Vancouver. The high-high crime cluster is primarily in the downtown peninsula of Vancouver. The areas to the east and immediately south of the downtown peninsula are Downtown Eastside and False Creek, respectively. These areas of Vancouver have high crime levels, particularly Downtown Eastside, and have not surprisingly been identified as such. The low-low local crime clusters are primarily on the west side of the city, historically the wealthiest area of Vancouver. As with the hot spots of crime, this effect is not a surprise. The negative spatial autocorrelation local crime clusters, high-low and low-high, are scattered around the city. Perhaps most interesting is the presence of low-high local crime clusters in the Downtown Eastside and False Creek areas. These clusters represent “havens” from automotive theft in a crime hot spot.

Exhibit 1

Percentages and Counts of Cluster Types, Dissemination Areas, Vancouver, British Columbia, Canada, 2001

	Cluster Type				
	Insignificant	High-High	Low-Low	Low-High	High-Low
Automotive theft	77.2 (764)	5.6 (55)	14.0 (139)	2.5 (25)	0.7 (7)
Burglary	76.7 (759)	8.6 (85)	8.3 (82)	3.9 (39)	2.5 (25)
Violent crime	82.9 (821)	4.0 (40)	10.7 (106)	1.6 (16)	0.7 (7)

Note: Counts are reported in parentheses.

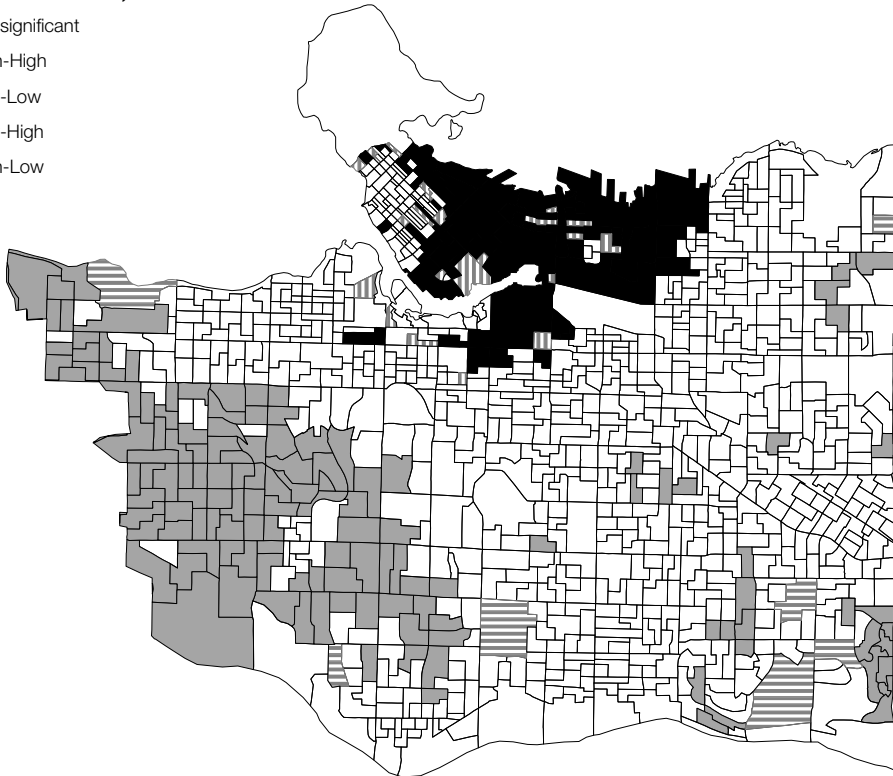
Source: Andresen (2011)

Exhibit 2

LISA Map, Dissemination Areas, Automotive Theft, Vancouver, British Columbia, Canada

LISA classifications, automotive theft

- Not significant
- High-High
- Low-Low
- ▨ Low-High
- ▨ High-Low








*LISA = local indicator of spatial association.
Source: Andresen (2011)*

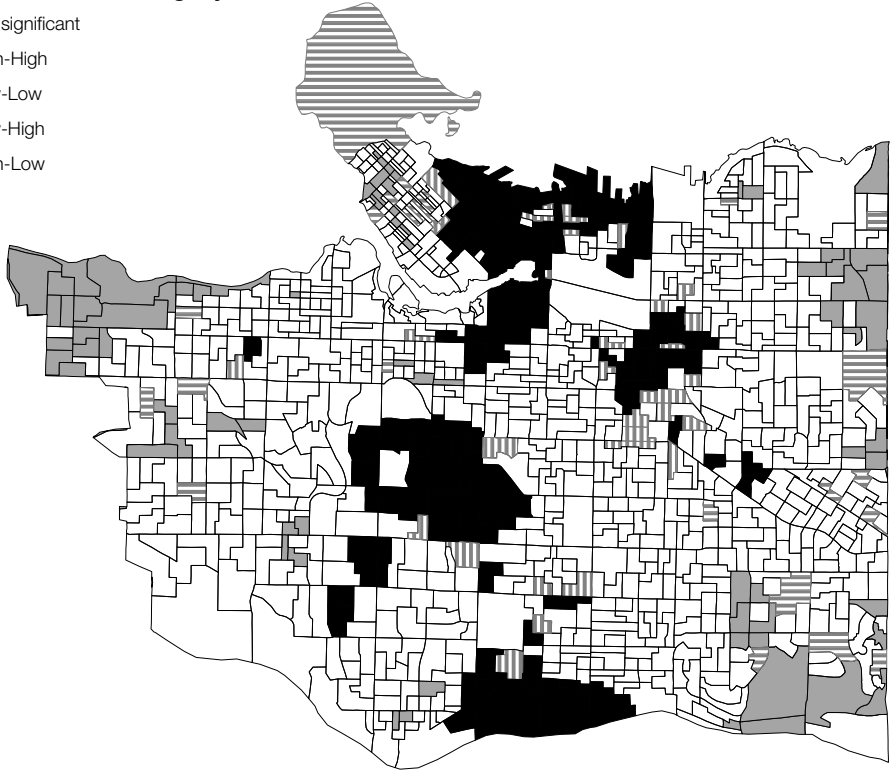
The local crime clusters for burglary in Vancouver, exhibit 3, are in quite different areas of the city, showing the importance of disaggregating crime types, particularly for spatial analyses (Andresen and Linning, 2012). Hot spots of burglary are still present in the poorer areas within the downtown peninsula and the Downtown Eastside. This result is hardly surprising. A hot spot of burglary is in the center of Vancouver, however, in an area that would not be expected. Although the far eastern portion of this hot spot is close to an area that traditionally has higher crime in Vancouver, much of the rest of this hot spot is not. The remaining hot spots are primarily on the east side of Vancouver, traditionally a lower income area of the city. Within the east side of the city, however, some areas have relative affluence; one of these areas has been shown to be a low-low local crime cluster in the southeast corner of the city.

Exhibit 3

LISA Map, Dissemination Areas, Burglary, Vancouver, British Columbia, Canada

LISA classifications, burglary

-  Not significant
-  High-High
-  Low-Low
-  Low-High
-  High-Low



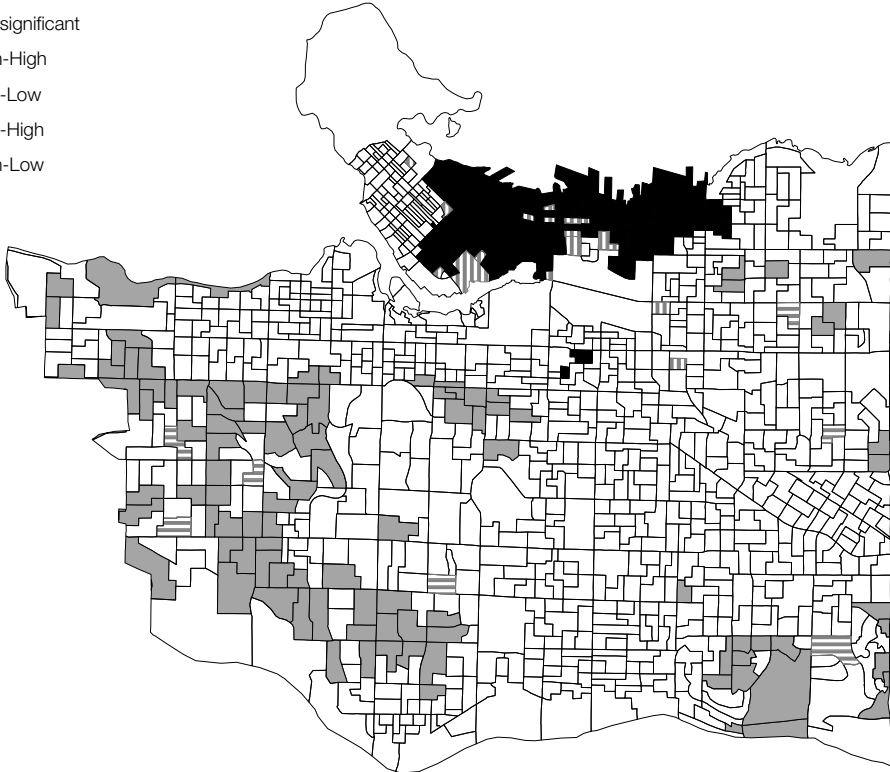
*LISA = local indicator of spatial association.
Source: Andresen (2011)*

Exhibit 4 shows the local crime clusters for violent crime in Vancouver. The overall pattern of automotive theft: low-low local crime cluster in the western portion of the city with a high-high local crime cluster in the downtown peninsula and the Downtown Eastside. One notable difference regarding the violent crime high-high local crime cluster is that it does not extend south like automotive theft but extends farther east along a main thoroughfare in the city—Hastings Street.

The local crime clusters are interesting on their own, but predicting their presence is of greater interest here. Exhibit 5 shows the estimation results from the multinomial logistic regression for automotive theft. We will not go into any detail regarding the magnitudes of the parameters and theoretical predictions here—the reader is referred to Andresen (2011) for such discussions. The most interesting comparisons for our purposes here are high-high versus low-low and low-high versus high-low. What is important to note is that each of these comparisons is not as expected. The most obvious is high-high and low-low, representing hot spots and cool spots of crime, respectively. One normally would expect the same variables to matter for each area but to have the opposite sign. This expectation does not occur here, although some overlap exists. Five explanatory variables

Exhibit 4**LISA Map, Dissemination Areas, Violent Crime, Vancouver, British Columbia, Canada****LISA classifications, violent crime**

- Not significant
- High-High
- Low-Low
- Low-High
- High-Low



LISA = local indicator of spatial association.

Source: Andresen (2011)

do not have a corresponding presence for the other local crime cluster type, and in two cases the signs on the estimated parameters are the same. This result is more pronounced for a comparison between low-high and high-low, having only one common estimated parameter.

Perhaps the most instructive aspects in exhibit 5, and in the subsequent tables as well, are the differences between the calculated marginal effects (middle number inside the brackets) and the odds-ratios (number to the right of the brackets), with the latter being more commonly reported in the literature. According to the odds-ratios, some of the explanatory variables have large impacts on the probability of local crime clusters. For example, in the case of high-high, young males, single parents, and the unemployment rate all appear to have a large impact on the probability: 16- to 18-percent change in probabilities. Remember, however, that these are relative changes that depend on the baseline probability of an event such as a local crime cluster occurring. The marginal effects show that the actual impacts on the probability that a high-high local crime cluster will occur are actually rather small, one-tenth of 1 percent. Such big differences in relative and absolute changes in probability are present for all other local crime cluster types as well. This analysis shows the importance of calculating marginal effects in a (multinomial) logistic regression context.

Exhibit 5

Multinomial Logistic Regression Results, Dissemination Areas, Automotive Theft, Vancouver, British Columbia, Canada

	High-High	Low-Low	Low-High	High-Low
Constant	- 3.848*	- 3.954*	- 6.191*	
Population change (%)		- 0.040 [- 0.240]* 0.96	0.052 [0.051] 1.05	- 0.189 [- 0.000]* 0.83
Males ages 15–24 (%)	- 0.180 [- 0.119] 0.84	0.252 [1.533]* 1.29		
Single-parent families (%)	- 0.198 [- 0.130]* 0.82	0.098 [0.597]* 1.10		
Ethnic diversity	0.029 [0.019] 1.03	- 0.028 [- 0.171]* 0.97		
Unemployment rate	0.159 [0.105]* 1.17		0.116 [0.114]* 1.12	
Post-secondary completion (%)	0.031 [0.021]* 1.03			
Average income (\$ thousands) ^a	0.027 [0.018]* 1.03	0.017 [0.105]* 1.02	0.022 [0.021]* 1.02	
Population density	- 0.006 [- 0.004]* 0.99	0.001 [0.001]* 1.00		- 0.127 [- 0.000]* 0.88
Dwelling value (\$ thousands) ^a	- 0.011 [- 0.007]* 0.99		- 0.005 [- 0.005] 0.99	
Rentals (%)		- 0.018 [- 0.106]* 0.98		
Housing in major repair (%)	- 0.091 [- 0.060]* 0.91	- 0.057 [- 0.349]* 0.95		
Probability of cluster	0.66	6.08	0.99	0.00
Pseudo- <i>R</i> ²	0.337			
Percent correct	82.32			

* $p < 0.05$.

^a In 2001 Canadian dollars.

Notes: Marginal effects, shown in brackets, are calculated using average values. Odds-ratios are below the marginal effect. All retained variables are statistically significant at least at the 10-percent level.

We can also see that the estimated probability, using the average values of the explanatory variables, of any of these local crime clusters is rather low. Only the cluster type low-low has a probability greater than 1 percent of occurring. Finally, the goodness-of-fit for the multinomial logistic regression model is moderate with a pseudo- R^2 of 0.337 and 82.32 percent of its estimates correctly identifying the appropriate local crime cluster type.

Turning to the results for burglary, exhibit 6, the same basic results are present, but with fewer variables: different local crime cluster types have different sets of statistically significant explanatory variables, aside from one case—the odds-ratios do not indicate large magnitude impacts on the prediction of local crime clusters. Notably different is that the probability that each local crime cluster will occur is greater than for automotive theft, except for low-low. In fact, both high-high and low-low have probabilities of occurring that are approximately 6 percent. The goodness-of-fit measures, however, are not as strong as for automotive theft: a pseudo- R^2 of 0.074 and 76.67 percent of its estimates correctly identifying the appropriate local crime cluster type.

Exhibit 6

Multinomial Logistic Regression Results, Dissemination Areas, Burglary, Vancouver, British Columbia, Canada

	High-High	Low-Low	Low-High	High-Low
Constant	- 5.26*		- 2.386*	
Population change (%)	0.026 [0.158]* 1.03			
Males ages 15–24 (%)				0.291 [0.379]* 1.34
Recent immigrants (%)		- 0.031 [- 0.180]* 0.97		- 0.060 [- 0.079]* 0.94
Ethnic diversity	0.025 [0.152]* 1.03	- 0.019 [- 0.110] 0.98		- 0.031 [- 0.040] 0.97
Unemployment rate	0.072 [0.435]* 1.08		0.067 [0.153]* 0.94	
Dwelling value (\$ thousands) ^a	0.003 [0.016]* 1.00	0.001 [0.007] 1.00	- 0.004 [- 0.009] 0.99	- 0.007 [- 0.009]* 0.99
Housing in major repair (%)		- 0.042 [- 0.244]* 0.96		
Probability of cluster	6.06	5.83	2.31	1.30
Pseudo- R^2	0.074			
Percent correct	76.67			

* $p < 0.05$.

^a In 2001 Canadian dollars.

Notes: Marginal effects, shown in brackets, are calculated using average values. Odds-ratios are below the marginal effect. All retained variables are statistically significant at least at the 10-percent level.

The results for violent crime are more similar to those for automotive theft. More statistically significant explanatory variables are retained for this model, particularly for high-high and low-low. Moreover, as with the other two crime types, each local crime cluster type retains a different set of explanatory variables and, when a variable is present in both high-high and low-low, the estimated parameters are opposite in sign, as would be expected. As with automotive theft, the odds-ratios for violent crime local crime clusters are, at times, indicating a large magnitude impact with the actual probability changes being small. The goodness-of-fit values are moderate with a pseudo- R^2 of 0.255 and 83.43 percent of its estimates correctly identifying the appropriate local crime cluster type.

Exhibit 7

Multinomial Logistic Regression Results, Dissemination Areas, Violent Crime, Vancouver, British Columbia, Canada

	High-High	Low-Low	Low-High	High-Low
Constant		- 1.869*	- 6.597*	
Males ages 15–24 (%)		0.116 [0.719]* 1.12		
Single-parent families (%)	- 0.195 [- 0.046]* 0.82			0.376 [0.000]* 1.46
Ethnic diversity	0.032 [0.008] 1.03	- 0.026 [- 0.158]* 0.97		
Unemployment rate	0.161 [0.038]* 1.18	- 0.048 [- 0.299]* 0.95	0.117 [0.066]* 1.12	
Average income (\$ thousands) ^a	0.022 [0.005]* 1.02			
Population density	- 0.006 [- 0.001]* 0.99	0.001 [0.002]* 1.00		- 0.052 [- 0.000]* 0.95
Dwelling value (\$ thousands) ^a	- 0.018 [- 0.004]* 0.98	0.002 [0.009] 1.00		
Rentals (%)		- 0.020 [- 0.125]* 0.98	0.031 [0.017]* 1.03	- 0.054 [- 0.000] 0.95
Housing in major repair (%)	- 0.078 [- 0.018]* 0.93			
Probability of cluster	0.24	6.20	0.56	0.00
Pseudo- R^2	0.255			
Percent correct	83.43			

* $p < 0.05$.

^a In 2001 Canadian dollars.

Notes: Marginal effects, shown in brackets, are calculated using average values. Odds-ratios are below the marginal effect. All retained variables are statistically significant at least at the 10-percent level.

Conclusion

This article investigates not only the use of local indicators of spatial association, specifically the local crime clusters of local Moran's I , but also their prediction. Although such spatial statistical techniques are useful and interesting on their own, it is also important to better understand which variables are associated with each local crime cluster so that we may be able to increase their presence (low-low) or decrease their presence (high-high).

The advantage of using this two-stage approach versus an ordinary least squares (OLS) or spatial regression of crime rates is that the local crime cluster technique specifically identifies the areas of interest: hot spots of crime and cool spots of crime. Then, the multinomial logistic regression can estimate the set of parameters that matter specifically for each local crime cluster type, whereas an OLS or spatial regression will estimate one parameter to be used for all areas. In addition, this estimation technique does not provide estimates for the places that do not exhibit spatial clustering, but these places are used as the base category for estimating the probability of the statistically significant local crime clusters.

Overall, this combination of techniques has proven instructive. This article found that different explanatory variables matter for the different local crime clusters, something that would not emerge in an OLS or spatial regression context. This result allows for a higher degree of understanding with the nuances of crime in an urban environment.

Acknowledgments

The author thanks Ronald Wilson for his discussions regarding this article.

Author

Martin A. Andresen is a professor in the School of Criminology and is the Associate Director of the Institute for Canadian Urban Research Studies at Simon Fraser University.

References

- Andresen, Martin A. 2011. "Estimating the Probability of Local Crime Clusters: The Impact of Immediate Spatial Neighbors," *Journal of Criminal Justice* 39 (5): 394–404.
- Andresen, Martin A., and Ehsan Jozaghi. 2012. "The Point of Diminishing Returns: An Examination of Expanding Vancouver's Insite," *Urban Studies* 49 (16): 3531–3544.
- Andresen, Martin A., and Shannon J. Linning. 2012. "The (In)appropriateness of Aggregating Across Crime Types," *Applied Geography* 35 (1–2): 275–282.
- Anselin, Luc. 1995. "Local Indicators of Spatial Association—LISA," *Geographical Analysis* 27 (2): 93–115.

- Boyce, Jillian, Adam Cotter, and Samuel Perreault. 2014. *Police-Reported Crime Statistics in Canada, 2013*. Ottawa, Ontario, Canada: Canadian Centre for Justice Statistics, Statistics Canada.
- Cohen, Lawrence, and Marcus Felson. 1979. "Social Change and Crime Rate Trends: A Routine Activity Approach," *American Sociological Review* 44 (4): 588–608.
- Greene, William H. 2000. *Econometric Analysis*, 4th ed. Upper Saddle River, NJ: Prentice Hall.
- Kennedy, Peter. 2003. *A Guide to Econometrics*, 5th ed. Malden, MA: Blackwell Publishing.
- McFadden, Daniel. 1981. "Econometric Models of Probabilistic Choice." In *Structural Analysis of Discrete Data Using Econometric Applications*, edited by Charles F. Manski and Daniel McFadden. Cambridge, MA: MIT Press: 198–272.
- . 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*, edited by Paul Zarembka. New York: Academic Press: 105–142.
- Ratcliffe, Jerry H. 2004. "Geocoding Crime and a First Estimate of a Minimum Acceptable Hit Rate," *International Journal of Geographical Information Science* 18 (1): 61–72.
- Shaw, Clifford R., and Henry D. McKay. 1942. *Juvenile Delinquency and Urban Areas: A Study of Rates of Delinquency in Relation to Differential Characteristics of Local Communities in American Cities*. Chicago: University of Chicago Press.
- Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Evaluation Tradecraft

Evaluation Tradecraft presents short articles about the art of evaluation in housing and urban research. Through this department of *Cityscape*, the Office of Policy Development and Research presents developments in the art of evaluation that might not be described in detail in published evaluations. Researchers often describe what they did and what their results were, but they might not give readers a step-by-step guide for implementing their methods. This department pulls back the curtain and shows readers exactly how program evaluation is done. If you have an idea for an article of about 3,000 words on a particular evaluation method or an interesting development in the art of evaluation, please send a one-paragraph abstract to marina.l.myhre@hud.gov.

Fair Housing Testing: Selecting, Training, and Managing an Effective Tester Pool

Claudia L. Aranda
Sarale H. Sewell
Urban Institute

Abstract

The paired-testing methodology originated as a tool for fair housing enforcement and has been used in the multiple housing discrimination research studies funded by the U.S. Department of Housing and Urban Development since the late 1970s. In a paired test, testers who are comparably matched on personal, financial, and homeseeking characteristics—except for the characteristic being investigated, such as race or ethnicity—independently record information received by a housing provider. Each tester in the pair collects data that can detect and document the incidence and forms of discrimination at multiple points in the homeseeking process. Whether a fair housing testing study is designed for enforcement or research purposes, its successful implementation requires an effective tester pool. This article highlights important steps in tester selection, training, and management, all of which have been executed by the Urban Institute’s Field Operations Team since the spring of 2011 while supervising the completion of more than 13,000 paired tests across multiple housing discrimination studies regarding race and ethnicity, familial status, disability, sexual orientation, and gender identity.

Introduction

Since the late 1970s, the paired-testing methodology has been used in housing discrimination studies funded by the U.S. Department of Housing and Urban Development (HUD) to measure patterns of adverse treatment across the housing market. The methodology has been adapted for both research and enforcement purposes to investigate differential treatment on the basis of specific characteristics. For example, in a test designed to estimate the level of discrimination against families with children, two comparably qualified homeseekers—one with children and one without children— inquire about available housing. Each tester documents the information he or she obtains and the level of service provided—from contacting an agent; securing an appointment; meeting with an agent to view available units; and learning about move-in dates, monthly rent, security deposits, utilities, and any required fees. The results of the paired tests are then compared to determine whether and how the treatment experienced by testers with children differs systematically from that experienced by testers without children. Since forms of discrimination can be less blatant than they once were, housing testing studies can reveal important insights into marketwide behaviors and uncover systemic practices that would otherwise go undetected.

Since the spring of 2011, the Urban Institute’s Field Operations Team has supervised the completion of more than 13,000 paired tests across multiple housing discrimination studies (HDS) on race and ethnicity (HDS2012), familial status (HDS-Families), disability (HDS-Disabilities), and sexual orientation and gender identity (HDS-LGT). During the course of these studies, the Urban Institute contracted with testing organizations based in more than 40 cities across the country to coordinate tests. Although most of these groups have been fair housing organizations with active testing programs, some have had limited or no previous testing experience.¹ For all these HDS studies, the Field Operations Team was led by a director of field operations and regional coordinators based at the Urban Institute who were responsible for training local test coordinators, overseeing tester recruitment, training testers, supervising testing and test report preparation, reviewing test reports, maintaining daily contact with test coordinators at each site, and monitoring incoming data (submitted via an online data collection system). Careful oversight and regular communication enabled the Field Operations Team to anticipate operational challenges and correct problems as soon as they developed at any study site. The implementation lessons of the many HDS studies can help illuminate the “best practices” in building and sustaining a tester pool capable of completing the meticulous work that paired testing requires. The successful completion of any fair housing testing study requires (1) a careful tester-selection process, (2) a rigorous training program, and (3) and effective management, all of which are discussed in the forthcoming sections.

Tester Selection

One of the first tasks any fair housing testing study must accomplish is the successful recruitment and selection of capable and committed testers. On each of the HDS studies, project staff have expended considerable effort recruiting testers who could be matched on age, gender, and other

¹ For the HDS-Disabilities study, more than one-third of the local testing organizations were disability advocacy groups and centers for independent living with no previous testing experience.

relevant characteristics to compose suitable tester pairs. Even organizations with robust testing programs have needed to recruit additional testers to complete the required number of HDS tests, particularly because some studies have required sites to conduct between 200 and 600 in-person paired tests. Selected testers ideally will have sufficient availability to complete multiple tests on a study.² On more recent HDS studies, the project team has established caps for the number of tests any single tester can conduct. A tester cap can help limit the extent to which the characteristics and behavior of any tester or tester pair can affect the study findings. Because of the amount of tester attrition that testing organizations experienced during HDS2012, groups participating in subsequent studies have been strongly encouraged to recruit 15 to 20 percent beyond the expected recruitment goals. Across all the recent HDS studies, the level of tester attrition has been attributed, in part, to the chronic underemployment of testers—when testers are offered permanent employment, they leave their short-term, part-time jobs as testers. Some level of tester attrition is also expected immediately after the tester training session (when testers learn how detailed the protocols are) and after testers conduct their first practice test (when some testers realize they are uncomfortable assuming a set of assigned but untrue characteristics). Because the recruitment and training of testers is laborious and costly, the assembly of a large tester pool early in the study can help forestall the need for subsequent recruitment drives and trainings, which can delay the completion of testing.

When conducting outreach for testers, HDS local testing organizations have been strongly encouraged to delve deeply within their existing networks of social service agencies, community groups, student associations, and nonprofit organizations to identify prospective tester candidates. To achieve the diverse tester pool that HDS required, testing organizations also expanded their typical recruitment efforts, forging new relationships with organizations and community leaders.³ For example, each HDS study established specific targets for racial and ethnic representation based on metropolitan area census data. Achieving these goals, particularly the goal numbers for Hispanic and Asian-American testers, proved difficult for many sites. As a result, the project team based at the Urban Institute provided support to local testing organizations by connecting with Hispanic and Asian-American community groups and national umbrella organizations to try to increase the diversity of the tester pool. Given the level of confidentiality that fair housing testing requires, project staff exercised vigilance at every step during the recruitment process to ensure that partner organizations helping identify prospective testers understood the protocols to guard against disclosure. General advertising was strongly discouraged, and anytime organizations prepared e-mails or flyers for targeted outreach to specific groups or communities, materials were submitted for approval to the project management team. The comprehensive communication strategy avoided the use of such terms as “testing,” “fair housing research,” and other terms that might disclose the sensitive nature of the work.

² Depending on the number of tests to be conducted and the size of the study’s budget, it may not be worthwhile to expend the time and resources necessary to train a tester who may have availability to conduct only a few tests.

³ Unlike some enforcement organizations, which only use testers who are obviously of their race or ethnicity, HDS testers have been diverse in skin color and accents. After the fieldwork for HDS2012 was completed, a team of coders assessed the racial and ethnic identifiability of each tester based on the tester’s name, recorded voice, and a standardized photograph. When overlaid with testing data, the identifiability analysis showed that minorities whose ethnicity is more readily identifiable experienced more discrimination than those who could be mistaken for White (Turner et al., 2013).

After outreach efforts identified a group of prospective testers, project staff conducted in-person interviews with candidates to determine which individuals were capable of fulfilling the responsibilities of the role. Because the testing process involves complex assignments and detailed protocols, testers must be selected carefully, according to their ability to perform the work. Project staff should consider the following criteria when assessing whether an applicant can be selected as a tester.

- **Affiliation with the housing industry.** Because of the sensitive nature of the work, applicants who wish to be testers should not work for or have immediate family who work for any segment of the housing industry, such as property management companies, insurance companies, appraisal companies, real estate firms, lending institutions, or other housing providers. When HDS testers notify project staff of such an affiliation, staff thank the applicant for his or her time and terminate the interview.
- **Confidentiality.** A successful testing program requires confidentiality—if housing providers were to learn that testing was under way, they could temporarily alter their practices or intensify efforts to identify potential testers. During the HDS interview process, project staff explain to potential testers that the information shared during the interview is to be kept strictly confidential, regardless of whether a candidate is ultimately selected to participate. After test coordinators determine the candidate meets the initial eligibility criteria, applicants are required to sign a confidentiality agreement.
- **Objectivity.** The importance of objectivity should be stressed during the initial stages of candidate interviews. Prospective testers must be able to conduct each test without making any assumptions about which housing providers are more likely to discriminate. Testers should be able to make fair and honest assessments of their experiences. Candidates who are unsure of their ability to remain objective throughout the testing process should not be considered.
- **Ability to be matched.** The selection of HDS applicants also hinges, to a large extent, on the ability to form tester pairs based on key characteristics, such as race and ethnicity, gender, and age. The personalities of testers also should be considered in the creation of pairs—individuals who are more outgoing should be matched with each other while more passive or reserved individuals should be established as matched pairs.
- **Ability to play a role.** Testers will be asked to assume certain personal and financial characteristics on tests that do not necessarily match their own. This set of characteristics includes an assigned household composition (marital/relationship status, number of children in the household, etc.), assigned employment (occupation, name and address of employer), and an assigned household income (the combined income of the tester and of any spouse/partner assigned that will be sharing the household). During the interview process, project staff will introduce the study and explain the role of the tester and expectations for the study. Testers are instructed to wear “clean and casual” attire and are expected to behave appropriately and credibly when playing the role of an interested homeseeker. If applicants are not willing or able to assume these characteristics on tests, then they should not be used for the study.

Training

Regardless of the testing experience of organizations participating in HDS, they all have participated in a comprehensive training program with in-person training sessions and supplemental webinars, designed in consultation with each study's expert advisors and conducted by Urban Institute staff. The training program helps prepare local project managers and test coordinators to skillfully coordinate tests while providing guidance and support for testers, including those who may require specific accommodations and modifications. For example, during the HDS-Disabilities test coordinator training sessions, project staff outlined the approved modifications to the study protocols. Because note-taking was mandatory for all site visits, testers in wheelchairs with manual limitations were permitted to use other means to write down important information; some testers in wheelchairs used tablets or other electronic devices (which some found easier than writing by hand) or they were permitted to ask the housing provider to help them take notes. By explicitly defining how such modifications could be implemented, the study's field team helped ensure procedures were consistently used among the 30 sites coordinating in-person tests.

As part of the HDS tester training program, all testers participate in an in-person session (typically 5 to 6 hours long) that establishes protocols and guidelines, explains data collection forms, and reviews particularly challenging scenarios that may arise in the field.⁴ The training delineates general codes of conduct and serves as a forum for any questions testers have before their first experience in the field. In addition, role-playing, watching short film clips, and taking pop quizzes increase tester engagement and reinforce key protocols. Testers also are strongly encouraged to read the entire tester manual, which includes a comprehensive index of everything covered during the training and sample test narratives that provide a chronological account of a tester's interaction with a housing provider. The HDS training program also requires that testers complete a practice test from start to finish. During practice tests, testers are briefed by test coordinators on their assignment, conduct a site visit, complete test forms, write a detailed narrative, and attend a debriefing session, exactly as they would during real tests. Testers also are trained on the use of the online data collection system. When training testers for a new study, test coordinators are encouraged to work closely with testers before and after their site visits to ensure that all protocols are followed. The practice test is an important opportunity for test coordinators to provide feedback to testers and to reinforce the high standards the work requires. Depending on the outcome of the initial practice test, some testers may be given a subsequent assignment before they can become an active member of the tester pool. By making this initial investment in time during the training phase of a study, testing coordinators can increase the quality of the tests to be completed.

⁴ Because many HDS study sites have required more than one tester training session, Urban Institute staff have conducted subsequent sessions remotely via webinar. Throughout data collection on various studies, Urban Institute staff also conducted "refresher" sessions, highlighting key protocols via webinar or conference call.

Management

Effective management is another key ingredient to maintaining a sufficient, capable tester pool. The HDS testing organizations that have consistently met the study goals on schedule and within the approved budget have excelled in working with and retaining their testers. The following tenets are important to a successful management strategy.

- **Communication.** Testing organizations that clearly communicate with testers about expectations, study goals and timelines, and payment and reimbursement processes—and any unexpected changes—have had the strongest relationships with their testers. The most successful test coordinators establish the expectation of regular communication during the completion of a test by texting, calling, or e-mailing testers to remind them of a scheduled site visit and requiring testers to contact them as soon as a test has ended. Such timely communication can make the difference between a failed test and one in which both testers successfully complete their site visits. In addition, before and at the conclusion of each test, staff conduct a one-on-one briefing and debriefing session, respectively. Testers are briefed in detail about their assignment to make sure they understand their profile and are reminded of the key protocols. After the tester completes the test, the test coordinator debriefs the tester about his or her experience, reviewing test forms, providing feedback, and answering any questions. If necessary, this session also will serve as an opportunity to review key test protocols if the tester made any errors; the complexity of the study protocols can affect the rate of error, which is usually highest on the initial tests that a tester completes. Even after a particular study has come to an end, organizations can notify testers of agency updates and upcoming testing opportunities, ascertaining any changes to testers' schedules and their ability to accept assignments.
- **Organization.** It is imperative that test coordinators remain highly organized, especially when their local testing organization has a large number of tests to complete. Test coordinators must stay informed about changes to testers' availability to ensure that tests will be completed according to schedule. In addition, test coordinators must perform a quality review of the test forms to address protocol errors with testers immediately after they occur, which will help minimize the number of repeated mistakes.
- **Efficient administrative procedures.** By establishing efficient and timely procedures for scheduling testers, processing invoices, and issuing payments, testing organizations help minimize tester frustration and increase the likelihood testers will accept assignments on an ongoing basis. HDS organizations that have experienced significant delays in issuing tester payments not coincidentally have also experienced a higher level of tester attrition. Groups that process payments according to an established payroll schedule help promote strong tester/test coordinator relations and maximize the use of available testers.
- **Setting expectations.** Throughout the recruitment and selection process of prospective testers, test coordinators can help set the tone for an entire study. For each of the HDS studies, testing organizations have been encouraged to set clear expectations at the outset, providing selected testers with all relevant information about the study, including training dates, compensation, and specific study requirements. By being explicit about the study's processes and rigorous

standards, test coordinators can identify testers who are most capable of completing test assignments and help reduce tester confusion and conflict. In addition, test coordinators must enforce any protocols or study requirements as necessary to ensure that tests are completed in a timely manner and meet the highest standards. Beyond setting expectations for testers' behavior, project staff also must exhibit the same high standards in their own conduct, adhering to study guidelines in test coordination and best practices in tester management.

Conclusion

As the experience of the recent HDS studies demonstrate, fair housing testing studies, whether they are designed for enforcement or research purposes, require the assembly of a capable pool of available testers. Testing organizations must meet the challenge of identifying individuals with the capacity of maintaining confidentiality and objectivity while successfully adhering to study protocols. Selected testers must then complete a rigorous training program, which includes conducting at least one practice test. By making a significant investment of time during the training phase of a study, test coordinators will help testers complete higher quality tests with fewer errors, reducing the time needed for corrections and minimizing tester frustration. Sustaining a robust pool of testers also requires a comprehensive management strategy. Testing organizations must practice effective communication, implement efficient administrative procedures, and set high expectations for both project staff and testers alike. By maintaining a pool of credible, committed testers, testing organizations can successfully complete fair housing testing studies, helping identify emerging industry trends and uncovering discriminatory patterns and practices in both rental and sales markets.

Acknowledgments

The authors thank the many individuals who have served as members of the HDS Field Operations team, including the extraordinary staff based at local organizations across the country. The perseverance of testers, including those who traveled significant distances (sometimes to other cities and states) to help conduct tests, provided an inspiring example for the entire project team based at the Urban Institute. The HDS studies could not have been successfully completed without their energy and dedication.

Authors

Claudia L. Aranda is a senior research associate in the Metropolitan Housing and Communities Policy Center at the Urban Institute.

Sarale H. Sewell is a research associate in the Metropolitan Housing and Communities Policy Center at the Urban Institute.

References

Turner, Margery Austin, Rob Santos, Diane K. Levy, Doug Wissoker, Claudia L. Aranda, and Rob Pitingolo. 2013. *Housing Discrimination Against Racial and Ethnic Minorities 2012*. Washington, DC: U.S. Department of Housing and Urban Development.

Contents

Symposium

Housing Discrimination Today	1
---	---

Guest Editors: Margery Austin Turner and Judson James

Guest Editors' Introduction

Discrimination as an Object of Measurement	3
--	---

What Have We Learned From Paired Testing in Housing Markets?

<i>by Sun Jung Oh and John Yinger</i>	15
---	----

Housing Discrimination Among Available Housing Units in 2012: Do Paired-Testing Studies Understate Housing Discrimination? *by Rob Pitigolo and Stephen L. Ross*

61

Changing Contexts and New Directions for the Use of Testing

<i>by Fred Freiberg and Gregory D. Squires</i>	87
--	----

Targeting Disability Discrimination: Findings and Reflections From the National Study on Housing Discrimination Against People Who Are Deaf and People Who Use Wheelchairs *by Claudia L. Aranda*

103

Other Protected Classes: Extending Estimates of Housing Discrimination

<i>by Margery Austin Turner</i>	123
---------------------------------------	-----

Commentary: Some Thoughts on Field Experiments on Housing Discrimination

<i>From a European View by Ali M. Ahmed</i>	137
---	-----

Commentary: Housing Discrimination Research in the 21st Century

<i>by Samantha Friedman</i>	143
-----------------------------------	-----

Commentary: Expanding the Fair Housing Testing Landscape *by James Perry*

151

Commentary: Testing Benefits Housing Providers and the Industry *by Fred Underwood*....

155

Refereed Papers

.....	159
-------	-----

What Happens to Housing Assistance Leavers?

<i>by Robin E. Smith, Susan J. Popkin, Taz George, and Jennifer Comey</i>	161
---	-----

Do the GSEs Meet the Credit Needs of Underserved Communities of Color?

<i>by Michela Zonta</i>	193
-------------------------------	-----

Departments

.....	219
-------	-----

Data Shop

Measuring Neighborhood Opportunity With AFFH Data <i>by Brent D. Mast</i>	221
---	-----

Graphic Detail

Civil Unrest and Marginalization in Baltimore <i>by John C. Huggins</i>	231
---	-----

Industrial Revolution

Rural America: Perceptions of Residential Energy Retrofits <i>by Nathan Barry</i>	233
---	-----

Foreign Exchange

Preparing Our Housing for the Transition to a Post-Baby Boom World: Reflections on Japan's May 26, 2015 Vacant Housing Law <i>by Peter Manda</i>	239
--	-----

SpAM

Predicting Local Crime Clusters Using (Multinomial) Logistic Regression <i>by Martin A. Andresen</i>	249
--	-----

Evaluation Tradecraft

Fair Housing Testing: Selecting, Training, and Managing an Effective Tester Pool <i>by Claudia L. Aranda and Sarale H. Sewell</i>	263
---	-----

