Housing Discrimination in Metropolitan America: Findings from the Latest National Paired Testing Study

Margery Austin Turner
Urban Institute

Stephen L. Ross
University of Connecticut

Working Paper 2003-08

February 2003
Abstract

During the summer and fall of 2000, local fair housing organizations in twenty major metropolitan areas nationwide conducted a total of 4,600 paired tests, directly comparing the treatment that African Americans and Hispanics receive to the treatment that whites receive when they visit real estate or rental offices to inquire about available housing. This study, which was sponsored by the U.S. Department of Housing and Urban Development and conducted by the Urban Institute, provides the most complete and up-to-date information available about the persistence of housing market discrimination against African American and Hispanic homeseekers in large urban areas of the United States today and about the progress we have made in combating discrimination over the last decade.
During the summer and fall of 2000, local fair housing organizations in twenty major metropolitan areas nationwide conducted a total of 4,600 paired tests, directly comparing the treatment that African Americans and Hispanics receive to the treatment that whites receive when they visit real estate or rental offices to inquire about available housing. This study, which was sponsored by the U.S. Department of Housing and Urban Development and conducted by the Urban Institute, provides the most complete and up-to-date information available about the persistence of housing market discrimination against African American and Hispanic homeseekers in large urban areas of the United States today and about the progress we have made in combating discrimination over the last decade.

Background

Paired testing originated as a tool for fair housing enforcement, detecting and documenting individual instances of discrimination. Since the late 1970s, this methodology has also been used to rigorously measure the prevalence of discrimination across the housing market as a whole (Fix and Struyk, 1992). When a large number of consistent and comparable tests are conducted for a representative sample of real estate and rental agents, the results control for differences between white and minority homeseekers, and directly measure patterns of adverse treatment based on a homeseeker’s race or ethnicity.

HDS2000 is the third national paired-testing study sponsored by HUD to measure patterns of racial and ethnic discrimination in urban housing markets.¹ Its predecessors, the 1977 Housing Market Practices Study (HMPS) and the 1989 Housing Discrimination Study (HDS) found significant levels of racial and ethnic discrimination in both rental and sales markets of urban areas nationwide (Wienk et al, 1979; Turner, Struyk, and Yinger, 1991). Moreover, although the 1989 HDS was not designed to yield precise estimates of change in discrimination, it found no evidence that overall levels of adverse treatment against African Americans had declined since 1977. Enforcement tests conducted over the intervening decade have also uncovered countless instances of illegal discrimination against minority homeseekers. Housing discrimination raises the costs of housing search, creates barriers to homeownership and housing choice, and contributes to the perpetuation of racial and ethnic segregation (Yinger, 1995).

¹ HDS2000 will ultimately involve three phases of paired testing, in as many as 60 metropolitan areas. HUD’s goals for the study include rigorous measures of change in adverse treatment against blacks and Hispanics nationwide, site-specific estimates of adverse treatment for major metropolitan
In HDS2000, the basic testing protocols replicated those implemented in the 1989 HDS in order to yield comparable measures of differential treatment. Random samples of advertised housing units were drawn from major metropolitan newspapers on a weekly basis, and testers visited the sampled offices to inquire about the availability of these advertised units. Both minority and white partners were assigned income, assets, and debt levels to make them equally qualified to buy or rent the advertised housing unit. Test partners were also assigned comparable family circumstances, job characteristics, education levels, and housing preferences. They visited sales or rental agents, and systematically recorded the information and assistance they received about the advertised unit and/or other similar units, including location, quality and condition, rent or sales price, and other terms and conditions. Test partners did not compare their experiences with one another or record any conclusions about differences in treatment; each simply reported the details of the treatment he or she experienced as an individual homeseeker.

The results presented here are based on a nationally representative sample of 20 metropolitan areas with populations greater than 100,000 and with significant black and/or Hispanic minorities. This sample of sites was selected from the 25-site sample of metropolitan areas covered by the 1989 Housing Discrimination Study. Black/white testing was conducted in sixteen of the twenty sites, and Hispanic/non-Hispanic testing was conducted in ten. Results are weighted to produce nationally representative estimates.

Methodology

HDS2000 was designed to replicate the 1989 Housing Discrimination Study, provide updated national estimates of adverse treatment against African Americans and Hispanics, and rigorously measure any changes that may have occurred during the 1990s. This section briefly summarizes the sampling procedures, testing protocols, and analysis techniques implemented in Phase I of HDS2000.

Sampling. The results presented here are based on a nationally representative sample of 20 metropolitan areas with population greater than 100,000 and with significant black and/or Hispanic minorities. This sample of sites was selected from the 25-site sample of metropolitan areas, estimates of adverse treatment for smaller metropolitan areas and adjoining rural communities, and new measures of adverse treatment against Asians and Native Americans.

2 HDS2000 is designed to measure the extent to which minority homeseekers experience adverse treatment when they look for housing in urban areas nationwide. The tests conducted for this study were not designed to assemble evidence of discrimination in individual cases. The question of when differential treatment warrants prosecution and the related question of whether sufficient evidence is available to prevail in court can only be resolved on a case-by-case basis, which might also consider other indicators of treatment than those reported here.

3 Selecting the phase I sites from the 1989 sample dramatically improves the precision of national estimates of changes in differential treatment between 1989 and 2000.

4 For a detailed description of the HDS2000 methodology, see Turner et al, 2002.
areas covered by the 1989 Housing Discrimination Study. Metro areas were included in the 1989 sampling frame if the proportion of blacks or Hispanics in their central cities were greater than their national average analogues. Five sites were chosen with certainty because they were major metropolitan areas with large minority populations. For the remaining sites, the probability of site selection depended on the minority population in the metro area. The HDS2000 sample includes all five of the metro areas that were selected with certainty in 1989 (Atlanta, Chicago, Los Angeles, New York, and San Antonio), while the remaining 15 sites were selected based on their minority population sizes. Black/white testing was conducted in 16 of the 20 sites, and Hispanic/non-Hispanic white testing was conducted in ten.

The basic objective of a paired testing study is to observe the relative treatment that housing agents provide to white and minority homeseekers in the private market. In order to measure this agent behavior, one would ideally draw a representative sample of rental and sales agents, where an agent’s probability of selection reflects his or her share of currently available housing units. In addition, however, the sampling methodology needs to incorporate information about the specific housing stock offered by an agent. The reason is that our field protocols require both members of a testing team to be assigned characteristics (such as household size and income) and preferences (such as housing type and location) that correspond to the agent’s available listings. Consistent with previous national testing studies, HDS2000 utilized classified advertisements in major metropolitan newspapers to generate samples of rental and sales agents. Within each metro area, paired tests were triggered using ads from a representative sample of housing units available for sale or rent, randomly selected from the Sunday classified advertisements of the major metropolitan newspaper.6 Specifically, a fresh sample of advertisements was selected from a site’s Sunday newspaper for each week in which testing was conducted.

The weekly ad-sampling methodology offers several important benefits. It yields a representative sample of housing agents who use the major metropolitan newspaper to advertise available units, where an agent’s probability of selection is proportionate to his or her share of all units advertised in this way. Because metropolitan newspapers are readily available (regardless of race, ethnicity, or other characteristics), this sampling frame includes agents who can realistically be accessed by any homeseeker. Secondly, the weekly sampling methodology provides a consistent and credible starting point for each test, tying the characteristics and preferences of testers to housing actually available from the sampled agent, and sending consistent signals from both members of a tester team. Finally, this methodology addresses one of the major ethical concerns about paired testing—that it imposes an unreasonable cost burden on housing agents who have to spend time responding to testers’ inquiries and

5 Selecting the Phase I sites from the 1989 sample dramatically improves the precision of national estimates of changes in differential treatment between 1989 and 2000 relative to drawing an independent sample at each period.

6 Samples were drawn from all pages containing real estate advertising in the major Sunday newspapers (including pages of “display ads,” often full-page ads bought by a single realty company, as well as pages containing actual classified ads).
potentially violates their expectations of privacy regarding these inquiries (Fix and Struyk 1993). By advertising in a widely available outlet, a housing agent is explicitly inviting inquiries from the general public and is implicitly declaring his or her compliance with federal fair housing laws.⁷

Target sample sizes were set at 72 tests per tenure and racial or ethnic group for each metropolitan area. This target was set to ensure that at the metropolitan level, differences between white-favored and minority-favored treatment as small as 5.3 to 7.1 percent could be discerned at a 90 percent level of statistical significance. At the national level, the large pooled sample sizes provide a much higher level of statistical precision, with differences between white-favored and minority-favored treatment as low as 3.0 to 4.8 percent discernible at a 95 percent level of statistical significance.⁸

The HDS2000 population of inference is composed of real estate and other housing agents who interact with households seeking to purchase or rent a home who use Sunday newspaper ads as their entryway into the housing market. In order to represent this population of inference, we developed a model-based weighting approach that balances the sample by stratum using Census 2000 data. The weights rely on a plausible “model,” which assumes that the percentage distribution of minority population across sampling strata (separately for each minority group) reflects the percentage distribution of agents who serve minority homeseekers across those strata. The sampling strata are defined as the sites that were chosen with certainty, the black/white sites, the Hispanic/Anglo sites, and the sites that cover both minority groups.

The stratum weights are based on the minority population represented by all sites in each stratum. Sites that were chosen with certainty are self-representing and have a weight based on their own population. Sites in the other strata represent all eligible metropolitan areas. (all metropolitan areas that have a central city population of 100,000 and minority population greater than the national average for one or both of the groups). All sites within a stratum are given equal weight. The intuition behind the equal weight decision arises from the traditional strategy for creation of finite sample analytic weights for a multi-stage sample. The finite sample weight is simply the product of the inverse selection probabilities for each stage, which in this case are the inverse probability of site selection (one over the minority population share)

⁷ Despite the many advantages of this sampling methodology, relying upon metropolitan newspapers to represent the housing market as a whole has some weaknesses. The 1989 HDS found that substantial geographic areas within metropolitan housing markets were under-represented in the Sunday newspaper advertisements that formed the sampling frame for the discrimination tests. Therefore, subsequent phases of HDS2000 are sampling from multiple advertising sources.

⁸ Not all of the sites achieved their targets in Phase I. In two of the sites in the national sample (Philadelphia and Pittsburgh), test coordinators were not successful in completing the high volume of tests required on a weekly basis, and a substantial number of tests failed to meet the Urban Institute’s quality control standards. Therefore, these sites were required to conduct additional testing during the summer of 2001 in order to yield sufficient tests to report results at the metro level.
and the probability of selection for a specific advertisement (the advertisement volume share).\textsuperscript{9} However, the population of advertisements is not our population of interest. Rather, we are interested in capturing treatment in a representative set of attempted housing transactions. As a result, we shift to a model-based approach where we assume that the minority population share is a good proxy for the importance of a given site in terms of representing housing market transactions for minorities. The first stage weight and the new second stage weight cancel yielding an equal weight for each site.\textsuperscript{10}

\textbf{Field Implementation and Paired Testing Protocols.} The HDS2000 field implementation was directed by two national sub-contractors (Abt Associates, Inc. and Progressive Management Resources) under the supervision of the Urban Institute’s Director of Field Operations. These two entities, in turn, subcontracted with a local fair housing organization in each sampled metro area to conduct the testing. Staff of these local testing organizations, designated as Test Coordinators, were responsible for the day-to-day testing activities, directing testers and ensuring that tests were completed according to established procedures and protocols.

For each advertised housing unit selected for testing, Urban Institute staff prepared a Test Authorization Form (TAF). Local testing organizations were required to use the TAFs they received each week in order, and to begin by making advance calls both to confirm the eligibility of the advertised units and to obtain information needed to make credible test assignments. Personal, household and financial characteristics, along with a detailed set of instructions, were provided to each tester prior to conducting a test. Financial characteristics assigned to testers and housing requests to be made by testers were based on the characteristics of the advertised housing unit to be tested, following detailed guides developed for the Test Coordinators. HDS2000 required testers to make appointment calls for all sales tests and some rental tests. While the standard approach for most rental tests was for the tester to “drop in” rather than making an appointment, appointment calls were required when the sampled advertisement did not provide the location of the available housing, when the advertisement indicated that an appointment was required; or when the advance call indicated that an appointment was required.

During their test visits, testers were trained to inquire about the availability of the advertised housing unit that prompted their visit, similar units (same size and price) that might be available, and other units that might meet their housing needs. They tried to inspect at least

\textsuperscript{9} This approach to generating the analytic weights was followed for the 1989 Housing Discrimination Study. In HDS 2000, we found that advertisement volume provided a very poor description of the underlying housing market. For example, Los Angeles has very low ad volumes and would have had virtually no impact on the underlying national estimates if we had used a finite sampling approach. On the other hand, Macon, GA. had unusually high ad volumes and that site alone comprised approximately 1/3 of information in the finite sample-based national estimates for black-white tests.

\textsuperscript{10} These results are comparable to both unweighted results and results from other model-based weighting schemes that were considered.
three housing units, making return visits or appointments with an agent if necessary, and in sales tests they recorded the address, size, and price of any other units that were recommended to them. In response to questions from the real estate or rental agent, testers provided information about their (assigned) household composition, financial characteristics, employment, and housing needs. They were trained to express no preferences for particular amenities or geographic locations, and they did not submit formal applications, agree to credit checks, or make offers to rent or buy available units. In conjunction with these basic testing protocols, testers were also trained to be convincing in the role of an ordinary homeseeker, obtain as much information as possible from the housing provider about available housing, and take notes in order to remember key information about what occurred during the test and what information was provided by the housing provider.

Following every test visit, each tester was required to complete a set of standardized reporting forms. Test partners did not compare their experiences with one another or record any conclusions about differences in treatment; each simply recorded the details of the treatment he or she experienced as an individual homeseeker. The site visit report forms record observations made by the tester and information provided by the housing provider. For sales tests, in addition to a site visit report form, each tester completed a log of recommended homes. In addition, for a randomly selected sub-set of tests (approximately 10 percent), testers were required to compose test narratives. The test narrative provided a detailed, chronological accounting of the test experience. Testers did not know prior to their conducting a test if a narrative would be required. This served both to ensure that testers were conducting all tests with equal attention to established protocols and procedures, including taking notes, and to ensure against fabrication of tests.

**Using Paired Tests to Measure Discrimination.** Data from a sample of standardized and consistent paired tests can be combined and analyzed to measure the incidence and forms of discrimination in urban housing markets. The remainder of this chapter describes the statistical techniques used to analyze data from Phase I of HDS2000 at both the national and metropolitan level. Specifically, we discuss basic measures of adverse treatment, the challenge of distinguishing systematic discrimination from random differences in treatment, rental and sales treatment indicators, summary indicators, and tests of statistical significance.

A paired test can result in any one of three basic outcomes for each measure of treatment: 1) the white tester is favored over the minority; 2) the minority tester is favored over the white; or 3) both testers receive the same treatment (which may be either favorable or unfavorable). The simplest measure of adverse treatment is the share of all tests in which the white tester is favored over the minority. This gross incidence approach provides very simple and understandable indicators of how often whites are treated more favorably than equally qualified minorities. However, there are instances in which minority testers receive better treatment than their white partners. Therefore, we report both the gross incidence of white-favored treatment and the gross incidence of minority-favored treatment.
Although these simple gross measures of white-favored and minority-favored treatment are straightforward and easily understandable, they may overstate the frequency of systematic discrimination. Specifically, adverse treatment may occur during a test not only because of differences in race or ethnicity, but also because of random differences between the circumstances of their visits to the real estate agency. For example, in the time between two testers’ visits, an apartment might have been rented, or the agent may have been distracted by personal matters and forgotten about an available unit. Or one member of a tester pair might meet with an agent who is unaware of some available units. Gross measures of white-favored and minority-favored treatment include some random factors, and therefore provide upper-bound estimates of systematic discrimination.

One strategy for estimating systematic discrimination, that is, to remove the cases where non-discriminatory random events are responsible for differences in treatment, is to subtract the incidence of minority-favored treatment from the incidence of white-favored treatment to produce a net measure. This approach essentially assumes that all cases of minority-favored treatment are attributable to random factors—that systematic discrimination never favors minorities—and that random white-favored treatment occurs just as frequently as random minority-favored treatment. Based on these assumptions, the net measure subtracts differences due to random factors from the total incidence white-favored treatment. However, it seems unlikely that all minority-favored treatment is the result of random factors; sometimes minorities may be systematically favored on the basis of their race or ethnicity. For example, a minority landlord might prefer to rent to families of his or her own race or a real estate agent might think that minority customers need extra assistance. Other instances of minority-favored treatment might reflect a form of race-based steering, in which white customers are discouraged from considering units in minority neighborhoods or developments. Therefore, the net measure subtracts not only random differences but some systematic differences, and therefore probably understates the frequency of systematic discrimination. Thus, net measures provide lower-bound estimates of systematic discrimination, and they reflect the extent to which the

---

11 We use the term “systematic discrimination” to mean differences in treatment that are attributable to a customer’s race or ethnicity, rather than to any other differences in tester characteristics or test circumstances. This term is not the same as “intentional” discrimination, nor is it intended to mean that these differences would necessarily be ruled as violations of federal fair housing law.

12 Note that it is conceivable that random factors might reduce the observed incidence of white-favored or minority-favored treatment, so that the gross-incidence measure is technically not an absolute upper-bound for systematic discrimination.

13 Even when no statistical pattern of race-based differential treatment is observed, individual cases of discrimination may occur. Specifically, even if the gross incidence of white favored treatment is statistically insignificant, this does not mean that discrimination never occurred, but only that the number of cases was too small to draw any conclusions about systematic patterns across the sample as a whole. Similarly, for variables where the net measure is close to zero, there may in fact be instances of race-based discrimination, even though the overall pattern does not systematically favor one group. See Annex 3 for a discussion on tests of statistical significance.
A visit with a rental or sales agent is a complex transaction, and may include many forms of favorable or unfavorable treatment. This report presents results for a series of individual treatment indicators that reflect important aspects of the housing transaction. Many, but not all, of these indicators are common to both rental and sales tests. Indicators of adverse treatment in rental housing transactions address four critical aspects of the interaction between a renter and a landlord or rental agent: housing availability, inspections, costs, and agent encouragement. Indicators of adverse treatment in sales housing transactions fall into five categories: housing availability, inspections, geographic steering, assistance with financing, and agent encouragement.

In addition to presenting results for all of the individual treatment indicators discussed above, this report combines these individual indicators to create composite measures for categories of treatment (such as housing availability or housing costs) as well as for the transaction as a whole. The first type of composite replicates the approach implemented in 1989. Specifically, tests are classified as white-favored if the white tester received favorable treatment on one or more individual items, while his or her partner received no favorable treatment. Tests are classified as “neutral” if one tester was favored on some individual treatment items and his or her partner was favored on even one item. This approach has the advantage that it identifies tests where one partner was unambiguously favored over the other. But it may incorrectly classify tests as neutral when one tester received favorable treatment on several items, while his or her partner was favored on only one. This approach also classifies tests as neutral if one tester was favored on the most important item while his or her partner was favored on items of lesser significance. Therefore, the 1989 composite methodology may underestimate the overall incidence of differential treatment across indicators, but nonetheless provides a very useful measure of the consistency of adverse treatment.

In addition to the consistency approach, hierarchical composites were constructed by considering the relative importance of individual treatment measures to determine whether one tester was favored over the other. For each category of treatment measures (and for the overall test experience), a hierarchy of importance was established independent of analysis of the testing results. For example, in the availability category, if the white tester was told that the advertised home was available, while the minority was told it was no longer available, then the white tester was deemed to be favored overall, even if the minority was favored on less

---

14 Ondrich, Ross and Yinger (2000) examine a structural approach to closing the gap between the upper and lower bounds provided by the net and gross measures, but their approach only reduced the bounds by a small amount.
15 Again, it is important to emphasize the difference between methods used for the statistical analysis of paired testing results and methods used to assemble or assess evidence of unlawful conduct in an individual case. No pre-determined set of decision criteria can substitute for case-by-case judgments about test results.
important items. The hierarchical composites offer the advantage of reflecting important differences in the treatment of minorities and whites. But because random differences on a single treatment indicator may cause a test to be classified as white-favored or minority-favored, the gross composite measures may over-state the incidence of systematic discrimination. Therefore, we report hierarchical composites for each category of treatment and both consistency composites and hierarchical composites for the overall testing experience.16

Findings and Implications

HDS2000 provides a wealth of new information and analysis about discrimination in rental and sales markets. Specifically, it provides up-to-date measures of the incidence of adverse treatment faced by African Americans and Hispanics in metropolitan housing markets nationwide, and rigorous estimates of change in adverse treatment over the last decade. In addition, estimates of adverse treatment are now available at the metropolitan-area level, allowing advocates and policymakers to assess local conditions. Finally, this study builds on past analyses to advance our understanding of geographic steering, sources of variation in patterns of adverse treatment, and the strengths and limitations of paired testing as a tool for measuring racial and ethnic discrimination.

African Americans and Hispanics Still Face Discrimination. In both rental and sales markets of metropolitan areas nationwide, black and Hispanic homeseekers experience significant levels of adverse treatment, relative to comparable white homeseekers (see Exhibit 1). Our “best estimate” of discrimination is the gross measure for the overall consistency composite which reflects the extent to which whites were consistently favored over their minority partners. It’s value ranges from 17.0 percent for African American homebuyers to 25.7 percent for Hispanic renters. “Upper-bound estimates,” which reflect gross, white-favored treatment according to the overall hierarchical composite, indicate that blacks and Hispanics experienced adverse treatment (compared to equally qualified whites) about half the times that they visited real estate or rental offices to inquire about the availability of housing advertised in the major metropolitan newspaper. Finally, our “lower-bound estimates” of systematic discrimination reflect the net (white-favored minus minority favored) measures for the hierarchical composite, and are statistically significant at about 8 percent for African American renters and homebuyers. Hispanic renters appear to face higher levels of systematic discrimination—15 percent, but the lower-bound estimate of discrimination against Hispanic homebuyers is not statistically significant.

Exhibit 1: National Estimates of Discrimination Against Blacks and Hispanics

16 For complete definitions of treatment measures and hierarchical composites, see Turner et al, 2002.
Discrimination appears to take different forms for African Americans and Hispanics, and in rental and sales markets (see Exhibit 2). For African American renters, the incidence of discrimination is highest for opportunities to inspect housing units, although significant discrimination also occurs for housing availability. Hispanic renters face the highest levels of discrimination for housing availability, with lower (but significant) levels for inspections. For African American homebuyers (like renters), levels of discrimination are highest for housing inspections, but significant discrimination also occurs with respect to geographic steering, assistance with financing, and overall encouragement. In contrast, Hispanic homebuyers only appear to face statistically significant levels of discrimination with respect to steering and financing assistance.
These findings show that African Americans and Hispanics still face serious barriers when they search for both rental and sales housing, and suggest a continued need for fair housing education and enforcement. Ongoing fair housing efforts should clearly focus on all aspects of the housing transaction, not just whether housing units are made available to minority customers. The ability to inspect available housing units represents an important source of adverse treatment for both blacks and Hispanic renters and for black homebuyers. In addition, differences in the assistance with financing that real estate agents provide represents the primary source of adverse treatment facing Hispanic homebuyers.

Although the results presented here provide convincing evidence that discrimination persists in metropolitan rental and sales markets, they do not necessarily measure all the forms of discrimination that may be occurring, nor do they represent all segments of the housing market. Like previous national paired testing studies, HDS2000 is limited in its coverage of urban housing markets and the experience of minority homeseekers. The sample of real estate and rental agents to be tested was drawn from newspaper advertisements, and the economic characteristics of tester teams were matched to the characteristics of the advertised units. However, not all housing units for sale or rent are advertised in major metropolitan newspapers, not all real estate and rental agents use newspaper advertising to attract customers, and not all homeseekers rely upon newspaper advertisements in their housing search. Therefore, results presented here do not necessarily reflect the experience of the typical minority homeseeker, but rather of homeseekers qualified to rent or buy the average housing unit advertised in a major metropolitan newspaper.

### Exhibit 2: Forms of Housing Discrimination

<table>
<thead>
<tr>
<th></th>
<th>Black Gross Upper-Bound</th>
<th>Black Net Lower-Bound</th>
<th>Hispanic Gross Upper-Bound</th>
<th>Hispanic Net Lower-Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rental Treatment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Housing Availability</td>
<td>31.5%</td>
<td>3.9%</td>
<td>34.0%</td>
<td>11.9%</td>
</tr>
<tr>
<td>Opportunities to Inspect Housing</td>
<td>27.5%</td>
<td>8.3%</td>
<td>24.4%</td>
<td>7.2%</td>
</tr>
<tr>
<td>Housing Costs</td>
<td>21.4%</td>
<td>--</td>
<td>21.7%</td>
<td>--</td>
</tr>
<tr>
<td>Agent Encouragement</td>
<td>31.3%</td>
<td>--</td>
<td>32.8%</td>
<td>--</td>
</tr>
<tr>
<td><strong>Sales Treatment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Housing Availability</td>
<td>46.2%</td>
<td>--</td>
<td>46.3%</td>
<td>--</td>
</tr>
<tr>
<td>Opportunities to Inspect Housing</td>
<td>42.9%</td>
<td>8.8%</td>
<td>38.3%</td>
<td>--</td>
</tr>
<tr>
<td>Geographic Steering</td>
<td>11.0%</td>
<td>3.5%</td>
<td>14.7%</td>
<td>5.0%</td>
</tr>
<tr>
<td>Assistance with Financing</td>
<td>36.6%</td>
<td>4.9%</td>
<td>38.6%</td>
<td>14.4%</td>
</tr>
<tr>
<td>Agent Encouragement</td>
<td>31.3%</td>
<td>5.2%</td>
<td>30.6%</td>
<td>--</td>
</tr>
</tbody>
</table>

All reported net measures are statistically significant at a 90% confidence level or higher.
Moreover, the results presented here do not encompass all phases of the housing market transaction. HDS2000, like most paired testing studies, focuses on the initial encounter between a homeseeker and a rental or sales agent. Additional incidents of adverse treatment may occur later in the housing transaction, when a renter submits an application or negotiates lease terms, or when a homebuyer makes an offer on a particular unit or applies for mortgage financing. Finally, HDS2000 is designed to measure the extent to which minority homeseekers experience adverse treatment when they look for housing in urban areas nationwide, not to assemble evidence of discrimination in individual cases. The question of when differential treatment warrants prosecution and the related question of whether sufficient evidence is available to prevail in court can only be resolved on a case-by-case basis, which might also consider other indicators of treatment than those reported here.

**Discrimination Against Blacks and Hispanics Has Generally Declined Since 1989.**

HDS2000 was explicitly designed to rigorously measure changes in levels and patterns of differential treatment since the last national paired testing study sponsored by HUD. The basic testing protocols replicated those implemented in the 1989 HDS. And testing was conducted in a sub-set of the metropolitan areas covered in 1989. Thus, we were able to produce comparable measures of differential treatment for both 1989 and 2000.¹⁷ These measures indicate that the nation has made progress in combating housing market discrimination, achieving significant reductions for black renters, and for both black and Hispanic homebuyers.¹⁸

The precise pattern of change in discrimination varies with tenure and race/ethnicity (see Exhibit 3). For African American renters, the gross incidence of white-favored treatment declined significantly overall and for most categories of treatment. Lower-bound (net) estimates of systematic discrimination also declined significantly for three of the four categories of treatment, although the change in the overall net incidence was not statistically significant. And

¹⁷ Note that these are not exactly the same treatment measures as reported in 1989; HDS2000 refined and strengthened measures to produce more rigorous estimates of differential treatment. Because real estate markets changed dramatically between 1989 and 2000, some changes in testing protocols were required to maintain the integrity of the testing process, and as a result some variables that were used in 1989 to assess the treatment of minority home seekers were no longer relevant. In addition, our understanding of the real estate market has evolved since 1989 leading to different decisions concerning the construction of specific treatment variables. Therefore, while the data generated by the 1989 and 2000 studies are comparable, no attempt was made for HDS2000 to replicate the precise results that were originally reported for HDS 1989. Moreover, the HDS2000 sample of sites does not include all of the metro areas covered by the original 1989 estimates, and the 1989 weights that are used for the analysis presented in this report were adjusted to ensure comparability between the estimated 1989 and 2000 incidences of adverse treatment.

¹⁸ As discussed earlier, however, it is important to recognize that these measures do not capture all the forms that discrimination may take or segments of the housing market. Discrimination might have increased in other parts of the market or for other steps in a housing transaction, while the measures presented here generally declined.
finally, the incidence of consistent white-favored treatment declined significantly. Thus, the findings show a consistent pattern of decline across measures of discrimination against African American renters. It appears that changing attitudes, education, and enforcement have combined to reduce (though not eliminate) the barriers that black renters face when they search for housing in metropolitan housing markets.

The picture is less clear-cut for Hispanic renters. Again, the gross incidence of adverse treatment declined significantly for three categories of treatment. But none of the overall indicators changed significantly between 1989 and 2000. Thus, we cannot conclude that discrimination against Hispanic renters has declined. In metropolitan rental markets, discrimination against Hispanics now appears to be more prevalent than discrimination against African Americans (according to both gross and net measures), suggesting that enhanced outreach, education, and enforcement efforts might be needed to address this problem.

For homebuyers—both black and Hispanic—the gross incidence of adverse treatment actually increased between 1989 and 2000 for several categories of treatment. However, the net measures dropped significantly for most forms of treatment, because the incidence of minority-favored treatment increased more than the incidence of white-favored treatment. The consistency measures also declined significantly, indicating that whites are less likely to be consistently favored across all forms of treatment in a test than they were in 1989. Thus,
although high levels of differential treatment still occur, they are less likely to systematically favor whites than a decade ago. This change may reflect systematic efforts by some real estate agents to favor minority customers, but it may also reflect—at least in part—an increase in random differential treatment.

Although the evidence from HDS2000 suggests that systematic discrimination against African American and Hispanic homebuyers has declined, the persistently high levels of differential treatment are troubling. They suggest that we need not only to learn more about the causes of minority-favored treatment, but also to continue educational and enforcement efforts aimed at promoting equal treatment of qualified customers. In addition, Hispanic homebuyers appear to face an increasing incidence of discrimination with respect to financing assistance, which may limit their ability to become homeowners. And, as discussed further below, the evidence suggests that geographic steering on the basis of race has increased significantly since 1989. Steering disadvantages both minority and white homeseekers, limiting their neighborhood options and perpetuating residential steering. Ongoing education and enforcement efforts should focus on these growing forms of discrimination.

_A Few Metropolitan Areas Exhibit Especially High or Low Levels of Adverse Treatment._ One of the innovations of HDS2000 was to conduct sufficient numbers of tests in each sample site to produce estimates of differential treatment for individual metropolitan areas. In the 1989 Housing Discrimination Study, samples large enough to report metro-level estimates were produced in only five “in-depth” sites. But in HDS2000 approximately 70 paired tests were conducted per tenure category (rental and sales) and per racial or ethnic minority in each of the metropolitan areas in the national sample. Therefore, all of the indicators of differential treatment reported at the national level are also reported for each metropolitan area where testing was conducted. However, the statistical precision of the metro-level estimates is not as high as for the national results, so these results need to be interpreted with some caution.

Although patterns of differential treatment vary across metropolitan areas, overall levels of white- or non-Hispanic white-favored treatment are generally not significantly different from the national average. African American renters appear to face the highest levels of consistent adverse treatment in Atlanta, Georgia and the lowest levels in Chicago, Illinois and Detroit, Michigan. Consistent adverse treatment of African American homebuyers is significantly higher than the national average in Austin, Texas and Birmingham, Alabama, while black homeowners face relatively low levels of consistent adverse treatment in Atlanta and Macon, in Georgia. For Hispanic renters, only Denver, Colorado exhibits levels of consistent adverse treatment significantly different from national results, with below-average levels of adverse treatment for Hispanics. On the sales side, Hispanics in Austin, Texas and New York, New York face relatively high levels of consistent adverse treatment, while Pueblo, Colorado and Tucson, Arizona exhibit relatively low levels. Multivariate analysis, which tested for differences across metro area results while controlling for other factors, also found no evidence of systematic variation in net estimates of discrimination.
These results suggest that discrimination against African American and Hispanic homeseekers remains a problem in large metropolitan areas nationwide—that no region of the country or group of metropolitan areas is exempt. Nonetheless, evidence of local variations in the forms of differential treatment may provide useful information for targeting education and enforcement activities. For example, in some metropolitan areas, minorities are highly likely to be denied information about available housing units, while in other areas, geographic steering or unequal assistance with financing play a bigger role.

**Paired Testing is a Powerful Tool for Measuring Differences in Treatment**

Paired testing is a tremendously powerful research tool because of its potential to directly observe differences in the treatment of equally qualified homeseekers, who differ only with respect to their race or ethnicity. However, as the research community gains more experience with paired testing, it is becoming increasingly clear that testing results are not as simple to interpret as they may seem (Foster, Mitchell, and Fienberg, 2002). Each of the national paired testing studies that HUD has sponsored has refined the paired testing methodology and contributed to a better understanding of how testing results can be used to measure systematic discrimination based on a homeseeker’s race or ethnicity.

As discussed earlier, the simplest measure of adverse treatment is the share of all tests in which the white tester is favored over the minority. Because there are also tests in which minority testers receive better treatment than their white partners, we report both the incidence of white-favored treatment and the incidence of minority-favored treatment. These gross measures of white-favored and minority-favored treatment include both random and systematic elements, and therefore provide upper-bound estimates of systematic discrimination. Net measures report the share of tests in which the white tester was favored minus the share of tests in which the minority was favored. This may understate the frequency of systematic discrimination, because not all minority-favored treatment is the result of random factors. Nevertheless, the net measure reflects the extent to which the differential treatment that occurs (some systematically and some randomly) is more likely to favor whites than minorities, and provides lower-bound estimates of systematic discrimination.

One strategy for estimating systematic discrimination is to use multivariate statistical methods to control for non-systematic factors. This report presents the results of multivariate analysis, controlling for the order of tester visits, whether both testers met with the same agent, and differences in the real-life characteristics of testers. But this strategy can only control for random factors that are observable, and provides no way of knowing whether or how much other random variation remains. Even after controlling for all the observable sources of random differences, an unknown amount of unobservable randomness remains. Therefore, multivariate estimates can be used to assess the robustness of basic gross and net measures, but it would be a mistake to interpret them as definitive measures of systematic discrimination.
We used multivariate statistical procedures to address four frequently asked questions about paired testing results:

1. **Are differences in treatment attributable to different agents?** Partly. The biggest measurable source of non-systematic differences in treatment was meeting with different rental or sales agents. When both testers meet with the same agent, the gross incidence of differential treatment (both white-favored and minority-favored) drops. The net measure, however, stays about the same. Between 1989 and 2000, the share of sales tests in which both partners met with the same agent dropped from about 50 percent to only 25 percent. This explains some of the increase in both white-favored and minority-favored treatment.

2. **Are differences in treatment attributable to the order or spacing of visits?** Not much. Controlling for which tester visited an agent first and for the amount of time that elapsed between their visits reduces the gross incidence of differential treatment only slightly. The net measures stay the same.

3. **Are differences in treatment attributable to differences in testers’ real characteristics?** No. Some critics of the paired testing methodology have suggested that observed differences in treatment are at least partly attributable to real-life differences in tester characteristics. We controlled for differences in testers’ actual education, income, homeownership experience, testing experience, and age, and found no evidence to support the argument that either gross or net measures of discrimination would be systematically lower if testers’ real-life characteristics were more similar.

4. **Are differences in treatment attributable to neighborhood characteristics?** No. One hypothesis about the increased incidence of minority-favored treatment is that it might occur in high-minority or high-poverty neighborhoods, where agents are actually *discouraging* whites from moving, while showing homes and apartments to minorities. However, when we controlled for characteristics of the neighborhoods in which advertised homes and apartments were located, we found little variation in the incidence of minority-favored treatment.

This sensitivity analysis confirms that the HDS2000 results are robust with respect to the order and spacing of tester visits, the real-life characteristics of testers, and the characteristics of neighborhoods in which advertised units are located.

An alternative strategy for eliminating the effects of non-systematic factors is to empirically observe differences in treatment between paired testers of the same race. If same-race testers are carefully matched and follow the protocols of a conventional paired test, any differences in treatment that are observed between them must reflect random factors (both observable and unobservable). HDS2000 experimented with three-part tests in two metropolitan areas, including tests involving visits by two whites and a minority as well as tests involving two minorities and a white. Analysis of these triad tests suggests that the incidence of same-race differences in treatment is generally not significantly different from the incidence of
minority-favored treatment. In other words, minority-favored treatment may be a reasonable proxy for random differences in treatment, and the net measure may provide a good estimate of systematic discrimination. However, because sample sizes are small, these preliminary results should be interpreted cautiously.

**Geographic Steering Represents an Increasingly Important Form of Discrimination.** For homebuyers, geographic steering constitutes an important form of discrimination that limits the housing and neighborhood choices available to both minority and white households, and may help perpetuate patterns of residential segregation. As discussed earlier, HDS2000 found that the incidence of steering for black and white homebuyers increased significantly, even though other measures of systematic discrimination were declining (see Exhibit 8-3). And as of 2000, white homebuyers were recommended and shown houses in more predominantly white neighborhoods than comparable black or Hispanic homebuyers.

In addition to these basic steering indicators, this report presents an expanded analysis of geographic steering for sales tests. This analysis explored three distinct types of steering—information steering, segregation steering, and class steering—that could occur through three techniques—recommendations, inspections, and editorializing. This analysis was conducted at three potentially important levels of geography—census tract, place, and school district. Tract-level results are summarized in Exhibit 4.
This analysis concludes that steering of all three types is occurring, especially at the tract level, when black and white homeseekers are involved. Black/white segregation and class steering occur more often when the advertised home and/or the agent’s office are located in neighborhoods with high percentages of white population. Roughly comparable incidences of segregation steering also are occurring in Hispanic/non-Hispanic white tests, though the other types of steering appear less prevalent. Hispanic/non-Hispanic white segregation and class steering are manifested more strongly when the advertised home is located in predominantly non-Hispanic white neighborhoods, though variations related to agent office location are less clear.

These findings suggest that geographic steering warrants increased attention in education and outreach efforts. Many local testing organizations have focused their efforts primarily on rental testing, and conducting rigorous tests of discrimination by sales agents is considerably more demanding for both testers and their supervisors. Moreover, to obtain credible evidence of geographic steering, testers need to avoid giving sales agents cues about where they want to live and may need to visit multiple homes and record any comments made about the surrounding neighborhoods.

Analysis of Variations Suggests Possible Causes of Discrimination. In addition to producing national and metropolitan estimates of discrimination, HDS2000 looked for possible patterns of variation in discrimination based on location, timing, tester characteristics, agent or agency characteristics, and neighborhood characteristics. This analysis tests three basic

<table>
<thead>
<tr>
<th></th>
<th>Black</th>
<th>Hispanic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gross Upper-Bound</td>
<td>Net Lower-Bound</td>
</tr>
<tr>
<td>Information Steering</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homes Recommended</td>
<td>14.1%</td>
<td>--</td>
</tr>
<tr>
<td>Homes Inspected</td>
<td>10.0%</td>
<td>--</td>
</tr>
<tr>
<td>Commentary</td>
<td>38.5%</td>
<td>15.0%</td>
</tr>
<tr>
<td>Segregation Steering</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homes Recommended</td>
<td>16.5%</td>
<td>3.8%</td>
</tr>
<tr>
<td>Homes Inspected</td>
<td>12.1%</td>
<td>3.8%</td>
</tr>
<tr>
<td>Commentary</td>
<td>37.1%</td>
<td>13.7%</td>
</tr>
<tr>
<td>Class Steering</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homes Recommended</td>
<td>6.9%</td>
<td>--</td>
</tr>
<tr>
<td>Homes Inspected</td>
<td>5.2%</td>
<td>--</td>
</tr>
<tr>
<td>Commentary</td>
<td>34.9%</td>
<td>11.5%</td>
</tr>
</tbody>
</table>

All reported net measures are statistically significant at a 90% confidence level or higher.
hypotheses about the causes of discrimination, assessing the extent to which it appears to stem from agent prejudice, from efforts by agents to protect their business with prejudiced white customers, and from stereotypes by agents about what minority and white customers want or can afford.

Many forms of discrimination clearly vary on the basis of housing agent characteristics. In every type of test, at least one regression indicates that older agents discriminate more than younger agents. Both prejudice and professional experience tend to increase with age, so this result could reflect either higher prejudice on the part of older agents or else some connection between discrimination and the types of firms where more experienced agents tend to work. In addition, we find some evidence that female agents discriminate more than male agents. In addition, discrimination is higher when white testers encounter more agents than their minority partners. This result suggests that discrimination may be more likely to occur in larger agencies that have more different units to offer customers.

Some results support the view that agent prejudice is a key cause of discrimination. Specifically, Hispanic agents discriminate less against Hispanic renters than do white agents, and black females encounter less discrimination in the sales market than do black males. A few results also support the view that housing agents discriminate on the basis of their perceptions about the preferences of black or Hispanic customers—and their desire to avoid spending time on transactions that are unlikely to be consummated. In particular, discrimination against Hispanics in the rental market is sometimes lower in largely Hispanic neighborhoods, and discrimination against blacks in the sales market is sometimes lower in largely black neighborhoods. Finally, one result supports the view that agents discriminate to avoid racial or ethnic tipping, which would undermine all the personal contacts they have developed in the white community. To be specific, discrimination against blacks in the sales market declines with the average value of housing in the advertised unit’s neighborhood.

**Conclusion**

Paired testing is a powerful tool for directly observing differences in the treatment that minority and white homeseekers experience when they inquire about the availability of advertised housing units. Despite the strengths of this methodology, HDS2000, like previous national paired testing studies, is limited in its coverage of metropolitan housing markets and the experience of minority homeseekers. Not all housing units for sale or rent are advertised in major metropolitan newspapers, not all real estate and rental agents use newspaper advertising to attract customers, and not all homeseekers rely upon newspaper advertisements in their housing search. Therefore, results presented here do not necessarily reflect the experience of the typical minority homeseeker, but rather of homeseekers qualified to rent or buy the average

---

19 This behavior represents a form of “statistical discrimination,” in which agents make assumptions about the preferences and resources of individual minority customers based upon information about minorities as a group. Ondrich, Ross, and Yinger (forthcoming) provide a careful
housing unit advertised in a major metropolitan newspaper. Moreover, HDS2000, like most paired testing studies, focuses on the initial encounter between a homeseeker and a rental or sales agent. Additional incidents of adverse treatment may occur later in the housing transaction, when a renter submits an application or negotiates lease terms, or when a homebuyer makes an offer on a particular unit or applies for mortgage financing. In spite of these important limitations, HDS2000 provides the most complete and up-to-date information available about the persistence of housing market discrimination against African American and Hispanic homeseekers in large urban areas of the United States today and about the progress we have made in combating discrimination over the last decade.
References


