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

Rob Pitingolo
Urban Institute

Stephen L. Ross
The University of Connecticut

Abstract

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

Introduction

In 2012, the U.S. Department of Housing and Urban Development launched the fourth major nationwide Housing Discrimination Study (HDS2012), with the goal of measuring housing discrimination in rental and owner-occupied housing for Black, Hispanic, and Asian homeseekers. The first major study, conducted in 1977,[1] found high levels of discrimination against Black homeseekers, including frequent occurrences of overt exclusion in which Black homeseekers could not complete or even schedule an appointment with real estate agents and often were summarily told that no housing was available (Weink et al., 1979). This study was followed by a nationwide study in 1989[2] of discrimination against Black and Hispanic homeseekers. This study was the first to document high levels of discrimination against Hispanic homeseekers, and it found no evidence of a decline in the measured levels of discrimination against Black homeseekers since HMPS1977, although most overt acts of exclusion, such as "door slamming," had declined considerably (Turner, Struyk, and Yinger, 1991; Yinger, 1995). The next study, conducted in 2000,[3] was designed to provide estimates of the changes in housing discrimination that Black and Hispanic homeseekers faced nationwide (Turner et al., 2002); a second phase in 2002 provided national estimates for Asian homeseekers (Turner and Ross, 2003a). Although meaningful levels of discrimination were detected for all three groups tested in both markets, HDS2000 found substantial declines in discrimination against Black homeseekers in the rental market and against Black and Hispanic homeseekers in the sales market, and the testers' narratives suggested a dramatic improvement in the environment that Black and Hispanic homeseekers experienced during their housing search (Ross and Turner, 2005). In 2012, the most recent nationwide housing discrimination study—HDS2012—continued to find persistent, but in many cases modest, levels of discrimination against Black, Hispanic, and Asian homeseekers that appear comparable to the levels detected during 2000 and 2002 (Turner et al., 2013).

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

---

[1] The Housing Market Practices Survey (HMPS1977).

[2] The Housing Discrimination Study (HDS1989).

[3] The Housing Discrimination Study (HDS2000).

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

Another article in this symposium summarizes the HDS2012 methodology and presents the study's core findings. This article reviews the findings and reanalyzes the data to consider three significant concerns raised regarding paired-testing studies of housing discrimination: (1) the exclusion of minority homeseekers during the process of setting up appointments, (2) the net measure of adverse treatment understating discrimination because some housing units are systematically not shown to White testers, and (3) the use of metropolitanwide advertisements that may systematically underrepresent neighborhoods where discrimination is higher. Our discussion mostly focuses on the net measure of adverse treatment in which one of the three concerns is that the net measure may understate discrimination relative to the gross measure. In the main body of the article, we first present the basic estimates of treatment patterns from HDS2012 rental tests and then present the sales test results. HDS2012 directly addresses the first concern raised previously in that the ability to make an appointment over the phone is tracked and measures of adverse treatment are developed based on this outcome. In rental markets, no differences are observed in the frequency of White-favored and minority-favored tests (net measure) on the ability to make an appointment for any group. In sales markets, statistically significant differences are observed for White-Black tests in the ability to make an appointment, but these differences are small—less than 3 percentage points—and no differences were observe for Anglo-Hispanic[4] or White-Asian tests (Turner et al., 2013). The study findings suggest that the inability of minority testers to obtain an appointment is not a major source of bias in measuring housing discrimination.

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

---

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

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

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

To partially address this final concern, we use data from the 2011 American Community Survey (ACS) to reweight the tests conducted in HDS2012 and so more accurately represent the population of available housing units. We specifically identify new residents in rental (or owner-occupied) housing within the past year as evidence that a rental (owner-occupied) housing unit turned over and was available for rent (sale) during that year. Using this proxy for available rental or owner-occupied housing, we estimate the number of available rental and owner-occupied housing units in each Public Use Microdata Area (PUMA) at each metropolitan site for HDS2012. For both rental and sales tests, we compare these shares with the shares of tests for each PUMA in each site and then develop weights that give higher weights to tests in PUMAs with more available rental housing. We use these weights to develop measures of adverse treatment that are representative of treatment in the sample of available housing units, as opposed to the housing stock or the sample of advertised units. We conduct this analysis for White-Black, Anglo-Hispanic, and White-Asian rental tests over the seven main measures[5] of adverse treatment in the rental market and for four key measures[6] for the owner-occupied market that were presented in HDS2012. This exercise

---

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

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

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

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

Methods

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

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

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

(1)

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

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

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

$$\widetilde{W}_{ips} = (W_{is}N_s)\frac{a_{ps}}{n_{ps}}\frac{1}{A_s} = \frac{a_{ps}}{n_{ps}}\frac{1}{P_sA_s} \quad , \tag{2}$$

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

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

The logic behind this equation is as follows.

1. Each test receives a weight based on its PUMA, which is the ratio of number of available units divided by number of tests.

2. This ratio is divided by the sum of the ratio over all tests in the site so that the sum of weights for each site is 1.

3. The resulting number is then multiplied by the original weight assigned to each site, 1 over the probability of selection.

Adverse treatment is then defined using the traditional net measure,

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

for discrete outcomes, or—

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

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

Data

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

Exhibit 1

HDS Site Selection and Test Count

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

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

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

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

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

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

Exhibit 2

Site PUMAs and Tests per PUMA

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

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

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

Exhibit 3

Test Subsamples and Within-Site Representativeness

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

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

Results

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

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

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

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

Exhibit 4

White-Black Measures of Adverse Treatment in Rental Housing

	Both	White	Black	Difference	p-Value
Rental market treatment measures (original weights)					
Appointment					
Tester(s) able to make an appointment	94.6%	1.0%	0.6%	0.4%	0.26
Information and availability *If able to meet with an agent:*					
Tester(s) told any units available	95.0%	3.0%	2.1%	0.9%	0.20
One tester told about more units than partner	54.1%	27.5%	18.5%	9.0%	0.00
Average number of units available (per visit)		1.83	1.63	0.20	0.00
Inspections *If available units recommended:*					
Average rent		$1,122	$1,126	– $4	0.06
One tester inspected more units than partner	69.1%	16.9%	14.1%	2.8%	0.03
Average number of units inspected (per visit)		1.28	1.23	0.04	0.04
Rental market treatment measures (available units)					
Appointment					
Tester(s) able to make an appointment	93.7%	1.0%	0.3%	0.6%	0.06
Information and availability *If able to meet with an agent:*					
Tester(s) told any units available	95.4%	2.8%	1.8%	1.0%	0.22
One tester told about more units than partner	55.7%	26.9%	17.5%	9.4%	0.00
Average number of units available (per visit)		1.78	1.59	0.19	0.00
Inspections *If available units recommended:*					
Average rent		$1,090	$1,097	– $7	0.04
One tester inspected more units than partner	69.0%	16.1%	14.9%	1.2%	0.64
Average number of units inspected (per visit)		1.24	1.21	0.03	0.35

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

Exhibit 5

Anglo–Hispanic Measures of Adverse Treatment in Rental Housing

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

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

Exhibit 6

White–Asian Measures of Adverse Treatment in Rental Housing

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

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

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

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

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

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

Exhibit 7

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

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

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

Exhibit 8

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

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

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

Exhibit 9

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

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

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

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

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

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

Exhibit 10

Ratio of Available Rental Units Weights to Original Weights

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

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

Exhibit 11

Ratio of Available Owner-Occupied Units Weights to Original Weights

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

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

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

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

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

Exhibit 12

Weighted Rental Sample Exposure to Census Tract Attributes

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

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

Exhibit 13

Weighted Owner-Occupied Sample Exposure to Census Tract Attributes

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

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

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

Conclusion

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

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

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

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

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

Acknowledgments

Authors

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

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

References

Baugh, John. 2007. "Attitudes towards Variation and Ear-Witness Testimony: Linguistic Profiling and Voice Discrimination in the Quest for Fair Housing and Fair Lending." In *Sociolinguistic Variation: Theory, Methods, and Applications*, edited by Robert Bayley and Ceil Lucas. Cambridge, United Kingdom: Cambridge University Press: 338–348.

Galster, George, and Stephen L. Ross. 2007. "Fair Housing Enforcement and Changes in Discrimination Between 1989 and 2000." In *Fragile Rights within Cities: Government, Housing, and Fairness*, edited by John Goering. Lanham, MD: Rowman & Littlefield: 178–202.

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

Ondrich, Jan, Stephen L. Ross, and John Yinger. 2000. "How Common Is Housing Discrimination? Improving on Traditional Measures," *Journal of Urban Economics* 47 (3): 470–500.

Ross, Stephen L. 2002. "Paired Testing and the 2000 Housing Discrimination Study." In *Measuring Housing Discrimination in a National Study: Report of a Workshop*, edited by Stephen E. Feinberg and Angela Williams Foster. Washington, DC: National Academy of Sciences, National Research Council.

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

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

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

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

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

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

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

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