

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.
