

What Does “Testing” Tell Us About the Incidence of Discrimination in Housing Markets?[†]

Michael J. Yelnosky *

Professor Paula Franzese, the organizer of this symposium, asked me to respond briefly to the following question: “What do we know about the incidence of discrimination in housing?” The question implies that there is a real-world phenomenon — housing discrimination — that is important to understand. Although academics often ignore reality,¹ I assume most would agree that it is important to determine whether housing discrimination exists, how much there is, and what form or forms it takes. This information can help determine whether intervention in housing markets is appropriate and, over time, help measure the efficacy of various regulatory approaches.² Thus, my first response to Professor Franzese is, “That is a good question.”

My second response is, “That is a hard question.” I have encountered it in studying the law’s prohibition of employment discrimination. The two phenomena, housing discrimination and employment discrimination, have something important in common for our purposes. They are difficult to measure because they are often

[†] Editor’s Note: This article is based upon a presentation given at the Housing and Hope Symposium at Seton Hall University School of Law.

* Associate Professor, Roger Williams University School of Law; Visiting Associate Professor, Seton Hall University School of Law (1998-1999). B.S., University of Vermont, 1982; J.D., University of Pennsylvania, 1987.

The author thanks Professor Paula Franzese for inviting him to participate in this symposium.

¹ See, e.g., Richard A. Posner, *Madison Lecture: Against Constitutional Theory*, 73 N.Y.U. L. REV. 1, 1-4 (1998) (criticizing the “academification” of law school professors — an increased emphasis on theory rather than empirical knowledge — which makes most legal scholarship relevant only to other academics and increases the isolation of law professors from practicing lawyers and judges); cf. Alan Hyde, *Employment Law After the Death of Employment*, 1 U. PA. J. LAB. & EMPL. L. 99, 100-01 (asserting that there is a great deal of denial in legal academia).

² See Michael Fix, et al., *An Overview of Auditing for Discrimination*, in CLEAR AND CONVINCING EVIDENCE: MEASUREMENT OF DISCRIMINATION IN AMERICA 11-12, 35-37, 41-42 (Michael Fix & Raymond J. Struyk eds., 1993) [hereinafter CLEAR AND CONVINCING EVIDENCE]; William R. Tisdale, *Going Beyond the Housing Discrimination Study: Comments*, in CLEAR AND CONVINCING EVIDENCE, *supra*, at 113-15.

hidden, covert, or otherwise difficult to identify.³ To be precise, by “discrimination” I am referring to what fair housing and fair employment law call disparate treatment — the intentional differential treatment of a person because of that person’s membership in a protected class — for example, because that person is black.⁴

Considering that definition of discrimination, the measurement problem becomes fairly obvious. We know, for example, that real estate agents regularly refuse to show housing units to potential purchasers or renters. An agent might refuse to show a unit because it is no longer available, because the agent is tired and wants to go home early, or because the agent thinks an agreement to sell or rent the unit is imminent. As long as the race of the prospective purchaser or renter did not influence the agent’s decision, an accurate measure of intentional discrimination will record these outcomes as non-discriminatory. To measure housing discrimination accurately, then, we need a method for sifting through the scores of adverse outcomes (those in which the prospective renter or purchaser is denied a unit) and distinguishing between those caused by intentional race discrimination and those caused by other factors.⁵ Without a method for making this distinction, it is difficult to know how much discrimination exists “on the ground”⁶ in a country where the signs that read “No Blacks” have been taken down.

Several approaches are possible. First, one could survey the actors — ask real estate agents in a particular market, for example, “Do you discriminate?” Or one could ask apartment hunters in that same

³ For a discussion of subtle forms of housing discrimination, see Teresa Coleman Hunter & Gary L. Fischer, *Essay: Fair Housing Testing — Uncovering Discriminatory Practices*, 28 CREIGHTON L. REV. 1127, 1128-29 (1995). With respect to employment discrimination, Title VII proof structures were formulated in part to compensate for the fact that evidence of intentional discrimination is hard to come by. See *Price Waterhouse v. Hopkins*, 490 U.S. 228, 271 (1989) (O’Connor, J., concurring).

⁴ See, e.g., *St. Mary’s Honor Ctr. v. Hicks*, 509 U.S. 502, 511 (1993) (interpreting Title VII, 42 U.S.C. §§ 2000e - 2000e-17 (1996)); *Gamble v. City of Escondido*, 104 F.3d 300, 304-05 (9th Cir. 1998) (interpreting the Fair Housing Act, 42 U.S.C. §§ 3601-3631 (1996)). For a discussion of critiques of the intent-based approach to Title VII enforcement see Michael J. Yelnosky, *Title VII, Mediation, and Collective Action*, 1999 U. ILL. L. REV. (forthcoming) (on file with author).

⁵ For purposes of enforcing Title VII the Supreme Court has created a proof structure for making that distinction. That structure, articulated in *Texas Department of Community Affairs v. Burdine*, 450 U.S. 248, 253-256, 256 n.6 (1981) and *McDonnell Douglas Corp. v. Green*, 411 U.S. 792, 802 (1973), has been adopted by the courts of appeals for use in litigation of Fair Housing Act disparate treatment cases. See, e.g., *Gamble*, 104 F.3d at 304-05.

⁶ “On the ground” appears to be the currently fashionable substitute for phrases like, “in the real world” or “out there.”

market, "Were you discriminated against in your search?" Because of the possibility of self-interested responses and imperfect information, relying on a survey to provide an accurate measure of housing discrimination would be unwise.⁷

A more widely accepted approach involves studying the aggregate results of transactions in a particular market.⁸ To determine the incidence of racial discrimination in home sales, for example, the starting point would be identifying the rates of home ownership among blacks and whites in that market. Discrimination against blacks would be suggested if a larger percentage of whites than blacks owns homes in that market. However, while the differential rates of home ownership might be the product of intentional racial discrimination, other factors, such as disparities between members of the two racial groups in income or desire to own a home, might be at work. Differences in the availability of homes for sale in the particular areas where members of the two groups chose to concentrate their searches might also explain the home ownership rates. Through the use of regression analysis, researchers try to control for these non-discriminatory factors and calculate a residual difference in outcomes that is not associated with them. That residual difference in outcomes is ascribed to discrimination. If the difference is statistically significant, the study concludes that discrimination exists, and it is roughly quantified.⁹

Although courts have admitted the results of these statistical analyses as probative of intentional discrimination,¹⁰ the methodology employed in a particular case is always subject to criticism. For example, researchers can be criticized for failing to control for all the possible non-discriminatory variables that might explain the differential outcomes.¹¹ If the two groups being compared are not identically situated to begin with, and of course they never are, then differential outcomes alone do not prove intentional discrimination. Moreover, wholly random factors can influence outcomes. It is difficult for a regression study to control for the agent who discourages a black apartment seeker not because of race, but because the agent simply

⁷ See Fix, et al., *supra* note 2 at 13.

⁸ See *id.* at 10; see also John Yinger, *Testing for Discrimination in Housing and Related Markets*, in A NATIONAL REPORT CARD ON DISCRIMINATION IN AMERICA: THE ROLE OF TESTING 27 (Michael Fix & Margery Austin Turner eds., 1999).

⁹ See Fix, et al., *supra* note 2, at 7.

¹⁰ See, e.g., *Teamsters v. United States*, 431 U.S. 324, 337, 340 n.20 (1977) (permitting use of statistics comparing the racial composition of an employer's work force to the racial composition of the labor pool from which it drew that work force to prove intentional race discrimination).

¹¹ See Fix, et al., *supra* note 2, at 7-8, 12.

wants to leave work early and the black apartment seeker appeared in the office at 4:30. The point is simply that this statistical method requires an inference — from differential outcomes to discriminatory treatment — and there are commonly reasonable arguments over whether that inference is sound.

A third technique — the one I want to focus on — is called “testing” or “auditing.” It grew out of the work of fair housing groups in the 1970s. Assume that such a group received a complaint of suspected discrimination from a twenty-five-year-old black male. More specifically, let us assume that he appeared at the office of the fair housing group and explained that he had just come from the office of a landlord who, the young man believed, refused to show him an apartment because he is black. The young man explained that the day before, he had responded by telephone to newspaper advertisement of an apartment for rent. When he called, the landlord said that he would be showing the apartment the following morning, beginning at 9:00 a.m. At 10:00 a.m. the next morning, when the young man arrived at the landlord’s office, he was told the apartment was no longer available.

At this point the fair housing group is in the same position as a researcher trying to assess discrimination based solely on adverse outcomes. The apartment might have been rented by the time the young man appeared at the landlord’s office. It is also possible, however, that it was still available, and the landlord refused to show it because of the young man’s race. If the fair housing group then sent its twenty-five-year-old male intern, who is white, to pose before the landlord as a prospective renter interested in the apartment, the group might be better able to determine whether the black male received discriminatory treatment. If the intern came back and reported that the apartment was still available and the landlord had shown it to him, the group would have powerful evidence to pursue a claim of discrimination on behalf of the black male. Many groups have used the evidence generated by “testers” in support of discrimination claims asserted by individuals such as our hypothetical black complainant.¹²

Fair housing groups also began to send matched pairs of testers — two individuals matched for all relevant personal characteristics other than the one that is presumed to lead to discrimination¹³ — to identify unlawful practices “in advance.”¹⁴ In such a scenario, the two

¹² See generally Hunter & Fischer, *supra* note 3, at 1131-34.

¹³ See Fix, et al., *supra* note 2, at 1.

¹⁴ See Hunter & Fischer, *supra* note 3, at 1136-39.

testers, one white and one black, both males or females of approximately the same age, are trained to behave similarly when interacting with the agent, and they are assigned similar incomes, occupations, and family characteristics for purposes of the "test."¹⁵ In 1982 the Supreme Court, in *Havens Realty Corp. v. Coleman*,¹⁶ held that a tester who was told, falsely, that an apartment was not available, had standing to sue for a violation of the Fair Housing Act, even though the tester had no interest in renting the apartment.¹⁷

The use of this technique to enforce the fair housing laws is not the focus of these remarks. To respond to Professor Franzese's question, I want to discuss the strengths and weaknesses of the technique as a device for measuring the incidence of discrimination in a particular housing market. In this research context, the technique is often referred to as "auditing," to distinguish it from its use as an enforcement tool, where it is referred to as "testing."¹⁸ To measure housing discrimination, coordinated audits of the behavior of real estate agents across the market being studied are required.¹⁹ These audits began their ascendancy as research tools in 1979, when the United States Department of Housing and Urban Development (HUD) sponsored a national study of race discrimination in home sales and rental markets based on more than 3,000 audits in forty metropolitan areas.²⁰ In 1989, HUD sponsored the second national testing study, which involved almost 4,000 audits with matched

¹⁵ See John Yinger, *Access Denied, Access Constrained: Results and Implications of the 1989 Housing Discrimination Study*, in CLEAR AND CONVINCING EVIDENCE, *supra* note 2, at 70.

¹⁶ 455 U.S. 363 (1982).

¹⁷ See *id.* at 373-74. The Court also held that Housing Opportunities Made Equal (HOME), the fair housing organization that hired the testers to investigate suspected housing discrimination, had standing to sue. See *id.* at 379. HOME claimed that it had to devote significant resources to identify and to counteract the defendant's practices, thereby frustrating its efforts to assist in establishing equal access to housing and to provide referral services. See *id.* The Court held that these allegations were sufficient to constitute injury for standing purposes because they were more than allegations of a setback to the organization's abstract social interests. See *id.* at 379.

For a thorough discussion of the issue of tester standing under the Fair Housing Act and Title VII, see generally Michael J. Yelnosky, *Filling an Enforcement Void: Using Testers to Uncover and Remedy Discrimination in Hiring for Lower-Skilled, Entry-Level Jobs*, 26 U. MICH. J.L. REFORM 403 (1993); Leroy D. Clark, *Employment Discrimination Testing: Theories of Standing and a Reply to Professor Yelnosky*, 28 U. MICH. J.L. REFORM 1 (1994); Michael J. Yelnosky, *Salvaging the Opportunity: A Response to Professor Clark*, 28 U. MICH. J.L. REFORM 151 (1994) [hereinafter Yelnosky, *Salvaging the Opportunity*].

¹⁸ See Fix, et al., *supra* note 2, at 1.

¹⁹ See *id.*

²⁰ See *id.* at 10-11.

white/black and white/Hispanic testers in twenty-five metropolitan areas who responded to newspaper advertisements of rental or sale properties.²¹ I review the results of that study below. Between 1995 and 1997, fair housing groups have conducted coordinated audits in five metropolitan areas.²²

The most attractive feature of auditing to measure housing discrimination is that it permits close observation of spontaneous reactions of real estate agents, sellers, and landlords to two prospective individual renters or buyers who are alike except for their race. Thus, auditing might permit researchers to identify and quantify more accurately discriminatory treatment.²³ Through careful training and matching of testers, the audit methodology, unlike regression analysis, can better control for the questions prospective renters or buyers ask, the way they behave when talking to the agent, including their responses to agent inquiries, and the timing of visits to agents.²⁴ Because audits permit close observation of agent behavior, they can help researchers identify the particular forms discriminatory treatment might take.²⁵ Finally, audit results have a powerful narrative force that regression studies lack. They are based not on inferences from observed outcomes, but on instances of actual, observed, treatment that are aggregated for study.²⁶

²¹ See *id.* at 18.

²² See Yinger, *supra* note 8, at 29.

²³ See Fix, et al., *supra* note 2, at 12.

²⁴ See Yinger, *supra* note 8, at 30.

²⁵ See Fix, et al., *supra* note 2, at 14-15.

²⁶ See *id.* at 13; Yinger, *supra* note 8, at 30. The narrative power of audits was made apparent to me in an episode of Michael Moore's short-lived television program, *TV Nation*. Moore conducted an informal but powerful audit of the behavior of cab drivers in New York City. A television camera documented repeated instances in which cab drivers passed a well-dressed black male (a lawyer at a prestigious New York law firm) in order to pick-up a casually dressed white male (who happened to be a thrice-convicted felon). See also *PrimeTime Live*, (ABC television broadcast, Sept. 26, 1991) (using undercover cameras and testers to document discrimination in employment, car sales, and housing).

Characteristic	1	2	3	4	5
	*	**	***	***	
	Opp'y Deny	Opp'y Deny	Opp'y Deny/ Diminish	Opp'y Deny/ Diminish	Opp'y Deny/ Diminish
	Minority (%)	White/Anglo (%)	Minority (%)	White/Anglo%	Minority(%)
	(Gross Measure)			(Net Measure)	
Black/White Rental	15	5	46	18	28
Black/White Sales	8	1	50	20	30
Hispanic/Anglo Rental	12	6	43	20	23
Hispanic/Anglo Sales	8	3	45	21	24

* The minority rental or home seeker was essentially denied any information while the information was given to the majority renter or home seeker.

** The majority rental or home seeker was essentially denied any information while the information was given to the minority rental or home seeker.

*** Opportunity diminishing acts are defined as follows: While either the minority or majority auditor was able to learn about the unit requested, he was treated less well than his counterpart. In housing, for example, one tester is offered fewer units, or the minority tester is "steered" to neighborhoods with more minorities.²⁷

With that simple explanation of the methodology employed, we can look at the results of the 1989 HUD study to see what they tell us about the incidence of discrimination in housing markets. Column 1 shows the percentage of audits in which the minority tester was denied access to the housing unit in question while the majority tester was given access. Those figures are reported separately depending on whether the testers were posing as individuals seeking to rent or purchase a housing unit. In the rental markets studied, in 12% to 15% of the audits the minority tester was denied access while access was granted to the majority tester. In the markets for home sales, more favorable treatment of the majority tester occurred in 8% of the audits.

Column 3 shows a dramatically higher rate of unfavorable treatment of minority testers when the authors of the study considered what they called "opportunity-diminishing behavior." A minority-group tester received opportunity-diminishing treatment when, as compared to the matched majority-group tester, the minority-group tester was shown fewer housing units, quoted less favorable rental or sales terms, steered to black or Hispanic neighborhoods, told to call back, or not offered help finding financing.²⁸ The HUD study results

²⁷ 1989 Housing Discrimination Study, United States Department of Housing and Urban Development, *reprinted in* CLEAR AND CONVINCING EVIDENCE, *supra* note 2, at 21.

²⁸ This treatment imposes higher search costs for black and Hispanic home

show that minority-group testers received less favorable treatment than their matched counterparts more than 40% of the time.

These results suggest that black and Hispanic home seekers are sometimes discriminated against because of their race or national origin. Thus, my next response to Professor Franzese's question is, "We know some individuals looking to rent or buy homes are being discriminated against in their search because of their race or national origin."

What does the HUD study tell us about how often this occurs? Can we conclude that 100% of the difference in treatment observed in the audits is attributable to intentional discrimination based on race or national origin? What about genuine variations in the circumstances facing the testers, such as when the unit in question is rented in the time between the visits of two matched testers? What about variable agent behavior that is not motivated by race, such as our agent who wants to go home early on a Friday afternoon? What about variable tester behavior, such as failure of one tester to follow the script, either because of a mistake or because she wants to influence the result? Finally, can we assume that testers report their experiences accurately?²⁹

Although there are several approaches to these measurement problems, discussing two will suffice both to highlight some important issues and refine my response to Professor Franzese's question. The first approach is referred to as the gross unfavorable treatment measure, and it essentially ignores these problems by assuming that differences in treatment observed in audits are caused only infrequently by non-discriminatory factors.³⁰ It thus treats all instances of less favorable treatment as discrimination. A gross measure proponent would assert that the more than 40% figures for less favorable treatment of minorities, listed in column 3 of the HUD chart, are the relevant and accurate measures of discrimination.

"Net measure proponents," on the other hand, warn that the gross measure may overstate the incidence of discrimination by ignoring the role non-discriminatory factors may play in less favorable treatment. Net measure proponents prefer a more conservative ap-

seekers, relative to whites, and even successful searches often result in less favorable transactions. See Yinger, *supra* note 8, at 36.

²⁹ See Fix, et al., *supra* note 2, at 20, 25-26, 29; James J. Heckman & Peter Siegelman, *The Urban Institute Studies: Their Methods and Findings*, in CLEAR AND CONVINCING EVIDENCE, *supra* note 2, at 191, 215-16; Ian Ayres, *Fair Driving: Gender and Race Discrimination in Retail Car Negotiations*, 104 HARV. L. REV. 817, 825-26; Yelnosky, *Salvaging the Opportunity*, *supra* note 17 at 161-65; Yinger, *supra* note 8, at 31-32.

³⁰ See Fix, et al., *supra* note 2, at 27; Yinger, *supra* note 8, at 29-32.

proach to reporting the incidence of discrimination. However, they recognize the difficulty in quantifying precisely the effect of non-discriminatory factors on less favorable treatment. The problem is similar to that encountered by researchers trying to refine a regression analysis to control for the effects of non-discriminatory factors on observed outcomes. Many net measure proponents suggest using the percentage of audits in which the minority tester was *avored* as an estimate of the extent to which non-discriminatory factors cause less favorable treatment of minority-group testers.³¹ Thus, for net measure proponents the relevant figures are in column 5 of the HUD chart. By this measure the 1989 HUD study still shows significant rates of housing discrimination: In approximately 25% of the audits the minority group tester was treated less favorably than her matched, white partner because of race or national origin.³²

Net measure critics point out that using the percentage of audits in which the white tester was favored as a proxy for the percentage of audits in which unfavorable treatment of minority testers resulted from non-discriminatory factors is too crude. They suggest more sophisticated statistical models to sort among audits in which the minority tester was treated less favorably.³³ Some critics also assert that use of the net measure can result in attributing a portion of the less favorable treatment of minorities to non-discriminatory factors where the more favorable treatment of minorities does not warrant an offset. For example, an offset is inappropriate when whites are shown fewer houses than their matched partners because whites are being steered into white neighborhoods with fewer units for sale.³⁴ This "opportunity-diminishing" behavior toward whites might actually make us more comfortable concluding that less favorable treatment of black testers observed in other audits was the product of intentional race discrimination. Net measure could also underestimate the incidence of discrimination, critics assert, because it would result in a conclusion that no discrimination exists in a housing market where both blacks and whites are discriminated against in 30% of audited housing searches. Instead of reporting that there is much

³¹ See Fix, et al., *supra* note 2, at 27-29. Heckman and Siegelman justify this approach on the ground that it is consistent with the view that whites experience little discrimination. Thus, less favorable treatment of white testers can safely be attributed to non-discriminatory factors. See Heckman & Siegelman, *supra* note 29, at 198.

³² The numbers in column 5 are arrived at by subtracting from the gross number of audits in which the minority tester received less favorable treatment (column 3) the number of audits in which the majority tester was favored (column 4).

³³ See Fix, et al., *supra* note 2, at 27; Yinger, *supra* note 8, at 32.

³⁴ See Yinger, *supra* note 8, at 32.

discrimination in that market, net measure would report that there is none.³⁵

Trying to resolve these issues is certainly beyond the scope of these remarks. However, assuming the integrity of the audits conducted by HUD in 1989, everyone seems to agree that the net measure likely *understates* the incidence of opportunity-diminishing behavior in the markets studied.³⁶ The results of the 1989 HUD study are similar to the results of HUD's 1979 study.³⁷ And the preliminary results of five studies conducted since 1995 show gross measures of opportunity-diminishing behavior against blacks and Hispanics of 50% in the rental markets of four metropolitan areas studied and 40% in the sales and rental markets of Washington, D.C.³⁸ However, no national testing study of housing discrimination has been conducted since 1989. Thus, my final response to Professor Franzese's question is, "Audit studies show housing discrimination exists. Conservative estimates suggest that in approximately 25% of searches for housing advertised in a major metropolitan area, blacks and Hispanics will be treated less favorably because of their race and ethnicity."³⁹

³⁵ See Fix, et al., *supra* note 2, at 28.

³⁶ See *id.* at 29; Yinger, *supra* note 8, at 38, 42.

³⁷ See Yinger, *supra* note 8, at 34.

³⁸ See *id.*

³⁹ See Fix, et al., *supra* note 2, at 35; Yinger, *supra* note 8, at 36.