What Have We Learned from Paired Testing in Housing Markets?

Sun Jung Oh, Syracuse University
John Yinger, Syracuse University*

May 2015

The authors are, respectively, a graduate student in social science and a professor of economics and public administration at the Maxwell School, Syracuse University. They are grateful to helpful comments from two anonymous referees.

*Corresponding author: Tel: +1 315 443 3114.
Email addresses: soh14@syr.edu (Sun Jung Oh), joyinger@maxwell.syr.edu (John Yinger),
I. Introduction

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

This paper begins with a detailed review of the evidence about discrimination against African-Americans, Hispanic-Americans, and Asian-Americans obtained from in-person paired testing in housing markets. This review is followed by an exploration of studies based on telephone or e-mail audits (some of which apply to other countries), an exploration of the link between paired testing and fair housing policy, and a brief review of the use of paired testing in some other markets. Our focus is on the use of audits (tests) for research purposes. Section II examines testing methodology. We review the basics of paired testing, discuss audits that do not
involve face-to-face contact (called correspondence audits), explain how testing results can be used to study the causes of discrimination, and describe key methodological issues in the four national housing audit studies. Section III turns to results from paired testing in the housing market. Based on the national studies, we explore the incidence of discrimination, trends in discrimination, results about the causes of discrimination, and results concerning racial and ethnic steering. Appendix A provides further evidence from smaller paired testing studies and from correspondence audits. Section IV addresses paired testing and public policy. We focus on the link between paired testing and fair housing policy, but also provide a brief review of paired testing research in markets other than the housing market.

II. Testing Methodology

1. The Basics of Paired Testing

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

Second, auditors must be trained about the role they should play during an audit. In most cases, they are instructed to inquire about an advertised unit and then to ask for additional suggestions from the housing provider. In some audit studies, the audits are “blind,” in the sense that the auditors are only told that they are helping with a study about the marketing of housing and are not told that they have a partner or that the study is investigating discrimination. In the 1989 and 2000 national audit studies reviewed below, however, the training was not “blind” in this sense. Auditors were told the purpose of the study, auditors were trained to provide
information as accurately as possible, and, to the extent possible, managers were given protocols
to check on the accuracy of the information provided. In these studies, auditors in different
racial or ethnic groups were trained together to ensure that they received the same training.

Blind auditing is not appropriate in this case because auditors can observe the group composition
of the trainees; without an explanation for this composition, auditors might make their own
guesses about the purpose of the study. ³

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

Fourth, auditors are matched for each test, with one member from a historically
disadvantaged group.⁴ Paired testers are assigned income and other household traits that make
them equally qualified for the sampled advertised unit about which they are inquiring.⁵ Audit
studies typically draw a random sample of housing advertisements from publicly available
sources such as the major local newspaper or widely used ad-listing websites, and one audit is
conducted for each advertisement.⁶ Even if the auditors are the same, the assigned income and
household traits vary from one audit to the next to match the associated advertised unit.

Teammates are assigned similar incomes and other traits for a given audit so that differences in
these traits do not lead to differences in treatment. Because housing market transactions are
relatively simple and because the people marketing housing do not usually ask potential
customers about very many traits, the qualifications of audit teammates are almost identical. It
follows that differences in the way teammates are treated can be attributed to random factors or
to discrimination. Because membership in a historically disadvantaged group cannot be
randomly assigned, this approach cannot fully rule out the possibility that some unassigned trait influences treatment, thereby biasing estimates of discrimination up or down; however, good management makes this outcome unlikely. As discussed below, some audit studies have also collected information on auditors’ actual traits, such as their income and education, to see if these traits affect measures of discrimination.⁷

Fifth, audit teammates separately contact the housing agent associated with one of the selected advertisements and attempt to schedule a visit. The initial contacts are completed during a short period, but not so short as to be suspicious to the agent. In most studies, the order of the visits is randomized. The visit, if it occurs, then follows the script that the auditors learned in training, with inquiries about the advertised unit and similar units.⁸ Auditors are generally encouraged to learn about and visit as many units as possible, while not stating preferences (beyond an interest in the advertised unit and units similar to it) that would guide this process.

Finally, once an audit is completed, each audit teammate is asked to record what he or she was told and how he or she was treated. These audit forms provide information on the number of houses or apartments shown to each auditor, as well as on many other aspects of housing agent behavior. Audit teammates have no contact with each other during an audit and they fill out their audit survey forms independently. Most audit studies then schedule de-briefing sessions when an audit manager reviews these forms with each auditor to ensure that all information on the forms is accurate.

Unlike alternative approaches that look for signals of discrimination in housing prices, housing quality, homeownership, and segregation patterns, in-person paired testing provides direct measures of discrimination by comparing the outcomes of equally qualified white and minority testers. Moreover, paired testing makes it possible to examine the multiple, complex
forms that discrimination can take by observing many types of housing agent behavior. This methodology yields a powerful narrative concerning the way people in different groups are treated. This narrative adds credibility to findings of discrimination in research, policy, and court settings. In addition, the results of the paired testing can shed light on the causes of discrimination because they provide information on the circumstances in which discrimination occurs. We return to research on the causes of discrimination in Section II.3.

One important feature of audits is that some of the unobservable factors are shared between audit teammates. This type of unobservable factor does not lead to bias in estimates of discrimination, but it does lead, if not accounted for, to an upward bias in standard errors. The studies discussed below provide a variety of ways to avoid this type of bias.

2. A New Development: Correspondence Audits

One of the disadvantages of paired testing is that it is expensive; a large management structure must be created and auditors must be hired, trained, sent into the field, and de-briefed. To address this practical problem, scholars have developed an alternative method, called a correspondence audit, that is less expensive and more precise, but addresses a narrower set of questions. This methodology uses e-mails or phone calls instead of visits and records the housing seller’s response to an e-mail or a phone call instead of his or her in-person treatment of an auditor. The rapid growth in the use of the internet for marketing housing has made this an appealing strategy.

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

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

3. Testing Hypotheses about the Causes of Discrimination

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

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

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

The statistical discrimination hypothesis states that discrimination occurs when agents treat people in different groups differently because they believe that group membership is correlated with unobserved characteristics that affect the profitability of their actions. In the
rental housing market, for example, a rental agent may use customers’ race or ethnicity as a signal about their preferences for housing type and/or neighbors or about constraints that are related to the probability of a transaction. The statistical discrimination hypothesis also predicts that discrimination against minorities will decrease with the minority population in a neighborhood if real estate agents believe that minority home seekers prefer living where minorities are concentrated. In addition, this hypothesis predicts that discrimination increases with the value of houses if real estate agents believe that minorities have a relatively high probability of financial difficulties.

4. The Housing Discrimination Studies

The largest paired-testing studies in the U.S. are the Housing Market Practices Survey (HMPS) in 1977 and the three Housing Discrimination Studies (HDS1989, 2000, and 2012) sponsored by the U.S. Department of Housing and Urban Development (HUD). These studies were designed to yield statistically reliable national estimates of discrimination against certain racial and ethnic groups in urban housing markets. Table 1 presents a few common features of the four studies (including race or ethnicity tested, scale, and locations where tests were conducted) along with some of their differences. In addition, several scholars have conducted smaller-scale in-person audit studies, and numerous phone and email correspondence audits have been conducted over the last two decades.

Outcomes measured in paired testing studies indicate the incidence and severity of unfavorable treatment experienced by minorities. Unfavorable treatment may arise, for example, in the probability of being told the advertised units is available, the probability of making an appointment when inquiring about advertised units, the probability of at least one in-person visit to an available apartment, the number of apartments suggested or shown, the characteristics of
Two types of discrimination measure have appeared in the literature: gross measures and net measures. Gross measures indicate the share of all audits in which the white auditor is favored over his or her minority teammate. Although gross measures are easily understandable, they may overstate the frequency of systematic discrimination because nondiscriminatory random events are responsible for some portion of observed treatment. A white auditor might appear to be favored, for example, because she went first and the apartment was rented before her minority teammate arrived. Net measures of discrimination are the proportion of audits in which the white auditor is favored minus the proportion of audits in which the minority auditor is favored. The net measure indicates the disadvantage minorities face in the housing market relative to whites. The net measure provides a lower-bound estimate of systematic discrimination in favor of whites. To the extent that minorities are systematically favored over whites in some share of housing inquiries, such as inquiries concerning apartments in largely minority neighborhoods, the net measure will understate the incidence of discrimination against minorities. Although the audit design makes it possible to calculate reasonable net and gross measures based on sample proportions, more precise measures of discrimination can be obtained with more advanced statistical procedures (See Ondrich, Ross, & Yinger, 2003; Ondrich, Stricker & Yinger, 1998, 1999; Page, 1995; Zhao, 2005).

III. The Results of Paired Testing Studies

This section first presents the findings from four nationwide paired-testing studies that began in the late 1970s and were sponsored by the U.S. Department of Housing and Urban
Development (HUD). In addition, we review the results of studies based on the data from the four national studies and of other studies that conducted paired tests. We also review the evidence on trends in housing discrimination based on the four HUD-sponsored studies.

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

1. The Incidence of Discrimination

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

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

First, when comparing the results between rental and sales tests for blacks, it is clear that the magnitude of discrimination against blacks is higher in sales tests than in rental tests (for the outcomes a, b, c, and d). For instance, blacks are told about fewer available units (outcome a) than whites in 13.4 percent of inquiries in the sales tests, compared to 9 percent of inquiries of the rental tests. Also, blacks are told about 0.5 fewer units available than whites (outcome b) in the sales tests, compared to 0.2 fewer units in the rental tests. Finally, blacks are shown fewer units than whites (outcome c) in 9.3 percent of inquiries of the sales tests, compared to 2.8 percent of inquiries of the rental tests. In terms of the number of units shown (outcome d), blacks are shown 0.30 fewer units than whites in the sales tests, compared to 0.04 fewer units in the rental tests.

Second, in rental markets, Hispanics experience more discrimination than blacks for all four outcomes that exhibit discrimination (outcomes a, c, d, and e). For instance, Hispanic renters are told about fewer available units than are white renters (outcome a) in 12.8 percent of inquiries, while black renters are told about fewer available units in 9 percent of inquiries. In
addition, Hispanic renters are shown fewer units than equally qualified whites (outcome c) in 6 percent of inquiries, compared to 2.8 percent of black renter’s inquiries. On average, Hispanic renters are shown 0.07 fewer units than white renters (outcome d), but black renters are shown 0.04 fewer units than white renters. Agents also quote slightly higher rents ($6 per month on average) to Hispanic renters than whites, and agents quote $4 higher rents to black renters than whites. In contrast to the results of the rental tests, however, Hispanic homebuyers are as favored as white homebuyers, while white homebuyers are favored over black homebuyers in five out of the eight outcomes.

Third, minority home seekers are rarely denied appointments (outcome f), and when both white and minority testers meet with an agent in person, they are rarely told that no unit is available (outcome g). Compared with comparable white customers, however, black homebuyers are slightly more likely to be denied an in-person appointment (in 2.8 percent of inquiries), and Hispanic renters are slightly more likely to be told that no homes or apartments are available (in 1.8 percent of inquiries). Overall levels of agent helpfulness to whites and minorities are not significantly different.

In addition to HDS2012, several smaller scale audits had been conducted in individual cities in the U.S. and in European countries. These studies traditionally used in-person paired tests, but recent studies generally use email or phone correspondence tests. Most correspondence tests determine whether ethnically linked names or speech patterns influence the probability that the agent responds to an inquiry or allows the home seeker to make an appointment.

Appendix A presents the summary results for other paired testing studies. We restrict our attention to the studies involving more than 100 tests. Although the location, time period, and minority groups in these studies are quite varied, they consistently find that, for various
outcomes, discrimination against racial and ethnic minorities is a common feature of housing markets in many countries. A number of studies also find discrimination against immigrants and people with low socioeconomic status.

2. Trends in Discrimination

Tracking discrimination over time can help to determine how successful anti-discrimination interventions have been. This section presents the trends in rental and sales discrimination using the results of the four national studies. The four national studies provide a reasonable approximation to national trends in housing discrimination because they were conducted about ten years apart using a similar methodology.\(^\text{17}\) Tables 3 and 4 summarize the findings of discriminatory treatment for blacks and Hispanics, based on types of behavior that are consistently measured in all four national studies.\(^\text{18}\) We present the results of rental tests and sales tests separately. The incidence of discrimination is drawn from the final reports of HMPS1977, HDS1989, HDS2000 and HDS2012.\(^\text{19}\) Two common outcomes presented for both rental and sales tests are whether the agent told only the white tester that the advertised unit was available and whether the white tester was shown more units. For sales tests, we present an additional outcome, namely, whether the real estate agent offered help with financing. To be more specific, this outcome indicates whether the agent had a general discussion with the home seeker about the mortgage process or offered to provide a mortgage pre-qualification for a maximum loan amount. The net and gross measures consistently show that minorities receive less favorable treatments than whites in both the rental and sales housing markets (Wienk et al., 1979; Turner, Struyk, & Yinger, 1991; Turner, Ross, Galster, & Yinger, 2002; Turner et al., 2013).
Tables 3 and 4 show that, based on gross measures, racial or ethnic minorities had a 5 to 30 percent lower probability than whites of being told that the advertised unit was available. Moreover, minorities inspected fewer housing units than their white teammates from 13 to 46 percent of the time. In sales tests, minorities had 16 to 29 percent lower probability of receiving financial help. The net measures indicate substantially lower levels of discrimination in both outcomes. Indeed, several of the results are not significantly different from zero. Nevertheless, the net measure indicates significant discrimination, as high as 19 percent, in several other cases. Tables 3 and 4 also show that the incidence of discrimination tends to be somewhat higher against African Americans than against Hispanics. Moreover, gross measures of discrimination tend to be higher in sales tests than in rental tests, with the notable exception of the availability of the advertised unit in 1977.

Charts 1 and 2 present the trends in rental and sales discrimination, respectively, using gross measures. In general, housing discrimination in the outcomes in Tables 3 and 4 has declined over time. In 1977, blacks were frequently denied access to advertised units that were available to equally qualified whites. For instance, one in three blacks in rental tests and one in every five blacks in sales tests were told that there were no houses available in 1977 (Wienk et al., 1979). In 2012, however, minority renters or homebuyers who call to inquire about advertised homes or apartments are rarely denied appointments that their white counterparts are able to make (Turner et al., 2013). The decline in discrimination is more apparent in rental tests than sales tests and is larger for blacks than for Hispanics or Asians. The decline also differs across outcomes. In both the rental and sales tests, for example, differential treatment in the number of inspected units has not declined very much. In the rental tests, discrimination against Asians has increased for the same outcome. Moreover, for the financial help offered outcome,
there is no clear evidence that the discrimination has declined over time. Although the most blatant forms of housing discrimination (such as refusing to show the advertised unit) have declined since the first national audit study in 1977, minorities’ housing opportunities are still limited in significant ways.

3. Racial and Ethnic Steering

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

Table 5 presents the results of steering based on the HDS studies for houses recommended and houses inspected separately. As seen in Table 5, each HDS finds evidence of steering. Black or Hispanic homebuyers had 4 to 25 percent higher probability of being
recommended or shown houses in neighborhoods with a lower percentage of white population
than whites or non-Hispanic whites, respectively. Steering is more evident for houses
recommended than for houses inspected by testers. Chart 3 illustrates the trends in steering
based on the incidence of steering of Table 5. This chart shows that the incidence of steering
becomes larger over time. These results indicate that steering plays a role in the overall pattern
of unfavorable treatment in the housing market. Despite the clear evidence of steering, the
housing discrimination studies also find that the composition of neighborhoods recommended to
minority homebuyers is similar to the composition of those recommended to equally qualified
whites (Turner et al., 1991; Turner et al., 2002; Turner et al., 2013). This apparent contradiction
arises because the differences in neighborhood ethnic composition between teammates are small
in magnitude; most of the houses shown and recommended to both minorities and majorities
located in predominantly white neighborhoods. Houses for sale in minority integrated
neighborhoods are underrepresented among advertisements in major metropolitan newspaper
(Galster, Freiberg, & Houk, 1987; Newburger, 1995; Turner, 1992), and consequently, these
neighborhoods are underrepresented in the HDS sample. Thus, results of HDS studies reflect the
incidence of steering in only one segment of the market.

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

4. The Causes of Discrimination
Table 6 summarizes how audit characteristics would affect discrimination based on the three hypotheses about the cause of discrimination discussed in Section II.3 and the findings of eight studies listed below. The characteristics tested for each hypothesis are not always mutually exclusive, and different hypotheses may predict the opposite effect of the same characteristic on discrimination. In addition, several audit characteristics interact with other in a complex way, and some predictions of the agent-prejudice and customer-prejudice hypotheses cannot be separated. For instance, either housing agents or their white customers may have stronger prejudice against younger minority than against older minorities (Choi, Ondrich, & Yinger, 2005).

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

Several results of the eight studies support the agent-prejudice hypothesis. Unless otherwise indicated, we focus on the results of black-white discrimination. For the advertised unit inspected or similar unit inspected outcome, there is more discrimination by older agents (OSY, 1999; COY, 2005), less discrimination by female agents (COY, 2005), and less discrimination against female auditors (COY, 2005; ZOY, 2006; HDS2012) and older auditor (OSY, 1999). Results concerning the effect of agent’s race on discrimination are inconsistent.
OSY (1998, 1999) find more discrimination against blacks when the agent is black, while Zhao (2005), COY (2005, 2008), ZOY (2006), and HDS2012 (only sales tests) find less discrimination when the agent is black. In the sales market, however, black agents are rare. According to HDS2012, testers met with a black agent in only 5 percent of the tests (p. 76).

The customer-prejudice hypothesis suggests that discrimination is likely to increase with the assigned income of auditors, the percent owner-occupied housing units in the neighborhood, and the share of white residents in the white-majority neighborhood. Some results support these predictions. For the advertised unit inspected or the number of houses shown outcomes, black home buyers face less discrimination in the neighborhoods with a significant share of blacks (Zhao, 2005) but more discrimination in the neighborhoods with higher percentage of owner-occupied housing units (OSY, 1999; Zhao, 2005). Page (1995) and Zhao (2005) find that blacks with higher income encounter less discrimination in the number of houses shown. However, the effect of having a white collar job or higher education on discrimination is unclear. While Ahmed and Hammarstedt (2010), Bosch, Carnero, and Farre (2010), and Carlsson and Eriksson (2014) find no effect of positive information on discrimination, Baldini and Federici (2011) and Hanson and Hawley (2011) find that minorities receive more email responses when they reveal positive information. Finally, ORY (2003) find that real estate agents discriminate more against higher-income black customers, and Carlsson and Eriksson (2013) find that racial discrimination is higher for ethnic minorities with a high-skill job than those with a low-skill job.

Some studies also find more discrimination against blacks in high-value neighborhoods (ORY, 2003; COY, 2005). For the advertised unit inspected outcome in sales audits, OSY (1999) find that discrimination decreases as the ratio of assigned auditor income to housing value increases. Several results related to tipping also support the customer-prejudice hypothesis.
Page (1995) and OSY (1999) find that discrimination increases as the percentage of minority representation approaches a neighborhood tipping point and decreases when the percentage exceeds the tipping point. Moreover, ORY (2003) find that discrimination is relatively high in central city integrated areas, which are the ones most likely to be threatened with tipping. ZOY (2006) also find that black homebuyers are less likely to encounter discrimination in Hispanic neighborhoods.

The customer-prejudice hypothesis also predicts that larger real estate agencies, which are less dependent on a particular neighborhood for their business, are less likely to discriminate. OSY (1998) find that discrimination is less likely in larger agencies for the advertised unit inspected outcome in rental audits, and OSY (1999) and ORY (2003) find similar results in sales audits. In contrast, COY (2008) and HDS2012 find the opposite result: larger agencies discriminate more against minority customers than do smaller agencies (for the advertised unit inspected, the number of units inspected, or rental incentive provided outcomes). Finally, ZOY (2006) find that real estate agents who use the internet, which allows them to steer black clients away from prejudiced white neighborhoods, are less likely to discriminate in terms of the number of units recommended.

ORY (2003) find strong evidence for statistical discrimination. They find that agents’ marketing efforts increase with asking price for whites, but not for blacks, an implication that is consistent with the hypothesis that agents practice statistical discrimination based on a preconception about the ability of black customers to purchase expensive homes. They also find that black customers, but not whites, are shown units that are cheaper than the advertised unit that is the basis for the audit. ORY (2003) and Zhao (2005) also find that real estate agents discriminate more in neighborhoods with higher house values, even controlling for the value of
the houses being shown. In addition, Page (1995) finds that blacks encounter more
discrimination when inquiring about more expensive houses, and a similar finding appears in
COY (2005). COY (2005) find a positive, significant coefficient for median house value in the
number of units inspected regression, but conclude that this result is also consistent with the
customer-prejudice hypothesis. OSY (1998), however, find that discrimination is not
significantly different when the advertised unit is in an integrated or a white neighborhood. Page
(1995) and ORY (2003) find less discrimination in integrated areas than in white areas in the
sales markets, which is consistent with the view that agents try to maximize the chances of a
successful match by making race-based assumptions about a customer’s preference. COY
(2008), however, find no evidence to support statistical discrimination hypothesis in the black-
white audits.

Ewens, Tomlin, and Wang (2014) design correspondence audits of rental housing to look
for landlord prejudice and statistical discrimination. If landlord prejudice is at work, they argue,
then discrimination should be lower in largely black than in largely white neighborhoods. Their
results do not support this prediction. They also argue that landlords practicing statistical
discrimination will find explicit signals about the “quality” of an applicant from a particular
racial or group to be more believable if they have more experience dealing with tenants from that
group. If so, landlords will respond more favorably to positive information (such as an
indication that the tenant is a nonsmoker or has a desirable job) about a black applicant in a
largely black neighborhood than in a largely white neighborhood. Their empirical results
support this prediction, and therefore support the conclusion that some landlords practice
statistical discrimination.
Finally, as Page (1995) recognized, Yinger (1995) shows that the level of discrimination depends on a broker’s opportunities to discriminate, defined as his access to available housing units. With controls for the opportunity to discriminate (an agent’s available units), Yinger (1995) finds evidence that real estate brokers discriminate to protect their business with prejudiced white clients and on the basis of stereotypes about black and Hispanic customers. These results indicate that the causes of discrimination in rental and sales housing are complex. The strongest results (from ORY, 2003) support statistical discrimination, but both the prejudice of agents and their responses to the prejudice of their white customers also appear to be at work in some cases.

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

1. Paired Testing and Fair Housing Enforcement

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

As of 2011, ninety-eight private non-profit agencies were engaged in fair housing enforcement. In 2006 alone, these agencies conducted more than 5,000 tests (Temkin, McCracken, & Liban, 2011). Since shortly after the Fair Housing Act was passed, these...
agencies have used tests to establish discrimination and to obtain settlements in hundreds of cases.\(^\text{32}\) In almost all cases, fair housing organizations obtain injunctive relief that includes a change in behavior and/or policies, training to prevent future discrimination, and monitoring to ensure compliance with the Fair Housing Act. Here are some examples.

- In 2013 and 2014, based on complaints, the Fair Housing Justice Center (FHJC) conducted audits in rental housing in the Woodlawn neighborhood of the Bronx (J.J.A Holding Corporation). The FHJC and three African American testers alleged that J.J.A Holdings engaged in racially discriminatory rental practices: Among other things, an agent told African American testers that no apartments were available while showing apartments to white testers on the same day. In 2015, the J.J.A agreed to change rental practices and pay the plaintiffs $200,000 for damages and attorneys’ fees (Gorman, 2015).

- In 2013, Latino home seekers did not receive a rental application from apartment managers of Bailey Properties in Arkansas (or did so after significant delay), while prospective white renters promptly received documents. The Arkansas Fair Housing Commission confirmed the discrimination against Hispanics based on six correspondence audits conducted by the National Fair Housing Alliance (NFHA, 2014).

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

The second connection between audits and fair housing enforcement is that the results of research audits and enforcement audits have provided powerful evidence in support of continued or expanded fair housing enforcement. Most importantly, perhaps, the audit results from HMPS were highlighted in Congressional Testimony about the Fair Housing Amendments Act, which greatly expanded the federal government’s powers to enforce fair housing laws. Testimony by HUD’s general council (Knapp, 1986b), for example, began by citing the HMPS results, but also explained that “HUD staff extrapolated from those findings to conclude that about 2 million instances of housing discrimination occurred every year. Experts in statistical methodology may
quibble over that extrapolation, but even if the estimate is wrong by half, it is nonetheless staggering and, to put it mildly, deeply disconcerting.” Testimony before the same subcommittee by the director of the Kentucky Commission on Human Rights (Martin, 1986b) cited not only the HMPS results for his state but also results from audits conducted by his agency in 1977 and 1985. In addition, several articles in professional journals cited evidence from HMPS in building a case for passage of the Fair Housing Amendments Act (Rice, 1984; James & Crow, 1986). This influence of research audits on enforcement policies is well summarized in the article by Freiberg (1993), which states that “Data showing widespread patterns of unlawful housing discrimination [from HMPS1977 and HDS1989] understandably evoke a response from well-meaning policymakers for more vigorous enforcement of fair housing laws” (p. 230).

The HMPS results also appear to have been influential in the 1987 passage of FHIP and in the 1984 passage of the Fair Housing Assistance program (FHAP), which is discussed below. See, for example, the testimony on FHIP by HUD’s general council (Knapp, 1986a), which includes the above quotation from his Fair Housing Amendments Act (FHAA) testimony. This hearing concerned amendments to restrict testing with FHIP funds. At that hearing, Martin (1986a) explicitly argued that testing by private fair housing groups in Kentucky had been very effective and that there was a “need for funding especially for private fair housing groups… without the restriction of the amendments.”

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

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

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

- Thanks to a complaint that a corporate owner and leasing agent discriminated based on race, in 2013, DOJ conducted a series of three tests at Baldwin Commons in Pittsburgh. The tests found that white testers were shown apartments and were offered the opportunity to rent them while black testers were told that the same apartments were unavailable. The court entered a consent decree in *United States v. S-2 Properties, Inc. (2014)*, and the defendants will pay a civil penalty to the United States of $15,000, develop and maintain non-discrimination housing policies, and attend fair housing training (DOJ, 2014).

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

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

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

2. Paired Testing in Other Markets

Audits are sufficiently advanced to conduct national-level tests in the areas of housing
sales and rentals. In-person audits, in which individuals are matched for all relevant
characteristics other than the one that is expected to lead to discrimination (e.g., race or
ethnicity), have also been used in several other markets, including entry-level hiring, inquiries
about home mortgages, house insurance, car sales, and selected areas of public accommodations
such as taxi service. The first in-person and correspondence audits to measure hiring
discrimination were conducted in Britain (Daniel, 1968; Jowell & Prescott-Clarke, 1970). In the
U.S., the Urban Institute conducted the first in-person audits of hiring discrimination against
Hispanic men applying for entry-level jobs in Chicago and San Diego in 1989 (Cross, Kenney,
Mell, & Zimmerman, 1989). Several in-person and correspondence tests of hiring discrimination
against minority groups have also been conducted in the U.S. and European countries since then
(Bendick, Jackson, & Reinoso, 1994; Bendick, Jackson, Reinoso, & Hodges, 1991; Bertrand & Mullainathan, 2004; Bursell, 2007; Carlsson & Rooth, 2007, 2012; Goldberg, Mourinho, & Kulke, 1995; Kaas & Manger, 2011; Nunley, Pugh, Romero, & Seals, 2014; Pager, 2003; Pager, Western, & Bonikowski, 2009; Wood et al., 2009; Turner, Fix, & Struyk, 1991). In addition, audits have been used to examine discrimination in automobile sales (Ayres, 1991; Ayres & Siegelman, 1995), taxicab service (Ridley, Baton, & Outtz, 1989), home mortgage inquiries (Galster, 1993; Turner & Skidmore, 1999), homeowners’ insurance (Wissoker, Zimmermann, & Galster, 1998), and the provision of medical care (Schulman et al., 1999).

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

V. Conclusion

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

Paired audits offer a uniquely effective tool for directly observing differential treatment of equally qualified home seekers. Because of their narrative power, these audits provide compelling evidence about discrimination for educating the public, for influencing fair housing policy, and for providing evidence in court. Nevertheless, audits also have some limitations. First, in-person audits are expensive and difficult to manage. Second, audits also only observe the marketing phase of a transaction and may miss discrimination that occurs in housing advertisements or after price negotiations begin, in the search for a mortgage, or, in rental markets, when the terms of the lease are specified. Third, audit studies are based on a sample of advertisements, usually from major metropolitan newspapers or ad-listing websites. These advertisements may not correspond to the actual housing experience of minority groups, who may use other means of identifying available housing or who may not be qualified for a share of
this advertised housing. The discrimination actually experienced by minority home seekers could therefore be higher or lower than the discrimination measured by an audit study. Evidence from HDS2000 addressed these issues by conducting some audits based on posted advertisements and by oversampling advertisements in neighborhoods with a high minority concentration. These steps did not lead to significant changes in measures of discrimination (See Turner et al., 2002).\textsuperscript{38}

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

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

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


<table>
<thead>
<tr>
<th>Study</th>
<th>Race/Ethnicity tested</th>
<th>Scale</th>
<th>Location</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMPS1977</td>
<td>Black</td>
<td>3,264 tests</td>
<td>40 metro areas</td>
<td>Individual real estate agents and apartment rental complexes were randomly selected from only one sample of newspaper advertisements for each metropolitan area. Auditors did not explicitly ask for the advertised unit. In rental audits, black auditors always preceded white auditors, but white auditors always preceded black auditors in all sales audits.</td>
</tr>
<tr>
<td>HDS1989</td>
<td>Black, Hispanic</td>
<td>3,800 tests</td>
<td>25 metro areas</td>
<td>Each audit began with a request for a specific, advertised unit randomly selected from the most recent Sunday newspaper (“anchoring” audit). It was the first study that measured racial and ethnic steering. For the purpose of steering analysis, auditors were instructed to ask about the availability of other homes similar in size and prices to the advertised unit. Order of initial call was randomized.</td>
</tr>
<tr>
<td>HDS2000</td>
<td>Black, Hispanic, Asian, Native American</td>
<td>4,600 tests</td>
<td>23 metro areas</td>
<td>Like HDS1989, the sample of housing units was randomly selected from the Sunday classified advertisements of major metropolitan newspapers. HDS2000 also used geographic over-sampling and supplemental sample from the secondary newspapers for areas that were under-represented in the newspaper advertisements. In addition, it recorded some of auditors’ actual characteristics such as income and education. Testers made appointment calls for sales and rental tests, and the order of initial call was randomized. On sales tests, testers were not to mention the advertised home during this call and were also to refrain from providing their personal and financial information. Testers inquired about the availability of the advertised housing unit that prompted their visit and about similar units.</td>
</tr>
<tr>
<td>HDS2012</td>
<td>Black, Hispanic, Asian</td>
<td>8,047 tests</td>
<td>28 metro areas</td>
<td>Testers attempted to make appointments for in-person visits by telephone or e-mail. Order of initial contact was randomized. On sales tests, testers were not to mention the advertised home during telephone conservation or email. If making an appointment was successful, testers used the in-person visit to learn about available homes or apartments. Testers inquired to view the home that was advertised. If told about at least one available housing unit, testers sought to inspect homes or apartments.</td>
</tr>
</tbody>
</table>
### Table 2. Results of HDS2012

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

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

Notes: Net measures are presented in the parentheses. * indicates statistical significance at 90%, 95%, or 99% level.
Table 3. Results of National Rental Tests

<table>
<thead>
<tr>
<th>Minority</th>
<th>Study</th>
<th>Advertised unit available</th>
<th>Inspected more units</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>% white favored</td>
<td>Net measure</td>
</tr>
<tr>
<td>Black</td>
<td>HMPS1977</td>
<td>30%</td>
<td>19%*</td>
</tr>
<tr>
<td></td>
<td>HDS1989</td>
<td>19%</td>
<td>7%*</td>
</tr>
<tr>
<td></td>
<td>HDS2000</td>
<td>12%</td>
<td>4%*</td>
</tr>
<tr>
<td></td>
<td>HDS2012</td>
<td>5%</td>
<td>0%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>HDS1989</td>
<td>17%</td>
<td>9%*</td>
</tr>
<tr>
<td></td>
<td>HDS2000</td>
<td>12%</td>
<td>7%*</td>
</tr>
<tr>
<td></td>
<td>HDS2012</td>
<td>5%</td>
<td>3%*</td>
</tr>
<tr>
<td>Asian</td>
<td>HDS2000</td>
<td>7%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>HDS2012</td>
<td>5%</td>
<td>2%</td>
</tr>
</tbody>
</table>

Source: Estimates of HMPS1977 come from Table 2 of Wienk et al. (1979); estimates of HDS1989 and HDS2000 (except for Asians) come from Exhibits 3-1 and 3-2 of Turner et al. (2002); and Asian estimates of HDS2000 and estimates of HDS2012 come from Exhibit V-1 of Turner et al. (2013).

Note: For net measures, * indicates statistical significance at 90%, 95%, or 99% level. Gross estimates (% white favored) are by definition statistically significant.
Table 4. Results of National Sales Tests

<table>
<thead>
<tr>
<th>Minority</th>
<th>Study</th>
<th>Advertised unit available</th>
<th>Inspected more units</th>
<th>Help with financing offered</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>% white favored</td>
<td>Net measure</td>
<td>% white favored</td>
</tr>
<tr>
<td>Black</td>
<td>HMPS1977</td>
<td>21%</td>
<td>10%*</td>
<td>38%</td>
</tr>
<tr>
<td></td>
<td>HDS1989</td>
<td>10%</td>
<td>4%*</td>
<td>24%</td>
</tr>
<tr>
<td></td>
<td>HDS2000</td>
<td>16%</td>
<td>1%</td>
<td>43%</td>
</tr>
<tr>
<td></td>
<td>HDS2012</td>
<td>13%</td>
<td>-1%</td>
<td>37%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>HDS1989</td>
<td>9%</td>
<td>4%*</td>
<td>27%</td>
</tr>
<tr>
<td></td>
<td>HDS2000</td>
<td>12%</td>
<td>-3%</td>
<td>36%</td>
</tr>
<tr>
<td></td>
<td>HDS2012</td>
<td>13%</td>
<td>0%</td>
<td>34%</td>
</tr>
<tr>
<td>Asian</td>
<td>HDS2000</td>
<td>16%</td>
<td>1%</td>
<td>46%</td>
</tr>
<tr>
<td></td>
<td>HDS2012</td>
<td>15%</td>
<td>3%</td>
<td>38%</td>
</tr>
</tbody>
</table>

Source: Estimates of HMPS1977 come from Table 25 of Wienk et al. (1979); estimates of HDS1989 and HDS2000 (except for Asians) come from Exhibits 3-11, 3-12, 3-14, 3-17, 3-18, and 3-20 of Turner et al. (2002); and Asian estimates of HDS2000 and estimates of HDS2012 come from Exhibit V-2 of Turner et al. (2013).

Note: For net measures, * indicates statistical significance at 90%, 95%, or 99% level. Gross estimates (% white favored) are by definition statistically significant.
Chart 1. Trends in Rental Discrimination

Advertised unit available (Gross measure)

Inspected more units (Gross measure)
Chart 2. Trends in Sales Discrimination

Advertised unit available (Gross measure)

<table>
<thead>
<tr>
<th>Year</th>
<th>Black</th>
<th>Hispanic</th>
<th>Asian</th>
</tr>
</thead>
<tbody>
<tr>
<td>'77</td>
<td>20</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>'89</td>
<td>15</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>'00</td>
<td>10</td>
<td>5</td>
<td>2.5</td>
</tr>
<tr>
<td>'12</td>
<td>5</td>
<td>2.5</td>
<td>1.25</td>
</tr>
</tbody>
</table>

Inspected more units (Gross measure)

<table>
<thead>
<tr>
<th>Year</th>
<th>Black</th>
<th>Hispanic</th>
<th>Asian</th>
</tr>
</thead>
<tbody>
<tr>
<td>'77</td>
<td>40</td>
<td>35</td>
<td>30</td>
</tr>
<tr>
<td>'89</td>
<td>35</td>
<td>30</td>
<td>25</td>
</tr>
<tr>
<td>'00</td>
<td>30</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>'12</td>
<td>25</td>
<td>20</td>
<td>15</td>
</tr>
</tbody>
</table>

Help with financing offered (Gross measure)

<table>
<thead>
<tr>
<th>Year</th>
<th>Black</th>
<th>Hispanic</th>
<th>Asian</th>
</tr>
</thead>
<tbody>
<tr>
<td>'89</td>
<td>20</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>'00</td>
<td>15</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>'12</td>
<td>10</td>
<td>5</td>
<td>2.5</td>
</tr>
<tr>
<td>'00</td>
<td>12</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>'12</td>
<td>5</td>
<td>2.5</td>
<td>1.25</td>
</tr>
</tbody>
</table>

Legend:
- HMPS1977
- HDS1989
- HDS2000
- HDS2012
Table 5. Steering Evidence from HDS Studies

<table>
<thead>
<tr>
<th>Minority</th>
<th>Study</th>
<th>Houses recommended in whiter tracts</th>
<th>Houses inspected in whiter tracts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>% white favored</td>
<td>Net measure</td>
</tr>
<tr>
<td>Black</td>
<td>HDS1989</td>
<td>6%</td>
<td>-6%</td>
</tr>
<tr>
<td></td>
<td>HDS2000</td>
<td>16%</td>
<td>4%*</td>
</tr>
<tr>
<td></td>
<td>HDS2012</td>
<td>25%</td>
<td>8%*</td>
</tr>
<tr>
<td>Hispanic</td>
<td>HDS1989</td>
<td>% non-Hisp. white favored</td>
<td>4%</td>
</tr>
<tr>
<td></td>
<td>HDS2000</td>
<td>17%</td>
<td>2%</td>
</tr>
<tr>
<td></td>
<td>HDS2012</td>
<td>23%</td>
<td>2%</td>
</tr>
</tbody>
</table>


Note: For net measures, * indicates statistical significance at 90%, 95%, or 99% level. Gross estimates (% white favored) are by definition statistically significant. Statistical significance of HDS1989 net measures is not available due to the lack of data.
Chart 3. Trends in Steering

Houses recommended in whiter tracts (Gross measure)

- HDS1989
- HDS2000
- HDS2012

Houses inspected in whiter tracts (Gross measure)

- HDS1989
- HDS2000
- HDS2012
Table 6. Predictions and Findings of Discrimination by Causal Hypotheses

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Agent-Prejudice Hypothesis</th>
<th>Customer-Prejudice Hypothesis</th>
<th>Statistical Discrimination Hypothesis</th>
</tr>
</thead>
</table>
| Agent           | Minority Agent of the Same Race (-)  
(Zhao, 2005; COY, 2005, 2008; ZOY, 2006)  
Age of Agent (+)  
(OSY, 1999; COY, 2005)  
Male Agent (+)  
(COY, 2005) | Size of Real-estate Agency (-)  
Use of Multi Real Estate Listing Service (-)  
(ZOY, 2006) | Size of Real-estate Agency (-)  
| Auditor         | Age of Auditor (-)  
(OSY, 1999)  
Male Auditor (+)  
(COY, 2005; ZOY, 2006) | Age of Auditor (-)  
(OSY, 1999)  
Male Auditor (+)  
(COY, 2005; ZOY, 2006)  
Assigned Marital Status (-)  
(ORY, 2003)  
Assigned Income (-)  
(Page, 1995; Zhao, 2005) | Assigned Income (-)  
(Page, 1995; Zhao, 2005)  
Asked Housing Value (+)  
(Page, 1995) |
| Neighborhood    | Percent Owner-occupied House (+)  
(OSY, 1999; Zhao, 2005)  
Percent White Residents (+)  
*Before tipping  
(Page, 1995; OSY, 1999; ORY, 2003; ZOY, 2006) | Percent White Residents (+)  
(Page, 1995; OSY, 1999; ORY, 2003; ZOY, 2006)  
Neighborhood Housing Value (+)  
(ORY, 2003; Zhao, 2005; COY, 2005) |

Notes: (+) indicates that discrimination against minority is positively correlated with the factor (for indicator variables, more discrimination if the factor is applicable).  
(−) indicates the discrimination against minority is negatively correlated with the factor (for indicator variables, less discrimination if the factor is applicable).  
The findings that support the predictions of each hypothesis are based on the advertised unit available/inspected, similar unit inspected, or the number of houses shown outcomes of black-white audits.
### Appendix A. Results of Housing Discrimination Audit Studies

<table>
<thead>
<tr>
<th>Authors</th>
<th>Data /Methodology</th>
<th>Scale</th>
<th>Other factors considered with race/ethnicity and gender</th>
<th>Location, Period, and market examined</th>
<th>Main findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ahmed, Andersson, &amp; Hammarstedt (2010)</td>
<td>Email correspondence tests Probit model</td>
<td>1,032 units (not paired: one email per unit)</td>
<td>Information on employment, education, marital status, age, smoking behavior, and rent payment issues</td>
<td>Sweden 2008 Rental tests</td>
<td>Arabic males have 15 to 23 percentage points and 17 to 20 percentage points lower probability of receiving a call back and being invited to further contacts than Swedish males. Providing additional information about themselves increases the probability of receiving a call back, being invited to further contacts, and being invited to showing by 12 to 19% for both Swedish and Arabic males. Increasing the amount of information in the application will not reduce discrimination in the housing market.</td>
</tr>
<tr>
<td>Ahmed &amp; Hammarstedt (2008)</td>
<td>Email correspondence tests Probit model</td>
<td>500 units (500*3=1,500 emails)</td>
<td></td>
<td>Sweden 2007 Rental tests</td>
<td>Arabic males have 21 to 26 and 7 percentage points lower probability of being invited to further contacts and being invited to a showing than Swedish males, respectively. Swedish males are almost 13 percentage points less likely to be invited to a flat showing than females.</td>
</tr>
<tr>
<td>Andersson, Jakobsson, &amp; Kotsadam (2012)</td>
<td>Email correspondence tests Probit model</td>
<td>950 units (not paired: one email per unit)</td>
<td>Socioeconomic status using occupation</td>
<td>Norway 2009-2010 Rental tests</td>
<td>Arabs have 13 percentage points lower probability of receiving a positive response compared to Norwegians. The probability of receiving a positive response is lowered by about 7 percentage points for males than females and by 7 percentage points for warehouse workers than economists.</td>
</tr>
<tr>
<td>Baldini &amp; Federici (2011)</td>
<td>Email correspondence tests Probit model</td>
<td>3,676 units (not paired: one email per unit)</td>
<td>Socioeconomic status using information on occupation and marital status</td>
<td>Italy 2010 Rental tests</td>
<td>Arabs have 22 percentage points lower probability of receiving a positive response than native Italians. East Europeans have 16 percentage points lower probability. The magnitude of discrimination is greater for men (24 percentage points) than for women (15 percentage points). Providing information reduces discrimination for foreign names, especially for males.</td>
</tr>
<tr>
<td>Bosch, Carnero, &amp; Farre (2010)</td>
<td>Email correspondence tests Probit, linear probability, and unit fixed-effects models</td>
<td>1,809 units (4,709 emails: 2 to 4 emails per unit)</td>
<td>Socioeconomic status using occupation</td>
<td>Spain 2009 Rental tests</td>
<td>Moroccan immigrants are 13 to 18 percentage points less likely to receive a response than native Spanish. Discrimination is much higher for Moroccan males (22 percentage points) than for Moroccan females (10 percentage points). Revealing positive information about occupation increases the contacts by about 6-8 percentage points. Positive information reduces discrimination only for males.</td>
</tr>
<tr>
<td>Country</td>
<td>Year</td>
<td>Test Type</td>
<td>Sample Size</td>
<td>Methodology</td>
<td>Ethnic Group</td>
</tr>
<tr>
<td>--------------</td>
<td>------</td>
<td>--------------------</td>
<td>-------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Spain</td>
<td>2008</td>
<td>In-person tests</td>
<td>135 units</td>
<td>Linear probability model</td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>1976</td>
<td>Rental tests</td>
<td>201 units</td>
<td>Linear probability model</td>
<td></td>
</tr>
<tr>
<td>UK</td>
<td>2011-2012</td>
<td>Rental tests</td>
<td>1,143 units (not paired: one email per unit)</td>
<td>Linear probability model</td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td>2010-2011</td>
<td>Rental tests</td>
<td>5,827 units (not paired: one email per unit)</td>
<td>Linear probability model</td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>2003</td>
<td>Email correspondence tests</td>
<td>1,115 emails (not paired: one email per unit)</td>
<td>Linear probability model</td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td>2010-2011</td>
<td>Rental tests</td>
<td>14,237 units (not paired: one email per unit)</td>
<td>Linear probability model</td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>1981</td>
<td>Rental tests</td>
<td>156 rental and 11.8 sales tests</td>
<td>Linear probability model</td>
<td></td>
</tr>
</tbody>
</table>

Note: The table includes studies conducted from 1976 to 2014, with a focus on telephone and email correspondence tests. The studies were conducted in different countries, including Spain, France, UK, Sweden, and France again, using various methodologies and sample sizes to measure discrimination in housing markets based on ethnic and racial groups.
<table>
<thead>
<tr>
<th>Study</th>
<th>Research Design</th>
<th>Tests/Emails</th>
<th>Socioeconomic Status</th>
<th>Cities</th>
<th>Rental and Sales Tests</th>
<th>Steering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hanson &amp; Hawley (2011)</td>
<td>Email correspondence tests</td>
<td>4,728 tests (4,728*2=9,456 emails)</td>
<td>Socioeconomic status using the prose quality of emails</td>
<td>10 U.S. cities 2009</td>
<td>Rental tests</td>
<td>There are no strong significant differences in the racial composition of neighborhoods in which white and African American auditors were shown houses. This steering result was based on the association of the housing characteristics and auditors’ race.</td>
</tr>
<tr>
<td></td>
<td>Probit model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>African Americans receive 4.5 percentage points (in the range between 3 and 26) lower email responses than whites. This difference ranged from over 8% in Boston and Los Angeles to less than 1% in Atlanta and Dallas. African Americans of higher social class experience small and not statistically distinguishable racial discrimination.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hanson, Hawley, &amp; Taylor (2011)</td>
<td>Email correspondence tests</td>
<td>3,153 tests (3,153*2=6,306 emails)</td>
<td>Socioeconomic status using the prose quality of emails</td>
<td>10 U.S. cities 2009</td>
<td>Rental tests</td>
<td>African Americans are treated less favorably than whites by landlords. Landlords reply faster, reply with an e-mail that is longer to inquiries made, make formal greetings, and use polite language when replying to e-mail inquiries from a white home seeker.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hanson &amp; Santas (2014)</td>
<td>Email correspondence tests</td>
<td>3,072 tests (3,072*2=6,144 emails)</td>
<td>Socioeconomic status using the prose quality of emails</td>
<td>21 large U.S. metropolitan areas 2011</td>
<td>Rental tests</td>
<td>Assimilated Hispanic Americans experience little discrimination, but recent Hispanic immigrants receive 2.9% lower response rates than whites. When email prose quality of the immigrants is low, discrimination doubles to 5.8% for non-response outcomes, and the incidence of discrimination is 6.89% for positive response outcomes.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hogan &amp; Berry (2011)</td>
<td>Email correspondence tests</td>
<td>1,124 tests (1,124*5=5,620 emails)</td>
<td></td>
<td>Toronto, Canada 2007</td>
<td>Rental tests</td>
<td>Arabs faced discrimination in 12 percent of experiments. The level of discrimination is modest but significant for Asian men (7 percent), blacks (5 percent), and Arabic women (5 percent). “Opportunity denying” discrimination (exclusion through nonresponse) was 10 times as common as “opportunity diminishing” discrimination (e.g., additional rental conditions).</td>
</tr>
<tr>
<td></td>
<td>Fixed-effects logit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>James, McCummings, &amp; Tynan (1987)</td>
<td>In-person tests</td>
<td>253 tests</td>
<td></td>
<td>Denver, CO 1982</td>
<td>Rental and sales tests Steering</td>
<td>African Americans and Hispanics have 1 to 11 percentage points higher probability of being told that either advertised or others similar houses are not available than whites (using net measures). Discrimination against minorities was higher in rental markets than sales markets. Minority homebuyers were offered less assistance with financing arrangements. In general, Hispanics experienced higher discrimination than African Americans. Discrimination in sales tests was evident in white neighborhoods, while discrimination against minority renters was evident in minority neighborhoods.</td>
</tr>
<tr>
<td>Massey &amp; Lundy (2001)</td>
<td>Telephone correspondence tests</td>
<td>79 units (79*6=474 calls)</td>
<td>Socioeconomic status using racially distinctive English (with</td>
<td>Philadelphia, PA 1999</td>
<td>Rental tests</td>
<td>African Americans with high social class have 1 to 5 percentage points lower probability of being told that a unit is available than whites. The gap for African Americans with low social class is 21 to 23 percentage points. Males have 4 to 10 percentage points higher probability of being told that a unit is available than</td>
</tr>
<tr>
<td>Author/Year</td>
<td>Methodology</td>
<td>Sample Size</td>
<td>Location/Time Period</td>
<td>Findings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>---------------------------------</td>
<td>--------------</td>
<td>----------------------</td>
<td>---------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>McIntosh &amp; Smith (1974)</td>
<td>In-person tests</td>
<td>178 tests</td>
<td>Five regions, UK</td>
<td>Much lower levels of net discrimination against <em>ethnic minorities</em> (<em>West Indians, Indian, Pakistani, or Greek</em>) were found in 1973 than in 1967.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(137 sales; 41 rental)</td>
<td>1967 &amp; 1973 Rental and sales tests</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearce (1979)</td>
<td>In-person tests</td>
<td>97 tests</td>
<td>Detroit, MI 1974-75</td>
<td><em>African Americans</em> have 44% [46%] lower probability of being shown houses by real estate agents. African Americans were shown houses in slightly higher percent black Census tracts and in communities with lower house values than whites.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purnell, Idsari, &amp; Baugh (1999)</td>
<td>Telephone correspondence tests</td>
<td>989 calls</td>
<td>San Francisco Bay Area, CA 1997</td>
<td><em>African Americans</em> have 8 to 41 percentage points lower rates of having a confirmed appointment to view an apartment in the white-majority areas, but 3 to 22 percentage points higher rate of having an appointment in black-majority areas. <em>Mexican Americans</em> have 7 to 17 percentage points lower rates of making an appointment than African Americans in all the areas.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roychoudhury &amp; Goodman (1992)</td>
<td>In-person tests Ordered probit</td>
<td>568 tests</td>
<td>Detroit 1980-1990 Rental tests</td>
<td>The probability that the agent does not suggest a unit to <em>African Americans</em> auditors increases by about 0.5. The probability that the African American auditors were refused an opportunity to inspect the unit increases by 0.58.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roychoudhury &amp; Goodman (1996)</td>
<td>In-person tests Ordered probit and bivariate probit</td>
<td>319 tests</td>
<td>Detroit 1980-1990 Sales tests Steering</td>
<td>Discrimination against <em>African Americans</em> was substantially less frequent among African American housing agents, and more frequent among older agents. An audit-by-audit comparison suggests that a considerable number of white home seekers were steered towards predominantly white and more affluent neighborhoods.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

Note: Gross measures are presented in brackets.
Appendix B. The Origins of Fair Housing Audits

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

To the best of our knowledge, the first published reference to fair housing audits appeared in The New York Times in 1956 (Rowland, 1956). This article describes audits conducted by the Committee on Civil Rights in East Manhattan (CCRM) in 1953 and 1954. These audits are also discussed in Frost (1958) and McEntire (1960).¹ This study “was supervised by a group of social scientists associated with the sponsoring Committee [CCRM]” (McEntire, 1960, p. 240n).² According to Frost (1958, p. 69), “CCRM was not seeking to duplicate the work of established organizations in the field of civil rights. Rather, it hoped to give to a varied group of individuals a more intensive experience as well as an opportunity to work cooperatively in producing changes. The technique selected for achieving this was that of the community audit.”

In the CCRM audits, a black person or a black couple “would visit the office of the real estate broker and inquire about the availability of apartments of a specified type. After the minority tester left, the control tester (white) would proceed to the same office with an identical

¹ These housing audits built on CCRM’s experience with restaurant audits, which were conducted as far back as 1950. These restaurant audits, like many of the housing audits that followed, involved a partnership between scholars and CCRM. According to Selltiz (1955), these scholars included Kenneth Clark (City College of New York), Dan Dodson (New York University), Herbert Hyman (Columbia University), Patricia Kendall (Queens College), Sophia M. Robison (New York School of Social Work), and Claire Selltiz (New York University). A detailed history of the restaurant audits is provided by Jou (2014).

² Although many scholars are listed in the preface to McEntire (1960), the particular social scientists who contributed to this audit study are not identified. It seems likely that Claire Selltiz was involved, because she was both a “technical consultant” to the CCRM restaurant audits (Selltiz, 1955, p. 19) and was listed in McEntire (1960, p. xi) as someone who prepared “Research memoranda on sociopsychological aspects of housing and minority groups.”
set of apartment specifications… [D]iscriminatory practice was to be the difference in treatment accorded the two testers” (p. 71). Testers were selected and trained, and “in November [1953] and March [1954] a total of 27 firms were visited and tested” (p. 72). In 22 out of the 27 cases, discrimination against African Americans was found. In addition, the CCRM conducted seventeen more tests in May 1954 based on advertisements in the Sunday *New York Times*. Discrimination was found in 10 of the tests.

McEntire (1960, p. 239) also reports on an audit study in Los Angeles in 1955, which is, so far as we know, the first fair housing audit study of the sales market. In this study, “a white couple, representing themselves as possible house buyers, called on twelve real estate brokers doing business in a new residential area of 12,000 homes, chiefly FHA- and VA-financed. The couple was followed after a brief interval by a Negro, also purporting to be looking for a house to buy.” Discrimination was found in every case. The report on this study was co-authored by James H. Kirk, a professor at Loyola University of Los Angeles.³

To the best of our knowledge, the first use of enforcement testing by a private fair housing group was in Brooklyn, New York in 1960 (Purnell, 2013). The first legislation in the country banning discrimination in private housing became effective in New York City in 1958, and the New York branch of the Congress of Racial Equality (CORE) decided to help enforce this act by conducting tests and, if necessary, giving the result to the relevant enforcement agency: the New York City Commission on Intergroup Relations. The first documented case of this strategy occurred in August 1960, when the New York CORE helped a black family who was told that a Brooklyn apartment they wanted was no longer available. Over the next five

³ The affiliation of the co-author, Lane D. Spane, is not indicated. This study appears to be one in “a series of related studies prepared by coöperating social scientists and other experts” upon which McEntire (1960) is “largely based” (p. x).
days, white testers were told over the phone that the apartment was still available, but testers who identified themselves as black were told that it had been rented. When the apartment was subsequently advertised in *The New York Times*, the black family was told once again that the apartment was no longer available, whereas a white tester who visited the rental agent’s office was told that he could rent the apartment. These techniques were then picked up by the Brooklyn branch of CORE. According to Purnell (2013), “Over the next year and a half, the chapter’s housing activists also improved on this basic model… and helped scores of African Americans move into apartments and homes in mostly white areas of Brooklyn.”

The first documented case of enforcement tests by private fair housing groups that were not linked to complaints comes from Chicago in the mid-1960s. The Coordinating Council of Community Organizations (CCCO) was formed in Chicago in April 1962; its first focus was on school segregation in the city (Cohen & Taylor, 2000, p. 285). Chicago passed a fair housing ordinance in 1963 and at some point CCCO initiated a testing program. This program came to light in 1966 when Martin Luther King, Jr. and the Southern Christian Leadership Council joined with the CCCO to initiate the Chicago Freedom Movement. As part of its efforts to combat discrimination, “The Freedom Movement had been sending testers into Gage Park,” one of the white neighborhoods in Chicago, and by the time this organization began a series of marches in white neighborhoods in July 1966, it “had already documented 121 cases of racial discrimination” (Cohen & Taylor, 2000, p. 392). An all-night vigil was held “at F. H. Halvorsen Realty in Gage Park… because, according to recent testing, it repeatedly discriminated against black applicants” (Cohen & Taylor, 2000, p. 392).

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

The passage of the Fair Housing Act and the U.S. Supreme Court’s resurrection of the Civil Rights Act of 1866, both of which occurred in 1968, greatly expanded the opportunity for audits to be used as an enforcement tool. Most importantly, this act gave private fair housing groups the standing to sue alleged discriminators. As a result, the use of audits by private fair housing groups quickly spread. The first testing-based case to appear in Federal Court relied on evidence from tests conducted in 1968 in Brown County, Ohio (Schwemm, 1992, p. 24). As pointed out by Yinger (1995), “how-to manuals for conducting audits were widely available” (p. 20) by the early 1970s. See, for example, Murphy (1972), Kovar (1974), or Leadership Council for Metropolitan Open Communities (1975). An assessment of the early use of audits by private fair housing groups is provided by Freiberg (1993). Eventually, governmental civil-rights enforcement agencies also started using audits, and they became a crucial tool in fair housing cases (Schwemm, 1992). In 1982, the use of audits as a fair housing enforcement tool was upheld by the U.S. Supreme Court.4

Scholars also recognized the power of audits to uncover discriminatory behavior, and more audit studies began to appear as early as the 1960s. Housing audits were conducted in Great Britain, for example, starting in 1967. See Daniel (1968) and McIntosh and Smith

In the United States, a study published in 1971 (Johnson, Porter, & Mateljan, 1971) used audits to examine discrimination against blacks and Mexican-Americans in a Southern California city. Another 1971 audit study, which examined the behavior of real estate brokers and landlords in Akron, Ohio, was “devised by the author [a professor at Kent State University] as part of the ongoing research program of the Fair Housing Contact Service of Akron (FHCS), a voluntary open housing group” (Saltman, 1975, p. 41). A study conducted in Detroit in 1974-5 used audits to examine racial steering by real estate brokers (Pearce, 1979). In 1977, scholars at the U.S. Department of Housing and Urban Development directed the first nation-wide audit study (Wienk et al., 1979).

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

Moreover, to the best of our knowledge, the first appearance of correspondence audits appeared in a study of racial discrimination in employment in Britain (Jowell & Prescott-Clarke, 1970).

Dr. Saltman was one of the founders of DHCS in 1965. See Walbeck (1974).

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

HUD had a partner for this study, namely, the National Committee against Discrimination in Housing, which carried out the audits. See Wienk et al. (1979).
tester evidence was used in a suit brought by the Attorney General” involved tests conducted in March 1970 (Schwemm, 1992, p. 40).

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

1 The origins of fair housing audits are described in Appendix B.

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

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

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

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

6 As discussed below, some audit studies supplement the sample from newspaper advertisements with audits from additional sources.

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

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

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

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

11 Another disadvantage of non-paired audits is that they do not yield gross measures of discrimination.

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

13 Several scholars have identified more specific hypotheses in this category. Galster (1987, 1990c) and Newberger (1989) argue that agents may or may not show houses to blacks in some neighborhoods because of “anticipated discrimination” against them there by white home sellers or mortgage lenders. Galster also argues that agents may discriminate on the basis of their beliefs about what customers in different groups prefer. Ondrich et al (2003) argue that agents may have stereotypes about the financial capabilities of people in certain groups that lead to inter-group differences in treatment despite the equal qualifications of audit teammates.

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

15 This point was clearly explained in Wienk et al. (1979, pp. 175-176).

16 For more on a review of earlier audit literature, see Yinger (1987).

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

18 The outcomes presented in Tables 3 and 4 that are similarly measured for the four national studies are different from the outcomes presented in Table 2.

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

20 This paper uses African American and black as synonyms.

21 The HDS 2012 report used seven summary outcome measures; differential denial of in-person meeting, differential denial of available units, differential number of units recommended, differential number of units shown, differences between testers in agent helpfulness, differential rent or sales price, and differential neighborhood racial/ethnic composition.
Steering can be also analyzed at other geographical levels such as municipalities and school districts (See Galster and Godfrey, 2005).

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

See also Table 3 of Zhao (2005) and Exhibit 1 of Choi et al. (2008).


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

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

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

Private fair housing groups also learned to use audits to shed light on the nature and extent of discrimination in their service area. Galster (1990a) reviews seventy-one audit studies by private fair housing groups in the 1980s.

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

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


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

Many of these scholars also point out that audits could increase the effectiveness of fair housing enforcement efforts by helping enforcement officials identify the circumstances under which discrimination is most likely to occur. It is not clear whether federal enforcement officials have taken this lesson to heart.

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


It is also possible to adjust discrimination measures so that they correspond to the actual income distribution of the minority group under study. Yinger (1995) finds that this type of calculation has little impact on the results of the 1989 HDS. However, some other research
suggests that less discrimination occurs when testers reveal that they have a more professional job (Baldini & Federici, 2011; Hanson & Hawley, 2011; Ewens et al., 2014). This possibility might be more common in actual searches than in audit studies. Minority home seekers also might avoid housing agents who are known to discriminate. This type of behavior imposes a cost on these households, of course, but it also might lower measured discrimination. For further discussion of this issue, see Ross and Yinger (2006).

39 The HDS1989 report on steering (Turner, Mikelsons, & Edwards, 1990) only considered a difference in percent white of more than five percentage points as discrimination, so we drew the estimates from the HDS2000 report, which used the same measurement for its analysis of both HDS1989 and HDS2000.