



Assessing Strategies to Combat Housing Discrimination in New York City: A Field Experiment

EVALUATION REPORT

DRAFT VERSION – NOT FOR PUBLIC CIRCULATION

Prepared for the New York City Commission on Human Rights

Albert H. Fang

Andrew M. Guess

Macartan Humphreys

Columbia University

May 2014

Contents

Acknowledgments and Front Matter	4
Executive Summary	5
List of Tables	10
List of Figures	11
1 Introduction	14
1.1 Study Overview and Aims	14
1.2 Summary of Main Findings	15
1.2.1 Baseline Levels of Racial Housing Discrimination	15
1.2.2 Effects of Messaging on Discrimination Incidence	15
1.2.3 Varieties of Racial Housing Discrimination	16
1.3 Organization of Report	17
2 Study Design and Methodology	18
2.1 The Timeline of a Housing Audit	18
2.2 Treatment Messages	20
2.2.1 Treatment Scripts	20
2.2.2 Theoretical Motivations for Treatment Design and Main Hypotheses	21
2.3 Randomization Procedure and Treatment Message Delivery	22
2.4 Varying Putative Signals of High versus Low Employment Stability	24
2.5 Data Collection	25
2.6 Measuring Discrimination	27
2.6.1 The Net Measure of Discrimination	27
2.6.2 Objective Measures of Discrimination (Post-Treatment Outcome)	28
2.6.3 Subjective Measures of Discrimination (Post-Treatment Outcome)	29
2.6.4 Measuring Early Stage Discrimination (Pre-Treatment)	30
3 Study Context and Sample Definitions	32
3.1 Study Context and Dates of Field Implementation	32
3.2 Sample Definitions	33
3.3 Distribution of Cases Across Boroughs, by Sample	37
4 Assessing Rental Housing Discrimination in New York City	39
4.1 Baseline Levels of Discrimination	39
4.1.1 Incidence of Early-Stage Discrimination	39
4.1.2 Incidence of Later-Stage Discrimination	42
4.2 Subjective Perceptions of Favorable Interactions during Appointments	42
4.2.1 Positive Editorializing	43
4.2.2 Sales Efforts	43
4.2.3 Praise About Qualifications	44
4.2.4 Positive Response to Background	44

4.2.5	Professionalism	45
4.3	Forms of Objective Discrimination Encountered During Appointments	45
4.3.1	Differences in the Number of Units Shown	46
4.3.2	Differences in Average Quoted Monthly Asking Rents Across Units Shown	49
4.3.3	Differences in Signaled Interest in the Housing Seeker	51
4.3.4	Differences in the Characteristics of Housing Units Shown	53
4.3.5	Differences in the Quoted Rental Price and Amenities-in-Rent for Listed Unit, if Shown	56
4.3.6	Differences in Move-In Requirements for Listed Unit, if Shown	57
4.3.7	Differences in Willingness to Negotiate on Fees	65
5	Impacts of Messaging on Discrimination	66
5.1	Estimated Effects of Messaging on Discrimination Incidence	68
5.1.1	Monitoring versus Control	68
5.1.2	Punitive versus Control	69
5.1.3	Punitive versus Monitoring	70
5.2	Estimated Effects of Message Content on Discrimination Incidence	71
5.2.1	Monitoring Message versus Control for Partial Compliers	74
5.2.2	Monitoring Message versus Control for Full Compliers	74
5.2.3	Punitive Message versus Control for Full Compliers	74
5.2.4	Punitive Message versus Monitoring Message for Full Compliers	74
5.2.5	Estimated Proportions of Compliance Types in the Experimental Sample .	75
5.3	Effects of Messaging Conditional on Putative Low Employment Stability Signals .	75
6	Conclusions	77
6.1	Summary of Main Findings	77
Appendices		79
A	Additional Implementation and Field Procedures	79
A.1	Sampling	79
A.1.1	Additional Details about the Sampling Procedure	79
A.1.2	“Scraping” Public Listings to Sample Available Units in the Housing Market	80
A.1.3	Best Practices for Future Approaches Using Online Listings	80
A.2	Manipulating Markers of Racial Identity	81
A.2.1	The Racial Soundingness of Names	81
A.2.2	Linguistic and Class-Correlated Signals of Race	82
A.3	Procedures for Screening and Hiring Testers to Pose as Interested Housing Seekers	83
B	Survey Instruments	84
C	Estimation	89
C.1	Framework	89
C.2	ITT Estimation	89
C.3	CACE Estimation	89
C.4	CATE Estimation	90

D Additional Results	92
D.1 Distributions of Indicators of Objective Discrimination during the Appointment Stage	92
D.1.1 Across Units Shown	92
D.1.2 Quoted Terms of Rent for Listed Unit, if Shown	100
D.1.3 Incidence of Quoted Fee Types for Listed Unit, if Shown	101
D.1.4 Quoted Fee Amounts for Listed Unit, if Shown	106
D.1.5 Willingness to Negotiate for Listed Unit, if Shown	108
D.2 Effects of Treatment Messaging on Net Discrimination: Unweighted Data	110
D.3 Full Results: Bayesian Estimates of Average Treatment Effects of Message Content by Complier Type	113
D.4 Estimated Effects of Messaging by Putative Low versus High Employment Stability Signal	125
Bibliography	131

Acknowledgments and Front Matter

This study was conducted with the generous support of the New York City Commission on Human Rights, the New York City Department of Housing Preservation and Development, the Russell Sage Foundation (RSF Award No. 87-13-02), and the Columbia Center for the Study of Development Strategies.

We thank especially Commissioner Patricia Gatling, Clifford Mulqueen, Lee Hudson, and the staff at the New York City Commission on Human Rights for their support. We extend special thanks to Joan Russell, Clare Wiseman, and Ruby Hlivko for their excellent work as project managers over the course of the study. We thank the following individuals for their dedicated work as testers for the project: Tania Aparicio, Evette Addai, Hannah Blume, Destinee Bowrin, John Cales, Jillian Cantwell, Courtney Cauthan, Xilonem Clarke, Reese Crispen, Dana DeBari, Cristina De La Rosa, Lucas Denton, Matthew Diaz, Kenneth Edusei, Inemesit Essien, Talisa Feliciano, Ellen Gagne, Bianca Garcia, Elizabeth Garcia, Emmanuel Garcia, Gabriel Garcia, Lais Gomes Duarte, Ruby Hlivko, Jevaun Joseph, Rumando Kelley, Keenan Lambert, Douglas Land, Vanesa Lauradin, John Long, Julia Mesler, Linden Miller, Richard Minaya, Ryan Mitchell, Shannon Murray, Luis Ortiz, Joe Palmisano, Christiaan Perez, Shanta Pamphile, Dale Reyes, Haley Rice, Julieta Salgado, Ayinde Samuel, Camillia Shofani, Rebecca Suldan, Fadumo Tahlil, Matthew Taylor, Joshua Thompson, Kirya Traber, Kenneth Turja, Steven Velez, Hannah-Sophie Wahle, and Jordan Woodsmall. We thank Yasmine Ergas from the Columbia University Institute for the Study of Human Rights for her support. We thank Caroline Peters for her valuable contributions to the project in its initial stages.

The authors serve a pro bono advisory role for the present study, by providing input into the experimental design and by conducting the analysis of the study data. The study and data are wholly owned by the New York City Commission of Human Rights. As such this study and the authors' involvement with the study have been reviewed and approved by Columbia IRB (Columbia Morningside IRB Protocols IRB-AAAI4007 and IRB-AAAN4059). A pre-analysis plan for this study was submitted to the Experiments in Governance and Politics (E-GAP) registry on April 3, 2013.

Executive Summary

This report describes the final results from an evaluation for the New York City Commission on Human Rights. The study has two aims: to measure baseline levels of discrimination in the rental housing market and to experimentally test the effectiveness of telephone messages from the city government urging landlords and brokers to comply with fair housing law.

Research Approach

Unlike previous studies focused on average discrimination rates nationally, this is the first comprehensive attempt to provide estimates specific to the largest rental housing market in the United States.

Our research strategy paired a matched audit design with a field experiment. In the audit portion, three testers—one black, one Hispanic, and one white—each responded to the same apartment listings and met with the associated agents. Before the visits, those landlords and brokers were randomly assigned to receive a punitive message informing them of the consequences (monetary and otherwise) of discrimination, a generic message about fair housing law intended to signal monitoring by the city, or no message at all.

Baseline Levels of Racial Housing Discrimination

First, we summarize baseline levels of discrimination in the housing market as determined by the audit portion of the study. These are average levels as measured in the control group and reflect the policy status quo:

- *There is strong evidence of discrimination, particularly against Hispanics.* We estimate that Hispanic testers were less likely than white testers to receive a callback from a landlord or broker—in 15.4% of cases compared to 21.5%, a difference of 6.1 percentage points ($p = 0.019$). They were also less likely to receive an offer for an apartment—in 6.1% of cases compared to 11.8%, a difference of 5.7 percentage points ($p = 0.011$).

More intuitively, we can present these absolute differences as percentage differences in the rate of favorable treatment for Hispanic testers, relative to the rate of favorable treatment for white testers. Interpreted this way, Hispanic testers were 28.4% less likely to receive a callback than white testers and 48.3% less likely to receive an offer for an apartment.

While we also found evidence of discrimination against black testers on the same outcomes, the magnitudes were lower than those for Hispanics and fell short of conventional levels of statistical significance.

Callbacks and offers were our primary objective indicators of post-visit discrimination. In addition, we collected outcomes based on testers' self-reports both before and during their visits.

- *There is mixed evidence for discrimination over the phone.* Counter to expectations, we found that white testers were more likely than black or Hispanic testers to report experiencing difficulty with qualifications to rent when speaking over the phone to schedule a potential visit with a landlord or broker.

- *There is some evidence, using measures of subjective treatment, of discrimination during visits.* Black testers were more likely than white testers to report sales efforts by landlords or brokers—in 50.9% of cases compared to 41.7%, a 9.2-percentage-point difference or 22.1% more than whites in terms of percent change ($p = 0.049$).

Do City Messaging Strategies Reduce Discrimination?

Below, we summarize the main findings of the field experiment portion of the study:

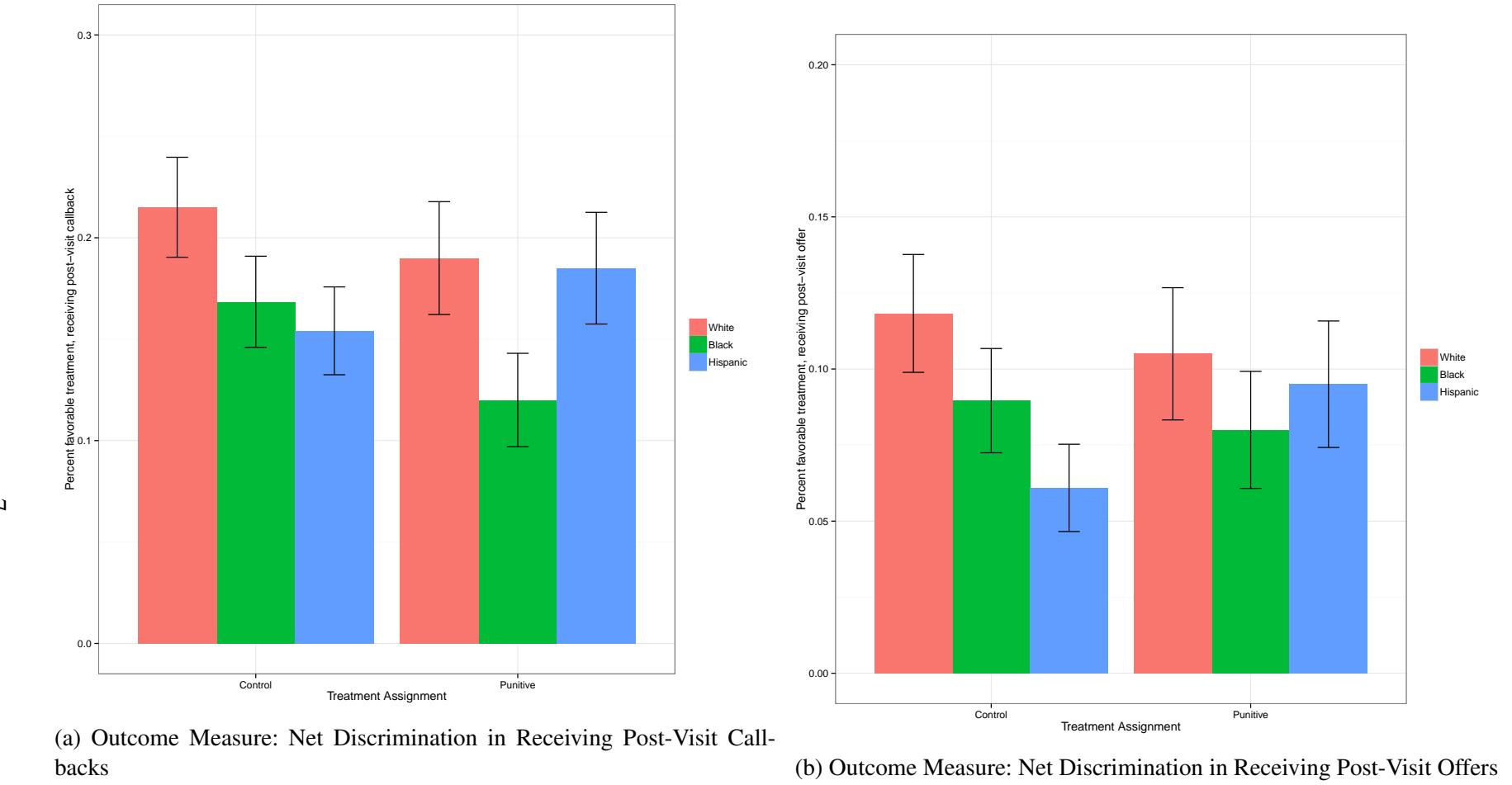
Effects of City Actions

- *When compared to sending no messages, sending punitive messages—regardless of whether they are received—reduces net discrimination against Hispanic testers (versus both black and white testers) in receiving callbacks and offers.*
 - Net discrimination rates against Hispanic testers (versus white testers) in receiving callbacks decreased by 5.6 percentage points ($p = 0.086$).
 - Net discrimination rates against Hispanic testers (versus white testers) in receiving offers for a housing unit decreased by 7.8 percentage points ($p = 0.012$).
 - Net discrimination rates against Hispanic testers (versus black testers) in receiving callbacks decreased by 7.6 percentage points ($p = 0.031$).
 - Net discrimination rates against Hispanic testers (versus black testers) in receiving offers for a housing unit decreased by 5.3 percentage points ($p = 0.048$).

Effects of Message Content

- *There is suggestive evidence that receiving punitive appeals consistently reduces net discrimination rates against black testers and against Hispanic testers when compared to only receiving a monitoring signal.*
 - Net discrimination rates against black testers (versus white testers) decreased between 4 and 7 percentage points.
 - Net discrimination rates against Hispanic testers (versus white testers) decreased between 4 and 5 percentage points.
 - While these findings were not statistically significant, they are suggestive and consistent with existing evidence.

The statistical evidence suggests that at least for Hispanics and possibly for African Americans, *discrimination in the housing market is a real phenomenon*, but one that can be *partially or wholly eliminated via a simple messaging campaign*. The findings are illustrated in Figure 0.1:



(a) Outcome Measure: Net Discrimination in Receiving Post-Visit Callbacks

(b) Outcome Measure: Net Discrimination in Receiving Post-Visit Offers

Figure 0.1: Effect of Punitive Messaging on Net Discrimination Against Hispanics (versus Whites) in Receiving Post-Visit Callbacks and Post-Visit Offers

In the next section, we highlight more qualitative accounts from testers that corroborate these findings.

Varieties of Racial Housing Discrimination

In addition to the systematic analyses summarized above, testers collected a wealth of information that we mined to paint a more subjective portrait of discrimination in the New York City rental housing market. For instance:

- *There is considerable variation in steering strategies in the New York City rental market.* Using data from paired tests, we found apparent differences by tester race in the average quoted price for rental units, the size of housing units shown, the number of amenities shown, and the stated number of amenities included in the quoted rental price.
- *There seem to be complex disparities in the structure and amount of up-front costs and fees required to secure and move into units.* Move-in costs packaged as several months' rent up-front appeared to be disproportionately levied on Hispanic testers as compared to their black and white counterparts. Hispanic and black testers reported that they were more likely than white testers to be told they needed to pay additional holding fees or good-faith deposits to secure units. White testers reported that they were more likely to be told they needed to pay broker fees, application fees, and administrative and processing fees than black and Hispanic testers. Black testers reported that they were more frequently told they had to pay for and pass a credit check or background check than white and Hispanic testers did.
- *White testers are more likely to receive offers to negotiate down fees than black or Hispanic testers.* Using data from paired tests, we found that more white testers reported being offered the possibility of negotiation than black or Hispanic testers. More black testers also reported being offered the possibility of negotiation than Hispanic testers.

Policy Recommendations

This evaluation finds that racial discrimination in the New York City rental housing market persists, and that the city possesses the tools to counteract it. Based on the findings summarized above, we offer the following policy recommendations:

- *Disseminate information about discrimination in the rental housing market:* A long-term solution will require a sustained engagement from citizens, policymakers, as well as landlords and brokers themselves. Publicizing the evidence for discrimination will send a signal that the city is committed to solving a genuine problem.
- *Continue the use of matched audits to uncover discrimination:* Many forms of differential treatment that occur are not easily identified within single interactions between any given housing seeker and a landlord or broker. This is especially true during the appointment stage with respect to subtle forms of racial discrimination such as steering, differences in the quoted terms of rent, and in the willingness to negotiate such terms. The continued use of matched or paired audits is essential for monitoring discrimination levels in the rental market.

- *Employ punitive appeals encouraging compliance with fair housing law:* Reminding landlords and brokers to comply with the city's Fair Housing Law and informing them of the pecuniary costs of violating it is effective at reducing the incidence of racial discrimination in rental housing, particularly as it affects Hispanic apartment seekers. This study cannot address whether modes of communication besides telephone calls would be equally effective. A cost-benefit analysis combined with further evaluation is recommended. Additional research should also investigate whether messaging has long-term effects on reducing discrimination.

List of Tables

2.1	Message Components and Appeals by Experimental Condition	21
2.2	Distribution of Experimental Subjects by Randomization Block	24
2.3	Comparing Net versus Gross Measures of Discrimination	27
3.1	Selected Characteristics of Housing Units in the Audit and Experimental Samples .	36
3.2	Distribution of Rental Units Across Boroughs, by Sample	37
4.1	Incidence of Early Stage Discrimination	40
4.2	Baseline Incidence of Discrimination: In-Person and Post-Visit	41
4.3	Indicators of Objective Discrimination during Appointments	47
4.4	Range of Net Differences in Objective Discrimination Indicators among Non-Missing Paired Test Reports	48
5.1	Estimated Effects of Messaging on Net Discrimination Levels	67
5.2	Observed Noncompliance with Treatment Assignment	71
5.3	Estimated Average Effects of Message Content on Net Discrimination Outcomes by Complier Type.	73
5.4	Summary of Differences in Marginal Treatment Effects between Low versus High Putative Employment Stability Signal Groups	76
D.1	Unweighted Data: The Effects of Assignment to Messaging Conditions on Net Discrimination in Landlords Honoring Scheduled Appointments	110
D.2	Unweighted Data: The Effects of Assignment to Messaging Conditions on Net Discrimination in Receiving a Post-Visit Callback	111
D.3	Unweighted Data: The Effects of Assignment to Messaging Conditions on Net Discrimination in Receiving a Post-Visit Offer for the Unit	112
D.4	Estimates of Principal Causal Effects on Net Discrimination against Blacks (vs. Whites) in Landlords and Brokers Honoring Scheduled Appointments	113
D.5	Estimates of Principal Causal Effects on Net Discrimination against Blacks (vs. Whites) in Testers' Subjective Perceptions of Favorable Interactions during the Appointment	114
D.6	Estimates of Principal Causal Effects on Net Discrimination against Blacks (vs. Whites) in Receiving Post-Visit Callbacks	115
D.7	Estimates of Principal Causal Effects on Net Discrimination against Blacks (vs. Whites) in Receiving Post-Visit Offers	116
D.8	Estimates of Principal Causal Effects on Net Discrimination against Hispanics (vs. Whites) in Landlords and Brokers Honoring Scheduled Appointments	117
D.9	Estimates of Principal Causal Effects on Net Discrimination against Hispanics (vs. Whites) in Testers' Subjective Perceptions of Favorable Interactions during the Appointment	118
D.10	Estimates of Principal Causal Effects on Net Discrimination against Hispanics (vs. Whites) in Receiving Post-Visit Callbacks	119
D.11	Estimates of Principal Causal Effects on Net Discrimination against Hispanics (vs. Whites) in Receiving Post-Visit Offers	120
D.12	Estimates of Principal Causal Effects on Net Discrimination against Hispanics (vs. Blacks) in Landlords and Brokers Honoring Scheduled Appointments	121

D.13 Estimates of Principal Causal Effects on Net Discrimination against Hispanics (vs. Blacks) in Testers' Subjective Perceptions of Favorable Interactions during the Appointment	122
D.14 Estimates of Principal Causal Effects on Net Discrimination against Hispanics (vs. Blacks) in Receiving Post-Visit Callbacks	123
D.15 Estimates of Principal Causal Effects on Net Discrimination against Hispanics (vs. Blacks) in Receiving Post-Visit Offers	124
D.16 Full Results: Estimated Conditional Effects of Messaging by Putative Low Employment Stability Signal: Net Discrimination against Blacks (versus Whites) . . .	125
D.17 Full Results: Estimated Conditional Effects of Messaging by Putative Low Employment Stability Signal: Net Discrimination against Hispanics (versus Whites) .	126
D.18 Full Results: Estimated Conditional Effects of Messaging by Putative Low Employment Stability Signal: Net Discrimination against Hispanics (versus Blacks) .	127

List of Figures

0.1 Effect of Punitive Messaging on Net Discrimination Against Hispanics (versus Whites)	7
2.1 Timeline of an Augmented Housing Audit	18
3.1 Cumulative Number of Cases Over Implementation Period	32
3.2 New York City Median Gross Rent Compared to Household Income, 2005-2011 . .	33
3.3 Flow Diagram of the Process Defining the Experimental Analysis Sample	34
3.4 Map of the Geographic Distribution of Housing Units Corresponding to Advertised Listings, Data Aggregated to Community Board Level	38
4.1 Average Quoted Price Across Units Shown, by Tester Race and by Group Differences	50
4.2 Distributions of Total Quoted Fees by Tester Race and by Group Differences . . .	61
4.3 Distributions of Quoted Sum of Up-Front Rent, Security Deposit, Holding Fees, and Broker Fees by Tester Race and by Group Differences	62
4.4 Distributions of Quoted Sum of Application and Credit/Background Check Fees by Tester Race and by Group Differences	64
4.5 Distributions of Quoted Administrative Fees and Other Net Costs by Tester Race and by Group Differences	65
D.1 Number of Units Shown, by Tester Race and by Group Differences	92
D.2 Average Quoted Price Across Units Shown, by Tester Race and by Group Differences	93
D.3 Average Number of Bedrooms in Units Across Units Shown, by Tester Race and by Group Differences	93
D.4 Number of Units Among Units Shown Where Rental Price Was Quoted in Person, by Tester Race and by Group Differences	94
D.5 Percentage of Units Among Units Shown Where Rental Price Was Quoted in Person, by Tester Race and by Group Differences	94
D.6 Number of Units Among Units Shown where Application Requirement was Quoted, by Tester Race and by Group Differences	95
D.7 Percentage of Units Among Units Shown where Application Requirement was Quoted, by Tester Race and by Group Differences	96

D.8 Number of Units Among Units Shown in Doorman Building, by Tester Race and by Group Differences	96
D.9 Percentage of Units Among Units Shown in Doorman Building, by Tester Race and by Group Differences	97
D.10 Number of Units Among Units Shown in Elevator Building, by Tester Race and by Group Differences	97
D.11 Percentage of Units Among Units Shown in Elevator Building, by Tester Race and by Group Differences	98
D.12 Number of Units Among Units Shown with Washer/Dryer in Building, by Tester Race and by Group Differences	98
D.13 Percentage of Units Among Units Shown with Washer/Dryer in Building, by Tester Race and by Group Differences	99
D.14 Number of Units Among Units Shown with Washer/Dryer in Unit, by Tester Race and by Group Differences	99
D.15 Percentage of Units Among Units Shown with Washer/Dryer in Unit, by Tester Race and by Group Differences	100
D.16 Number of Quoted Amenities Included in Rent for Listed Unit, by Tester Race and by Group Differences	100
D.17 Quoted Monthly Asking Rental Price for Listed Unit, by Tester Race and by Group Differences	101
D.18 Incidence of a Quoted Requirement to Pay of Any Number of Months of Rent Up Front for the Listed Unit, by Tester Race and by Group Differences	102
D.19 Incidence of a Quoted Security Deposit Requirement for the Listed Unit, by Tester Race and by Group Differences	103
D.20 Incidence of a Quoted Broker Fee Requirement for the Listed Unit, by Tester Race and by Group Differences	103
D.21 Incidence of a Quoted Application Fee Requirement for the Listed Unit, by Tester Race and by Group Differences	104
D.22 Incidence of a Quoted Administrative or Processing Fee Requirement for the Listed Unit, by Tester Race and by Group Differences	104
D.23 Incidence of a Quoted Holding Fee Requirement for the Listed Unit, by Tester Race and by Group Differences	105
D.24 Incidence of a Quoted Credit or Background Check Fee Requirement for the Listed Unit, by Tester Race and by Group Differences	105
D.25 Incidence of a Quoted Other Fees Requirement for the Listed Unit, by Tester Race and by Group Differences	106
D.26 Quoted Sum of All Fees for the Listed Unit, by Tester Race and by Group Differences	106
D.27 Quoted Sum of Application Fees and Other Net Costs for the Listed Unit, by Tester Race and by Group Differences	107
D.28 Quoted Sum of Up-Front Rent, Security Deposit, Holding Fees, and Broker Fees for the Listed Unit, by Tester Race and by Group Differences	107
D.29 Quoted Sum of Administrative Fees and Other Net Costs for the Listed Unit, by Tester Race and by Group Differences	108
D.30 Incidence of a Offer to Negotiate Fees, by Tester Race and by Group Differences	109

D.31 Estimated Marginal Effects of Messaging Treatments by Putative Low Employment Stability Signal on Net Discrimination against Blacks (versus Whites)	128
D.32 Estimated Marginal Effects of Messaging Treatments by Putative Low Employment Stability Signal on Net Discrimination against Hispanics (versus Whites) . .	129
D.33 Estimated Marginal Effects of Messaging Treatments by Putative Low Employment Stability Signal on Net Discrimination against Hispanics (versus Blacks) . . .	130

1 Introduction

1.1 Study Overview and Aims

Housing discrimination plays a central role in the creation and perpetuation of inequities between racial and ethnic minority groups and neighborhoods. In turn, federal, state, and local governments have a keen interest in knowing and in reducing the incidence of housing discrimination. Considerable resources have been channeled into the study of housing discrimination incidence since the late 1970s, when the United States Department of Housing and Urban Development (HUD) began using paired testing methods to identify discriminatory practices in housing and to monitor unlawful behavior in the housing market. This method has been replicated widely in government-led housing audits, such as the large-scale housing audits conducted by HUD every decade since, as well as in academic studies of housing discrimination and discrimination in other economic contexts such as employment.

While vast academic and policy literatures exist on the incidence of housing discrimination, relatively little is known on how governments may effectively reduce housing discrimination and encourage otherwise unlawful landlords and brokers to comply with fair housing law. Recent HUD audit reports find that housing discrimination levels have gradually decreased over time (Turner et al. 2013), but the trajectory of discrimination rates since the 1980s have developed alongside changes in public attitudes toward racial and ethnic minorities.¹ Governments with a vested interest in promoting fairness in economic opportunities and individuals who are the targets of discrimination do not have the luxury of time and instead seek policy solutions that are able to reduce discrimination in the near term. Public appeals and messaging campaigns have provided a popular means by which governments seek to encourage “good behavior” within short time horizons. Such campaigns have been widely theorized by academics and deployed in practice (Thaler and Sunstein 2008; Kahan 2000), but whether these efforts yield meaningful changes in behavior has yet to be rigorously evaluated across a wide swath of policy designs and issue domains.

This evaluation assesses these questions in the context of the New York City rental housing market. Using an audit study design, we assess discrimination levels against blacks and Hispanics (versus whites). Additionally, we augment the audit study with a field experiment to assess the effectiveness of different governmental messaging campaigns encouraging landlords and brokers in the New York City rental housing market to comply with the City’s Fair Housing Law. We focus in particular on the average effects of two different messaging strategies on discrimination: a monitoring message and a punitive message. Is discrimination reduced? And if so, by how much?

¹The General Social Survey (GSS) has documented that the percentage of Americans who support fair housing laws has risen from 35% in 1973 to more than 70% today.

1.2 Summary of Main Findings

1.2.1 Baseline Levels of Racial Housing Discrimination

- *There is mixed evidence for discrimination at the pre-visit stage.* Counter to expectations, we found that white testers were more likely than black testers to experience difficulty from landlords or brokers about their qualifications to rent when speaking over the phone to schedule a potential visit. Hispanics were also more likely to encounter skepticism or negative comments about their qualifications than black testers. (Note that these findings are based on self-reports and that we cannot distinguish between differences in tester perceptions and differences in actual experiences.)
- *There is evidence of discrimination against both black and Hispanic testers after scheduled visits.* We estimate that black testers were less likely than white testers to receive a callback from a landlord or broker — in 14.7% of cases compared to 19.9%, a difference of 5.2 percentage points ($p=0.005$). For Hispanics, we found a significant difference in receiving an offer: 7.2% compared to 10.4% for white testers, a 3.2-percentage-point difference ($p=0.04$). More intuitively, we can interpret these absolute differences as percentage differences in the rate of favorable treatment for minority testers, relative to the rate of favorable treatment for white testers. When compared to the rate of favorable treatment experienced by white testers, black testers receive favorable treatment at a lower rate—26% for receiving callbacks. Hispanic testers receive favorable treatment at a lower rate as well — nearly –31% for receiving offers.
- *There is some evidence of discrimination from measures of subjective treatment during visits.* White testers were more likely than black testers to report instances of positive editorializing by landlords or brokers — 80.6% vs. 75.3%, a -5.3 percentage point difference or –6.6% less in terms of percent change ($p=0.038$). These subjective measures are based on testers’ open-ended field notes and could reflect differences in reporting styles as well as actual treatment during visits.

1.2.2 Effects of Messaging on Discrimination Incidence

- *When compared to sending no messages, sending punitive messages reduces net discrimination against Hispanic testers (versus both black and white testers) in receiving callbacks and offers.*
 - Net discrimination rates against Hispanic testers (versus white testers) in receiving callbacks decrease by 5.6 percentage points ($p=0.086$).

- Net discrimination rates against Hispanic testers (versus white testers) in receiving offers for a housing unit decrease by 7.8 percentage points ($p=0.012$).
- Net discrimination rates against Hispanic testers (versus black testers) in receiving callbacks decrease by 7.6 percentage points ($p=0.031$).
- Net discrimination rates against Hispanic testers (versus black testers) in receiving offers for a housing unit decrease by 5.3 percentage points ($p=0.048$).
- *There is suggestive evidence that receiving punitive appeals from the city encouraging compliance with fair housing law consistently reduces net discrimination rates against black testers (versus white testers) and against Hispanic testers (versus white testers) when compared to only receiving a monitoring signal from the City.*
 - Net discrimination rates against black testers (versus white testers) decrease between 4 and 7 percentage points.
 - Net discrimination rates against Hispanic testers (versus white testers) decrease between 4 and 5 percentage points.
 - While these findings are not statistically significant since all 95% credible intervals contain zero, these results are substantively significant.

1.2.3 Varieties of Racial Housing Discrimination

We also explore and describe variation in subtle forms of racial housing discrimination.

- *There is considerable variation in steering strategies in the New York City rental market.* Using data from paired tests, we find considerable differences by tester race in the average quoted price for rental units across units shown to testers; in the size of housing units and in number of unit or building amenities among units shown to testers; and in the stated number of amenities included in the quoted rental price.
- *When controlling for the unit viewed, we document complex disparities in the structure and amount of up-front costs and fees required to secure and move into units, by tester race.* Move-in costs packaged as several months' rent up-front seem to be disproportionately levied on Hispanic testers as compared to their black and white counterparts. Hispanic and black testers were more likely than white testers to be told they needed to pay additional holding fees or good faith deposits to secure units. White testers were more likely to be told they needed to pay broker fees, application fees, and administrative and processing fees than black and Hispanic testers. Black testers were more frequently told they had to pay for and pass a credit check or background check than white and Hispanic testers.

- *White testers are more likely to receive offers to negotiate down fees than black and Hispanic testers.* Using data from paired tests, we find that white testers are offered the possibility of negotiation at a rate 2.3 percentage points higher than the rate reported by black testers and at a rate 1.8 percentage points higher than the rate reported by Hispanic testers. Black testers are offered the possibility of negotiation at a rate 0.5 percentage points less than the rate reported by Hispanic testers.

1.3 Organization of Report

The remainder of this report is organized as follows. Section 2 describes the study design and methodology. Section 3 describes the study context and defines the evaluation sample. Section 4 provides a descriptive summary of the varieties of racial discrimination in the New York City rental housing market as revealed by the matched audits, and the baseline levels of racial discrimination that occur over the timeline of finding a rental housing unit in New York City. Section 5 presents the impact findings from the field experimental study and answers the question of whether government messaging reduces racial discrimination in rental housing. Section 6 concludes.

2 Study Design and Methodology

In this section, we describe the timeline, field protocols, measurement procedures, and analysis plan in detail.

2.1 The Timeline of a Housing Audit

We investigate the effects of different messaging strategies on discrimination by using a matched audit design, which is used to measure differential treatment by race for a given vacant rental unit, augmented with a field experiment that randomly assigns a targeted message to the landlord who is associated with a particular unit and interacts with a team of testers. Figure 2.1 summarizes the timeline of the augmented housing audit.

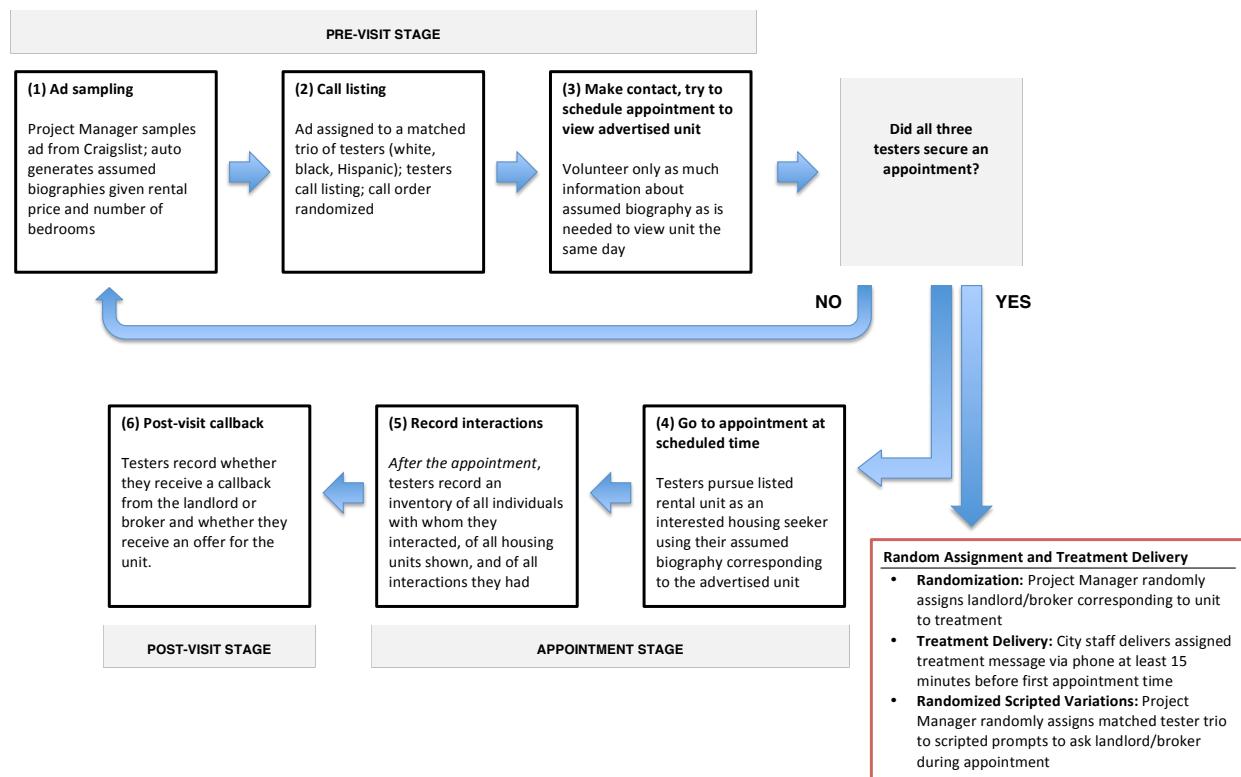


Figure 2.1: Timeline of an Augmented Housing Audit

On each day of the study’s implementation, a set of rental housing ads from the past 24 hours was selected from Craigslist using an automated script.² First, a list of “likely discrimination” ads were non-randomly drawn; these ads were identified using a search procedure that looked for words and phrases in ad text signaling class-based preferences for potential tenants that could imply racial

²See Appendix A for technical details on the methods used to scrape public listings and best practices for similar sampling procedures using online ad listings.

preferences in housing seekers. Second, among the remainder of ads posted on Craigslist that day (excluding those identified in the “likely discrimination” search), a representative sample of ads was randomly drawn in a way that was representative of the distribution of advertised vacant rental housing stock by borough. Only advertisements that invited housing seekers to reply by phone were pursued.³

Testers responded to sampled ads posing as individuals interested in renting the listed apartment. Each ad was pursued by a matched team of three testers of the same gender but varying by race: one white, one black, and one Hispanic. The features of testers’ assumed biographies that were matched included their credit score range, income level, household composition, occupation, perceived stability of employment and income, gender, and perceived age. Call order was randomized across testers in a matched trio.

Assigned assumed biographies were automatically generated at the time of case scraping. Biographical information was pre-defined in a master database and extracted such that assigned biographical information for a particular case was logically consistent with the monthly asking rental price and the number of bedrooms in the unit listed on the advertisement. To ensure that no two testers in a matched team had the same biographical information, there was a set of possible biographies for each characteristic by rental price level and by the number of bedrooms. Then, for a given case, attributes corresponding to that rental amount and type of housing stock were randomly sampled and assigned to testers without replacement.⁴ A Project Manager assigned vacant rental ads to tester teams available to work so long as new pairings did not duplicate previous tester-landlord, tester-broker, or tester-unit interactions.⁵

Upon reaching an individual when replying to a housing ad, testers were instructed to provide *limited* information about themselves and their racial identities in this initial stage, and were instructed to schedule an appointment to view the unit and meet with the person on the phone on the same day at the earliest convenience of the individual with whom they were speaking. What was revealed over the phone typically included the tester’s assumed first name,⁶ their interest in pursuing and renting the unit, their availability for an in-person appointment to view the unit, and their financial qualifications to rent, which were commonly asked by landlords and brokers during this stage of the housing search process. If testers were asked about other aspects of their biographies,

³In an earlier pilot study, we found that the contact rate and appointment scheduling rate are significantly higher when replying to Craigslist ads by phone than when replying to ads via email.

⁴Ad-specific information scraped at baseline included the listing URL, the borough, the listed monthly asking rental price, the listed neighborhood, and the number of bedrooms.

⁵The Project Manager checked for duplicates in case assignments using a master database continuously updated with information about landlords, brokers, and housing units previously pursued by any tester during the study.

⁶Testers mentioned their first name when introducing themselves on the call; if landlords or brokers asked for the tester’s full name, this was provided. Testers had assumed first and last names, each of which was randomly drawn from a database of names tagged with racial and ethnic identifiers and stayed with them through the course of the study. For additional details on the name assignment procedure, see Appendix A.

they volunteered this information accordingly.

When all three testers in a matched team successfully secured an appointment to view an advertised rental unit, the landlord associated with that case was admitted into the experiment and was randomly assigned to a treatment condition by the Project Manager. Testers then made their individually scheduled appointments, viewed the unit, interacted with the subject, and recorded their interactions afterwards. Testers were blind to treatment, but they were not blind to the interest in assessing discrimination as one of multiple characteristics of the rental housing market.⁷ As such, they were extensively instructed and coached not to fish for particular reactions or to let their personal opinions about landlord behavior interfere with their ability to continue interactions during the audit. Under the supervision of the Project Manager, matched testers also coordinated to ensure that testers, who were trained to inquire about the housing stock, terms of rent, and the neighborhood—questions one would normally ask when looking for a place to live—did not use the same language so that the landlord would not become suspicious of an audit.

2.2 Treatment Messages

2.2.1 Treatment Scripts

Landlords were randomly assigned to one of three possible experimental groups, each corresponding to a different messaging condition:

- **Control:** No message was sent.
- **Monitoring:** The target landlord or broker received a call, was told that the call was from the NYC Commission on Human Rights and was provided the Commission’s webpage address for more information. This condition allowed us to distinguish the effects of specific appeals from a monitoring effect on discrimination. This was the script of the monitoring message:

Hello, could I speak with [First Name of Landlord/Broker] please? *If prompted for identifying info:* I'm calling with a message from the New York City Commission on Human Rights. This will take less than a minute.

Once the targeted recipient is on the line: Good [morning/afternoon/evening]. I'm calling from the New York City Commission on Human Rights as part of an ongoing informational campaign to remind landlords and brokers of their obligations under fair housing law.

⁷Testers were only told that this was a participant-observation and ethnographic study about the dynamics occurring in the rental housing market that leveraged an audit design.

Please take a moment to visit nyc.gov/cchr to learn how fair housing law protects individuals from discrimination. Thank you very much for your time.

- **Punitive:** The target landlord or broker received a call, received the monitoring message, and received a punitive appeal which stated that discrimination is illegal and highlights the costs of violating fair housing law. This was the script of the punitive message:

Hello, could I speak with [First Name of Landlord/Broker] please? *If prompted for identifying info:* I'm calling with a message from the New York City Commission on Human Rights. This will take less than a minute.

Once the targeted recipient is on the line: Good [morning/afternoon/evening]. I'm calling from the New York City Commission on Human Rights as part of an ongoing informational campaign to remind landlords and brokers of their obligations under fair housing law. It is illegal to discriminate against a person seeking housing due to their membership in a protected class.

If you are found to have broken the law, you may be ordered to pay damages, provide reasonable accommodation, or incur civil penalties of up to \$250,000.

For more information, please visit nyc.gov/cchr. Thank you very much for your time.

Table 2.1 summarizes the message components and appeals made by each experimental condition.

Message Component	Experimental Condition		
	1 Punitive	2 Monitoring	3 Control
State call is from NYCCHR and provide reference for more information	Yes	Yes	No
State that housing discrimination is illegal and describe possible sanctions	Yes	No	No

Table 2.1: Message Components and Appeals by Experimental Condition

2.2.2 Theoretical Motivations for Treatment Design and Main Hypotheses

Each of these appeals was theoretically motivated such that their effects measured different causal mechanisms that explain resulting discrimination levels. We explicate these motivations for the

monitoring and punitive conditions in turn, and formally present pre-specified hypotheses about the expected effects of each messaging condition.

Monitoring condition: Comparing outcomes under different conditions helps us disaggregate the effects of *contact* by government from the effects of the *content* of appeals government sends.⁸ Contact on the issue of housing discrimination can both increase the salience of fair housing considerations in the landlord's thinking and can signal to landlords that they are being monitored by the city, even though the *costs* associated with monitoring, in case of discrimination, are not highlighted. Insights due to Mendelberg (2001) and others suggest that this *explicit* priming of discrimination considerations can result in less discrimination (relative to implicit priming) due to an invoking of social nondiscrimination norms. The first quantity we seek to estimate then is the causal effect of monitoring messages on discriminatory behavior. Formally we will test the hypothesis that monitoring messages decrease rental housing discrimination rates against blacks and Hispanics.

Punitive condition: In contrast to simple monitoring, a *punitive* message highlights the actual pecuniary cost of violating the law, which include civil penalties and the potential cost of damages to the complainant. This in turn may prime individuals to behave as self-interested, profit-maximizing actors. Under some conditions it is possible that such punitive messages can have *weaker* effects on discrimination than simple monitoring messages. This could arise for example if increasing the salience of discrimination induced compliance with non-discriminatory social norms but priming pecuniary considerations focuses landlords on cost-benefit calculations only. If the effects of norm inducement is stronger than the effect of the punitive message on cost calculations, then observed effects from the simple monitoring treatment may be stronger. The second quantity we seek to assess is the causal effect of punitive messages, relative both to the control condition (thus the combined effect of monitoring and punitive content) and to the monitoring condition (thus the additional effect of punitive content). Formally we will assess the hypothesis that punitive messages decrease rental housing discrimination rates against blacks and Hispanics relative to the baseline condition and to the monitoring condition.

2.3 Randomization Procedure and Treatment Message Delivery

Only after testers in a matched trio secured an appointment to view the same housing unit in person, the Project Manager randomly assigned the landlord to receive either a punitive message priming the costs of violating fair housing law, a monitoring message providing an information about the law, or no message. The randomization procedure was as follows. From the start of the study in April 2012, through September 10, 2012, we employed a five-group design where subjects

⁸Monitoring effects like these are akin to Hawthorne effects (Mayo 1949), although we note that in this setting the monitoring is by a politically relevant actor, not by the experimenter and the effects are thus of substantive interest.

could be randomly assigned with equal probability to one of five conditions: to one of the three experimental conditions of interest (control, monitoring, or punitive conditions) or to one of two other treatment groups (receiving a “normative” appeal, or receiving both normative and punitive appeals). We employed a blocked randomization procedure where a landlord’s treatment assignment probabilities depended on the sampling frame from which they are drawn (Bronx, Brooklyn, Manhattan, Queens, Staten Island, or the “likely discrimination” oversample). On September 10, 2012, the experiment was reduced to a three-group design including only the relevant treatment arms. For this second randomization regime lasting between September 2012, and May 7, 2013, we reset the treatment probabilities using the same definition of experimental blocks. By block, landlords could be assigned to the control group with 50% probability or to either the punitive or monitoring conditions with 25% probability each. On May 8, 2013, we altered the randomization procedure a final time by eliminating the “likely discrimination” sampling frame and experimental stratum because the search terms used were generating no hits. We reset the treatment probabilities such that landlords could be assigned to each condition with an equal probability. Across the three randomization regimes, there are a total of 17 blocks. The distribution of cases across blocks and treatment assignments is summarized in Table 2.2.

The process of matching a subject to an experimental block was automated to minimize the possibility of human error. At the time of case scraping and sampling, the block membership of each case was automatically determined by information from the ad; this information was saved to a master database. At the time of randomization, the Project Manager simply entered a unique case identifier into a Python-based user interface, which cross-referenced the block membership of that case and selected the next treatment assignment from stored treatment assignment vectors we generated by block given the design.

The staffer at the Commission in charge of administering treatment delivered the assigned message by phone. Information about the landlord or broker with whom the testers would meet was collected from the initial set of calls. This information included every subject’s name and phone number and was compiled by the Project Manager, who then forwarded it to a designated Treatment Administrator. This individual was a dedicated staff member working for the city agency in charge of enforcing fair housing law, the New York City Commission on Human Rights (CCHR). The Treatment Administrator, calling on behalf of the City’s Commission on Human Rights, delivered the assigned treatment message by phone before the first scheduled appointment time and recorded the components of the assigned message that were successfully delivered.

Experimental Block	Assigned to Control		Assigned to Monitoring		Assigned to Punitive	
	N	Percent	N	Percent	N	Percent
Regime 1: Apr 13, 2012 to Sep 9, 2012						
Brooklyn	14	2.1	11	1.7	17	2.6
Bronx	4	0.6	3	0.5	2	0.3
Manhattan	13	2	12	1.8	18	2.8
Queens	2	0.3	3	0.5	6	0.9
Staten Island	1	0.2	1	0.2	0	0
Likely Discrimination Frame	8	1.2	8	1.2	9	1.4
Regime 2: Sep 10, 2012 to May 7, 2013						
Brooklyn	63	9.6	24	3.7	29	4.4
Bronx	21	3.2	10	1.5	11	1.7
Manhattan	50	7.7	32	4.9	23	3.5
Queens	28	4.3	10	1.5	14	2.1
Staten Island	8	1.2	2	0.3	3	0.5
Likely Discrimination Frame	11	1.7	4	0.6	4	0.6
Regime 3: May 8, 2013 to Dec 20, 2013						
Brooklyn	13	2	23	3.5	14	2.1
Bronx	6	0.9	3	0.5	8	1.2
Manhattan	25	3.8	18	2.8	25	3.8
Queens	8	1.2	7	1.1	14	2.1
Staten Island	4	0.6	3	0.5	3	0.5
Total	279	42.7	174	26.6	200	30.6

Table 2.2: Distribution of Experimental Subjects by Randomization Block. Cells contain counts of the number and share of experimental subjects (i.e., landlords and brokers) randomly assigned to the control group, the monitoring condition, and the treatment group, by block. Each regime denotes a different randomization procedure used. See text in Section 2.3 for a description of the randomization protocol corresponding to each regime. Percentages may not sum to 100 due to rounding.

2.4 Varying Putative Signals of High versus Low Employment Stability

In addition to the primary manipulation of treatment messages, we also manipulated whether tester teams assigned to a given case all provided a putative signal of high versus low employment stability.⁹ This additional manipulation allows us to examine the conditional effects of the messaging intervention by the signaled financial stability of housing seekers.

The script guidelines given to a matched team are detailed below:

1. **High employment stability.** If assigned to this condition, all testers in a triple indicate full-time employment in a firm with high job security and/or tenure (e.g., company has a wide range of stable clients; if nonprofit, large endowment).

⁹The messaging treatment and the scripted variation are cross-cutting treatments.

2. Low employment stability. If assigned to this condition, testers indicate that they work freelance, at a small firm dependent on a limited number of key clients, or a risky start-up venture. None of these situations should imply that testers can't make the stated monthly rent amount.

2.5 Data Collection

Testers record their interactions with landlords and brokers at multiple stages of the housing search process: *prior to the visit (pre-treatment)*, when replying to rental listings, making contact with a landlord or broker, and arranging an appointment to view the unit; *during a housing unit visit (post-treatment)*, when the tester meets with a landlord or broker to view the unit of interest; and *after the visit (post-treatment)*, when the tester follows up with or receives a followup call from the landlord or broker and, conditional on making contact, receives an offer or rejection for the unit of interest. The survey instruments used by testers are shown in Appendix B.

In the pre-visit stage of the housing search process, testers document information about how difficult it was to successfully schedule an appointment to view the unit of interest, including: whether they were able to schedule an appointment, the number of call attempts made before scheduling an appointment, the time when the appointment was made, the appointment date and time, who they interacted with, and if an appointment could not be made the reasons why. Testers also document the aspects of their assumed biography that came up during pre-visit interactions over the phone and how landlords and brokers with whom they interacted reacted to the information provided.¹⁰

During the appointment stage, testers collect detailed information about all the primary individuals with whom they interacted during the visit¹¹ and information about the units they were shown.¹² In addition to accounting for the people and housing units they encountered during their

¹⁰The survey instrument prompts testers to indicate whether the following aspects of their biography arose in conversation: name, personal income, household income, occupation, employer, credit score, marital/partner status, children/dependents, reason for moving, location (neighborhood) of current residence, location (neighborhood) of workplace, gender, educational background/pedigree, race, ethnicity or national origin, sexual orientation, linguistic or speech-related traits, age, phone number, and employment stability or source of income. testers may also report additional attributes about their assumed biographies that are questioned.

¹¹Testers record each individual's name, firm affiliation, job description, and whether each individual was the same person with whom they spoke to set up the appointment. Testers also record their perceptions of each individual's age range, race, and ethnicity. To verify this information, testers are also instructed to ask for and collect business cards.

¹²Relevant fields include whether a particular unit shown is the sampled unit; the unit's street address, borough, and neighborhood as described by the landlord or agent; the monthly rental price (quoted in person); the number of bedrooms and bathrooms; whether the building has a doorman and an elevator; whether the unit or building includes a washer/dryer; whether the landlord or broker claimed the unit would be renovated before move-in; the amenities included in the rent; the length of the lease; the security deposit required; any additional fees required to secure the apartment and their respective amounts; and whether an application is required (if yes, a copy of the application is requested). Testers also report their subjective assessments of the unit interior and the building exterior.

visit, testers also provide open-ended responses about their interactions with landlords, brokers, and agents. Testers record the general demeanor (including but not limited to their body language; professionalism; and instances of expressed interest, lack of interest, skepticism, attentiveness, repulsion) of the landlords and brokers of interest toward them at the beginning, middle, and end of the visit. Testers record the sales efforts landlords and brokers make during the visit, which include rental incentives and extra amenities offered such as waived fees, discounted rent, discounts on local goods and services, gifts, or other “perks” meant to persuade testers to sign a contract soon; attempts to editorialize about the neighborhood, its residents, amenities, and/or character; attempts to editorialize about the building, its residents, amenities, and/or character; offers to follow up after the appointment; and attempts to editorialize about the housing search process or the housing market in general.

Testers also document the other-regarding beliefs and group perceptions revealed by landlords and brokers during the appointment. Testers record whether landlords or brokers suggest either explicitly or implicitly that the presence of persons of any particular group in the area may result in an increase or decrease of property values, directly or indirectly; that the presence of persons of any particular group in the area may result in an increase or decrease of criminal or anti-social behavior in the neighborhood/area, either directly or indirectly; if landlords or brokers express judgment toward the tester based on their revealed perceptions of the tester; if landlords question their qualifications to rent; if landlords or brokers reveal prejudices or beliefs in stereotypes about any economic or social group, including the group to which the tester belongs, during the visit. Testers are to record their reactions in these interactions as well.

Post-visit stage interactions that testers document include the callback date and time, if any; if the tester was offered the unit and if not, whether the unit was already rented out; whether the landlord or broker offered to show the tester other vacant rental units; and other interactions that occurred during post-visit correspondence.

We employ multiple data collection methods – specifically closed- and open-ended survey questions and qualitative participant-observation field notes – that help us unambiguously measure discrimination given ancillary information about the context of social interactions. This also allows to construct more stable composite measures of discrimination by utilizing more information from both qualitative and quantitative data records of tester interactions with landlords and brokers. The added information also allows us to capture more nuanced forms of discrimination that may be implicitly indicative of discrimination against minorities, such as steering. Posing open-ended questions to testers provides an inductive mode of data collection where we are receptive to the many possible forms of discrimination that occur in the rental housing market today that may not be easily defined *a priori*.

2.6 Measuring Discrimination

2.6.1 The Net Measure of Discrimination

Differential treatment by race is measured by comparing observed objective and subjective indicators of favorable treatment across matched testers within a case. Following the methods employed by prior housing audit studies, for each outcome indicator of discrimination measured at the case-tester level, we construct a case-level measure of **net discrimination**, defined as the difference in the rate of favorable treatment for the majority and favorable treatment for the minority. To illustrate, consider Table 2.3, which explains the construction of the net measure and contrasts it with a common, alternative gross measure of discrimination. Suppose there are four cases for which we measure favorable treatment indicators for a majority and minority tester pair. For Case 1, both testers are treated favorably. For Case 2, the majority tester is treated favorably, but the minority tester is not. For Case 3, only the minority tester is treated favorably. For Case 4, neither tester is treated favorably. Across all four cases, we capture all possible combinations of treatment toward testers in the pair. These combinations are described in columns (A) and (B) in Table 2.3, where favorable treatment toward a tester is coded as “1”, and unfavorable treatment toward a tester is coded as “0”.

Case ID	(A) Majority Tester Treated Favorably? (1=Yes, 0=No)	(B) Minority Tester Treated Favorably? (1=Yes, 0=No)	Measure of Discrimination	
	Net (Equals A-B)	Gross (1 only if A=1 & B=0)		
1	Yes (1)	Yes (1)	0	0
2	Yes (1)	No (0)	1	1
3	No (0)	Yes (1)	-1	0
4	No (0)	No (0)	0	0
Average level of discrimination			0%	25%

Source: Authors' representation of net and gross measures of discrimination.

Table 2.3: Comparing Net versus Gross Measures of Discrimination

The net measure is constructed by subtracting column (B) from column (A). The gross measure is coded as a 1 only when column (A) equals 1 and column (B) equals 0. The main difference then is in the coding of Case 3. The net measure captures the average level of discrimination such that on the margins, a case where a minority tester is treated favorably but the majority tester unfavorably effectively “counteracts” a case where the majority tester is treated favorably but the minority tester unfavorably. The gross measure only counts up the share of cases where the majority tester is treated favorably but the minority tester unfavorably. The implication is most evident when describing the average level of discrimination across cases. In this example, the average level of

net discrimination is 0%, whereas the average level of gross discrimination is 25%.

Both measures may contain bias in the measurement of differential treatment, but the net measure is preferable so long as it is interpreted as a lower bound on the level of discrimination that exists. As Stephen Ross (2002) explains in the context of measuring racial housing discrimination:

The net measure is constructed under the assumption that adverse treatment against the white tester occurs only because the testers' visits differed, and so adverse treatment against the white tester provides an accurate measure of the number of instances of minority adverse treatment that arose because the testers' visits differed (55).

Using this measurement strategy, the study focuses on the effects of the experimentally-assigned treatment on a range of objective and subjective measures of discrimination in the New York City rental housing market.

2.6.2 Objective Measures of Discrimination (Post-Treatment Outcome)

We measure three forms of post-treatment objective discrimination which are included in the set of outcome variables for the experimental analysis. We are concerned with whether there is differential treatment by race in landlord or broker efforts:

1. to honor their scheduled appointments;
2. to call back and follow up with testers after the appointment; and
3. to offer the unit to the tester

All of these measures are computed for each majority-minority pairing (white-black, white-Hispanic, and black-Hispanic) and can take three values:

- -1 if only the minority tester is treated favorably,
- 0 if both the minority and majority testers are treated equally, and
- 1 if only the majority tester is treated favorably.

When examining average levels of the net measure, 1 means 100% net discrimination against the minority group, and -1 means 100% net discrimination against the majority group. A value of 0 means that the two groups are treated equally.

2.6.3 Subjective Measures of Discrimination (Post-Treatment Outcome)

Our measurement strategy is informed by well-known macro-level shifts in Americans' racial attitudes, as well as the impact of these shifts on behavior. First, the vast majority of Americans oppose discrimination against minorities in private transactions such as buying and selling a house, a finding that mirrors other gradual yet decisive changes in public opinion surrounding race relations in the United States. Second, social scientists have documented the emergence and persistence of "new racism," or discriminatory attitudes that manifest themselves not in overt behavior or socially desirable survey responses but in subtler attitudes about minorities' competence, abilities, or cultural tendencies.

These findings, suggesting some divergence between public norms and private behavior, have motivated continued research and debate on how best to measure individuals' so-called "implicit attitudes." Our contribution is to demonstrate one method of addressing this challenge in the context of a field experiment: marshaling testers' open-ended survey responses to shed light on subtler interactions that might not be captured via traditional quantitative outcome measures. One potential approach would be to hand-code testers' subjective observations in the field, a lengthy process that might uncover fresh insight but at the cost of introducing inconsistencies between coders. Another option would be to utilize unsupervised learning methods to reveal latent meaning, but this approach could overlook or misinterpret useful contextual information such as responses to specific neighborhoods. We opt for the middle path: using supervised learning algorithms trained with a set of hand-coded survey responses. This allows us to maintain focus on essential features of the text while ensuring uniformity during the classification stage.

Our procedure was as follows. First, we randomly selected 300 open-ended tester survey responses (roughly 15% of the total) to be manually coded. Based on a protocol we inductively developed, a pair of research assistants who were blind to treatment independently evaluated each case along 15 potential dimensions. The assistants then worked together to agree on a final set of codes where their individual assessments diverged. Once this adjudication process was complete, we took the subset from our list of 15 available codes with the highest intercoder reliability in the first step.¹³ The resulting five codes are: sales efforts by landlords/brokers; praise about a tester's qualifications; positive response to a tester's background; positive editorializing about an apartment or neighborhood; and professionalism of landlords/brokers.

These codes cover a range of potential responses to testers' presence during the course of interactions with landlords and brokers in the field. Sales efforts can include inducements to rent an apartment. "Positive editorializing" captures instances in which landlords or brokers express their opinions about aspects of the neighborhood or apartment that a prospective tenant might find

¹³Cohen's kappa for this subset ranged from 0.23-0.61, which reflects "fair" agreement on the low end to "substantial" agreement on the high end (Viera and Garrett 2005).

appealing. Such attempts to “talk up” a neighborhood present a favorable picture of the inhabitants, character, safety, and other features of an area. And so on. Each of the five responses was coded as either present or not present in the selected open-ended case-tester-level data.

Once the “training set” of documents was coded in this manner, we used natural language processing algorithms to classify the remainder of the cases. For each of the five codes, we took the classifications generated via *maximum entropy* as our dichotomous measures of subjective treatment due to its superior performance compared to *support vector machines*, another well-known algorithm. As a validity check, we found that maximum entropy returned approximately the same proportion of codes as the original training set.

Nearly all of the classifications were made with greater than 90% confidence. The share of total case-tester-level responses classified as containing instances of positive editorializing was the highest at 79.4%. Just over half of the data, 50.4%, exhibited landlord/broker professionalism as indicated by testers’ open-ended responses and the subsequent automated text analysis. Nearly half, 42.9%, of responses contained evidence of sales efforts. The other two codes do not indicate widespread behavior: 5.8% of responses showed evidence of praise for testers’ rental qualifications, and 0.7% showed positive responses to an aspect of a tester’s background.

We use these five indicators to construct a single composite index measure of differential treatment in testers’ in-person interactions with landlords during the appointment.¹⁴ The resulting index measure is a continuous scale that ranges from -2 to 2, where 2 means 100% discrimination against the minority tester, and -2 means 100% discrimination against the majority tester. (Alternatively, these values may be interpreted as 100% favorable treatment toward the majority tester and 100% favorable treatment toward the minority tester, respectively). A value of 0 means that both testers were treated equally. The index measure we will construct is useful for dealing with multiple outcome measures – both to dampen the noise surrounding each individual indicator, and to mitigate issues with inference arising from multiple comparisons.

2.6.4 Measuring Early Stage Discrimination (Pre-Treatment)

We also measure differential treatment between testers prior to randomization during the pre-visit stage, between the point in time when they make first contact with the landlord and the time of random assignment. During this period, signaled markers of testers’ identity include information about their assumed biographies that landlords inquire about over the phone, their name, and their voice.

¹⁴We follow the index construction method used by Kling, Liebman, and Katz (2007), where for a set of outcome indicator variables Y_1, \dots, Y_k the value of each indicator variable for a given observation is differenced by the control group mean value of the corresponding variable; this difference is then divided by the standard deviation of the corresponding variable among the control group. The transformed indicator variables are then summed and divided by the total number of indicator variables to create a standardized summary index measure for that observation.

We measure the following types of early stage discrimination that occurs in the rental housing market: differences in the ability of matched testers to make contact with the landlord; differences in the ability to schedule an appointment; the number of biographical attributes landlords inquire about over the phone while screening potential tenants prior to the appointment; differences in the number and percentage of attributes for which testers receive positive, negative, neutral, and skeptical responses from landlords after providing relevant information; and differences in whether testers receive any negative or any skeptical feedback on any aspect of their biography communicated over the phone. These measures are used to describe baseline levels of discrimination and as covariates in the experimental analysis.

3 Study Context and Sample Definitions

3.1 Study Context and Dates of Field Implementation

The study was in the field from April 13, 2012 to December 20, 2013. Figure 3.1 summarizes the cumulative number of cases admitted into the experimental sample over this period.

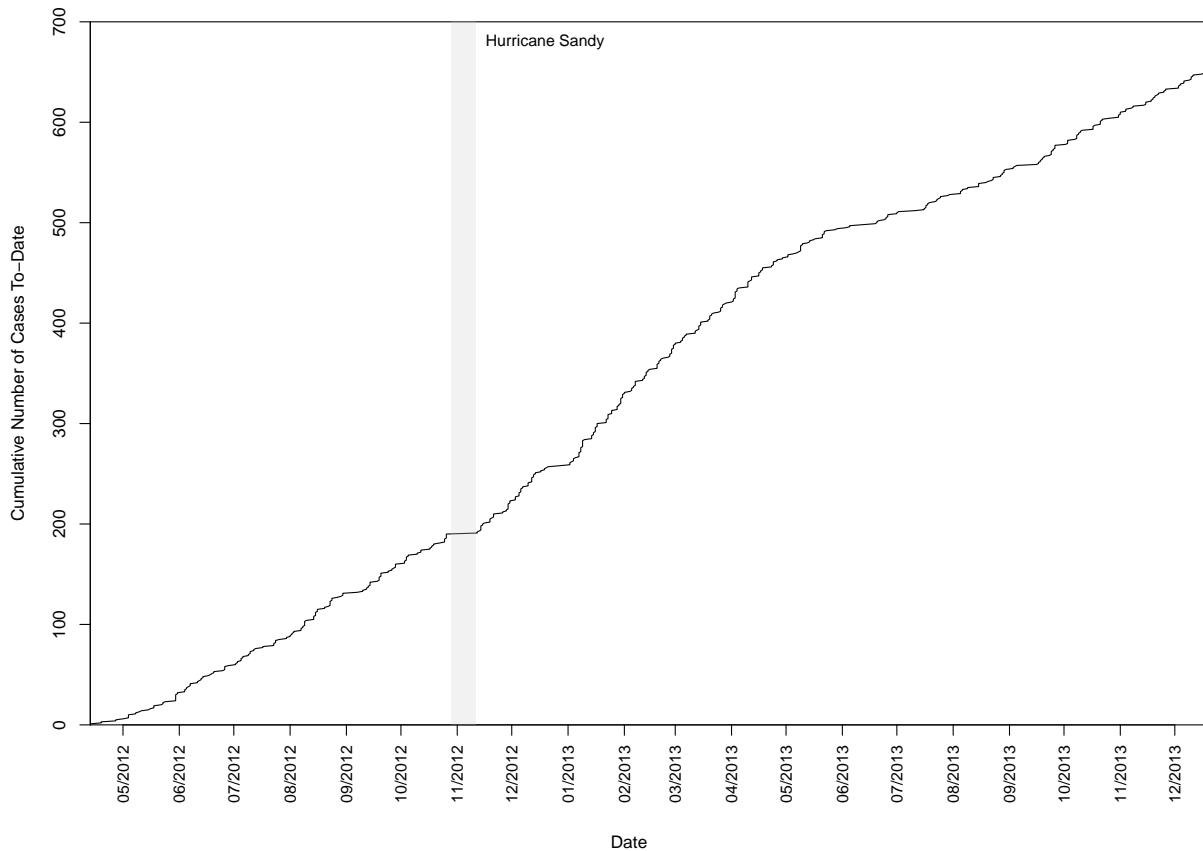


Figure 3.1: Cumulative Number of Cases Over Implementation Period between April 13, 2012 and December 20, 2013.

During that time, the rental market continued its trend of rising monthly rates (except Staten Island) and rapid high-end construction. According to the Furman Center at New York University, this was also occurring as median household incomes were declining in New York City due in part to the recession (see Figure 3.2). From 2005 to 2012, the median rent rose 11 percent across all boroughs (to \$1,216). During the same time period, household income rose 2 percent (to \$41,000). Data is not available beyond 2012, so we cannot extrapolate to much of the period during which the study was in the field.

The study also experienced an exogenous shock: Hurricane Sandy, which struck the city on

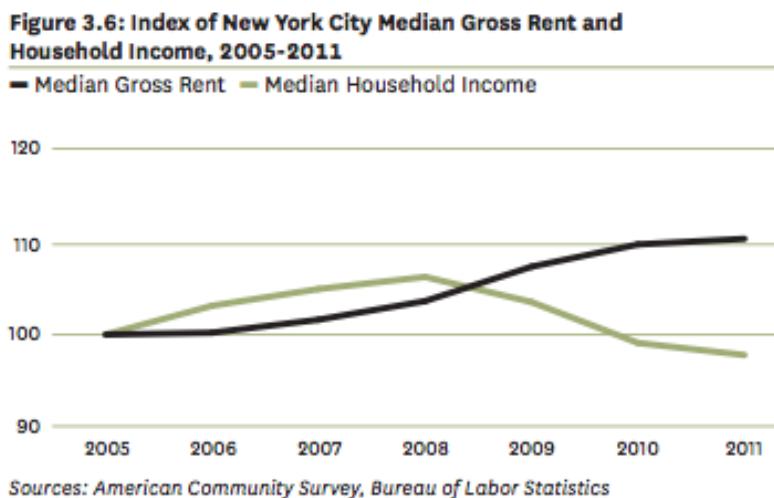


Figure 3.2: New York City Median Gross Rent Compared to Household Income, 2005-2011. Lines show median gross rent and median household income as a percentage of constant, inflation-adjusted 2005 dollars. Source: Furman Center for Real Estate and Urban Policy, New York University.

October 29, 2012 and caused major damage and disruption. CCHR's operations were suspended and did not resume for weeks. Since the rental market was not operating under normal conditions at that time (and public transportation between the boroughs was difficult), we suspended implementation while the city recovered from the storm.

3.2 Sample Definitions

We focus on two key samples for the analysis: the **audit sample**, which contains 2711 case-level observations, and the **experimental sample**, which contains 653 case-level observations. Each case includes information about three matched testers' interactions with the same landlord or broker corresponding to a sampled listing. A flow diagram of the process by which the experimental analysis sample is defined is shown in Figure 3.3.

Table 3.1 summarizes the characteristics of the advertised housing units in the audit sample (Panel I), in the experimental sample (II), in the subset of cases in the experimental sample assigned to any treatment condition (III), and in the subset of cases in the experimental sample assigned to the control condition (IV). We briefly describe the characteristics of housing units in each of these samples.

- **Audit Sample** – For advertised housing units in the audit sample (N=2711), the mean advertised monthly asking rental price is \$2,238 and the median advertised monthly asking rental

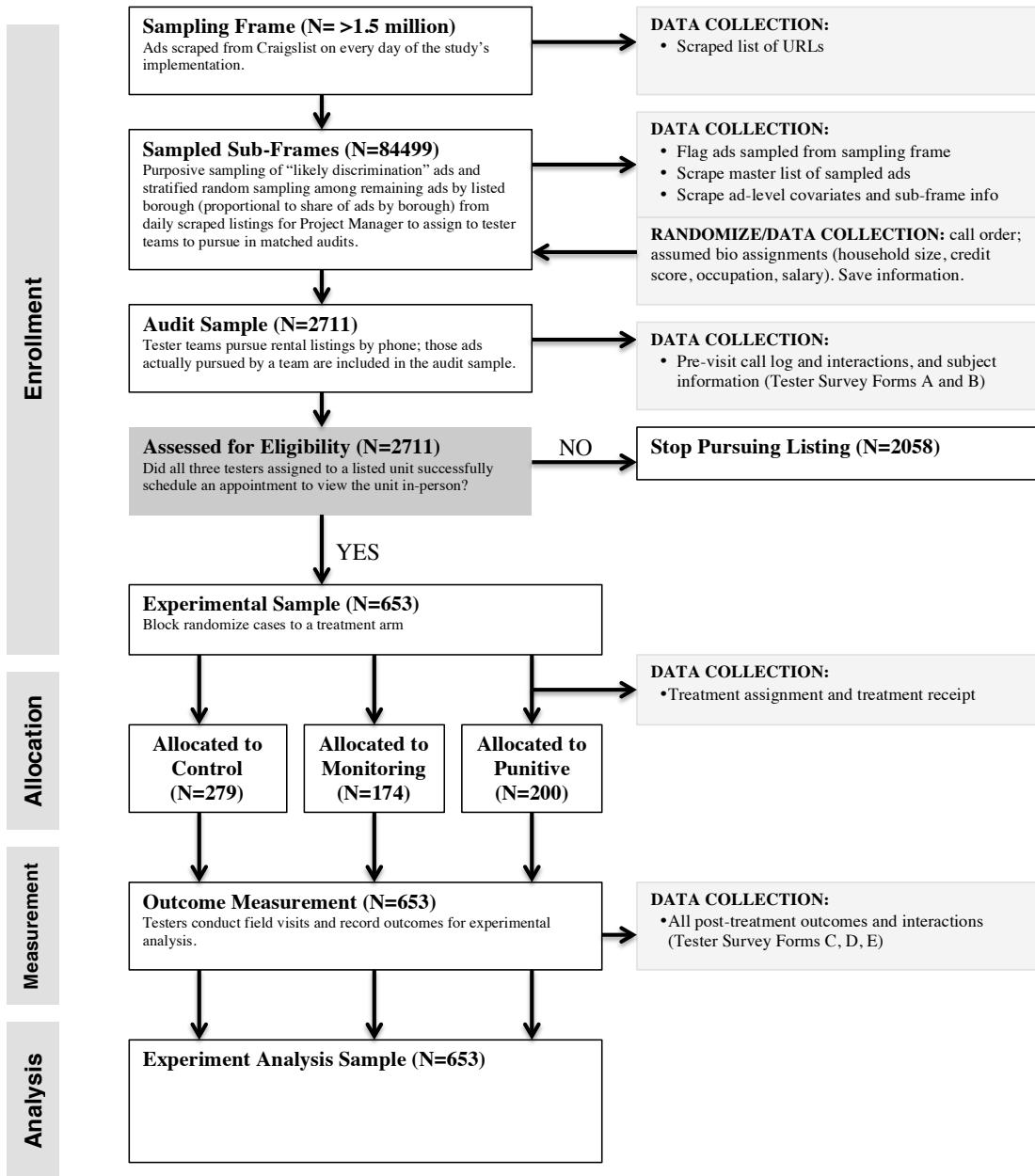


Figure 3.3: Flow Diagram of the Process Defining the Experimental Analysis Sample. This flow diagram summarizes the phases of the randomized field experiment – enrollment, intervention allocation, outcome measurement, and data analysis – and how the final analysis sample for the experiment is defined over the course of this process.

price is \$1,850. The advertised monthly asking rental price ranges from \$400 per month to \$15,000 per month. The mean advertised number of bedrooms is 0.94 and the average advertised square footage of a listed unit is 1,021.35 square feet. Of the listed units in the audit sample, 57.29% were listed by brokers.

- **Experimental Sample** – For advertised housing units in the experimental sample (N=653), the mean advertised monthly asking rental price is \$2,435 and the median advertised monthly asking rental price is \$2,200. The advertised monthly asking rental price ranges from \$750 per month to \$9,495 per month. The mean advertised number of bedrooms is 0.88 and the average advertised square footage of a listed unit is 1,017.49 square feet. Of the listed units in the audit sample, 84.53% were listed by brokers.
- **Cases Assigned to Any Treatment Group** – For advertised housing units corresponding to cases in the experimental sample assigned to any treatment condition (N=374), the mean advertised monthly asking rental price is \$2,420 and the median advertised monthly asking rental price is \$2,200. The advertised monthly asking rental price ranges from \$750 per month to \$9,495 per month. The mean advertised number of bedrooms is 0.85 and the average advertised square footage of a listed unit is 915 square feet. Of the listed units in the audit sample, 83.42% were listed by brokers.
- **Cases Assigned to the Control Group** – For advertised housing units corresponding to cases in the experimental sample assigned to the control condition (N=279), the mean advertised monthly asking rental price is \$2,456 and the median advertised monthly asking rental price is \$2,025. The advertised monthly asking rental price ranges from \$850 per month to \$8,900 per month. The mean advertised number of bedrooms is 0.92 and the average advertised square footage of a listed unit is 1,116.68 square feet. Of the listed units in the audit sample, 86.02% were listed by brokers.

Variable	I. Audit Sample		II. Experimental Sample		III. Any Treatment Group		IV. Control Group	
	N	% of Audit Sample	N	% of Audit Sample	N	% of Exp. Sample	N	% of Exp. Sample
Panel A								
Number of Units								
Total	2711	(100%)	653	(24.09%)	374	(57.27%)	279	(42.73%)
Bronx	337	(100%)	68	(20.18%)	37	(54.41%)	31	(45.59%)
Brooklyn	801	(100%)	208	(25.97%)	118	(56.73%)	90	(43.27%)
Manhattan	668	(100%)	216	(32.34%)	128	(59.26%)	88	(40.74%)
Queens	495	(100%)	92	(18.59%)	54	(58.7%)	38	(41.3%)
Staten Island	254	(100%)	25	(9.84%)	12	(48%)	13	(52%)
Likely Discrimination Frame	156	(100%)	44	(28.21%)	25	(56.82%)	19	(43.18%)
Panel B								
Monthly Asking Rental Price (\$)								
Total	2238	(1295)	2435	(1204)	2420	(1157)	2456	(1272)
Bronx	1419	(512)	1578	(662)	1563	(628)	1597	(719)
Brooklyn	2194	(1047)	2319	(957)	2285	(855)	2370	(1096)
Manhattan	3252	(1600)	3163	(1447)	3134	(1404)	3206	(1520)
Queens	1718	(562)	1885	(558)	1847	(450)	1941	(693)
Staten Island	1383	(511)	1336	(596)	1369	(655)	1293	(559)
Likely Discrimination Frame	2479	(1452)	2321	(736)	2336	(778)	2302	(697)
Panel C								
Monthly Asking Rental Price (\$)								
Total	1850		2200		2200		2025	
Bronx	1325		1400		1400		1400	
Brooklyn	1992		2200		2299		2050	
Manhattan	2950		2898		2900		2872	
Queens	1600		1850		1850		1850	
Staten Island	1300		1100		1195		1100	
Likely Discrimination Frame	2272		2250		2200		2300	
Panel D								
Number of Bedrooms								
Total	0.94	(1.22)	0.88	(1.19)	0.85	(1.17)	0.92	(1.21)
Bronx	0.82	(1.2)	0.85	(1.16)	0.86	(1.23)	0.84	(1.1)
Brooklyn	1	(1.23)	1.01	(1.3)	0.97	(1.26)	1.07	(1.36)
Manhattan	0.85	(1.15)	0.69	(1.04)	0.63	(0.98)	0.76	(1.13)
Queens	0.81	(1.15)	0.59	(0.9)	0.56	(0.9)	0.63	(0.91)
Staten Island	0.89	(1.29)	0.8	(1.5)	1.08	(1.83)	0.54	(1.13)
Likely Discrimination Frame	1.88	(1.23)	1.95	(1.01)	1.92	(1)	2	(1.05)
Panel E								
Square Footage								
Total	1021.35	(618.52)	1017.49	(448.09)	915	(449.36)	1116.68	(430.93)
Bronx	999.59	(352.42)	1328.75	(453.84)	1100	(453.84)	1405	(523.52)
Brooklyn	1041.87	(462.63)	1125.94	(638.07)	1085.11	(661.49)	1171.88	(652.64)
Manhattan	1029.82	(495.06)	926.79	(350.32)	777.43	(334.36)	1135.9	(262.42)
Queens	914.92	(777.31)	917.6	(260.29)	866.67	(288.68)	994	(291.33)
Staten Island	1203.25	(981.19)	1233.33	(305.51)	1300	(305.51)	1200	(424.26)
Likely Discrimination Frame	1011.5	(568.51)	885	(272.45)	900	(141.42)	880	(315.91)
Panel F								
Listed by Broker								
Total	1553	(57.29%)	552	(84.53%)	312	(83.42%)	240	(86.02%)
Bronx	172	(51.04%)	51	(75%)	27	(72.97%)	24	(77.42%)
Brooklyn	452	(56.43%)	178	(85.58%)	102	(86.44%)	76	(84.44%)
Manhattan	439	(65.72%)	192	(88.89%)	112	(87.5%)	80	(90.91%)
Queens	256	(51.72%)	75	(81.52%)	43	(79.63%)	32	(84.21%)
Staten Island	133	(52.36%)	14	(56%)	5	(41.67%)	9	(69.23%)
Likely Discrimination Frame	101	(64.74%)	42	(95.45%)	23	(92%)	19	(100%)

Table 3.1: Selected Characteristics of Housing Units in the Audit and Experimental Samples

3.3 Distribution of Cases Across Boroughs, by Sample

Table 3.2 summarizes the distribution of cases across boroughs by sample. Focusing on the sampling blocks corresponding to the five boroughs, the distribution of the audit sample (including all assigned cases, whether or not all testers in a group were able to schedule an appointment) is very close to the distribution of known rental units in New York City. Using the latest New York City Housing and Vacancy Survey data from 2011, we can see that the proportion of cases in each borough in the audit sample closely tracks the overall proportions, with the exception that Staten Island units appear to be overrepresented in our sample at the expense of apartments in the Bronx. In the final experimental sample, Manhattan and Brooklyn are overrepresented at the expense mainly of Queens.

	Citywide 2011		Audit Sample		Experimental Sample		Control Group Only	
Borough	# Units	%	# Units	%	# Units	%	# Units	%
Bronx	388,022	17.86	337	13.19	68	11.17	31	11.11
Brooklyn	691,178	31.81	801	31.35	208	34.15	90	32.26
Manhattan	587,313	27.03	668	26.14	216	35.47	88	31.54
Queens	449,108	20.67	495	19.37	92	15.11	38	13.62
Staten Island	57,013	2.62	254	9.94	25	4.11	13	4.66
Total	2,172,634	100	2,555	100	609	100	279	100

Table 3.2: Distribution of Rental Units Across Boroughs, by Sample. The full audit and experimental samples are compared to the totals citywide (from the 2011 New York City Housing and Vacancy Survey). Cases from the likely-discrimination block not included.

According to the city survey, the overall net vacancy rate across all boroughs was 3.12%: 3.23% in the Bronx, 2.61% in Brooklyn, 2.80% in Manhattan, 3.79% in Queens, and 6.65% in Staten Island.

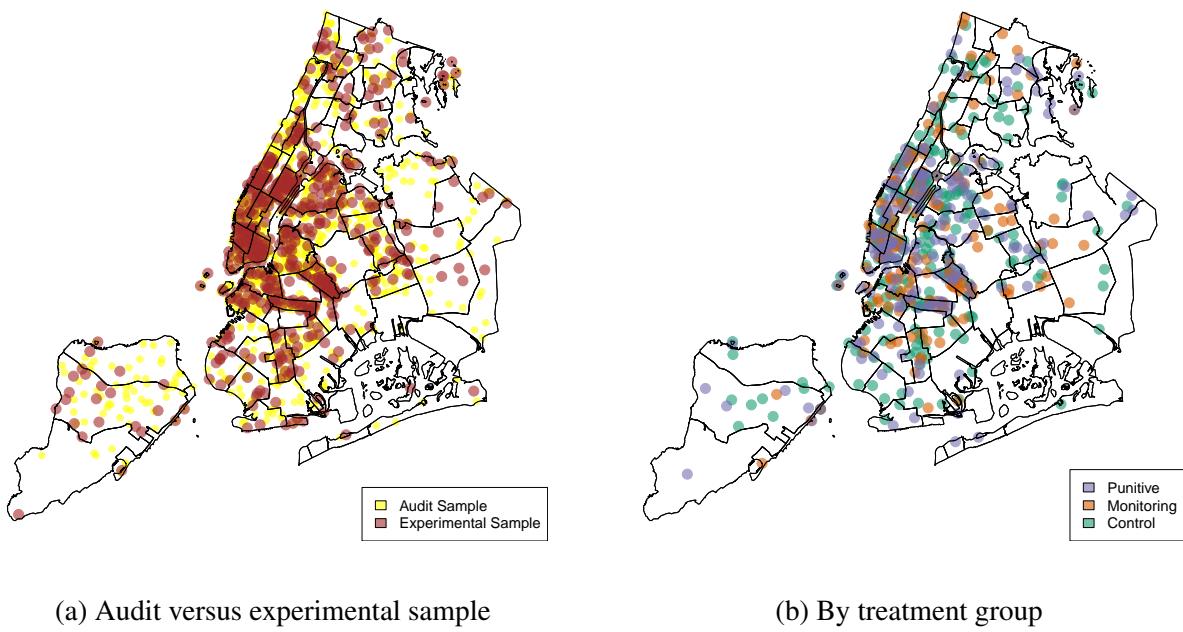


Figure 3.4: Map of the Geographic Distribution of Housing Units Corresponding to Advertised Listings, Data Aggregated to Community Board Level. The exact locations are not reported to maintain the anonymity of study subjects. The geographic location of advertised units is aggregated to the Community Board level. The number of housing units within each Community Board is then randomly distributed within the boundaries of the Community Board.

4 Assessing Rental Housing Discrimination in New York City

4.1 Baseline Levels of Discrimination

4.1.1 Incidence of Early-Stage Discrimination

First, we summarize our measures of discrimination toward testers before they made an appointment with a landlord or broker—that is, before random assignment occurred. Among all pursued cases in the audit sample, there were a number of statistically significant differences in the way testers of differing racial and ethnic groups were treated during their over-the-phone interactions with landlords and brokers. A conventional threshold for statistical significance is 5%, meaning that if there were no real difference between groups, we would expect to observe, by chance, a significant difference in 5% of hypothetical replications of the experiment. Thus, if we find that more than 5% of observed differences are significant at the 5% level, we would hesitate to attribute the results to anything but chance. This happened to be the case when looking at the 39 pre-treatment measures of net discrimination across the three possible pairings (white-black, white-Hispanic, and black-Hispanic), of which 3 were significant at only the 5% level (i.e., $0.01 < p < 0.05$).

However, there were several differences that were statistically significant at the 0.1% ($p < 0.001$) level or better, giving us substantially greater confidence that the results were not due to random chance. These results can be grouped into two sets of differences. First, white testers appeared to encounter more skepticism about their attributes or qualifications to rent a given apartment than black testers—about twice as much (white testers in 3.2% of pursued cases experienced skeptical mentions of an attribute by landlords or brokers over the phone compared to 1.6% of those cases for black testers). The same was true of negative comments about their attributes or qualifications (3.5% vs. 1.6%). Other measures intended to capture the same difference in treatment (such as the percentage of attributes or the number of attributes mentioned) showed broadly the same pattern. This was counter to expectations and could reflect real differences in treatment as well as differential perceptions in treatment.

The second set of statistically significant differences showed that Hispanic testers appeared to experience substantially more negative and skeptical comments about their attributes and qualifications to rent than African-American testers. For example, in 3.6% of pursued cases, Hispanic testers encountered skeptical responses as compared to 1.6% of cases for white testers. The corresponding percentages for outright negative reactions were 3% and 1.6%, respectively.

As seen in the tables (left column), the average levels of unfavorable treatment as captured by these measures were fairly low across all groups. Finding meaningful differences between them was therefore difficult given our sample size. However, the significant differences that remained were relatively large in magnitude and may be of substantive interest.

Measure	I. All Pursued Cases in Audit Sample					II. All Cases in Experimental Sample					III. Cases Assigned to Control Group				
	Mean Level		Difference			Mean Level		Difference			Mean Level		Difference		
	Majority	Minority	(Maj.-Min.)	P-value	[N]	Majority	Minority	(Maj.-Min.)	P-value	[N]	Majority	Minority	(Maj.-Min.)	P-value	[N]
A. White vs. Black Testers															
Any contact	0.512	0.524	-0.012	(0.184)	[2711]	—	—	—	—	—	—	—	—	—	—
Scheduling appointment	0.348	0.361	-0.013	(0.035)	[2711]	—	—	—	—	—	—	—	—	—	—
No. of attributes brought up by landlord/broker	1.101	1.041	0.06	(0.136)	[2711]	2.023	1.936	0.087	(0.334)	[653]	2.151	1.961	0.19	(0.145)	[279]
No. attributes - skeptical response	0.053	0.027	0.026	(0.001)	[2711]	0.077	0.041	0.035	(0.073)	[653]	0.097	0.039	0.057	(0.106)	[279]
No. attributes - positive response	0.132	0.125	0.007	(0.637)	[2711]	0.262	0.256	0.006	(0.876)	[653]	0.269	0.258	0.011	(0.852)	[279]
No. attributes - neutral response	0.918	0.894	0.024	(0.501)	[2711]	1.703	1.668	0.035	(0.675)	[653]	1.821	1.685	0.136	(0.259)	[279]
No. attributes - negative response	0.051	0.022	0.029	(0)	[2711]	0.058	0.012	0.046	(0.001)	[653]	0.061	0.018	0.043	(0.064)	[279]
Pct. of attributes - skeptical response	0.012	0.008	0.005	(0.021)	[2711]	0.018	0.015	0.003	(0.602)	[653]	0.022	0.011	0.012	(0.176)	[279]
Pct. of attributes - positive response	0.033	0.033	0	(0.933)	[2711]	0.07	0.07	0.001	(0.925)	[653]	0.071	0.066	0.005	(0.743)	[279]
Pct. of attributes - neutral response	0.35	0.328	0.022	(0.026)	[2711]	0.706	0.646	0.06	(0.011)	[653]	0.745	0.666	0.078	(0.028)	[279]
Pct. of attributes - negative response	0.013	0.006	0.007	(0)	[2711]	0.014	0.003	0.011	(0.002)	[653]	0.016	0.003	0.013	(0.027)	[279]
Responded skeptically for any attribute	0.032	0.016	0.017	(0)	[2711]	0.049	0.025	0.025	(0.018)	[653]	0.054	0.018	0.036	(0.025)	[279]
Responded negatively for any attribute	0.035	0.016	0.019	(0)	[2711]	0.04	0.009	0.031	(0)	[653]	0.039	0.011	0.029	(0.032)	[279]
B. White vs. Hispanic Testers															
Any contact	0.512	0.512	0	(0.968)	[2711]	—	—	—	—	—	—	—	—	—	—
Scheduling appointment	0.348	0.354	-0.006	(0.283)	[2711]	—	—	—	—	—	—	—	—	—	—
No. of attributes brought up by landlord/broker	1.101	1.07	0.031	(0.434)	[2711]	2.023	2.072	-0.049	(0.549)	[653]	2.151	1.989	0.161	(0.19)	[279]
No. attributes - skeptical response	0.053	0.065	-0.012	(0.241)	[2711]	0.077	0.081	-0.005	(0.831)	[653]	0.097	0.039	0.057	(0.074)	[279]
No. attributes - positive response	0.132	0.14	-0.008	(0.532)	[2711]	0.262	0.27	-0.008	(0.832)	[653]	0.269	0.24	0.029	(0.533)	[279]
No. attributes - neutral response	0.918	0.884	0.035	(0.32)	[2711]	1.703	1.75	-0.047	(0.524)	[653]	1.821	1.728	0.093	(0.383)	[279]
No. attributes - negative response	0.051	0.046	0.005	(0.543)	[2711]	0.058	0.052	0.006	(0.706)	[653]	0.061	0.022	0.039	(0.07)	[279]
Pct. of attributes - skeptical response	0.012	0.016	-0.004	(0.114)	[2711]	0.018	0.023	-0.005	(0.39)	[653]	0.022	0.012	0.01	(0.218)	[279]
Pct. of attributes - positive response	0.033	0.037	-0.004	(0.27)	[2711]	0.07	0.077	-0.007	(0.514)	[653]	0.071	0.069	0.002	(0.898)	[279]
Pct. of attributes - neutral response	0.35	0.34	0.01	(0.305)	[2711]	0.706	0.714	-0.009	(0.656)	[653]	0.745	0.733	0.011	(0.699)	[279]
Pct. of attributes - negative response	0.013	0.013	0	(0.929)	[2711]	0.014	0.015	-0.002	(0.742)	[653]	0.016	0.007	0.009	(0.204)	[279]
Responded skeptically for any attribute	0.032	0.036	-0.004	(0.435)	[2711]	0.049	0.051	-0.002	(0.898)	[653]	0.054	0.029	0.025	(0.145)	[279]
Responded negatively for any attribute	0.035	0.03	0.004	(0.324)	[2711]	0.04	0.038	0.002	(0.876)	[653]	0.039	0.018	0.022	(0.109)	[279]
C. Black vs. Hispanic Testers															
Any contact	0.524	0.512	0.012	(0.189)	[2711]	—	—	—	—	—	—	—	—	—	—
Scheduling appointment	0.361	0.354	0.007	(0.284)	[2711]	—	—	—	—	—	—	—	—	—	—
No. of attributes brought up by landlord/broker	1.041	1.07	-0.029	(0.47)	[2711]	1.936	2.072	-0.136	(0.107)	[653]	1.961	1.989	-0.029	(0.813)	[279]
No. attributes - skeptical response	0.027	0.065	-0.038	(0)	[2711]	0.041	0.081	-0.04	(0.045)	[653]	0.039	0.039	0	(1)	[279]
No. attributes - positive response	0.125	0.14	-0.015	(0.327)	[2711]	0.256	0.27	-0.014	(0.749)	[653]	0.258	0.24	0.018	(0.772)	[279]
No. attributes - neutral response	0.894	0.884	0.01	(0.775)	[2711]	1.668	1.75	-0.083	(0.298)	[653]	1.685	1.728	-0.043	(0.703)	[279]
No. attributes - negative response	0.022	0.046	-0.024	(0)	[2711]	0.012	0.052	-0.04	(0.002)	[653]	0.018	0.022	-0.004	(0.782)	[279]
Pct. of attributes - skeptical response	0.008	0.016	-0.008	(0)	[2711]	0.015	0.023	-0.008	(0.196)	[653]	0.011	0.012	-0.001	(0.849)	[279]
Pct. of attributes - positive response	0.033	0.037	-0.004	(0.278)	[2711]	0.07	0.077	-0.008	(0.498)	[653]	0.066	0.069	-0.003	(0.831)	[279]
Pct. of attributes - neutral response	0.328	0.34	-0.013	(0.202)	[2711]	0.646	0.714	-0.069	(0.003)	[653]	0.666	0.733	-0.067	(0.036)	[279]
Pct. of attributes - negative response	0.006	0.013	-0.007	(0.001)	[2711]	0.003	0.015	-0.012	(0.001)	[653]	0.003	0.007	-0.005	(0.302)	[279]
Responded skeptically for any attribute	0.016	0.036	-0.02	(0)	[2711]	0.025	0.051	-0.026	(0.011)	[653]	0.018	0.029	-0.011	(0.367)	[279]
Responded negatively for any attribute	0.016	0.03	-0.014	(0)	[2711]	0.009	0.038	-0.029	(0)	[653]	0.011	0.018	-0.007	(0.415)	[279]

Table 4.1: Incidence of Early Stage Discrimination

Net Measure of Discrimination	I. Cases Assigned to Control Group					II. Cases Assigned to Any Treatment Group					III. All Cases in Experimental Sample													
	Mean level of favorable treatment		Difference (Maj.-Min.)	P-value	[N]	Mean level of favorable treatment		Difference (Maj.-Min.)	P-value	[N]	Mean level of favorable treatment		Difference (Maj.-Min.)	P-value	[N]									
	Majority	Minority				Majority	Minority				Majority	Minority												
A. White vs. Black Testers																								
<i>Making the Appointment</i>																								
Landlord/broker honored appointment	0.993	0.996	-0.004	(0.318)	[279]	0.995	0.987	0.008	(0.18)	[374]	0.994	0.991	0.003	(0.415)	[653]									
<i>Subjective Evaluations of Interaction Quality</i>																								
Perceived sales efforts	0.417	0.509	-0.091	(0.049)	[253]	0.436	0.421	0.015	(0.715)	[334]	0.428	0.46	-0.032	(0.294)	[587]									
Received praise about rental qualifications	0.061	0.068	-0.008	(0.743)	[253]	0.052	0.045	0.008	(0.652)	[334]	0.056	0.055	0.001	(0.947)	[587]									
Positive reactions to testers' background	0.017	0.004	0.013	(0.174)	[253]	0.007	0.003	0.003	(0.587)	[334]	0.011	0.004	0.007	(0.161)	[587]									
Positive editorializing	0.817	0.765	0.052	(0.165)	[253]	0.797	0.743	0.054	(0.121)	[334]	0.806	0.753	0.053	(0.038)	[587]									
Professionalism	0.522	0.483	0.039	(0.404)	[253]	0.498	0.507	-0.008	(0.836)	[334]	0.508	0.496	0.012	(0.691)	[587]									
<i>Post-Visit Follow-Up</i>																								
Received post-visit callback	0.215	0.168	0.047	(0.107)	[279]	0.187	0.131	0.056	(0.018)	[374]	0.199	0.147	0.052	(0.005)	[653]									
Received post-visit offer for unit	0.118	0.09	0.029	(0.239)	[279]	0.094	0.08	0.013	(0.467)	[374]	0.104	0.084	0.02	(0.178)	[653]									
B. White vs. Hispanic Testers																								
<i>Making the Appointment</i>																								
Landlord/broker honored appointment	0.993	0.996	-0.004	(0.318)	[279]	0.995	0.995	0	(1)	[374]	0.994	0.995	-0.002	(0.705)	[653]									
<i>Subjective Evaluations of Interaction Quality</i>																								
Perceived sales efforts	0.417	0.45	-0.033	(0.486)	[252]	0.436	0.394	0.042	(0.295)	[334]	0.428	0.418	0.01	(0.737)	[586]									
Received praise about rental qualifications	0.061	0.045	0.015	(0.467)	[252]	0.052	0.081	-0.028	(0.164)	[334]	0.056	0.066	-0.01	(0.512)	[586]									
Positive reactions to testers' background	0.017	0.009	0.008	(0.441)	[252]	0.007	0.007	0	(0.979)	[334]	0.011	0.008	0.003	(0.56)	[586]									
Positive editorializing	0.817	0.786	0.031	(0.411)	[252]	0.797	0.838	-0.042	(0.186)	[334]	0.806	0.816	-0.011	(0.66)	[586]									
Professionalism	0.522	0.591	-0.069	(0.14)	[252]	0.498	0.458	0.04	(0.321)	[334]	0.508	0.515	-0.006	(0.843)	[586]									
<i>Post-Visit Follow-Up</i>																								
Received post-visit callback	0.215	0.154	0.061	(0.019)	[279]	0.187	0.171	0.016	(0.503)	[374]	0.199	0.164	0.035	(0.099)	[653]									
Received post-visit offer for unit	0.118	0.061	0.057	(0.011)	[279]	0.094	0.08	0.013	(0.476)	[374]	0.104	0.072	0.032	(0.04)	[653]									
C. Black vs. Hispanic Testers																								
<i>Making the Appointment</i>																								
Landlord/broker honored appointment	0.996	0.996	0	(NaN)	[279]	0.987	0.995	-0.008	(0.18)	[374]	0.991	0.995	-0.005	(0.18)	[653]									
<i>Subjective Evaluations of Interaction Quality</i>																								
Perceived sales efforts	0.509	0.45	0.059	(0.213)	[251]	0.421	0.394	0.027	(0.501)	[339]	0.46	0.418	0.042	(0.169)	[590]									
Received praise about rental qualifications	0.068	0.045	0.023	(0.292)	[251]	0.045	0.081	-0.036	(0.069)	[339]	0.055	0.066	-0.011	(0.472)	[590]									
Positive reactions to testers' background	0.004	0.009	-0.005	(0.532)	[251]	0.003	0.007	-0.003	(0.572)	[339]	0.004	0.008	-0.004	(0.403)	[590]									
Positive editorializing	0.765	0.786	-0.021	(0.586)	[251]	0.743	0.838	-0.095	(0.004)	[339]	0.753	0.816	-0.063	(0.013)	[590]									
Professionalism	0.483	0.591	-0.108	(0.021)	[251]	0.507	0.458	0.049	(0.235)	[339]	0.496	0.515	-0.018	(0.555)	[590]									
<i>Post-Visit Follow-Up</i>																								
Received post-visit callback	0.168	0.154	0.014	(0.587)	[279]	0.131	0.171	-0.04	(0.079)	[374]	0.147	0.164	-0.017	(0.329)	[653]									
Received post-visit offer for unit	0.09	0.061	0.029	(0.17)	[279]	0.08	0.08	0	(1)	[374]	0.084	0.072	0.012	(0.359)	[653]									

Table 4.2: Baseline Incidence of Discrimination: In-Person and Post-Visit

4.1.2 Incidence of Later-Stage Discrimination

Turning to the experimental sample of cases to which all three testers in a team were able to successfully schedule a visit, we examine average differences in favorable treatment between groups. We focus only on cases assigned to the control group in order to capture the policy status quo (i.e., no messaging). Here we find statistically significant differences. We estimate that Hispanic testers were less likely than white testers to receive a callback from a landlord or broker—in 15.4% of cases compared to 21.5%, a difference of 6.1 percentage points ($p = 0.019$). They were also less likely to receive an offer for an apartment—in 6.1% of cases compared to 11.8%, a difference of 5.7 percentage points ($p = 0.011$).

We are also interested in whether there are systematic differences in treatment not captured by relatively blunt objective measures. In particular, are there subtler behaviors and interactions that can be measured via field notes, tabulated, and analyzed? The open-ended observations recorded by testers contained rich detail about day-to-day interactions in the New York City rental market. The one statistically significant difference we found was in sales efforts by landlords and brokers, e.g. offering incentives, following up with testers after a visit, etc. In 50.9% of cases, black testers reported experiencing sales efforts compared to 41.7% for white testers—a difference of nearly 9.2 percentage points in favor of black testers ($p = 0.049$).

We also employ these measures in the experimental analysis. Below, we delve into the kinds of interactions uncovered by testers in a more qualitative fashion, grouped under the categories we used to construct our subjective measures.

4.2 Subjective Perceptions of Favorable Interactions during Appointments

We constructed five indicators of testers' subjective perceptions of their treatment in the field by landlords and brokers. Those indicators were taken from a list of 15 that we generated both inductively, from speaking with testers and reading over preliminary field notes, and deductively, by taking into account behaviors and actions that may be of interest to CCHR. Since the interactions captured by these measures were subjective by design, we wanted to first group them by domain and then build an overall index of treatment. This makes it more likely that errors and subjective biases will cancel out in the aggregate.

During the initial coding process (as detailed in Section 2.6.4), we used pairs of research assistants to gauge each set of open-ended tester responses along the 15 dimensions that we outlined. However, some measures generated more agreement between coders than others, which we took to mean that there was variation in whether the indicators were validly capturing something real about the testers' interactions. Using all indicators to generate our overall index measure regardless of coder agreement would have risked introducing a great deal of noise and obscuring any possible

signal. For this reason, we kept the five indicators with the highest agreement as measured by the Cohen's kappa statistic (Viera and Garrett 2005).

Below we go through each of these five indicators in turn, illustrated by actual examples from the testers' field notes.

4.2.1 Positive Editorializing

As mentioned above, positive editorializing was defined as favorable commentary about an apartment building, residents, or neighborhood in an attempt to "talk up" the apartment to prospective tenants. Entries that were coded to indicate that positive editorializing was present in a tester's interaction for a given case exhibited certain commonalities:

- The phrase "up and coming" frequently recurred in landlords' or brokers' comments to testers, and they were eager to describe an area as desirable for young renters in terms of amenities, nightlife, etc.
- Landlords and brokers wanted to convey the trendiness of a neighborhood. Another tester wrote that the agent "described the residents as young, successful, and attractive. He'd repeat: '*Do you see what all the buzz and hype is about?*'"

Some testers did not experience this type of editorializing, and as the results above suggest, this varied in part by race. For instance, a Hispanic tester noted that "when I asked about the neighborhood, [the agent] replied in a very neutral and matter-of-fact tone: '*It's pretty quiet, but the front room facing 1st Avenue gets some noise on Friday and Saturday nights, I mean it's Manhattan, I'm not going to sugarcoat it.*'"

4.2.2 Sales Efforts

Sales efforts encompassed a variety of behaviors. These include expressing interest in following up, being responsive to testers' questions, being informative, offering rental incentives, etc. Some examples:

- Landlords and brokers sometimes offered contact information and applications to encourage prospective tenants: "Agent gave me her business card along with an application. She asked that I keep her posted about my decision. Agent also made mention that I shouldn't wait too long before committing to a decision as the unit will not remain on the market for too long."
- Testers were easily able to pick up on agents' level of interest and responsiveness: "P1 remained friendly throughout the visit answering the questions I posed to her and giving me specifics about the application process."

- Offering to follow up was also a definite signal of interest and a common sales tactic: “P1 definitely seemed interested in continued contact by the fact that she offered me her business card and urged me to contact her ASAP if I was interested in any of the units she had showed me. This coupled with the fact that P1 conducted the visit in a very attentive and engaged manner made me feel that P1 would be initiating continued contact with me.”

4.2.3 Praise About Qualifications

This indicator was intended to capture positive comments about testers’ qualifications for renting, such as credit score, income, etc. Illustrative examples from the data:

- Sometimes, testers simply described landlords or brokers as pleased with their qualifications: “Agent did appear to be interested in conducting further business with me. She seemed very impressed with my qualifications. Agent also expressed that I will make a great fit to the area.”
- Agents’ concern was often about renters’ income: “When I mentioned what my career was, a government job making 80k a year, he became more friendly in his tone. He said ‘*with that you can get a good apartment.*’ [His] tone became elated.”
- And sometimes, positive assessments came only after pointed questioning: “After drilling me with questions about my renting qualifications, he felt more positive about me. I tried to remain stoic while I answered.”

4.2.4 Positive Response to Background

Aside from judging qualifications, landlords or brokers could also offer spontaneous positive assessments of testers’ backgrounds—such as household composition, age, race/ethnicity, etc.—conveying their perceived suitability or “fit” with the apartment or neighborhood. This particular category encompasses responses of many types. For example:

- Landlords and brokers sometimes expressed interest in young tenants: One tester wrote that an agent “suggested that students and young professionals would make it a ‘*good neighborhood for you*’.”
- They sometimes attempted to make a connection with testers on the basis of personal attributes: The same tester wrote that the agent said “‘*I like your accent, it’s very cool*’ though he made no other comments about where I might be from.”

- Sometimes positive responses were at the expense of others, possibly on the basis of race. A white tester, on a visit together with another tester, wrote: “P1 favored me over the black tester, taking me aside and encouraging me to rent the better apartment.”

4.2.5 Professionalism

Professionalism encompasses aspects such as punctuality, courtesy, respectfulness, and attention to detail:

- Some testers noted the diligence of particular agents: “P1 remained friendly and engaged throughout the visit. P1 was very thorough and detailed in her descriptions of the units and building. P1 asked me how I liked each unit after showing it to me.”
- Professionalism can also involve basic components of etiquette: “P1 ended the visit in a very friendly manner with a handshake and smile.”
- Testers sometimes simply noted that agents were professional: “P2 and P3 were friendly and professional.”

Professionalism is perhaps most easily seen when it is absent. This could be evidenced by tardiness, distraction, or inattentiveness. On one visit, a tester noted: “P1 shook my hand and said I should contact her if I was interested. Her friends stood there staring at their phones and ignoring me. I got the feeling that I was just an errand on the way for them to go do something more fun.”

4.3 Forms of Objective Discrimination Encountered During Appointments

We also examine forms of objective discrimination that occur in-person during appointments. While these outcomes occur after treatment is assigned and delivered, we observe missing values in these measures because they are conditional on the occurrence of other outcomes. For example, a net measure of discrimination in the quoted monthly rental amount for a given rental unit is available only if the landlords and brokers showing the unit provide that information to both a majority group tester and a minority group tester. If such information is not provided to both testers in a matched pair, it is not possible to leverage the paired testing design to construct a net measure. We are concerned about egregious biases in treatment effect estimates that arise when outcome data are missing because missingness may be due to treatment. As such, we focus our analysis of these indicators to understanding the types of objective discrimination revealed in the data among audits where data are not missing. As a caveat, we stress that reported levels are not representative of the actual incidence of discrimination on these dimensions due to missing data, and differences should not be interpreted as baseline levels of discrimination on these dimensions in the policy status quo.

We examine patterns in multiple forms of objective discrimination that occur during the appointment stage: differences in the number of units shown; differences in the average quoted monthly asking rents across units shown; differences in signaled interest in the housing seeker as revealed by the landlord or broker's behavior; differences in the characteristics of housing units shown to testers; differences in the quoted rental price and amenities-in-rent for the listed unit, if shown; difference in move-in requirements for the listed unit, if shown; and difference in the willingness of the landlord or broker to negotiate on fees. The main findings reported in this section are summarized in Tables 4.3 and 4.4 and in histograms contained in Appendix D.

4.3.1 Differences in the Number of Units Shown

First we examine the differences in the number of units shown to white, black, and Hispanic testers. As shown in Table 4.3, on average, black testers reported viewing 1.252 units during their appointments with landlords and brokers, compared to 1.283 units for Hispanic testers and 1.341 units for white testers. Among valid paired tester reports, white testers were shown on average 0.077 more housing units than black testers and on average 0.047 more housing units than Hispanic testers. Black testers were shown on average 0.052 fewer housing units than Hispanic testers. These means provide a sense of the average levels of differences in the number of units shown for valid data from paired tests, but do not provide information about the distribution of differences that exist in the observed sample. Table 4.4 summarizes the range of valid difference measures from the audit and shows that the differences in the number of units shown to paired testers ranges from -3 to 4.¹⁵

¹⁵Histograms summarizing the distribution of the variables discussed in this section may be found in Appendix D.

Measure	Among Tester Reports						Among Valid Paired Tester Reports					
	Black Testers		Hispanic Testers		White Testers		White v. Black		White v. Hisp.		Black v. Hisp.	
	Mean	N	Mean	N	Mean	N	Difference	N	Difference	N	Difference	N
A. Across Units Shown												
Number of units shown	1.252	532	1.283	520	1.341	540	0.077	482	0.047	472	-0.052	460
Average quoted monthly rental price (\$)	2438.10	521	2461.51	516	2429.53	535	10.09	471	5.22	465	-9.22	449
Average number of bedrooms per unit	1.737	518	1.752	516	1.69	533	-0.033	465	-0.026	464	0.003	447
Number of units with quoted rent provided	1.245	526	1.269	517	1.325	536	0.069	475	0.047	466	-0.042	453
Number of units requiring application	1.002	526	1.046	517	1.179	536	0.16	475	0.129	466	-0.088	453
Number of units in building with doorman	0.272	526	0.304	517	0.332	536	0.055	475	0.019	466	-0.051	453
Number of units in building with elevator	0.546	526	0.586	517	0.593	536	0.042	475	0.004	466	-0.062	453
Number of units with washer and dryer in unit	0.183	526	0.199	517	0.192	536	0.008	475	-0.004	466	-0.026	453
Number of units with washer and dryer in building	0.437	526	0.493	517	0.519	536	0.078	475	0.017	466	-0.088	453
Percent of units with quoted rent provided	98.6	526	99.2	517	98.8	536	0.1	475	-0.1	466	-0.3	453
Percent of units requiring application	77.5	526	77.8	517	85.6	536	7.3	475	8.1	466	-2.4	453
Percent of units in building with doorman	18.9	526	17.9	517	18.6	536	-0.3	475	0.1	466	0.2	453
Percent of units in building with elevator	38.8	526	39.9	517	36.8	536	-1.5	475	-2.3	466	-1.1	453
Percent of units with washer and dryer in unit	13.4	526	13.6	517	13.4	536	0.4	475	0	466	-0.2	453
Percent of units with washer and dryer in building	30.3	526	33.3	517	32.1	536	2.1	475	-1.5	466	-4.5	453
Willing to negotiate on fees (Yes=1, No=0)	0.012	506	0.024	499	0.039	513	0.023	442	0.018	436	-0.005	425
B. For Advertised Unit, if Shown												
Quoted monthly rental price (\$)	2453.66	355	2449.62	346	2449.97	356	0	316	0	305	0	302
Stated number of amenities included in rent	1.587	506	1.95	499	2.287	513	0.676	442	0.374	436	-0.348	425
Quoted requirement to secure and move into unit (Yes=1, No=0)												
At least one month's rent	0.273	506	0.413	499	0.298	513	0.029	442	-0.119	436	-0.139	425
Security deposit	0.937	506	0.942	499	0.922	513	-0.011	442	-0.023	436	-0.009	425
Broker fees	0.322	506	0.337	499	0.345	513	0.009	442	0.018	436	-0.026	425
Application fees	0.243	506	0.228	499	0.349	513	0.09	442	0.096	436	0.007	425
Administrative and processing fees	0.002	506	0.006	499	0.008	513	0.007	442	0.002	436	-0.005	425
Additional holding fees or good faith deposits	0.071	506	0.076	499	0.058	513	-0.009	442	-0.023	436	-0.007	425
Credit or background check fees	0.178	506	0.132	499	0.142	513	-0.052	442	0.009	436	0.045	425
Other fees	0.016	506	0.016	499	0.008	513	-0.007	442	-0.011	436	0.002	425
Total fees and up-front costs (\$)	4047.89	526	4602.45	517	3865.42	536	-98.50	475	-657.28	466	-425.37	453
Quoted up-front costs, by fee category (\$)												
Upfront rent, deposit, holding, broker fees	4012.76	526	4565.10	517	3823.89	536	-104.07	475	-660.16	466	-424.11	453
Application and credit/background check fees	28.67	526	32.73	517	33.54	536	3.71	475	-0.12	466	-4.74	453
Administrative fees and other net costs	6.46	526	4.61	517	7.99	536	1.86	475	3.00	466	3.48	453

Table 4.3: Indicators of Objective Discrimination during Appointments

Measure	White vs. Black		White vs. Hispanic		Black vs. Hispanic		All Net Measures	
	Min	Max	Min	Max	Min	Max	Min	Max
A. Across Units Shown								
Number of units shown	-3	3	-3	4	-3	2	-3	4
Average quoted monthly rental price (\$)	-1400	2100	-1800	1400	-1800	895	-1800	2100
Average number of bedrooms per unit	-3	3	-3	3	-2	2	-3	3
Number of units with quoted rent provided	-3	3	-3	4	-3	3	-3	4
Number of units requiring application	-4	3	-3	5	-5	4	-5	5
Number of units in building with doorman	-2	3	-3	4	-3	3	-3	4
Number of units in building with elevator	-3	3	-3	4	-3	2	-3	4
Number of units with washer and dryer in unit	-3	3	-2	3	-3	2	-3	3
Number of units with washer and dryer in building	-3	3	-3	5	-3	3	-3	5
Percent of units with quoted rent provided	-1	1	-0.5	1	-1	1	-1	1
Percent of units requiring application	-1	1	-1	1	-1	1	-1	1
Percent of units in building with doorman	-1	1	-1	1	-1	1	-1	1
Percent of units in building with elevator	-1	1	-1	1	-1	1	-1	1
Percent of units with washer and dryer in unit	-1	1	-1	1	-1	1	-1	1
Percent of units with washer and dryer in building	-1	1	-1	1	-1	1	-1	1
Willing to negotiate on fees (Yes=1, No=0)	-1	1	-1	1	-1	1	-1	1
B. For Advertised Unit, if Shown								
Quoted monthly rental price for listed unit	0	0	0	0	0	0	0	0
Stated number of amenities included in rent for listed unit	-6	6	-4	4	-6	5	-6	6
Quoted requirement to secure and move into unit (Yes=1, No=0)								
At least one month's rent	-1	1	-1	1	-1	1	-1	1
Security deposit	-1	1	-1	1	-1	1	-1	1
Broker fees	-1	1	-1	1	-1	1	-1	1
Application fees	-1	1	-1	1	-1	1	-1	1
Administrative and processing fees	0	1	0	1	-1	0	-1	1
Additional holding fees or good faith deposits	-1	1	-1	1	-1	1	-1	1
Credit or background check fees	-1	1	-1	1	-1	1	-1	1
Other fees	-1	1	-1	1	-1	1	-1	1
Total fees	-10710	10080	-13375	9150	-15700	15900	-15700	15900
Quoted up-front costs, by fee category (\$)								
Upfront rent, deposit, holding, broker fees	-10710	10080	-13400	9150	-15700	15900	-15700	15900
Application and credit/background check fees	-400	1950	-1500	1850	-1400	150	-1500	1950
Administrative fees and other net costs	-600	1950	-700	2000	-450	2000	-700	2000

Table 4.4: Range of Net Differences in Objective Discrimination Indicators among Non-Missing Paired Test Reports

Among valid paired tests involving a white and black tester, the black tester was shown 4 or more units than the white tester in zero cases, 3 more units than the white tester in 2 cases, 2 more units than the white tester in 3 cases, and 1 more unit than the white tester in 26 cases. In this same set of paired tests, the white tester was shown 4 or more units than the black tester in zero cases, 3 more units than the black tester in 1 case, 2 more units than the black tester in 13 cases, and 1 more unit than the black tester in 46 cases.

Among valid paired tests involving a white and Hispanic tester, the Hispanic tester was shown 4 or more units than the white tester in zero cases, 3 more units than the white tester in 2 cases, 2 more units than the white tester in 5 cases, and 1 more unit than the white tester in 35 cases. In this same set of paired tests, the white tester was shown 5 or more units than the Hispanic tester in zero cases, 4 more units than the Hispanic tester in one case, 3 more units than the Hispanic tester in 2 cases, 2 more units than the Hispanic tester in 6 cases, and 1 more unit than the Hispanic tester in 51 cases.

Among valid paired tests involving a black and Hispanic tester, the Hispanic tester was shown 4 or more units than the black tester in zero cases, 3 more units than the black tester in 4 cases, 2 more units than the black tester in 10 cases, and 1 more unit than the black tester in 38 cases. In this same set of paired tests, the black tester was shown 3 or more units than the Hispanic tester in zero cases, 2 more units than the Hispanic tester in 6 cases, and 1 more unit than the Hispanic tester in 34 cases.

4.3.2 Differences in Average Quoted Monthly Asking Rents Across Units Shown

Next, we examine whether, on average over units shown, there are differences in the quoted monthly asking price for white, black, and Hispanic housing seekers as an indicator of steering in the rental housing market. On average, black testers reported an average quoted monthly asking rental price of \$2,438.10 for a rental unit, compared to \$2,461.51 for Hispanic testers, and \$2,429.53 for white testers. Comparing mean differences among valid paired tests, white testers were quoted monthly asking rental prices that were on average \$10.09 more per unit than black testers, and \$5.22 more per unit than Hispanic testers. Black testers were quoted monthly asking rental prices that were on average \$9.22 less per unit than for Hispanic testers. While these mean differences do not suggest evidence of large disparities in average quoted rental prices per unit, the distribution of average quoted monthly rents reveals more heterogeneity in the range of differential experiences paired testers confronted.

As shown in Figure 4.1, we find that among the same majority-minority tester pairing, there is a substantively significant disparities in the average quoted monthly rental price per unit. Among valid paired tests involving a white and black tester, white testers were quoted average monthly rents up to \$1,400 less than the average price per unit quoted to black testers. Black testers were

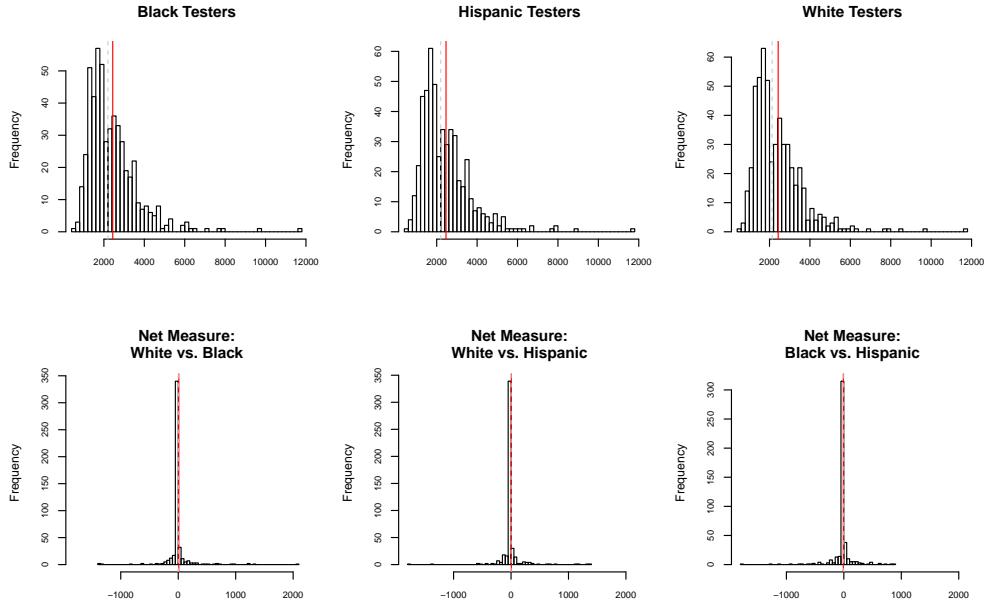


Figure 4.1: Average Quoted Price Across Units Shown, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

quoted average monthly rents up to \$2,100 less than the average price quoted to white testers. Among valid paired tests involving a white and Hispanic tester, we observe a similarly wide range. At times, white testers were quoted average monthly rents up to \$1,800 less than the average prices quoted to Hispanic testers. Other times, Hispanic testers were quoted average monthly rents up to \$1,400 less than the average price quoted to white testers. Similarly, among valid paired tests involving a black and Hispanic tester, black testers were quoted average monthly rental prices up to \$1,800 less than the average price quoted to black testers. Hispanic testers were sometimes quoted average monthly rental prices up to \$895 less than the average price quoted to black testers.

Several examples illustrate this phenomenon concretely:

- One team of testers responded to an advertisement for a 1-bedroom apartment in the West Village listed at \$4,200/month. Upon arriving for the appointment, the black and white testers were both shown two units; across these units the average quoted rent was \$6,100/month. In contrast, the Hispanic tester was shown only one unit that cost \$7,900/month.
- Another team of testers responded to an advertisement for a 3-bedroom apartment in the Upper East Side listed at \$5,500/month. The black tester was shown one unit listed at \$5,995/month. The Hispanic tester was shown two units; the average quoted price across

these two units was \$5,845/month. The white tester was shown only one unit listed at \$6,995/month.

- When responding to an advertisement for a \$1,450/month studio in Brooklyn's Prospect-Lefferts Gardens neighborhood, the black and Hispanic testers were both shown the listed unit and were quoted the listed monthly asking rental price of \$1,450/month. In contrast, the white tester was shown two units with an average quoted rent of \$2,800/month.
- When pursuing an advertised listing for a \$3,200/month studio in Manhattan's Union Square neighborhood, the white tester was shown two apartments with the highest average quoted rent of \$7,850/month. The black tester was shown three apartments with the next highest average quoted rent of \$7,131.67/month. The Hispanic tester was shown only one unit with a quoted rental price of \$6,675/month.
- A matched tester trio pursued a 3-bedroom apartment in Bushwick listed at \$2,800/month. Both the black and Hispanic testers were shown one apartment each: the black tester was shown a \$1,695/month apartment whereas the Hispanic tester was shown a \$2,800/month apartment. The white tester was shown two apartments with an average quoted rent of \$2,900/month.

4.3.3 Differences in Signaled Interest in the Housing Seeker

We examine two features of the behavior of landlords and brokers which may reveal differences in signaled interest between housing seekers: differences in their willingness to quote the exact price for units shown and differences in whether they cite application requirements as a signal of initiating procedures to execute a lease.

- **Willingness to quote prices across rental units shown**

In a few instances, landlords and brokers did not disclose the monthly asking rental price in person to testers during the appointment. We conjecture that the lack of willingness of a landlord or broker to disclose the rental price to a housing seeker may suggest that the landlord or broker is not seriously considering the housing seeker as a potential tenant.

On average, black testers reported receiving a quoted rental price for 1.245 units per appointment, Hispanic testers reported receiving a quoted rental price for 1.269 units per appointment, and white testers reported receiving a quoted rental price for 1.325 units per appointment. Among paired tests with valid data on this measure, we observe very small differences in the average number of units per appointment for which landlords and brokers are willing to disclose the rental price. On average, white testers receive a quoted rental price for 0.069 more units than black testers and

for 0.047 more units than Hispanic testers. On average, black testers receive a quoted rental price for 0.042 fewer units than Hispanic testers.

To make these apparent levels and differences comparable, we calculate the share of units viewed and difference in shares for which testers are quoted a monthly asking rental price. On average, black testers reported receiving a quoted rental price in 98.6% of the units they viewed, Hispanic testers reported receiving a quoted rental price in 99.2% of the units they viewed, and white testers reported receiving a quoted rental price in 98.8% of the units they viewed. For valid paired tester reports, taking the difference between these percentages yield percentage point differences in the rates of being quoted rental prices for units shown. Within this sample, the rate at which white testers are quoted rental prices is 0.1 percentage points greater than the rate for black testers but 0.1 percentage points lower than the rate for Hispanic testers. The rate at which black testers are quoted rental prices is 0.3 percentage points lower than the rate for Hispanic testers.

- **Citing application requirements**

Where applications are required, differences in whether landlords and brokers cite application requirements across housing seekers may signal differences in their level of interest in housing seekers as potential tenants. This signal is also critical because it is necessary to begin the process of securing a vacant rental unit and executing a lease.

On average, black testers view 1.002 housing units where they are informed that an application is required, compared to 1.046 housing units for Hispanic testers and 1.179 housing units for white testers. In percentage terms of the number of housing units viewed, black testers are informed about application requirements 77.5% of the time, compared to 77.8% for Hispanic testers and 85.6% for white testers.

Apparent differences between these average counts and levels may be due to differences across reported cases and differences in housing units within a case. We therefore examine differences between testers by race among valid paired tests. On average, white testers are informed about application requirements for 0.16 more units than their black tester partners and for 0.129 more units than their Hispanic tester partners in valid paired tests. Black testers are informed about applications requirements for 0.088 fewer units than Hispanic testers in valid paired tests. When comparing rates among units shown, white testers are informed about application requirements at a rate 7.3 percentage points higher than the rate for black testers and 8.1 percentage points higher than the rate for Hispanic testers in valid paired tests. Black testers are informed about application requirements at a rate 2.4 percentage points lower than the rate for Hispanic testers in valid paired tests.

4.3.4 Differences in the Characteristics of Housing Units Shown

Next, we compare differences in the characteristics of housing units shown to the testers within a matched pair. We examine several indicators of housing stock quality among the units testers were shown: the average number of bedrooms per unit across units shown; and the number and share of units shown with a doorman, an elevator, a washer/dryer unit in the unit or in the building.

- **Average number of bedrooms per unit among housing units shown**

On average, black testers reported being shown units with 1.737 bedrooms, Hispanic testers reported being shown units with 1.752 bedrooms, and white testers reported being shown units with 1.69 bedrooms. Among the set of paired tests between white and black testers with valid data, white testers reported viewing units with on average 0.033 fewer bedrooms than black testers. Among paired tests between white and Hispanic testers with valid data, white testers reported viewing units with on average 0.026 fewer bedrooms than Hispanic testers. Among paired tests between black and Hispanic testers with valid data, black testers reported viewing units with on average 0.003 more bedrooms than Hispanic testers.

These apparent differences in means seem negligible. When examining the distribution of differences in average bedroom counts per unit, we observe greater variation in differences, suggesting that in some instances steering by housing stock characteristics matters considerably. Among paired tests with white and black testers and valid data, there are 9 cases where the white tester is shown housing units with 0.5 fewer bedrooms on average than the black tester; 26 cases where the white tester is shown housing units with 1 fewer bedrooms on average than the black tester; one case where the white tester is shown units with 1.5 fewer bedrooms on average than the black tester; two cases where the white tester is shown units with 2 fewer bedrooms on average than the black tester, and two cases where the white tester is shown units with 3 fewer bedrooms on average than the black tester. In this same subset of paired tests between white and black testers, there are 11 cases where the white tester is shown units with 0.5 more bedrooms on average than the black tester, 15 cases where the white tester is shown units with one more bedroom on average than the black tester, 2 cases where the white tester is shown units with 2 more bedrooms on average than the black tester, and one case where the white tester is shown 3 more bedrooms on average than the black tester.

Similar patterns exist when we examine the set of paired tests between white and Hispanic testers with valid data. There are 6 cases where the white tester in a matched pair is shown housing units with 0.5 fewer bedrooms on average than the Hispanic tester, 30 cases where the white tester is shown units with 1 fewer bedroom on average than the Hispanic tester, 3 cases where the white tester is shown units with 2 fewer bedrooms on average than the Hispanic tester, and 1 case where the white tester is shown units with 3 fewer bedrooms on average than the Hispanic tester. The

opposite pattern also exists. There are 2 cases where the white tester is shown units with 0.25 more bedrooms on average than the Hispanic tester, 13 cases where the the white tester is shown units with 0.5 more bedrooms on average than the Hispanic tester, 12 cases where the white tester is shown units with 1 more bedroom on average than the Hispanic tester, 3 cases where the white tester is shown units with 2 more bedrooms on average than the Hispanic tester, and one case where the white tester is shown units with 3 more bedrooms on average than the Hispanic tester.

The range of differences in the average number of bedrooms in units shown to paired black and Hispanic testers is less extreme but similar patterns manifest. Among paired tests between black and Hispanic testers with valid data, there are 2 cases where black testers are shown units with 2 fewer bedrooms on average than Hispanic testers, 26 cases where black testers are shown units with 1 fewer bedroom on average than Hispanic testers, 1 case where black testers are shown 0.75 fewer bedrooms on average than Hispanic testers, and 10 cases where black testers are shown units with 0.5 fewer bedrooms on average than Hispanic testers. There are 2 cases where black testers are shown units with 2 more bedrooms on average than Hispanic testers, 26 cases where black testers are shown units with 1 more bedroom on average than Hispanic testers, 12 cases where black testers are shown units with 0.5 more bedrooms on average than Hispanic testers, and 4 cases where black testers are shown units with 0.25 more bedrooms on average than Hispanic testers.

- **Selected unit and building amenities among housing units shown**

We now examine the average incidence of selected unit and building amenities among the housing units landlords and brokers chose to show testers: units in a building with a doorman, units in a building with an elevator, units where the washer and dryer are in the unit, and units where the washer and dryer are in the building. We selected these amenities as housing stock characteristics for two reasons. First, these amenities are often used as a marker of the quality of housing units. Second, observed variation on these dimensions by tester race would suggest the presence of steering.

- *Number/percent of units in a building with a doorman:* White testers were shown more rental housing units that were in buildings with doormen than Hispanic and black testers on average. On average, black testers were shown 0.272 units in a building with a doorman, Hispanic testers were shown 0.304 units in a building with a doorman, and white testers were shown 0.332 units in a building with a doorman. Among paired tests with valid data, white testers were shown on average 0.055 more units in doorman buildings than their paired black tester counterparts and on average 0.019 more units in doorman buildings than their paired Hispanic tester counterparts. Black testers were shown on average -0.051 fewer units than their Hispanic tester partners.

When we account for the total number of units each tester was shown and examine the average percent of units shown that were located in a doorman building, we observe that black testers were shown units in doorman buildings 18.9% of the time, as compared to 18.6% for white testers and 17.9% for Hispanic testers. Among paired tests with valid data, white testers were shown units in doorman buildings at a rate 0.3 percentage points less than the rate for black testers but 0.1 percentage points greater than the rate for Hispanic testers. Black testers were shown units in doorman buildings at a rate 0.2 percentage points greater than the rate for their paired Hispanic tester counterparts.

- *Number/percent of units in a building with an elevator:* White testers were shown more rental housing units that were in buildings with elevators than Hispanic and black testers on average. On average, black testers were shown 0.546 units in a building with an elevator, compared to 0.586 units for Hispanic testers and 0.593 units for white testers. Among paired tests with valid data, white testers were shown on average 0.042 more units in elevator buildings than black testers and on average 0.004 more units in elevator buildings than Hispanic testers. Black testers were shown on average 0.062 fewer units in elevator buildings than Hispanic testers in valid paired tests.

When we account for the total number of units each tester was shown and examine the average percent of units shown that were located in an elevator building, we observe that black testers were shown units in elevator buildings 38.8% of the time, as compared to 39.9% for Hispanic testers and 36.8% for white testers. Among paired tests with valid data, white testers were shown units in elevator buildings at a rate 1.5 percentage points less than the rate for black testers and a rate 2.3 percentage points less than the rate for Hispanic testers. Black testers were shown units in elevator buildings at a rate 1.1 percentage points less than the rate for Hispanic testers in valid paired tests.

- *Number/percent of units with washer/dryer in unit:* Hispanic testers were shown more rental units that had a washer and dryer system in the unit than black and white testers. On average, Hispanic testers were shown 0.199 units that included a washer and dryer, compared to 0.192 units for white testers and 0.183 units for black testers. Among paired tests with valid data, white testers were shown on average 0.008 more units that included a washer and dryer in the apartment than black testers but 0.004 fewer units than Hispanic testers. Black testers were shown on average 0.026 fewer units that included a washer and dryer in the apartment than Hispanic testers in valid paired tests.

When we account for the total number of units each tester was shown and examine the average percent of units shown that includes a washer and dryer system in the unit, we observe that both black and white testers were shown units with a washer and dryer in the

apartment 13.4% of the time. Hispanic testers were shown units with a washer and dryer in the apartment 13.6% of the time. Among paired tests with valid data, white testers were shown units with a washer and dryer in the apartment at a rate 0.4 percentage points higher than the rate for black testers and at the same rate as Hispanic testers. Black testers were shown units with a washer and dryer in the apartment at a rate 0.2 percentage point lower than the rate for Hispanic testers in valid paired tests.

- *Number/percent of units with washer/dryer in building:* White testers were shown more rental units that had a washer and dryer in the building than black and Hispanic testers. On average, white testers were shown 0.519 housing units where a washer and dryer was located in the building, compared to 0.493 housing units for Hispanic testers and 0.437 housing units for black testers. Among paired tests with valid data, white testers were shown on average 0.078 more housing units with a washer and dryer in the building as compared to black testers and 0.017 more housing units as compared to Hispanic testers. On average, black testers were shown 0.088 fewer housing units with a washer and dryer in the building as compared to Hispanic testers in valid paired tests.

When we account for the total number of units each tester was shown and examine the average percent of units shown with a washer and dryer in the building, we observe that black testers were shown units in a washer and dryer in the building 30.3% of the time on average, compared to 33.3% for Hispanic testers and 32.1% for white testers. Among valid paired tests, white testers were shown housing units in buildings with a washer and dryer at a rate 2.1 percentage points higher than the rate for black testers but 1.5 percentage points lower than the rate for Hispanic testers. On average, black testers were shown housing units in buildings with a washer and dryer at a rate 4.5 percentage points lower than the rate for Hispanic testers in valid paired tests.

4.3.5 Differences in the Quoted Rental Price and Amenities-in-Rent for Listed Unit, if Shown

Next, we describe differences in the quoted monthly rental price and amenities included in the rental price for the listed unit, if shown. Comparing these quantities for the same listed unit is useful because it allows us to control for physical characteristics of the housing unit of interest.

- **Quoted monthly rental price for listed unit, if shown**

On average, white, black and Hispanic testers reported approximately the same quoted rental price for the advertised unit. Black testers reported an average quoted monthly rent of \$2,453.66 for the listed unit, compared to \$2,449.62 for Hispanic testers and \$2449.97 for white testers. Among

valid paired tests, there was no difference in quoted rental prices for the listed unit for white-black, white-Hispanic, or black-Hispanic comparisons.

- **Stated number of amenities included in rental price for listed unit, if shown**

We observed differences in the set of amenities that were quoted as being included in the rental price for the listed unit. Testers recorded whether each of the following amenities as quoted as being included in the rental price: parking, gym or health club membership or access, cleaning or laundry service, Internet or cable, heating, electric, gas, sewage, and water. For each case-tester observation, we calculated the total number of amenities quoted as being included in the rent for the listed unit, and compare patterns in this composite measure.

White testers were offered more amenities as part of their rent than black and Hispanic testers. On average, white testers reported offered 2.287 amenities included in rent compared to 1.95 amenities for Hispanic testers and 1.587 amenities for black testers. Among paired tests with valid data, white testers were offered on average 0.676 more amenities included in rent than black testers and 0.374 more amenities included in rent than Hispanic testers. Black testers were offered 0.348 fewer amenities included in rent than Hispanic testers in valid paired tests.

4.3.6 Differences in Move-In Requirements for Listed Unit, if Shown

We now turn to examining patterns in move-in requirements for the listed unit, if the listed unit was shown to testers. We focus in particular on the structure of quoted costs required to secure and move into the listed unit and the quoted amount of costs required.

- **Structure of costs required to secure and move into unit, for listed unit if shown**

In the field visit survey corresponding to the appointment stage of the housing audit, testers were instructed to complete an open-ended question describing the *structure* and *amount* of quoted fees and costs required to secure and move into every housing unit viewed. These open-ended responses were coded to reveal eight categories of fees: (a) several months of rent: this amount is to be paid up-front to secure and/or move-in to the unit; (b) application fees; (c) a security deposit; (d) administrative and processing fees on top of the application fee; (e) broker fees, usually equal to several months rent (1-3 months) or a percentage of the annual rent (often 10-15 percent); (f) additional “holding fees” (on top of the security deposit) or a “good faith deposit” to take the unit off the market and to force the housing seeker to signal a credible commitment to execute the lease; (g) credit check or background check fees (separate from application fees); and (h) other fees, which usually means unspecified fees or non-refundable “move-in/move-out fees.”

We assess variation in the structure of fees quoted to testers by race, and among valid paired tests by race.

- *At least one month's rent:* On average, Hispanic testers reported that at least one month's rent was required up front more frequently (41.3%) than was reported by white testers (29.8%) and by black testers (27.3%). Among valid paired tests, white testers reported that at least one month's rent was required up front at a rate 2.9 percentage points greater than the rate for black testers but 11.9 percentage points less than the rate for Hispanic testers. Black testers reported that at least one month's rent was required up front at a rate 13.9 percentage points less than rate for Hispanic testers in valid paired tests.
- *Security deposit:* On average, Hispanic testers reported that a security deposit was required more often (94.2%) than was reported by black testers (93.7%) and by white testers (92.2%). Among valid paired tests, white testers reported that a security deposit was required at a rate 1.1 percentage points less than the reported rate for black testers and 2.3 percentage points less than the reported rate for Hispanic testers. Black testers reported that a security deposit was required at a rate 0.9 percentage points less than the rate for Hispanic testers in valid paired tests.
- *Broker fees:* On average, white testers reported that broker fees were required more frequently (34.5%) than was reported by Hispanic testers (33.7%) and by black testers (32.2%). Among valid paired tests, white testers reported that broker fees were required at a rate 0.9 percentage points higher than the reported rate for black testers and 1.8 percentage points higher than the reported rate for Hispanic testers. Black testers reported that broker fees were required at a rate 2.6 percentage points less than the rate reported by Hispanic testers in valid paired tests.
- *Application fees:* On average, white testers reported that application fees were required more frequently (34.9%) than was reported by black testers (24.3%) and by Hispanic testers (22.8%). Among valid paired tests, white testers reported that application fees were required at a rate 9 percentage points higher than the reported rate for black testers and 9.6 percentage points higher than the reported rate for Hispanic testers. Black testers reported that application fees were required at a rate 0.7 percentage points higher than the reported rate for Hispanic testers in valid paired tests.
- *Administrative and processing fees:* Overall across groups, administrative and processing fees were rarely quoted. On average, white testers reported that administrative and processing fees were required more frequently (0.8%) than was reported by Hispanic testers (0.6%) and by black testers (0.2%). Among valid paired tests, white testers reported that administrative and processing fees were required at a rate 0.7 percentage points higher than the reported rate for black testers and 0.2 percentage points higher than the reported rate for Hispanic

testers. Black testers reported that administrative and processing fees were required at a rate 0.5 percentage points less than the reported rate for Hispanic testers in valid paired tests.

- *Additional holding fees or good faith deposits:* On average, Hispanic testers reported that additional holding fees or good faith deposits were required more frequently (7.6%) than was reported by black testers (7.1%) and by white testers (5.8%). Among valid paired tests, white testers reported that such holding fees and good faith deposits were required at a rate 0.9 percentage points lower than the rate reported by black testers and 2.3 percentage points lower than the rate reported by Hispanic testers. Black testers reported that holding fees and good faith deposits were required at a rate 0.7 percentage points less than the rate reported by Hispanic testers in valid paired tests.
- *Credit or background check fees:* Black testers reported that credit or background check fees were required at higher rates (17.8%) on average than was reported by white tester (14.2%) and by Hispanic testers (13.2%). Among valid paired tests, white testers reported that credit or background check fees were required at a rate 5.2 percentage points less than the rate reported by black testers but 0.9 percentage points higher than the rate reported by Hispanic testers. Black testers reported that credit or background check fees were required at a rate 4.5 percentage points higher than the rate reported by Hispanic testers in valid paired tests.
- *Other fees:* Finally, black and Hispanic testers reported that other fees (often unspecified) were required at higher rates (1.6% for both) than was reported by white testers (0.8%). Among valid paired tests, white testers reported that other fees were quoted at rates 0.7% percentage points lower than the rates reported by black testers and 1.1 percentage points lower than the rates reported by Hispanic testers. Black testers reported that other fees were required at a rate 0.2 percentage points higher than the rates quoted to Hispanic testers in valid paired tests.

We observe considerable variation in the way fees are incurred on rental housing seekers by race. Move-in costs packaged as several months' rent up-front seem to be disproportionately levied on Hispanic testers as compared to their black and white counterparts. Hispanic and black testers were more likely than white testers to be told they needed to pay additional holding fees or good faith deposits. Hispanic and black testers were also more likely than white testers to be told they needed to pay a security deposit, although apparent differences are small magnitude. White testers were more frequently asked to pay broker fees, application fees, and administrative and processing fees than black and Hispanic testers. Black testers were more frequently levied fees for credit or background checks than both Hispanic and white testers.

- **Quoted amount of costs required to secure and move into unit, for listed unit if shown**

We now examine patterns in the quoted amount of up-front costs required to secure and move into the listed unit. We examine total amounts quoted, as well as fees organized into three sub-categories. These categories are collections of the eight types of fees mentioned above. The first of these categories groups application fees and credit or background check fees together because, as revealed through inductive analysis of the qualitative open-ended data, these labels were used interchangeably both within and across cases. The second category of fees sums all up-front rent, security deposits, holding fees, and broker fees because they constitute the bulk of up-front costs. We necessarily have to combine these because in several cases only a total amount for these costs was quoted and how these fees were decomposed by fee type was not provided to testers. The third category of fees is comprised of administrative fees and other net costs. The inclusion of all other net costs (such as discounts) in this measure allows us to sum the three subtotals to compute an accurate net total cost quoted to testers during the appointment.

- *Total fees:* On average, Hispanic testers were quoted \$4,602.45 in total fees and up-front costs, compared to \$4,047.89 for black testers and \$3,865.42 for white testers. Among paired tests with valid data, white testers were quoted on average \$98.50 less in total fees than black testers and on average \$657.26 less in total fees than Hispanic testers. Black testers were quoted on average \$425.37 less in total fees than Hispanic testers in valid paired tests.

We examine the distribution of these fees to get a better sense of variation in the range of quoted total fees by the tester's race and differences in fees by race. Figure 4.2 summarizes the distribution of total quoted fees by tester race and differences in total quoted fees between paired tests with valid data. Among valid paired tests, differences between white and black testers in quoted total costs required to secure and move into the listed unit ranged from -\$10,710 (i.e., the white tester was quoted that amount less than the black tester) to \$10,080 (i.e., the white tester was quoted that amount more than the black tester). Among valid paired tests, the difference in quoted total costs ranged from -\$13,375 to \$9,150 when comparing white and Hispanic testers, and from -\$15,700 to \$15,900 when comparing black and Hispanic testers.

- *Sum of up-front rent, security deposit, holding fees, and broker fees :* On average, Hispanic testers were quoted \$4565.10 in total up-front rent, deposits, holding fees and broker fees, compared to \$4,012.76 for black testers and \$3823.89 for white testers. Among paired tests with valid data, white testers were quoted on average \$104.07 less in these costs than black testers and on average \$660.16 less in these costs than Hispanic testers. Black testers were quoted on average \$424.11 less in up-front rent, deposits, and holding fees than Hispanic testers.

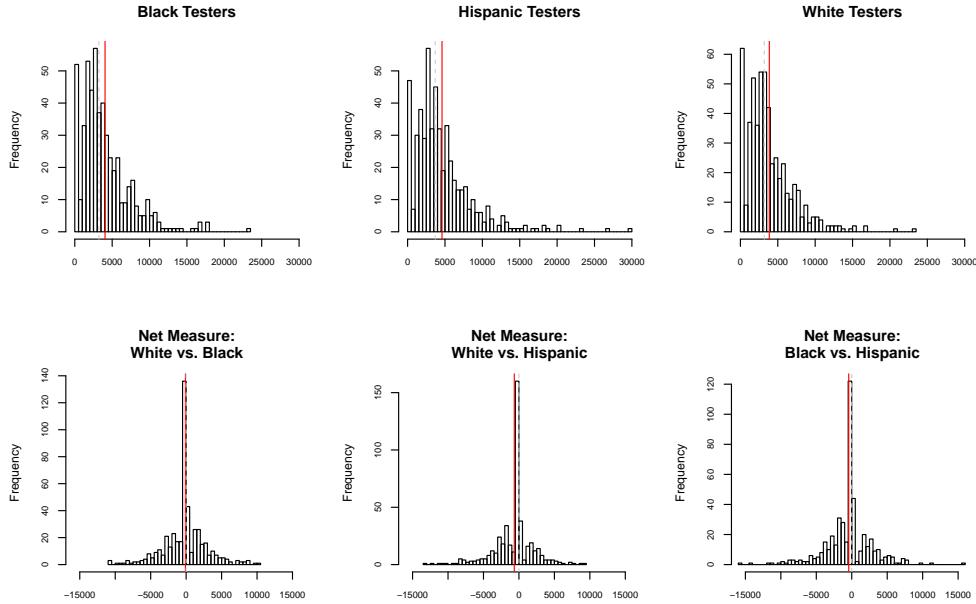


Figure 4.2: Distributions of Total Quoted Fees by Tester Race and by Group Differences. The first row presents the distribution of quoted fees by tester race, using valid case-tester-level reports. The second row presents the distribution of differences in quoted fees by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

Figure 4.3 summarizes the distribution of quoted up-front rent, security deposit, holding fees, and broker fees by tester race using data from valid tester reports, and by group differences using data from valid paired tests. Among valid paired tests, differences in this quoted sum between white and black testers range from -\$10,710 to \$10,080 are fairly symmetrically distributed around zero. These differences range from -\$13,400 to \$9,150 when comparing white and Hispanic testers in valid paired tests. The range of differences in up-front rent, security deposits and holding fees is greatest when comparing black and Hispanic testers from paired tests; this difference ranges from -\$15,700 to \$15,900.

The following illustrative examples fix ideas about the way different fee amounts and fee structures are quoted to different testers in a paired audit setting:

- One team of testers visited a one-bedroom unit in Brooklyn’s Midwood neighborhood near Ocean Avenue that was advertised at \$1,250/month. All three testers were shown the advertised unit and were quoted the advertised monthly rental price. However, the black tester was informed that a security deposit equal to two months rent (\$2,500) was required, whereas the Hispanic and white testers were both told that a security deposit equal to only one month’s rent (\$1,250) was required. The black tester was told that

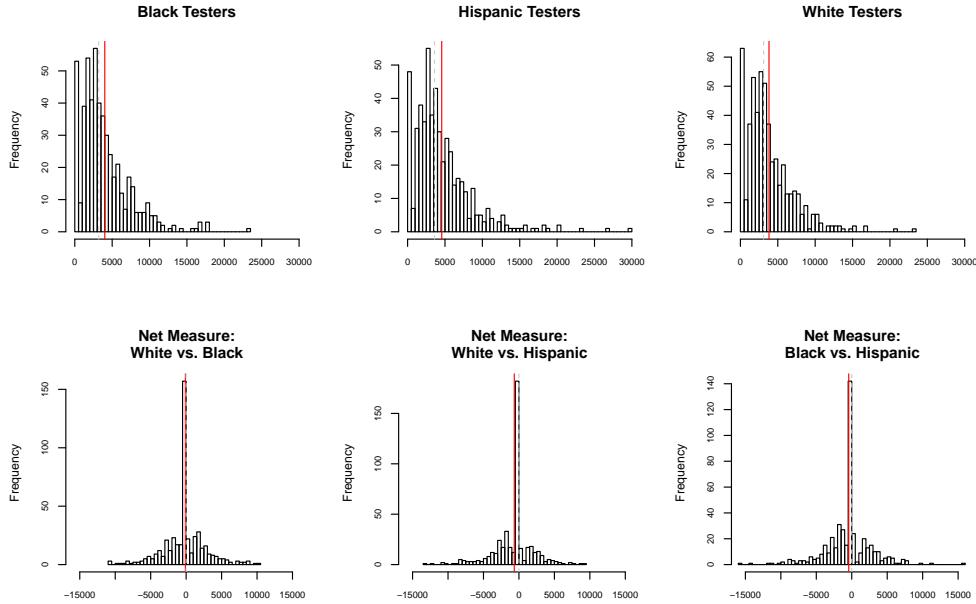


Figure 4.3: Distributions of Quoted Sum of Up-Front Rent, Security Deposit, Holding Fees, and Broker Fees by Tester Race and by Group Differences. The first row presents the distribution of quoted fees by tester race, using valid case-tester-level reports. The second row presents the distribution of differences in quoted fees by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

an additional \$500 deposit was required up-front, which would then go toward the first month's rent; neither the Hispanic nor white tester was required to pay this additional \$500 deposit up-front.

- Another tester team viewed a one-bedroom apartment in Astoria advertised at \$2,000 per month. All three testers were shown the advertised unit and were quoted the advertised monthly rental price. None of the testers were told that they needed to pay any months of rent up front. Both the black and white testers were told that a \$2,000 security deposit was required up front; this was not required for the Hispanic tester. However, the Hispanic and white testers were told they would need to pay \$500 “to take the unit off the market” as a holding fee.
- For a \$1,950/month one-bedroom apartment in Sunnyside, all three testers were quoted the advertised monthly rent and were told that one month's rent and a security deposit equal to one month's rent would be required. Only the Hispanic tester was told that an additional \$200 deposit was required to secure the unit. Of the three testers, only the black and Hispanic testers were told that a broker fee equal to one months' rent would be required.

- For a listed 3-bedroom apartment in Brooklyn’s Williamsburg neighborhood advertised at \$2,300/month, the white and Hispanic testers were quoted the advertised monthly rental price in person but the black tester was quoted a higher rental price of \$2,490 per month. According to the white tester’s field notes, the \$2,300/month price was a “special teaser rent” after a discount was applied to the “true” rent of \$2,490. Both the black and white testers were told that \$4,980 was needed up-front, plus a security deposit of \$2,490, or \$7,470 total. The Hispanic tester was told that only a security deposit of \$4,600 and no months of rent was required up-front. Broker fees also varied by tester. All three testers were told that the broker fee was equal to a quarter of one month’s rent. Because the monthly quoted rental price differed by testers, the white and Hispanic testers were told that a \$575 broker fee was required, whereas the black tester was told that a \$622.50 broker fee was required.
- A one-bedroom Bushwick apartment listed at \$2,000/month was pursued by a tester team. All three testers were quoted the advertised monthly rental price during the appointment. All testers were told that a \$2,000 security deposit was required, and that an additional \$1,000 fee was needed to hold the apartment. The Hispanic tester was told that the \$1,000 fee was needed to hold the apartment for three days, whereas the white tester was told that the fee would allow them to hold the apartment for two days. No additional information was provided to the black tester about the terms associated with the holding fee. In addition, only the black tester was required to pay two additional months of rent, or \$4,000, up front.
- For a two-bedroom apartment in the Financial District listed at \$1,900/month, the Hispanic and white testers were told that a broker fee equal to 15% of the annual rent would be required, whereas the black tester was told that the broker fee was equal to 12% of the annual rent.
- *Application and credit/background check fees:* On average, white testers were quoted \$33.54 in application and credit or background check fees, compared to \$32.73 for Hispanic testers and \$28.67 for black testers. Among paired tests with valid data, white testers were quoted on average \$3.71 more in these fees than black testers but \$0.12 less on average than Hispanic testers. Hispanic testers were quoted on average \$4.74 more in application and credit or background check fees than black testers in valid paired tests.

Figure 4.4 summarizes the distribution of the quoted sum of application and credit or background check fees by tester race using data from valid tester reports, and by group differences using data from valid paired tests. Among valid paired tests, differences in this quoted sum range from -\$400 to \$1,950 for white versus black testers, between -\$1,500 to \$1,850

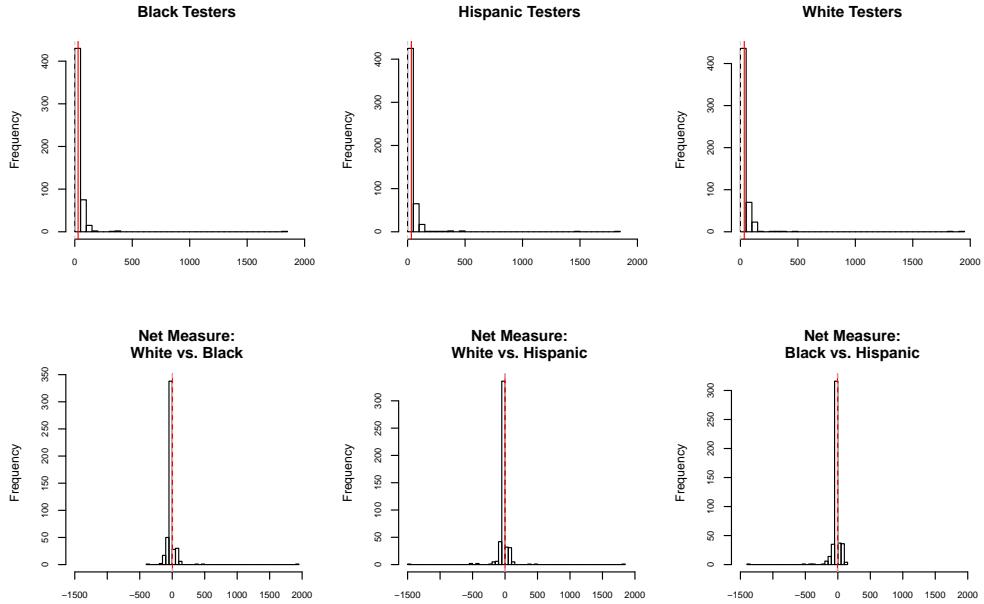


Figure 4.4: Distributions of Quoted Application and Credit/Background Check Fees by Tester Race and by Group Differences. The first row presents the distribution of quoted fees by tester race, using valid case-tester-level reports. The second row presents the distribution of differences in quoted fees by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

for white versus Hispanic testers, and between -\$1,400 and \$150 for black versus Hispanic testers.

- *Administrative fees and other net costs:* On average white testers were quoted \$7.99 in administrative fees and other net costs, compared to \$6.46 for black testers and \$4.61 for Hispanic testers. Among paired tests with valid data, white testers were quoted on average \$1.86 more in administrative fees and other net costs than black testers and on average \$3.00 more than Hispanic testers. Black testers were quoted on average \$3.48 more than Hispanic testers in valid paired tests.

Figure 4.4 summarizes the distribution of the quoted sum of application fees and other net costs by tester race using data from valid tester reports, and by group differences using data from valid paired tests. Among valid paired tests, differences in this quoted sum range from -\$600 to \$1,950 for white versus black testers, between -\$700 to \$2000 for white versus Hispanic testers, and between -\$450 and \$2000 for black versus Hispanic testers. We find that both differentials in the quoted sum of application fees and in other net costs (such as move-in/move-out fees and additional discounts) are fairly rare in the sample we examine.

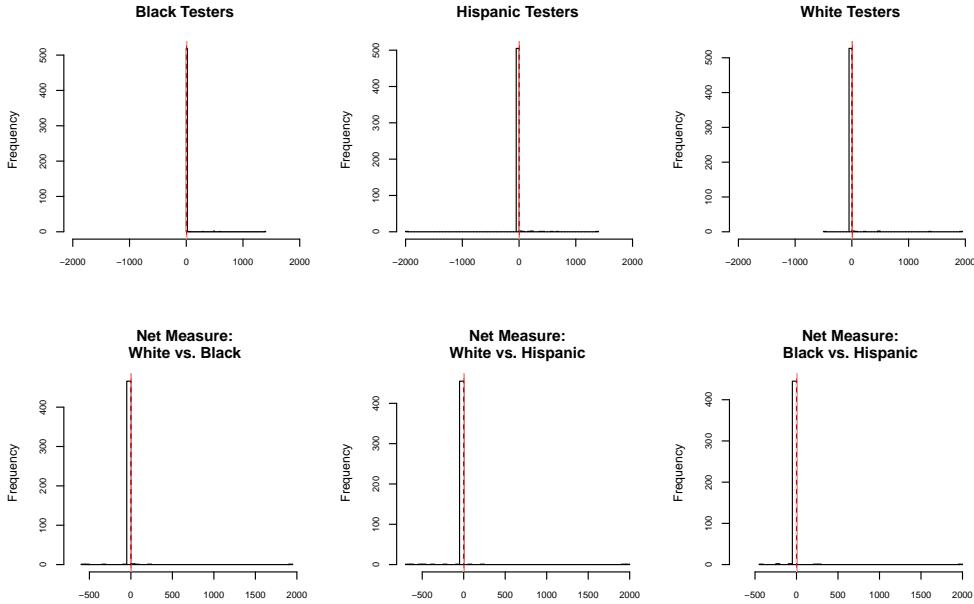


Figure 4.5: Distributions of Quoted Administrative Fees and Other Net Costs by Tester Race and by Group Differences. The first row presents the distribution of quoted fees by tester race, using valid case-tester-level reports. The second row presents the distribution of differences in quoted fees by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

4.3.7 Differences in Willingness to Negotiate on Fees

Finally, we examine whether testers confronted differences in the willingness of landlords and brokers to negotiate on up-front costs and fees. Testers were instructed not to ask about the possibility of negotiation but rather allowed the individuals with whom they interacted to volunteer the possibility of negotiating fees and up-front costs. We measured whether, for a given case, a tester was offered the possibility of negotiating down fees.

Landlords and brokers were most willing to negotiate on fees with white testers than black and Hispanic testers. On average, we observe that black testers reported being offered the possibility of negotiating fees 1.2% of the time, as compared to 2.4% of the time for Hispanic testers and 3.9% of the time for white testers. Among paired tests with valid data, white testers received negotiation offers at a rate 2.3 percentage points higher than the rate for black testers and 1.8 percentage points higher than the rate for Hispanic testers. Black testers were offered the possibility of negotiation at a rate 0.5 percentage points lower than the rate for Hispanic testers in valid paired tests.

5 Impacts of Messaging on Discrimination

This section summarizes the results from the experimental analysis assessing the impacts of government messaging strategies on discrimination incidence. The analyses we conduct are as follows:

- **Effects of messaging on discrimination:** First, we estimate the average effect of *assignment* to a messaging condition on discrimination rates against blacks and Hispanics. We estimate this quantity by calculating the mean outcome among the treatment group minus the mean outcome among the control group. These estimates address the policy question of how different city actions affect discrimination rates.
- **Effects of message content on discrimination:** Second, we estimate the average effect of different *message content* (i.e., message receipt) on discrimination rates among landlords and brokers who actually receive their assigned message. We estimate this quantity by calculating the mean outcome among the treatment group minus the mean outcome among the control group, scaled by the share of treated landlord/broker subjects who receive their assigned treatment. These estimates address the question of how the content of different appeals sent by the city affect discrimination rates.
- **Effects of messaging conditional on putative low employment stability signaling:** Third, we are interested in how the marginal effect of messaging on discrimination might vary conditional on different putative signals of high versus low employment stability. The idea here is to assess whether reductions in discrimination incidence attributable to government messaging might vary depending on the perceived financial stability of the housing seeker.

For all quantities of interest, we estimate the statistical uncertainty around each estimate. The key statistic we report is the *p*-value. For the treatment effect estimates, the *p*-value is interpreted as the probability of obtaining an effect at least as large (in absolute value) as the one observed in the actual experiment for the punitive-monitoring (monitoring-control or punitive-control) comparisons. By convention, estimated effects are statistically significant if the *p*-value is less than 0.05. The following sections presents the results from each of these analyses in turn.

Outcome Measure of Net Discrimination	I. Monitoring vs. Control				II. Punitive vs. Control				III. Punitive vs. Monitoring			
	Effect	SE	t	P-Value	Effect	SE	t	P-Value	Effect	SE	t	P-Value
A. White vs. Black												
Landlord/broker honored appointment	0.015	0.009	1.661	(0.951)	0.008	0.008	0.971	(0.834)	-0.007	0.01	-0.697	(0.486)
Index measure of favorable in-person interactions	0.001	0.059	0.017	(0.507)	0.01	0.052	0.194	(0.577)	0.009	0.066	0.138	(0.89)
Received post-visit callback	-0.01	0.048	-0.21	(0.417)	0.02	0.043	0.454	(0.675)	0.03	0.054	0.554	(0.58)
Received post-visit offer for unit	-0.054	0.039	-1.383	(0.084)	-0.025	0.035	-0.71	(0.239)	0.029	0.044	0.66	(0.51)
B. White vs. Hispanic												
Landlord/broker honored appointment	0.006	0.008	0.697	(0.757)	0.004	0.008	0.487	(0.687)	-0.002	0.01	-0.228	(0.82)
Index measure of favorable in-person interactions	-0.079	0.062	-1.278	(0.101)	0.078	0.054	1.426	(0.923)	0.157	0.069	2.287	(0.023)
Received post-visit callback	-0.028	0.045	-0.612	(0.271)	-0.056	0.041	-1.37	(0.086)	-0.028	0.051	-0.561	(0.575)
Received post-visit offer for unit	-0.062	0.038	-1.625	(0.052)	-0.078	0.034	-2.276	(0.012)	-0.016	0.042	-0.388	(0.698)
C. Black vs. Hispanic												
Landlord/broker honored appointment	-0.009	0.008	-1.16	(0.123)	-0.004	0.007	-0.589	(0.278)	0.005	0.009	0.559	(0.576)
Index measure of favorable in-person interactions	-0.097	0.057	-1.705	(0.044)	0.059	0.051	1.167	(0.878)	0.156	0.064	2.431	(0.015)
Received post-visit callback	-0.018	0.045	-0.395	(0.346)	-0.076	0.04	-1.874	(0.031)	-0.058	0.05	-1.161	(0.246)
Received post-visit offer for unit	-0.008	0.035	-0.221	(0.413)	-0.053	0.032	-1.666	(0.048)	-0.045	0.039	-1.149	(0.251)

Table 5.1: Estimated Effects of Messaging on Net Discrimination Levels, by Majority-Minority Group and Treatment Comparisons. Cells contain estimates of the Intent-to-Treat effect estimated using ordinary least squares regression models with inverse probability weighting to account for the randomization procedure. Models are estimated without covariate adjustment. P-values correspond to a one-sided test of the null hypothesis of equality of means for the monitoring-control and punitive-control comparisons, and to a two-sided test of the null hypothesis of equality of means for the punitive-monitoring comparison.

5.1 Estimated Effects of Messaging on Discrimination Incidence

First, we examine estimates of the average effect of different governmental messaging signals on the incidence of racial discrimination in the rental market. We find that sending punitive messages reduces net discrimination against Hispanic testers (versus both black and white testers) in two critical outcomes in the later stages of the housing search: receiving a callback and receiving an offer for a unit. These findings are statistically significant. Sending monitoring messaging consistently reduces discrimination against Hispanics (versus both blacks and whites) in objective post-visit outcomes and subjective perceptions of the favorability of interactions during appointments, but of these effects, only the effect of monitoring messaging on net discrimination against Hispanics (versus blacks) in subjective perceptions is statistically significant. Table 5.1 summarizes the main results.¹⁶

5.1.1 Monitoring versus Control

- *Net discrimination against blacks (versus whites):* When compared to a pure control condition, sending a monitoring message has mixed effects on net discrimination against blacks (versus whites) across the four outcome measures since the direction of the estimated effects is inconsistent. Whereas assignment to a monitoring signal increases net discrimination against blacks in honoring scheduled appointments (estimated change in net discrimination = 0.015; $p=0.951$) and net discrimination against blacks in testers' subjective perception of favorable in-person interactions during the appointment (0.001; $p=0.507$), the monitoring signal decreases net discrimination against blacks in receiving a callback (-0.01; $p=0.417$) and in receiving an offer for the unit (-0.054; $p=0.084$). None of these estimated effects are statistically significant at the 0.05 level; only the effect of monitoring on net discrimination in receiving a post-visit offer is statistically significant but at the 0.1 level.¹⁷
- *Net discrimination against Hispanics (versus whites):* When compared to a pure control condition, sending a monitoring message has mixed effects on net discrimination against Hispanics (versus whites) across the four outcome measures. Sending the monitoring message increases net discrimination against Hispanics in honoring scheduled appointments (0.006, $p=0.757$) but decreases net discrimination against Hispanics in the subjective index measure

¹⁶For the sake of transparency and to present findings closest to the data given the research design, we focus on treatment effect estimates generated using ordinary least squares with inverse probability weights to account for each unit's probability of assignment to the observed treatment assignment. See Appendix D for additional details on the estimation and inference procedures used.

¹⁷This means that there is less than a 10% probability of obtaining an effect of monitoring (versus control) on net discrimination in honoring the scheduled appointment at least as large in absolute value as the one observed in the experiment.

of favorable interactions (-0.079, $p=0.101$) and in the objective measures of receiving a callback (-0.028, $p=0.271$) and receiving an offer for the unit (-0.062, $p=0.052$). None of these estimated effects are statistically significant at the 0.05 level. However, the estimated effect of monitoring messaging on net discrimination against Hispanics (versus whites) in receiving a post-visit offer approaches significance since the p -value is just above the significance threshold ($p=0.052$).

- *Net discrimination against Hispanics (versus blacks):* When compared to a pure control condition, sending a monitoring message consistently decreases net discrimination against Hispanics (versus blacks) on all four outcome measures of discrimination: whether the landlord honored the scheduled appointment (-0.009; $p = 0.123$), the tester's subjective perception of favorable in-person interactions during the appointment stage (-0.097; $p=0.044$), whether the tester received a post-visit callback (-0.018; $p=0.346$), and whether the tester was offered the unit (-0.008; $p=0.413$). Of these estimated effects, the effect of monitoring on net discrimination in testers' subjective perceptions of favorable interactions during the appointment is statistically significant at the 0.05 level ($p=0.044$).

5.1.2 Punitive versus Control

- *Net discrimination against blacks (versus whites):* When compared to a pure control condition, sending a punitive message has mixed effects on net discrimination against blacks (versus whites) across the four outcome measures. Sending the punitive message increases net discrimination against blacks in whether landlords honor their scheduled appointment (0.008, $p=0.834$), in testers' subjective perceptions of favorable interactions during appointments (0.01, $p=0.577$), and in receiving a post-visit follow-up callback from the landlord (0.02; $p=0.675$). However, the punitive message decreases net discrimination against blacks in receiving a post-visit offer (-0.025; $p=0.239$). None of these estimated effects are statistically significant at the 0.05 level.
- *Net discrimination against Hispanics (versus whites):* When compared to a pure control condition, sending a punitive message has mixed effects on net discrimination against Hispanics (versus whites) across the four outcome measures. Punitive messaging increases net discrimination against Hispanics (versus whites) in whether landlords honor their scheduled appointment (0.004, $p=0.687$) and in the testers' subjective perceptions of favorable interactions during appointments (0.078; $p=0.923$). However, punitive messaging decreases net discrimination against Hispanics in receiving a post-visit callback (-0.056; $p=0.086$) and in receiving a post-visit offer for the unit (-0.078; $p=0.012$). The effect of punitive messaging on net discrimination in receiving a post-visit offer is statistically significant at the 0.05 level

($p=0.012$); the effect of punitive messaging on net discrimination in receiving a post-visit callback is statistically significant at the 0.1 level ($p=0.086$).

- *Net discrimination against Hispanics (versus blacks)*: When compared to a pure control condition, sending a punitive message has mixed effects on net discrimination against Hispanics (versus blacks) across the four outcome measures. Punitive messaging decreases net discrimination against Hispanics in whether landlords honor scheduled appointments (-0.004, $p=0.278$), in receiving callbacks following the appointment (-0.076; $p=0.031$), and in receiving offers for units (-0.053, $p=0.048$). In contrast, punitive messaging increases net discrimination against Hispanics in testers' subjective perceptions of favorable treatment during the appointment (0.059, $p=0.878$). Of these estimated treatment effects, two are statistically significant at the 0.05 level: the effects of punitive messaging on the post-visit outcomes of receiving a callback and on receiving an offer ($p=0.031$ and 0.048, respectively).

5.1.3 Punitive versus Monitoring

- *Net discrimination against blacks (versus whites)*: When compared to a monitoring condition, sending a punitive message has mixed effects on net discrimination against blacks (versus whites) across the four outcome measures. Relative to a monitoring signal, punitive messaging decreases net discrimination against blacks in whether landlords honor scheduled appointments (-0.007, $p=0.486$). However, punitive messaging also increases net discrimination against blacks in testers' subjective perceptions of favorable treatment during appointments (0.009, $p=0.89$), in receiving a post-visit callback (0.03, $p=0.58$), and in receiving a post-visit offer for the unit (0.029, $p=0.51$). None of these estimated effects are statistically significant at the 0.05 level.
- *Net discrimination against Hispanics (versus whites)*: When compared to monitoring condition, sending a punitive message has mixed effects on net discrimination against Hispanics (versus whites) across the four outcome measures. Punitive messaging decreases net discrimination against Hispanics (versus whites) in whether landlords honor scheduled appointments (-0.002, $p=0.82$), in receiving callbacks (-0.028, $p=0.575$), and in receiving offers for units (-0.016, $p=0.698$). However, punitive messaging increases net discrimination against Hispanics in testers' subjective perceptions of favorable interactions during the appointment (0.157, $p=0.023$); this effect is statistically significant at the 0.05 level.
- *Net discrimination against Hispanics (versus blacks)*: When compared to a monitoring condition, sending a punitive message has mixed effects on net discrimination against Hispanics (versus blacks) across the four outcome measures. Relative to a monitoring signal, punitive

messaging increases net discrimination against Hispanics (versus blacks) in whether landlords honor scheduled appointments (0.005, $p=0.576$) and in testers' subjective perceptions of favorable interactions during the appointment (0.156, $p=0.015$); the latter effect is statistically significant at the 0.05 level. Punitive messaging also decreases net discrimination against Hispanics in receiving a post-visit callback (-0.058, $p=0.248$) and in receiving a post-visit offer for the unit (-0.045, $p=0.251$).

5.2 Estimated Effects of Message Content on Discrimination Incidence

Next, we present estimates of the average effects of message content on discrimination incidence. This question is different than the prior question of the policy effects of sending messages on discrimination incidence, because they concern the effects of receiving different message content on discriminatory behavior.

The content of the treatment message assigned to an individual is not necessarily the same as the content received. Individuals may not pick up the phone or may hang up on the caller and only receive a subset of the assigned messages. It is therefore imperative to track differences in which messages individuals are sent and the content that is successfully delivered to each individual. The study protocol allows us to do this. Since treatments were all delivered by phone, the Treatment Administrator recorded this information for every call sent. Table 5.2 summarizes the distribution of treatment noncompliance in the experimental sample. Among the 279 landlords and brokers assigned to the control condition, all 279 (100%) received no message. Of the 174 landlords and brokers assigned to the monitoring condition, 143 (82.2%) successfully received the full monitoring message, but 31 (17.8%) could not be reached and effectively received the control condition. Among the 200 landlords and brokers assigned to the punitive message, 145 (72.5%) successfully received the full punitive message; 38 (19%) never picked up the phone and effectively received the control condition; and 17 (8.5%) picked up the phone and received the contents of the monitoring message before hanging up. For these 17 landlords and brokers, the punitive appeal was never successfully delivered.

Treatment Assigned	Treatment Receipt		
	Control	Monitoring	Punitive
Control	279	0	0
Monitoring	31	143	0
Punitive	38	17	145

Table 5.2: Observed Noncompliance with Treatment Assignment. Table entries are the number of subjects corresponding to the observed treatment assigned and received.

Given this information, we then define the quantities of interest as the average effects of different message content for different complier types. By this, we mean a different effect for different subpopulations defined by how they would receive treatments for each possible treatment that could be assigned to them. We conceptualize the individuals in our study as being comprised of three types:

- **Full Compliers** are individuals who receive their full assigned message, regardless of which message they are assigned to receive. This means if they are assigned to the control condition and are meant to receive no message, they do not receive a message. If they are assigned to the monitoring treatment, they receive the full monitoring message and nothing more. If they are assigned to the punitive treatment, they receive the full punitive message.
- **Partial Compliers** are individuals who receive no message if assigned to the control condition. If assigned to the monitoring treatment, they would successfully receive the full monitoring message. However, if assigned to the punitive message, they only receive a portion of this message – the monitoring signal. This occurred when the City staffer tasked with delivering treatment calls and the targeted landlord or broker hung up before the punitive appeal was read.
- **Never-Takers** are individuals who would receive the control condition regardless of which treatment they were assigned.

We then formally define four effects of interest:

- The effect of the monitoring message versus no message (control) for Partial Compliers,
- The effect of the monitoring message versus no message (control) for Full Compliers,
- The effect of the punitive message versus no message (control) for Full Compliers, and
- The additional effect of the punitive appeal on top of receiving a monitoring message for Full Compliers.

A summary of the main results from this analysis are shown in Table 5.3.¹⁸ Each horizontal panel summarizes the estimated effects for each of the aforementioned effects by net discrimination outcome measure. The estimate of the average treatment effect of interest is presented in the column labeled “Estimate” and is accompanied by a 95% credible interval which, if bounded away from zero, means that the estimated mean effect is statistically distinguishable from zero.

¹⁸To estimate these effects we used a Bayesian inference method explicated by Long, Little, and Lin (2010); for details see Appendix D. A full set of estimation results are available in Appendix D.

Effect	I. White vs. Black					II. White vs. Hispanic					III. Black vs. Hispanic				
	Estimate	95% Credible Interval		\hat{R}	Eff. N	Estimate	95% Credible Interval		\hat{R}	Eff. N	Estimate	95% Credible Interval		\hat{R}	Eff. N
A. Effect of Monitoring Message vs. Control for Partial Compliers															
Landlord/broker honored appointment	-0.023	-0.207	0.151	1.001	5000	-0.022	-0.204	0.147	1.001	5000	-0.022	-0.208	0.154	1.001	5000
Index of favorable in-person interactions	-0.022	-0.205	0.154	1.001	5000	-0.025	-0.209	0.147	1.001	5000	-0.023	-0.2	0.146	1.001	5000
Received post-visit callback	0.026	-0.171	0.221	1.001	5000	-0.006	-0.205	0.187	1.002	3000	-0.027	-0.219	0.159	1.001	4000
Received post-visit offer	-0.011	-0.204	0.176	1.001	5000	-0.045	-0.233	0.142	1.001	4000	-0.025	-0.215	0.159	1.002	2300
B. Effect of Monitoring Message vs. Control for Full Compliers															
Landlord/broker honored appointment	0.053	-0.014	0.132	1.001	5000	0.052	-0.018	0.128	1.001	5000	0.05	-0.016	0.127	1.001	5000
Index of favorable in-person interactions	0.047	-0.021	0.121	1.002	1800	0.047	-0.021	0.121	1.001	4900	0.047	-0.019	0.122	1.001	5000
Received post-visit callback	0.043	-0.06	0.15	1.001	3300	0.045	-0.059	0.153	1.001	5000	0.029	-0.065	0.125	1.001	5000
Received post-visit offer	0.028	-0.06	0.123	1.001	5000	0.033	-0.058	0.13	1.001	5000	0.007	-0.074	0.092	1.001	5000
C. Effect of Punitive Message vs. Control for Full Compliers															
Landlord/broker honored appointment	0.006	-0.051	0.062	1.002	1800	0.006	-0.047	0.062	1.001	3400	0.006	-0.051	0.062	1.001	4000
Index of favorable in-person interactions	0.005	-0.05	0.061	1.001	5000	0.006	-0.047	0.063	1.001	5000	0.006	-0.049	0.062	1.001	5000
Received post-visit callback	-0.029	-0.112	0.053	1.003	1400	-0.004	-0.088	0.078	1.001	5000	-0.014	-0.088	0.064	1.001	5000
Received post-visit offer	-0.015	-0.09	0.057	1.001	4000	0.013	-0.063	0.087	1.001	5000	0.011	-0.066	0.085	1.002	2200
D. Effect of Punitive Message vs. Monitoring Message for Full Compliers															
Landlord/broker honored appointment	-0.048	-0.127	0.022	1.002	2300	-0.046	-0.124	0.024	1.001	5000	-0.044	-0.125	0.026	1.001	4500
Index of favorable in-person interactions	-0.042	-0.117	0.025	1.002	3000	-0.041	-0.116	0.027	1.001	3300	-0.041	-0.116	0.029	1.001	5000
Received post-visit callback	-0.072	-0.175	0.028	1.001	5000	-0.049	-0.155	0.047	1.001	5000	-0.042	-0.139	0.05	1.001	5000
Received post-visit offer	-0.043	-0.138	0.047	1.002	2500	-0.02	-0.118	0.073	1.001	5000	0.003	-0.084	0.087	1.001	5000

Table 5.3: Estimated Average Effects of Message Content on Net Discrimination Outcomes by Complier Type.

5.2.1 Monitoring Message versus Control for Partial Compliers

Panel A in Table 5.3 summarizes the estimated average effect of receiving a monitoring message from the City (versus no message) on net discrimination levels for Partial Compliers (i.e., individuals who would receive *at most* the contents of a monitoring message if assigned to a monitoring or punitive message, and who would comply with the control condition if assigned to that). The monitoring message consistently reduces net discrimination levels against Hispanics (versus both blacks and whites) across all four outcome measures but the results are not statistically significant since the 95% credible interval includes zero. The effects of receiving the monitoring message on net discrimination against blacks (versus whites) are mixed and lack statistical significance.

5.2.2 Monitoring Message versus Control for Full Compliers

Panel B in Table 5.3 summarizes the estimated average effect of receiving a monitoring message from the City (versus no message) on net discrimination levels for Full Compliers (i.e., individuals who would receive the full contents of any treatment message they are assigned). Receiving a monitoring message seems to increase net discrimination against each minority group as suggested by the consistently positive estimates of the mean effects. However, none of these estimates are statistically distinguishable from zero because the 95% credible intervals bracket zero.

5.2.3 Punitive Message versus Control for Full Compliers

Panel C in Table 5.3 summarizes the estimated average effect of receiving a punitive message from the City (versus no message) on net discrimination levels for Full Compliers (i.e., individuals who would receive the full contents of any treatment message they are assigned). Receiving a punitive message seems to have mixed effects on net discrimination levels against blacks (versus whites) and against Hispanics (against both blacks and whites). None of these estimates are statistically distinguishable from zero because the 95% credible intervals bracket zero.

5.2.4 Punitive Message versus Monitoring Message for Full Compliers

Panel D in Table 5.3 summarizes the estimated average effect of receiving a punitive appeal from the City (versus just receiving a monitoring signal from the City) on net discrimination levels for Full Compliers (i.e., individuals who would receive the full contents of any treatment message they are assigned). Receiving the additional punitive appeal seems to consistently reduce net discrimination against blacks (versus whites) and against Hispanics (versus whites) on all four outcome measures. The punitive appeal also seems to reduce net discrimination against Hispanics (versus blacks) on all outcome measures except receiving an offer for the unit. None of these estimates

are statistically distinguishable from zero because the 95% credible intervals bracket zero, but the magnitudes of these estimated mean effects seem to be consistently large than the effects of monitoring messages (versus control) on net discrimination.

5.2.5 Estimated Proportions of Compliance Types in the Experimental Sample

Lastly, we estimate the proportion of compliance types in the experimental sample to place these estimated effects in context. We estimate that approximately 30% of the sample is comprised of Never-Takers — or people who would always receive no message regardless of which message the City sends; approximately 7.7% of the sample is comprised of Partial Compliers — or people who would pick up the phone if assigned to call, receive the monitoring message at most, but would hang up if additional content were to be provided; and approximately 62.4% of the sample is comprised of Full Compliers — or people who would receive the full treatment messages the City assigns and sends them.

5.3 Effects of Messaging Conditional on Putative Low Employment Stability Signals

We now turn to estimates of the marginal effects of messaging strategies conditional on whether the tester team provided a putative high or low employment stability signal during the appointment. This analysis examines two post-treatment outcome measures: net discrimination in receiving a post-visit callback, and in receiving a post-visit offer for the unit.¹⁹

We assess whether the marginal effect of a given treatment on discrimination is higher or lower among landlords who received a putative low employment stability signal from testers as compared to landlords who received a putative high employment stability signal from testers. If the estimated marginal treatment effect among landlords receiving a low employment stability signal is *greater* than the estimated marginal treatment effect among landlords receiving a high employment stability signal, this means the treatment message of interest is *less* effective than the comparison message at reducing discrimination in market interactions where potential tenants seeking housing signal greater financial instability than not. Conversely, if the estimated marginal treatment effect among landlords receiving a low employment stability signal is *less* than the estimated marginal treatment effect among landlords receiving a high employment stability signal, this means the treatment message of interest is *more* effective than the comparison message at reducing discrimination in market interactions where potential tenants seeking housing signal greater financial instability

¹⁹Net discrimination in whether the landlord honored the scheduled appointment is not examined as an outcome measure because it occurs prior to testers delivering the scripted signal. Similarly because the measure of net discrimination in testers' perceptions of favorable interactions during the appointment is not strictly causally prior to the scripted signal, we do not examine this measure as an outcome of interest for this analysis.

than not. Second, we also assess whether the difference in the marginal treatment effects between the low and high signal groups is statistically significant.

Net Discrimination Outcome Measure	Monitoring vs. Control	Punitive vs. Control	Punitive vs. Monitoring
A. White vs. Black			
Received Callback	-	-	-
Received Offer	+	+	+
B. White vs. Hispanic			
Received Callback	+	-	-
Received Offer	+	+	+
C. Black vs. Hispanic			
Received Callback	+	+	-
Received Offer	-	-	+

Table 5.4: Summary of Differences in Marginal Treatment Effects between Low versus High Putative Employment Stability Signal Groups. A ‘+’ symbol denotes that the marginal treatment effect on the net measure is larger in number in the low employment stability group than in the high employment stability group, which means that the treatment is *less* effective at reducing net discrimination in the low employment stability setting. Conversely, a ‘-’ symbol denotes that the marginal treatment of effect on the net measure is smaller in number in the low employment stability group than in the high employment stability group, which means that the treatment is *more* effective at reducing net discrimination in the low employment stability setting.

Table 5.4 summarizes the sign of the difference between estimated marginal effects between the low versus high employment stability groups for each treatment-comparison pairing and for each outcome variable. (Full estimation results are shown in Appendix D). A ‘+’ symbol denotes that the marginal treatment of effect on the net measure is larger in number in the low employment stability group than in the high employment stability group, which means that the treatment is *less* effective at reducing net discrimination in the low employment stability setting. Conversely, a ‘-’ symbol denotes that the marginal treatment of effect on the net measure is smaller in number in the low employment stability group than in the high employment stability group, which means that the treatment is *more* effective at reducing net discrimination in the low employment stability setting.

As shown in Table 5.4, we find that in general, messaging effects are not consistently magnified or attenuated when housing seekers provide a putative signal of low employment stability instead of high employment stability. However, the effects of both monitoring and punitive messaging on reducing net discrimination against Hispanics (versus both blacks and whites) is magnified when testers signal low employment stability as opposed to high employment stability. Apparent differences in marginal effects between low and high employment stability groups are not statistically significant at the 0.05 level.

6 Conclusions

This section summarizes the main findings from this study and suggests policy recommendations based on these findings.

6.1 Summary of Main Findings

This evaluation assessed baseline levels of racial discrimination against blacks and Hispanics (versus whites) in the New York City rental housing market, examined the effectiveness of government messaging strategies to reduce racial discrimination in rental housing, and explored variation in subtler forms of discriminatory behavior that occur in person during rental unit viewings between landlords and brokers on the one hand and rental housing seekers on the other.

- **Baseline levels of racial housing discrimination:** There is mixed evidence for discrimination in the earlier stages of the rental housing search process. White testers were more likely than black or Hispanic testers to report experiencing difficulty with qualifications to rent when speaking over the phone to schedule a potential visit with a landlord or broker. Hispanic testers were more likely to encounter skepticism or negative comments about their qualifications or negative comments about their rental qualifications than black testers. In later stages of the housing search process, Hispanic testers were less likely than white testers to receive callbacks or offers for units after the appointment.
- **Effects of government messaging on racial discrimination incidence:** Sending punitive appeals to comply with the city's Fair Housing Law reduced net discrimination levels against Hispanic testers (versus both black and white testers) in receiving callbacks and in receiving offers following the appointment. These findings are statistically significant. We also find suggestive evidence that receiving the content of punitive appeals from the city *on top of the contents of the monitoring message* consistently reduces net discrimination rates against black testers and against Hispanic testers (versus white testers). These findings are not statistically significant but are substantively significant.
- **Variation in forms of discriminatory behavior:** The study also inductively explored variation in forms of discriminatory behavior occurring during in-person interactions between landlords or brokers and testers, by tester race. Testers are differentially steered toward different housing units by race. There is considerable variation in the forms of steering, whether by price, housing stock quality, or the quoted average number of amenities included in the rental price. In addition, we documented complex disparities by tester race in the structure and amount of up-front costs and fees required to secure and move into units. Higher amounts

of up-front rent are quoted for Hispanic testers than for black and white testers. More Hispanic and black testers were required to pay holding fees and good-faith deposits than white testers. White testers were told to pay broker fees, application fees, and administrative and processing fees more frequently than Hispanic and black testers. Black testers were told to pay for and pass a credit check or background check more frequently than white and Hispanic testers. Lastly, we found that white testers are more likely to receive offers to negotiate down fees than black and Hispanic testers.

Policy Recommendations

This evaluation finds that racial discrimination in the New York City rental housing market persists, and that the city possesses the tools to counteract it. Based on the findings summarized above, we offer the following policy recommendations:

- *Disseminate information about discrimination in the rental housing market:* A long-term solution will require a sustained engagement from citizens, policymakers, as well as landlords and brokers themselves. Publicizing the evidence for discrimination will send a signal that the city is committed to solving a genuine problem.
- *Continue the use of matched audits to uncover discrimination:* Many forms of differential treatment that occur are not easily identified within single interactions between any given housing seeker and a landlord or broker. This is especially true during the appointment stage with respect to subtle forms of racial discrimination such as steering, differences in the quoted terms of rent, and in the willingness to negotiate such terms. The continued use of matched or paired audits is essential for monitoring discrimination levels in the rental market.
- *Employ punitive appeals encouraging compliance with fair housing law:* Reminding landlords and brokers to comply with the city's Fair Housing Law and informing them of the pecuniary costs of violating it is effective at reducing the incidence of racial discrimination in rental housing, particularly as it affects Hispanic apartment seekers. This study cannot address whether modes of communication besides telephone calls would be equally effective. A cost-benefit analysis combined with further evaluation is recommended. Additional research should also investigate whether messaging has long-term effects on reducing discrimination.

Appendices

A Additional Implementation and Field Procedures

A.1 Sampling

A.1.1 Additional Details about the Sampling Procedure

Vacant rental housing advertisements (which we call “cases”) are sampled using a stratified random sampling procedure with proportional allocation by New York City borough from a widely used online classified listings website, Craigslist.²⁰ We restrict borough-specific listings to the “All Apartments” category available when browsing by borough-specific sub-sections of Craigslist. Stratified sampling by borough is important to increase homogeneity in potential outcomes across subjects to increase statistical power. In addition we construct a sampling frame of vacant rental housing ads that we designate as ads that contain language that suggest possible discrimination; this is done to create an oversample of “likely discrimination” ads to increase statistical power.²¹ Since explicitly discriminatory ads are flagged and removed, searches for words and phrases that are explicitly discriminatory generally yield no hits. We therefore limit our search to identify ads containing words and phrases that implicitly suggest markers of racial prejudices signaled through class preferences, given the strong relationship between race and class in the United States.²² These ads are scraped from Craigslist and a random draw is sampled from this set prior to borough-stratified random sampling. These “likely discrimination” ads are excluded from the borough-stratified sampling frames so that they are not double-sampled.

Sampling occurs each day the study is implemented, which is limited to weekdays when the City is open for business. The sampling frame for each draw on a given day is the set of advertisements listed on Craigslist during that day up until the time of the draw, and any advertisements listed on previous business days during which the study was not conducted.²³ Sampling daily or near-daily ensures that vacant rental housing advertisements pursued by testers are recent ads that a real person looking for rental housing would likely pursue. Only ads containing landlord or broker telephone numbers are pursued by testers; the rest are discarded.²⁴ For sampled ads, copies of the

²⁰The proportions are 35% Manhattan, 30% Brooklyn, 20% Queens, 10% Bronx, and 5% Staten Island; these shares reflect the rough distribution of ads by borough on Craigslist as identified in the pilot study.

²¹We are aiming for the final sample to have 60% “regular” ads and 40% oversampled “likely discrimination” ads; each of these is treated as a block in which we randomize or further sub-block.

²²Search terms used are: “hip,” “up and coming,” “yuppie,” and “qualified.”

²³The study is implemented five business days every week; project staff do not work on City holidays.

²⁴From the pilot study we learned that response rates, the probability of reaching a landlord or broker, and the probability of scheduling an appointment to view a rental unit are significantly higher among ads containing phone numbers in comparison to ads that request replies by email.

original ads as they appear on Craigslist are saved.

A.1.2 “Scraping” Public Listings to Sample Available Units in the Housing Market

Both audit studies and field experiments in the housing market require sampling methods that are replicable and easy to implement. We were able to take advantage of two features of the study context: CCHR’s interest in publicly listed units and the fact that the vast majority of such listings can be found online. In particular, by regularly sampling Craigslist, we were able to assemble a representative set of listings covering the range of units available in the New York City rental housing market.

Before turning to best practices for future research and enforcement efforts, a short technical note. The Python programming environment proved to be well suited to the task of “scraping” Craigslist on a daily or almost-daily basis, sampling from the universe of listings, and saving the appropriate information to a secure location. Python can be installed on Windows-based PCs and comes included with Mac OS X, and our project managers were able to run the scripts for the most part without trouble. In particular, we highly recommend the BeautifulSoup screen-scraping library for its flexibility and straightforward implementation. Craigslist is famous for its plain, no-frills layout, and this was a major advantage when developing the script. However, there were several occasions when the layout subtly changed without warning, which caused numerous errors and hasty revisions. We would recommend building in robust error-catching routines in addition to notification systems (i.e. automated emails sent to the primary administrator) in order to minimize the risk of this kind of change. Having a backlog of available cases (perhaps one or two days old) also helped when there were technical issues impeding the usual sampling procedure.

Finally, while password-protected cloud-based storage services such as Dropbox are vital for data management in studies of this kind, we found that a system based on writing a large number of small files in embedded directory structures can greatly slow down the syncing process. One solution is to regularly move data files from completed cases to a secure location separate from the active operation of the scraping and sampling mechanism.

A.1.3 Best Practices for Future Approaches Using Online Listings

- *Random sampling.* Since the entire universe of relevant listings can be scraped (for example within a given time period, borough, or neighborhood), discretion at this stage can be eliminated. While hundreds of thousands of listings will be posted to Craigslist on a typical day in New York City, random samples can provide representative snapshots that are more manageable for a given purpose. However, it is important to note that scraping before the day is over may be necessary for studies or investigations requiring engagement with active

listings. This introduces the possibility of bias due to the types of listings that may be posted at given times of the day. The issue can be minimized by scraping sufficiently backwards in time.

- *Search terms.* Researchers or enforcement officials may be interested in pursuing suspicious listings by using search terms rather than an open-ended scrape. We attempted to incorporate a version of this procedure into an earlier version of the study but found no systematic differences in the sample. Since Craigslist actively pulls listings containing certain words, explicitly discriminatory language may be difficult to find. Moreover, the possibility of false positives using this kind of directed search is real.
- *Handling duplicates.* A major difficulty with Craigslist (at least in the New York City rental housing market) is that some brokers post bulk listings for duplicate or even nonexistent units. Our solution was to keep a running list of phone numbers and broker names from completed cases which the scraping program used to automatically remove listings from the sample. However, even this procedure was far from perfect as names and even numbers seemed to change frequently. Project managers had to devote a significant amount of time to handling this problem.

A.2 Manipulating Markers of Racial Identity

There have been several debates on how to clearly signal racial identity in field experimental studies about discrimination. Audit studies studying discrimination employ matched pair (or triple) audit designs where the trait or marker of auditors' group membership, which is used by the landlord or broker to assess the auditor and affects discrimination, is manipulated. All other characteristics of the testers that affect potential outcomes are fixed. In this section we review how we contribute to three debates on how to manipulate markers of racial identity in field experimental research.

A.2.1 The Racial Soundness of Names

Most field experimental work examining the effects of race on disparate treatment in employment and housing has manipulated the racial soundness of names (most notably Bertrand and Mullainathan 2004) since the racial identities signaled by testers' assumed names are an important signal of race (Fryer and Levitt 2004). Much of the use of names to signal racial identity has employed researcher discretion in choosing names to maximize the size of the expected effect of the racial signal on discrimination.

While this does not pose an internal validity problem, this is problematic with respect to external validity. When researchers employ discretion in choosing names to maximize the expected

effect size of names on discrimination, the estimand of interest is a quantity that is not generalizable to the population since the distribution of names does not match the distribution of names in the population.

To address this issue, we turned to a rare publicly available data set of real names tagged with racial and gender information.²⁵ The data set consists of the names of children between the third and tenth grades tested in the Colorado state assessment system from 2007 to 2010. Approximately 400,000 students are tested each year. We randomly sampled (with replacement) from this database, separately drawing four first and last names for each of six race-gender groups (white male, white female, black male, black female, Hispanic male, Hispanic female). We then paired together sampled first and last names within each group. We then had a list of representative names from the given population such that more common names were more likely to be drawn.

The Colorado student database population is distinct from the population of interest in this study, but we argue that it may be employed under the assumption that, since New York City attracts many people from across the country, it is reasonable to assume that names associated with particular regions of the United States will be encountered in New York.

In general, there was a final concern that a name signaling a particular racial identity also signals a particular ethnic identity. It would compromise the study if a tester was assigned a racial-sounding name that is incongruent with the tester's *actual* ethnic background, which landlords and brokers may be able to detect from the tester's physical attributes. Thus, if an obvious incongruence was detected, the name was discarded and another name from the list was chosen and assigned to the tester.

A.2.2 Linguistic and Class-Correlated Signals of Race

As Pager and Shepherd (2008) note in their review of the housing discrimination literature, “research using telephone audits further points to a gender and class dimension of racial discrimination in which black women and/or blacks who speak in a manner associated with a lower-class upbringing suffer greater discrimination than black men and/or those signaling a middle-class upbringing (Massey & Lundy 2001, Purnell et al. 1999)” (189).

To maximize the probability that minority testers are able to make an appointment to view a housing unit and to control for between-tester variation in class signaled through race and linguistic patterns, we account for linguistic and verbal markers of racial background they demonstrate during the phone conversation they have when replying to advertisements. All testers hired for the study are able to speak in a manner associated with a middle-class upbringing so as to not prime extreme class associations that drive racial perceptions.

²⁵The database of names is contained in a package for the R statistical programming language, **randomNames**, written by Damian W. Betebenner (2012).

A.3 Procedures for Screening and Hiring Testers to Pose as Interested Housing Seekers

Matched teams of three testers – one white, one black, and one Hispanic – are assigned vacant rental housing ads sampled from Craigslist to pursue. The effective composition and matching of testers to conduct in-person audits is therefore a major concern. The project seeks to compose a final team of 24 testers (or 24 FTE equivalents) with equal shares of testers for each race by sex combination. The city implements the following procedure to ensure the quality of testers employed in the study.²⁶

Successful applicants are subject to two lengthy interviews. In the first round interview, conducted via a video chat client (e.g. Skype or Google Video Chat), applicants are required to articulate their interest in the study to assess overall fit; articulate concrete work experiences that demonstrate experience working in groups and working individually on detailed tasks; and demonstrate familiarity with multiple neighborhoods across New York City’s five boroughs.

The second round interviews are conducted in-person. Applicants are required to participate in four simulated landlord/broker-tester interactions in which they take on the role of both the landlord/broker and tester given real ads pulled from Craigslist. Those playing the part of the tester are given an assumed biography and are evaluated on their ability to convincingly act out the part of an interested renter with that biography.

Testers are also asked complex questions for which they know little but that they are likely to encounter in the field, including: requests to elaborate reasons for moving to a particular neighborhood given one’s current neighborhood of residence; elaborations on what one does at work; follow-up questions commonly asked by landlords and brokers about whether one is being truthful about one’s income and source of income; detailed questions about “what’s going on” in one’s assumed neighborhood of residence for which a tester may actually know little to nothing. This test is done to see how adeptly applicants can ad lib without falling “out of character” or compromising the audit.

Finally, applicants are required to recall interactions from a simulated landlord/broker-tester interaction and quickly produce a set of detailed field notes in 10 minutes. This exercise is used to evaluate testers’ ability and capacity to conduct participant-observation research and record detailed observations about verbal interactions, non-verbal behavior, and contextual information about social interactions. Lastly the city assesses applicants’ attention to detail and their ability to use online data entry interfaces by observing how successfully they follow nuanced application instructions and interview scheduling instructions.

Each tester hired for the study is required to complete a standard training and a training period.

²⁶Methods used by Pager, Western, and Bonikowski (2009) serve as a benchmark.

Ongoing spot checks for quality control by the Project Manager and quality control checks of the data collected will be regularly conducted by the research team.

B Survey Instruments

This appendix includes copies of the following data collection instruments used by the testers:

- Form A: Pre-Visit Call Log
- Form B: Pre-Visit Interactions Log
- Form C: Field Visit Survey
- Form D: Post-Visit/Case Closure Log

In addition to these four forms, testers were also able to submit other open-ended qualitative field notes about their interactions for each audit conducted. Testers used hardcopy paper forms, then entered the data into an electronic form that compiled the data across testers and audits into a machine-readable database format.

PRE-VISIT CALL LOG

FORM A

ID1 Case ID

ID2 Tester ID

Tester Certification Yes No Initials: _____

By checking YES, I certify that I am the only person who prepared this report and attest that it is, to the best of my recollection, a true and accurate account of the events that took place during the case study in which I participated.

PART A Replying to the Assigned Ad

The process of replying to the ad:

→ A1 Phone Number Used (include area code)

			—				—			
--	--	--	---	--	--	--	---	--	--	--

→ A2 FIRST AND LAST NAME of Person Contacted (from ad or call)

→ A3 Firm Affiliation of Person in A2

If you did not reach anyone, go to B13; otherwise proceed to A4

The primary person you spoke with to schedule the appointment:

A4 Is this the person named in Question A2? Yes No

IF YES, skip to A7

IF NO, continue to A5

→ A5 FIRST AND LAST NAME of Person You Spoke with to Schedule Appointment (if not the person named in A2)

→ A6 Firm Affiliation of person named in A5

A7 What is the job description of the person you spoke with to schedule the appointment???

- Landlord (Individual Property Owner)
- Landlord (Firm or Agent of Firm)
- Real Estate Broker, Agent, or Salesperson
- Management Company Representative
- Superintendent or Handyman
- Other (specify): _____

PART B Appointment Scheduling and Location

→ B1 Did you successfully schedule an appointment to view the unit? Yes No

IF YES, continue to B2

IF NO, skip to B13

B2 Number of Call Attempts before Scheduling Appointment

B3 Time of Call when Appt. was Made:

HH	MM	AM/PM

MM DD YYYY

→ B4 Appointment Date (Scheduled)

HH	MM	AM/PM

HH MM AM/PM

→ B5 Appointment Time (Scheduled)

HH	MM	AM/PM

HH MM AM/PM

B6 Street Address of Assigned Rental Housing Unit

[a] Number	[b] Street Name	[c] Unit/Apt #
------------	-----------------	----------------

B7 Borough (check one)

- Bronx
- Brooklyn
- Manhattan
- Queens
- Staten Island

B8 Neighborhood (as described by landlord/agent in call)

Who you will be meeting at the appointment?

B9 I am meeting the same person I spoke with to set up the visit: Yes No

IF YES, go to FORM B

IF NO, continue to B10

B10 FIRST AND LAST NAME of Person You Will Meet at the Appointment: ←

B11 Firm Affiliation of person named in B10 ←

B12 What is the job description of the person named in B10? ←

- Landlord (Individual Property Owner)
- Landlord (Firm or Agent of Firm)
- Real Estate Broker, Agent, or Salesperson
- Management Company Representative
- Superintendent or Handyman
- Other (specify): _____

*If appointment scheduled, skip rest of form & go to FORM B.
IF NO APPOINTMENT SCHEDULED, go to B13*

Reasons why an appointment was not successfully secured:

B13 Were you unable to connect with a landlord or agent (check all that apply)

Regardless of contact method:

- [a] NOT APPLICABLE; successfully connected

- [b] Reached person, told to call back, unable to reconnect
- [c] Reached person, was told would be called back, no call received
- [d] Wrong number
- [e] Line always busy after max. attempts
- [f] No phone pickup after max. attempts
- [g] Listed number disconnected
- [h] Other (specify): _____

B14 Were you unable to schedule an appointment during work hours? (check all that apply)

- [a] NOT APPLICABLE; there was no scheduling conflict
- [b] Unable to schedule appointment during business hours (9 am – 6 pm, weekdays)
- [c] Unable to schedule appointment during my available work hours
- [d] Other (specify): _____

B15 Did the Project Manager close the case for some other reason not listed above? Yes No

B16 If you reached someone, were you told the unit was already rented out? Yes No NA

B17 If you reached someone and the unit was rented out, did the person you spoke with offer to show you other units? Yes No NA

B18 If you replied YES to B17, describe what they said when they offered to show you other units and how they said it. Use the back of this sheet if you need more space.

END OF FORM A

PRE-VISIT INTERACTIONS
FORM B

ID1 Case ID ID2 Tester ID

Tester Certification Yes No Initials: _____

By checking YES, I certify that I am the only person who prepared this report and attest that it is, to the best of my recollection, a true and accurate account of the events that took place during the case study in which I participated.

Of the following aspects of your assigned biography, which aspect of your biography came up during the pre-visit phase and how did the primary person you spoke with to schedule the appointment receive this information?

Aspect of Assumed Biography	1 Did this aspect of your biography come up?		2 Assess their overall reaction to this information			3 Did they challenge or seem skeptical of this information?		4 If they had a generally positive reaction to this information, describe what they said and how they said it		5 If they had a generally negative reaction to this information, describe what they said and how they said it	
	Yes	No	Positive	Neutral	Negative	Yes	No	INSTRUCTIONS: Record exact quotes and describe their demeanor			
[A] Name	1	0	1	0	-1	1	0				
[B] Personal Income	1	0	1	0	-1	1	0				
[C] Household Income	1	0	1	0	-1	1	0				
[D] Occupation	1	0	1	0	-1	1	0				
[E] Employer	1	0	1	0	-1	1	0				
[F] Credit Score	1	0	1	0	-1	1	0				
[G] Marital/Partner Status	1	0	1	0	-1	1	0				
[H] Children/Dependents	1	0	1	0	-1	1	0				
[I] Reason for Moving	1	0	1	0	-1	1	0				
[J] Location of Current Residence	1	0	1	0	-1	1	0				
[K] Location of Current Workplace	1	0	1	0	-1	1	0				

FIELD VISIT SURVEY

FORM C

PART A People With Whom You Interacted In Person During Field Visit

A0. Number of people with whom you interacted (check one):

1 2 3 4 5

List and describe the individuals with whom you interacted during this case (up to 5)

	A1 Person 1	A2 Person 2	A3 Person 3	A4 Person 4	A5 Person 5
[a] Is this the same person you spoke with to set up the appointment?	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Yes <input type="checkbox"/> No
[b] Primary individual(s) showing you housing unit(s)	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Yes <input type="checkbox"/> No	<input type="checkbox"/> Yes <input type="checkbox"/> No
If there was more than one individual showing the unit, please explain their apparent roles in the process:					
CHECK ALL THAT APPLY					
[c] First Name					
[d] Last Name					
[e] Firm Affiliation					
[f] Job Position	<input type="checkbox"/> Landlord (Individual Owner) <input type="checkbox"/> Landlord (Firm or Agent of Firm) <input type="checkbox"/> Real Estate Broker/Agent <input type="checkbox"/> Management Co. Representative <input type="checkbox"/> Superintendent or Handyman <input type="checkbox"/> Other (specify): _____				
[g] Age Category	<input type="checkbox"/> 18-34	<input type="checkbox"/> 45-64	<input type="checkbox"/> 18-34	<input type="checkbox"/> 45-64	<input type="checkbox"/> 18-34
CHECK ONE	<input type="checkbox"/> 35-44	<input type="checkbox"/> 65+	<input type="checkbox"/> 35-44	<input type="checkbox"/> 65+	<input type="checkbox"/> 35-44
Based on your perceptions	<input type="checkbox"/> White <input type="checkbox"/> Black or African American <input type="checkbox"/> Asian <input type="checkbox"/> Native American/American Indian <input type="checkbox"/> Hispanic, Latino, or Spanish origin				
[h] Census-Based Race and/or Hispanic Origin	<input type="checkbox"/> White <input type="checkbox"/> Black or African American <input type="checkbox"/> Asian <input type="checkbox"/> Native American/American Indian <input type="checkbox"/> Hispanic, Latino, or Spanish origin				
CHECK ALL THAT APPLY					
[i] Ethnicity and/or National Origin					
WRITE-IN & EXPLAIN					
Based on your perceptions					
<p>Note Sheet Included? (Initials): _____ Note Sheet Included? (Initials): _____</p> <p>By checking YES, I certify that I am the only person who prepared this report and attest that it is, to the best of my recollection, a true and accurate account of the events that took place during the case study in which I participated. A completed field visit note sheet I used during this case study is attached to the end of this form.</p>					

Based on your perceptions

GO TO NEXT PAGE ➔ ➔ ➔

CASE CLOSURE/POST-VISIT FORM D

ID1	Case ID	ID2	Tester ID

Tester Certification Yes No Initials: _____

By checking YES, I certify that I am the only person who prepared this report and attest that it is, to the best of my recollection, a true and accurate account of the events that took place during the case study in which I participated.

PART A Receiving a Post-Visit Callback

A1 Did you receive a callback from the person with whom you met during the housing unit visit **within 48 hours of the visit?**

Yes No Not Applicable

IF YES, go to PART B

IF NO or NOT APPLICABLE, go to PART C

PART B Post-Visit Callback Details

B1 Callback Date

MM	DD	YYYY

B2 Callback Time

HH	MM	AM/PM

B3 Who contacted you?

Record the **PERSON NUMBER (1-5)** from the Person list from the Field Survey you completed for this case.

--

B4 Did this person offer you the unit? Yes No

B5 If they did not offer you the rental unit, did they indicate that the unit had already been rented out?

Yes, unit already rented out
 No, unit not yet rented out
 Unclear; they did not volunteer info

B6 If they did not offer you the rental unit, did they offer to show you other vacant rental units or share additional listings with you?

Yes

No

B7 In the space below, describe any additional conversation and interactions that occurred during this call.

PART C Case Closure Procedures

COMPLETE THIS SECTION IF YOU COMPLETED A PRE-VISIT CALL BUT WE DID NOT PROCEED WITH THE CASE (ALL 3 TESTER DID NOT GET AN APPOINTMENT), OR IF YOU WENT ON A FIELD VISIT BUT DID NOT RECEIVE A CALLBACK WITHIN 48 HOURS AFTER THE VISIT

Please note the date and time when you called them back to decline continued interest and to close the case.

C1 Case Closure Date

MM	DD	YYYY

C2 Case Closure Time

HH	MM	AM/PM

C3 In the space below, describe any additional conversation and interactions that occurred during this call. **Use the back of sheet if needed.**

C Estimation

This appendix details the methods used to estimate the treatment effects central to the experimental analysis. We are interested in the following primary sample quantities of interest:

1. the Intent-to-Treat (ITT) effect of assignment to a treatment message on discrimination incidence,
2. the Principal Causal Effect (also known as the Complier Average Causal Effect, or CACE) of a treatment on discrimination incidence, and
3. the Conditional Average Treatment Effect (CATE) of treatment assignment on discrimination incidence conditional on a tester team providing a randomly assigned putative signal of low employment stability during the appointment.

C.1 Framework

Let treatment assignment $Z = \{0, 1, 2\}$ where 0=control, 1=monitoring signal, 2=punitive signal. Let treatment received $D = \{0, 1, 2\}$, with values equivalent to those for Z . Assume the ignorability of Z due to random assignment, non-interference, and the exclusion restriction. Let Y denote an outcome variable and $Y(d)$ is the potential outcome that responds only to the received treatment under the exclusion restriction. For the sake of exposition, let the treatment-comparison contrasts of interest be denoted (d_T, d_C) where $d_T, d_C \in D; d_T > d_C$. Thus we care about the effect of monitoring versus control (1,0); punitive versus control (2,0); and punitive versus monitoring (2,1). We estimate the aforementioned quantities separately for each treatment contrast and for each majority-minority group pairing (white-black, white-Hispanic, and black-Hispanic).

C.2 ITT Estimation

Because this is a randomized experiment, estimation of the ITT is straightforward. The primary model we use for the policy report is an ordinary least squares model with inverse probability weights to account for different assignment probabilities due to the randomization procedure and experimental design. Subject-specific weights are calculated as the inverse of the probability of assignment to the subject's observed treatment assignment, accounting for blocking in the design.

C.3 CACE Estimation

Complier average causal effects are of interest because we care about the effects of the message content – i.e., receiving the message – on discriminatory behavior. We observe one-way noncom-

pliance with treatment assignment as shown in Table 5.2. Noncompliance occurred for a range of reasons. Most commonly, landlords assigned to a treatment arm did not pick up the targeted phone call, even after multiple call attempts. Some landlords also hung up the phone once they realized they were receiving a call from the Commission on Human Rights.

Under noncompliance of this form, standard estimators of treatment effects are biased. Thus we redefine estimands as CACEs, or the average treatment effect among different complier types, or population subgroups defined by how they would comply with different treatments if they were hypothetically assigned to these groups. We define three types of compliers in this study:

- Never-Takers: Individuals who would receive the control condition regardless of which treatment they were assigned to
- Partial Compliers: Individuals who would take the control condition if assigned to control, but would receive at most the monitoring condition if assigned to the monitoring or punitive messages.
- Compliers: Individuals who receive the full treatment to which they are assigned, regardless of the experimental group to which they are allocated.

We define four principal causal effects of interests: (a) the average causal effect of the monitoring message (versus no message) among Partial Compliers; (b) the average causal effect of the monitoring message (versus no message) among Full Compliers; (c) the average causal effect of the punitive message (versus no message) among Full Compliers; and (d) the average causal effect of the punitive message (versus the monitoring message only) among Full Compliers.

Following Long, Little, and Lin (2010), we estimate these effects with a Bayesian data augmentation procedure implemented using a Gibbs sampler with uninformative priors, under standard structural assumptions for experimental analysis²⁷, monotonicity assumptions that yield the aforementioned definition of principal strata, and a perfect blindness assumption which posits that compliance status is orthogonal to treatment. The MCMC procedure used involves 5 chains with 12,000 iterations per chain, discarding the first 10,000 burn-in iterations in each chain. We report the following statistics from the posterior distribution: the posterior mean, median, 95% credible interval, and standard deviation. We also report the \hat{R} and effective N to demonstrate model convergence. Full results are reported in Appendix D.

C.4 CATE Estimation

To estimate the conditional effects of messaging on net discrimination incidence by putative high versus low employment stability signals, we estimate ordinary least squares models that include

²⁷Ignorability of treatment assignment, non-interference, the exclusion restriction

a main effect for the low employment stability treatment (1=low, 0=high) and interaction terms between each of the messaging treatments and the low employment stability treatment. We then estimate marginal effects of messaging by low versus high employment stability signal and report 95% confidence intervals around each estimate.

D Additional Results

This appendix presents results from additional analyses.

D.1 Distributions of Indicators of Objective Discrimination during the Appointment Stage

This section presents the distribution of indicators of objective discrimination during the appointment stage. For each indicator presented, we show the distribution by tester using valid case-tester level reports. We also show the distribution of difference measures by majority-minority group comparisons, using valid paired test data.

D.1.1 Across Units Shown

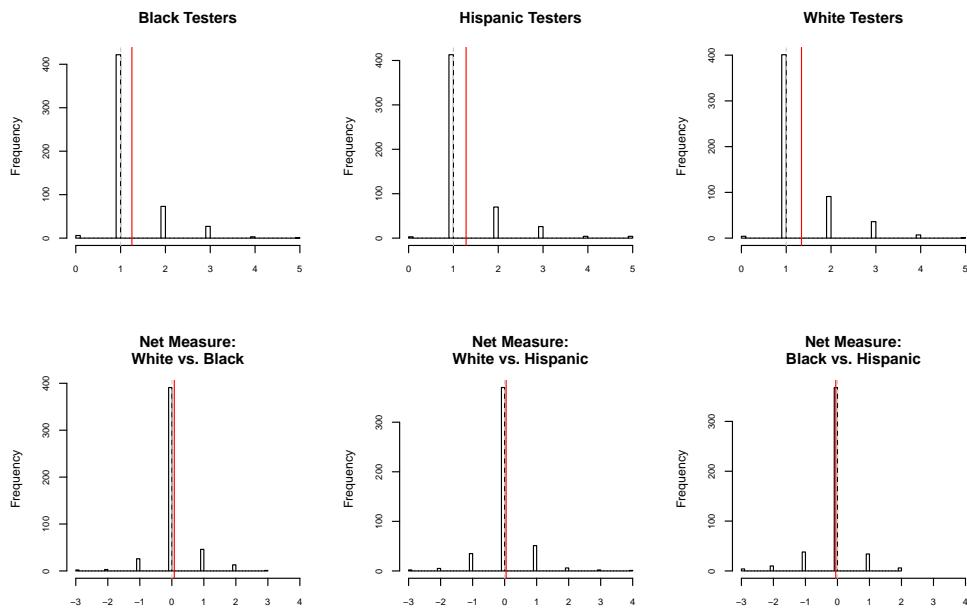


Figure D.1: Number of Units Shown, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

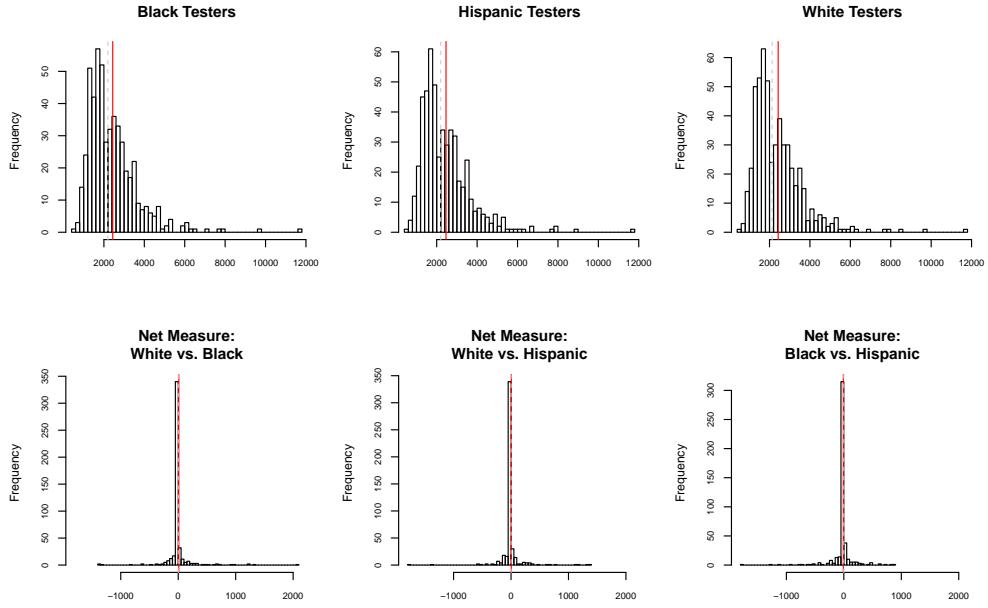


Figure D.2: Average Quoted Price Across Units Shown, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

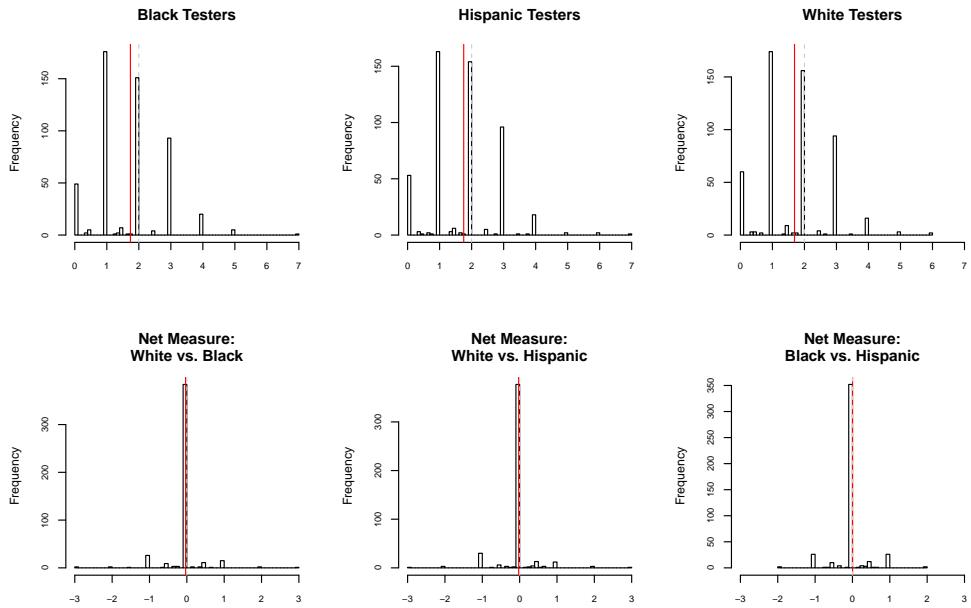


Figure D.3: Average Number of Bedrooms in Units Across Units Shown, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

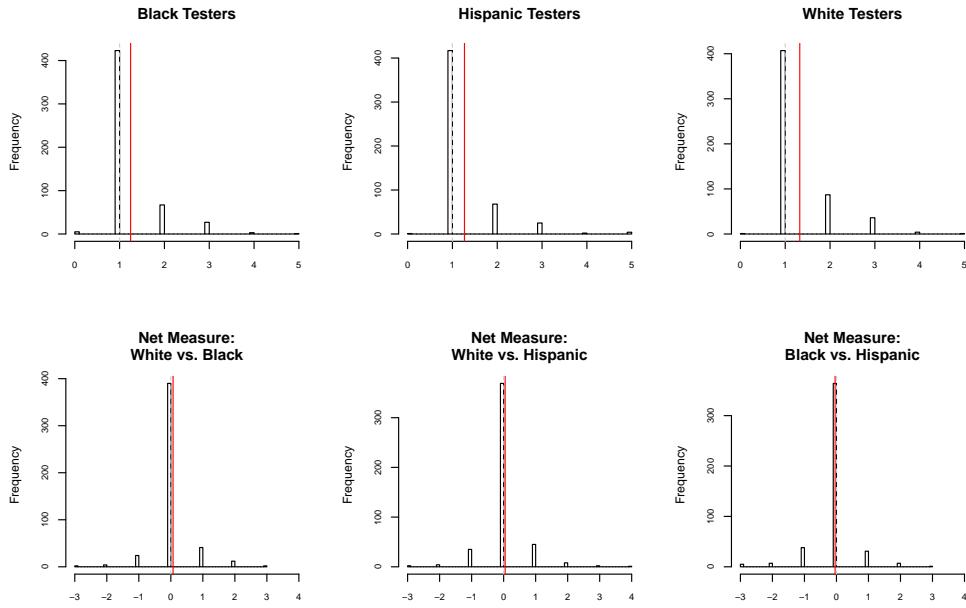


Figure D.4: Number of Units Among Units Shown Where Rental Price Was Quoted in Person, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

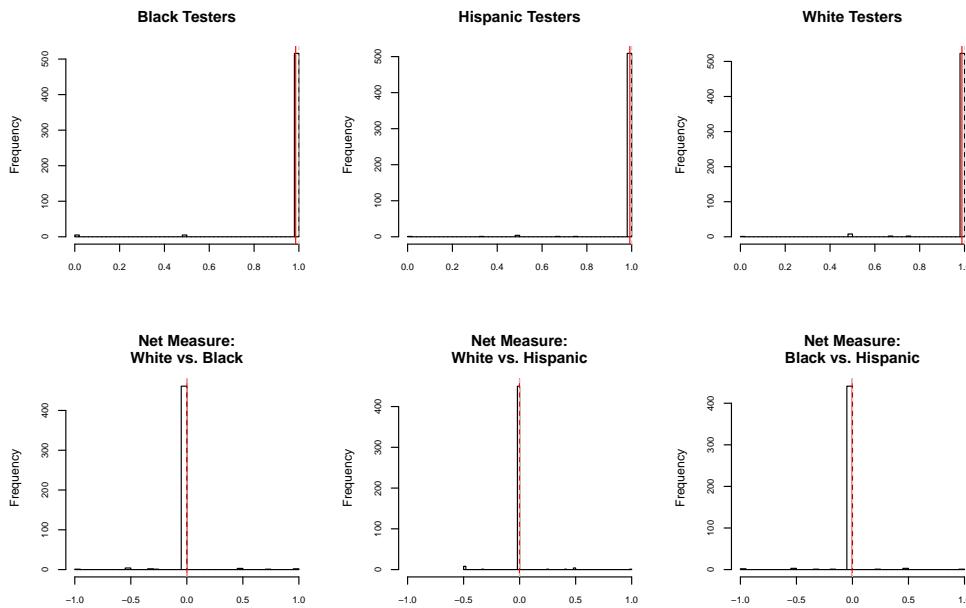


Figure D.5: Percentage of Units Among Units Shown Where Rental Price Was Quoted in Person, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

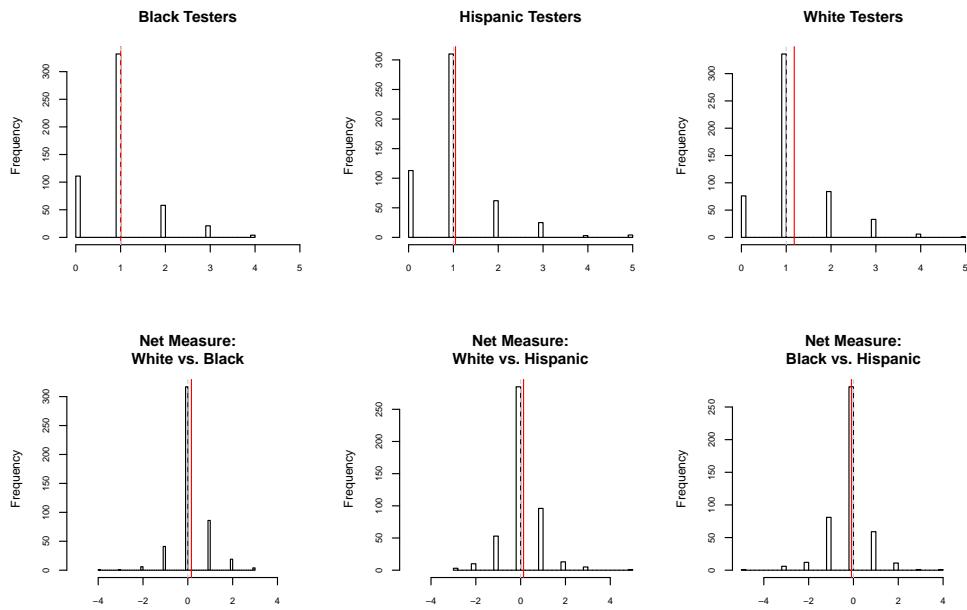


Figure D.6: Number of Units Among Units Shown where Application Requirement was Quoted, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

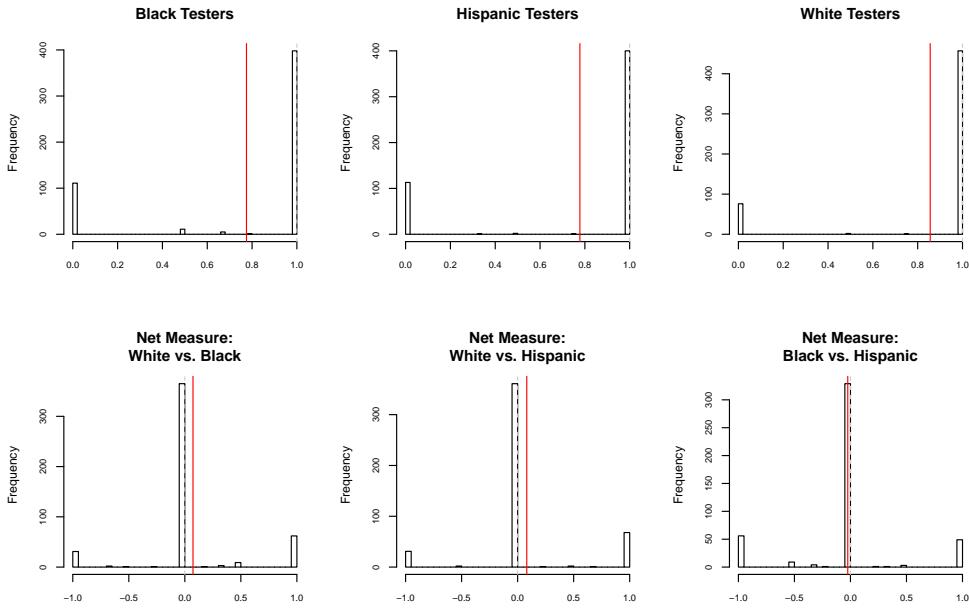


Figure D.7: Percentage of Units Among Units Shown where Application Requirement was Quoted, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

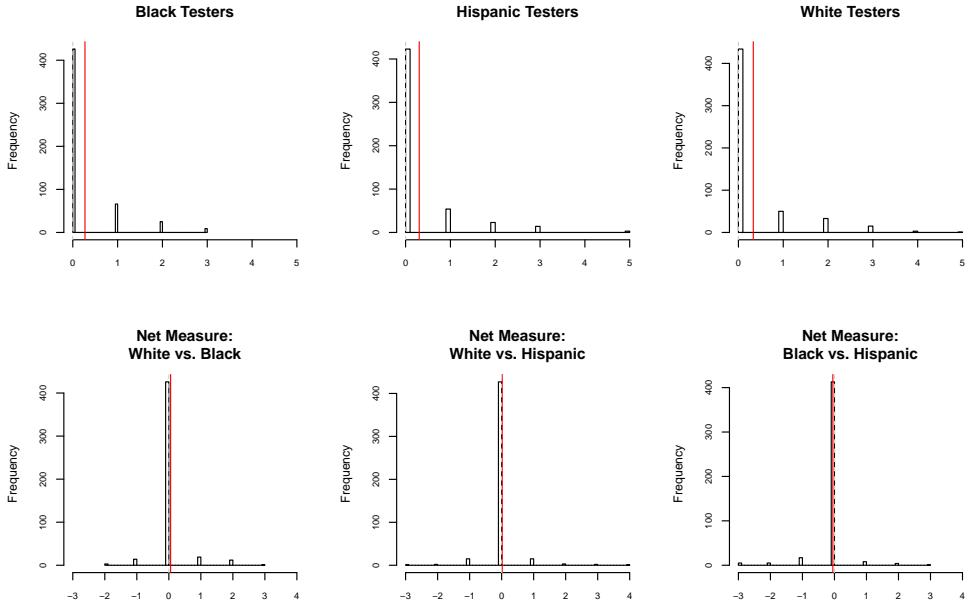


Figure D.8: Number of Units Among Units Shown in Doorman Building, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

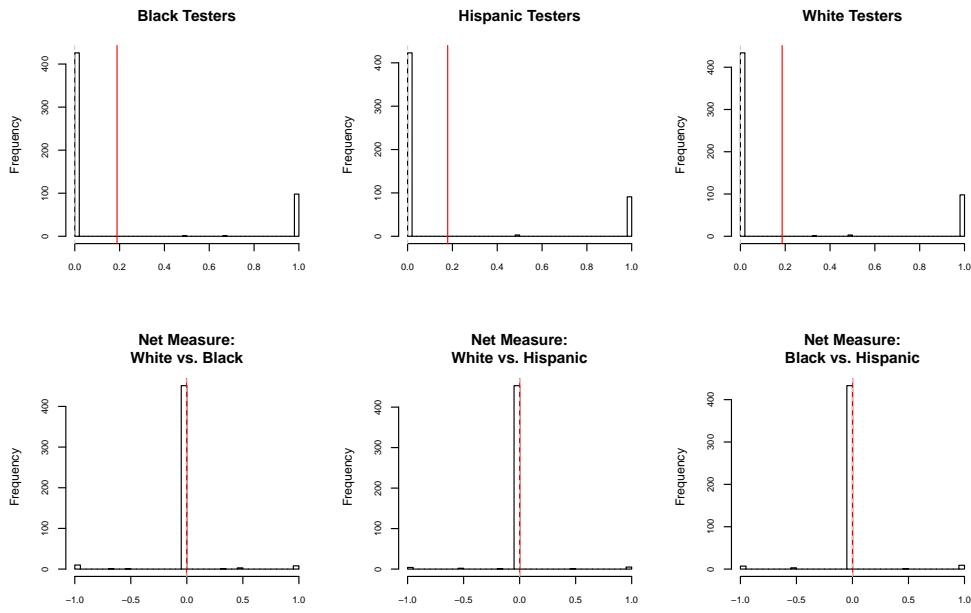


Figure D.9: Percentage of Units Among Units Shown in Doorman Building, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

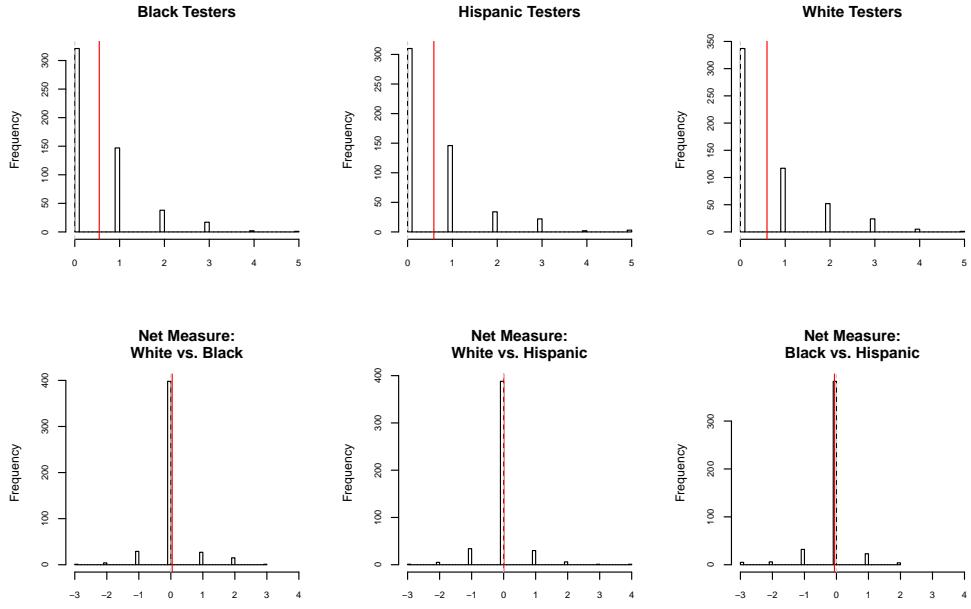


Figure D.10: Number of Units Among Units Shown in Elevator Building, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

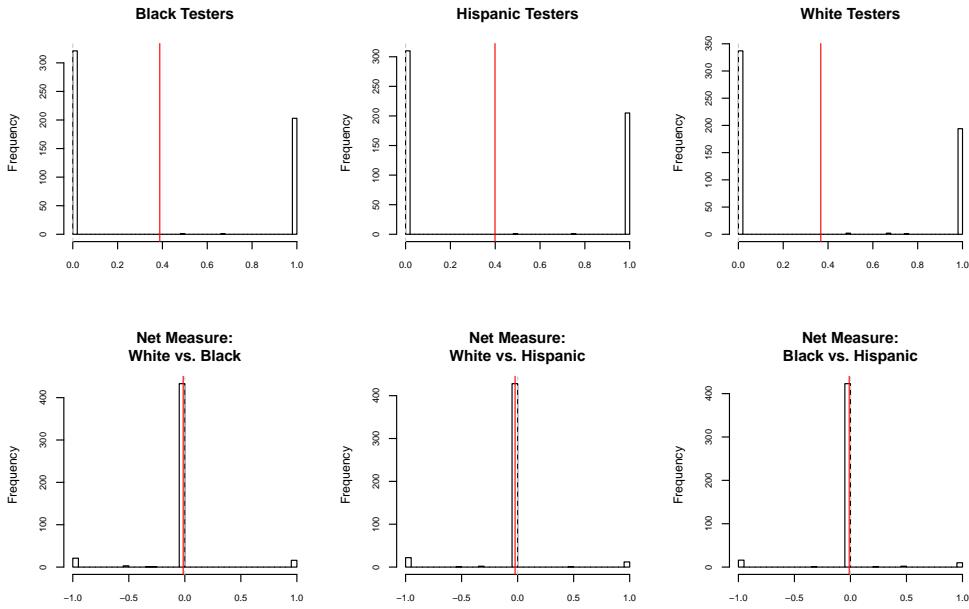


Figure D.11: Percentage of Units Among Units Shown in Elevator Building, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

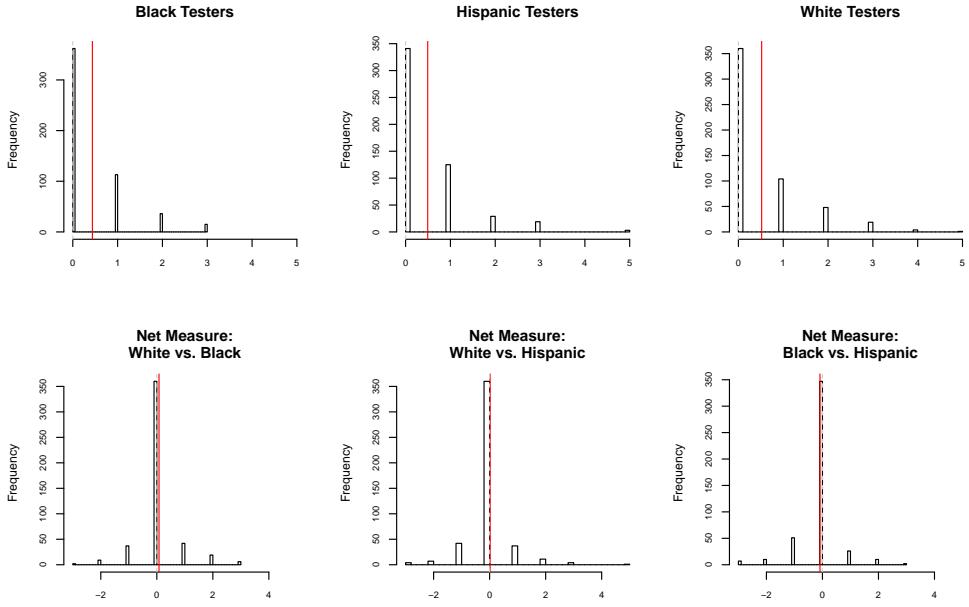


Figure D.12: Number of Units Among Units with Washer/Dryer in Building, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

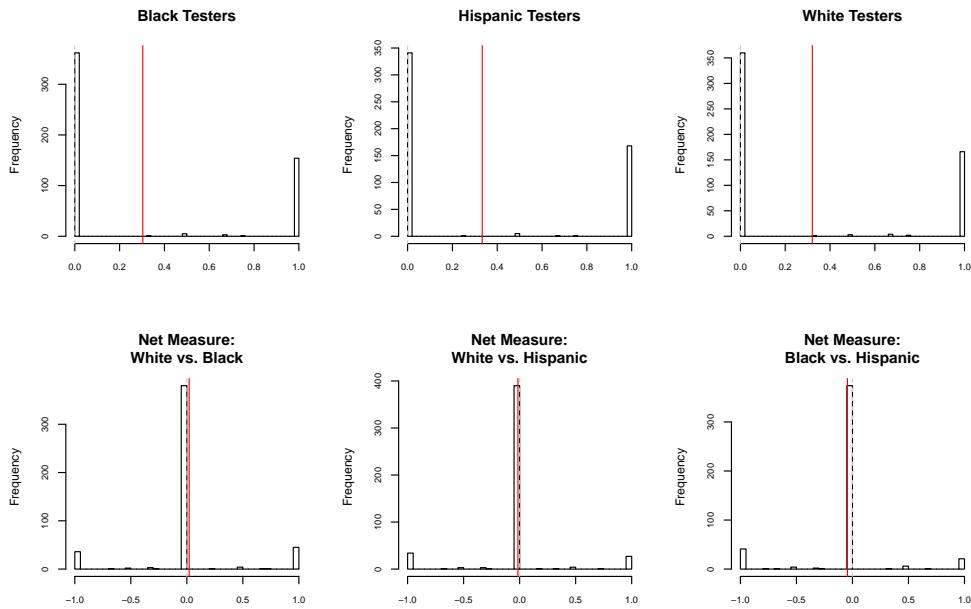


Figure D.13: Percentage of Units Among Units with Washer/Dryer in Building, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

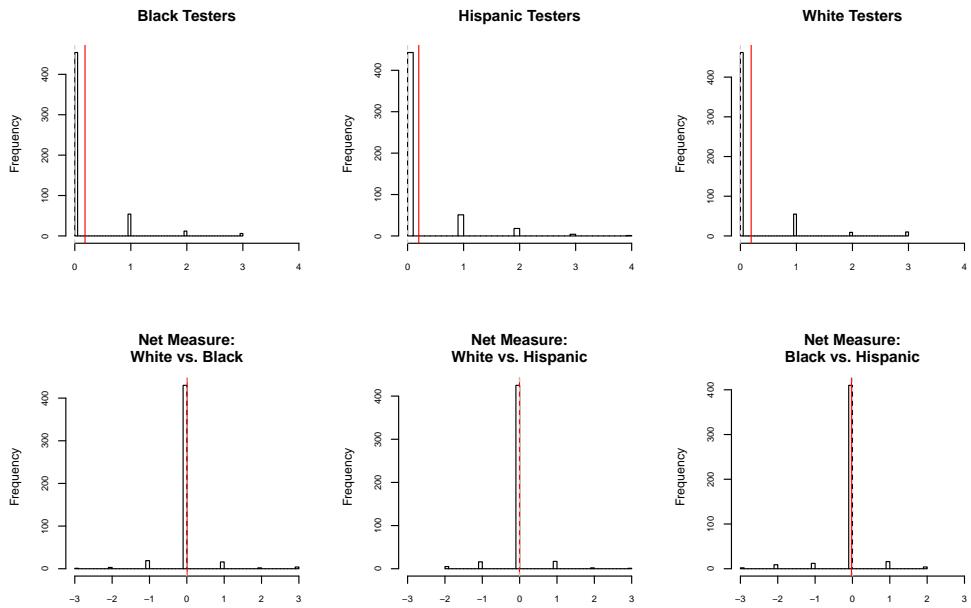


Figure D.14: Number of Units Among Units with Washer/Dryer in Unit, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

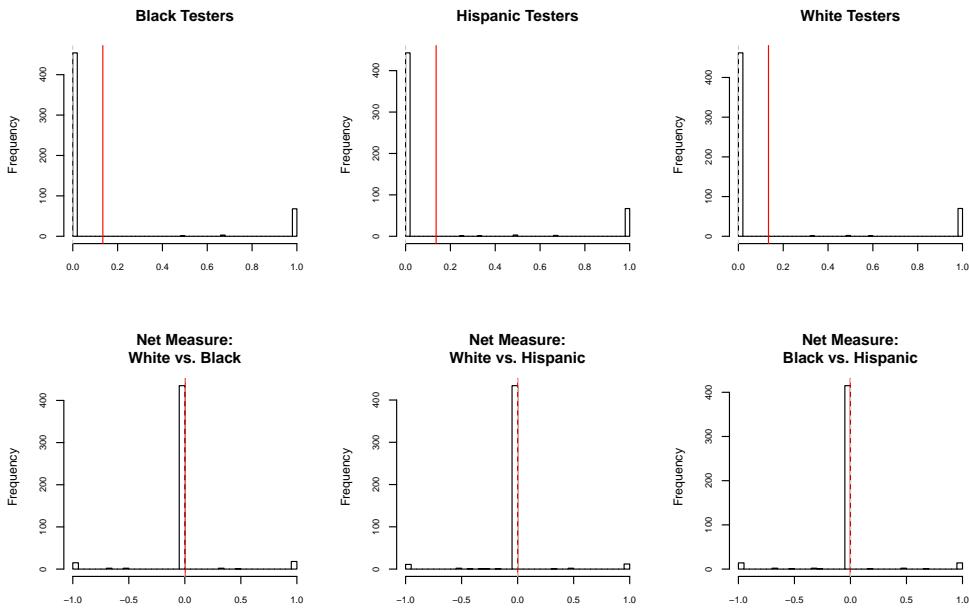


Figure D.15: Percentage of Units Among Units with Washer/Dryer in Unit, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

D.1.2 Quoted Terms of Rent for Listed Unit, if Shown

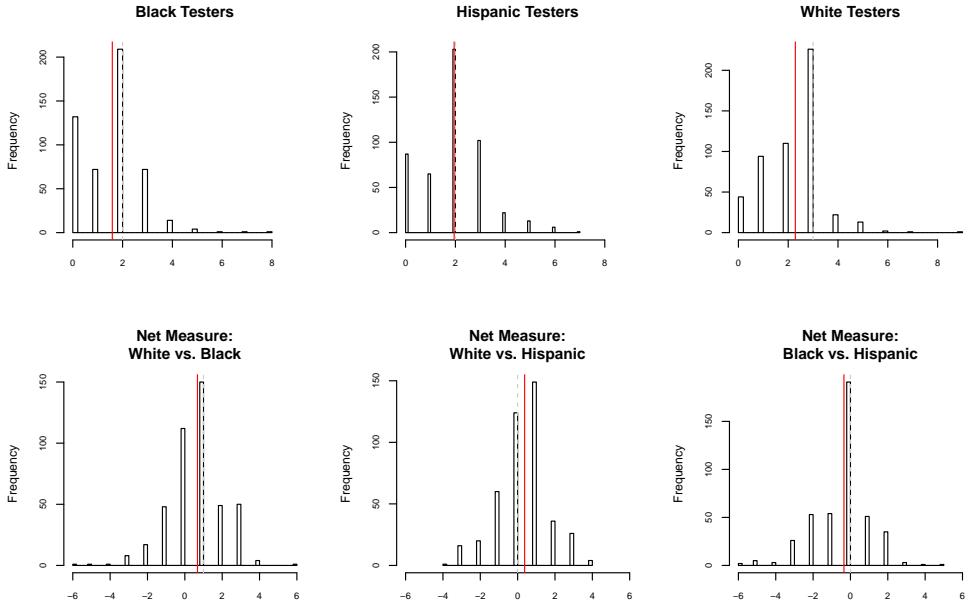


Figure D.16: Number of Quoted Amenities Included in Rent for Listed Unit, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

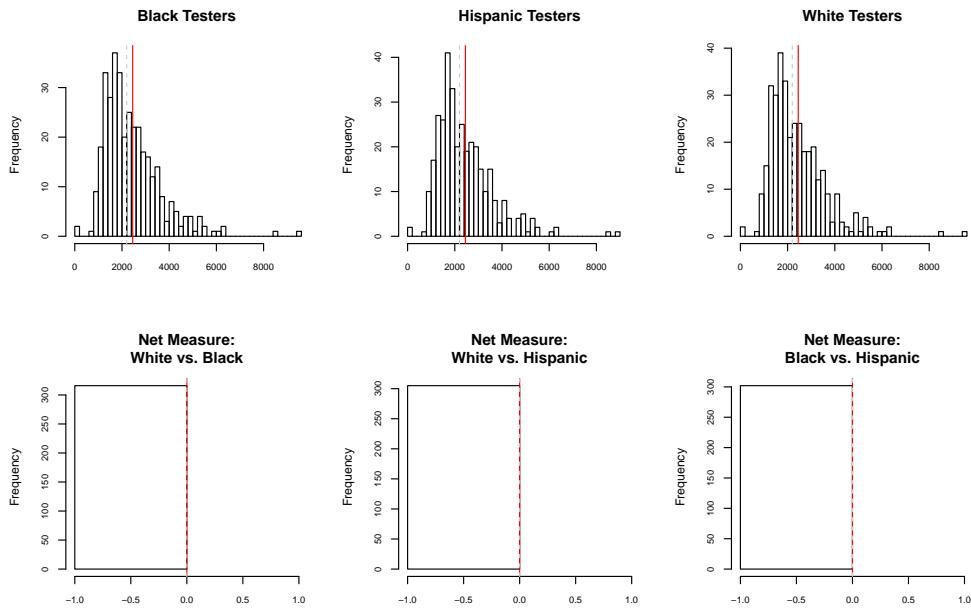


Figure D.17: Quoted Monthly Asking Rental Price for Listed Unit, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

D.1.3 Incidence of Quoted Fee Types for Listed Unit, if Shown

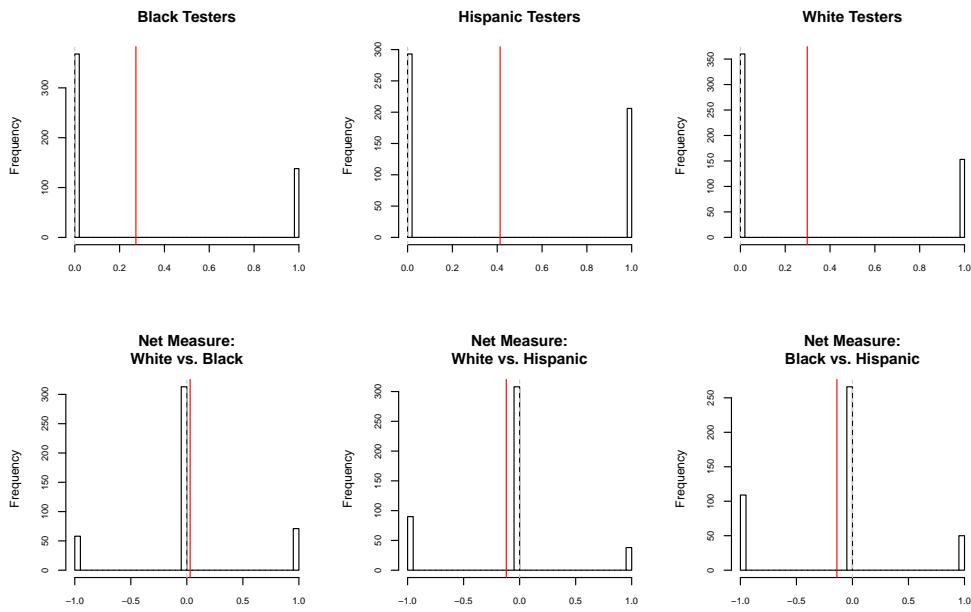


Figure D.18: Incidence of a Quoted Requirement to Pay of Any Number of Months of Rent Up Front for the Listed Unit, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

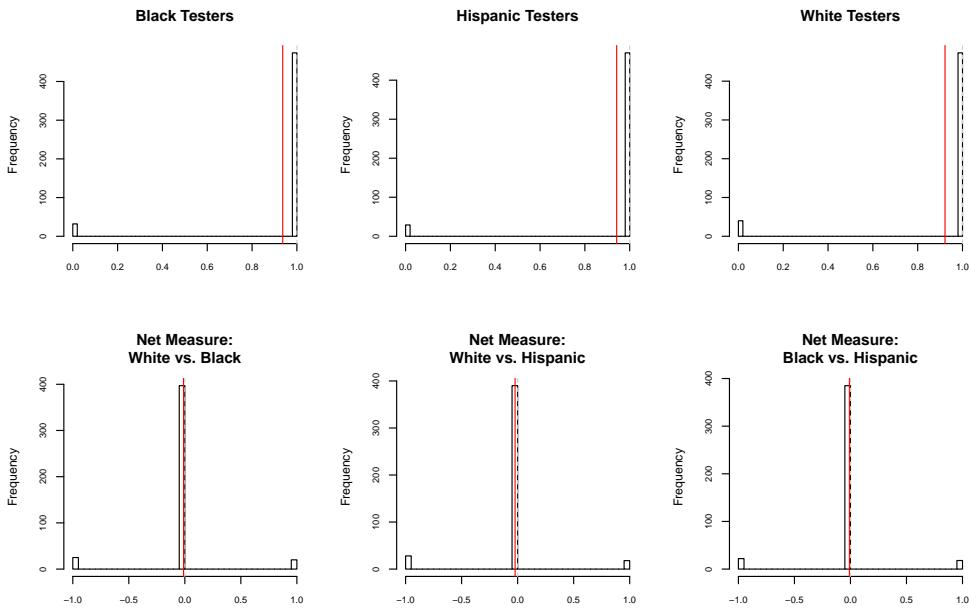


Figure D.19: Incidence of a Quoted Security Deposit Requirement for the Listed Unit, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

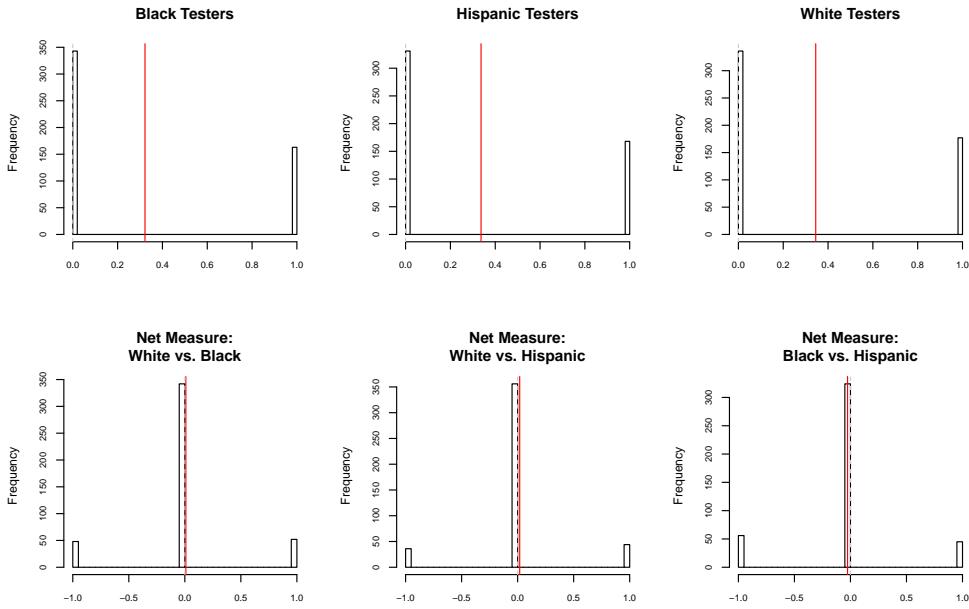


Figure D.20: Incidence of a Quoted Broker Fee Requirement for the Listed Unit, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

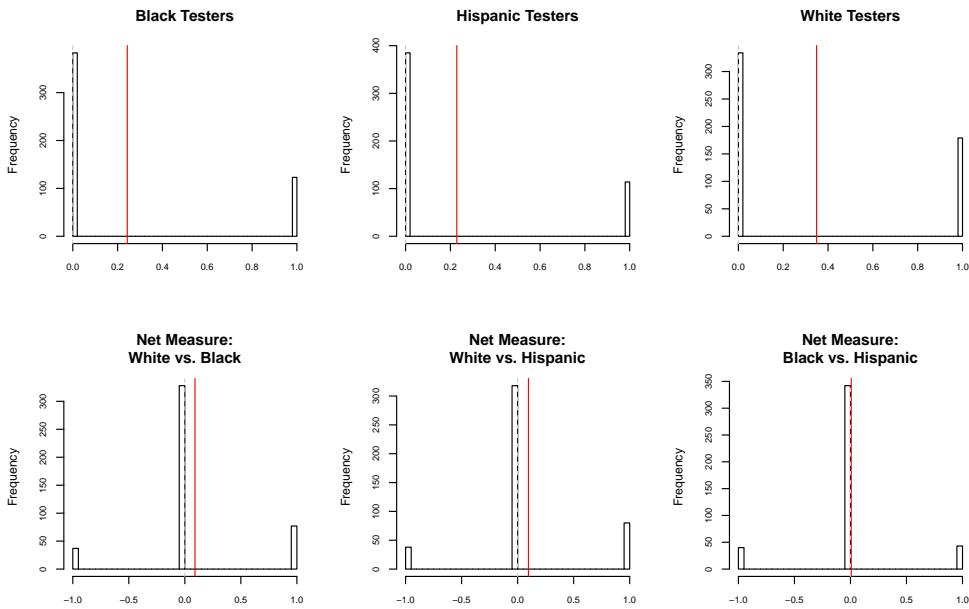


Figure D.21: Incidence of a Quoted Application Fee Requirement for the Listed Unit, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

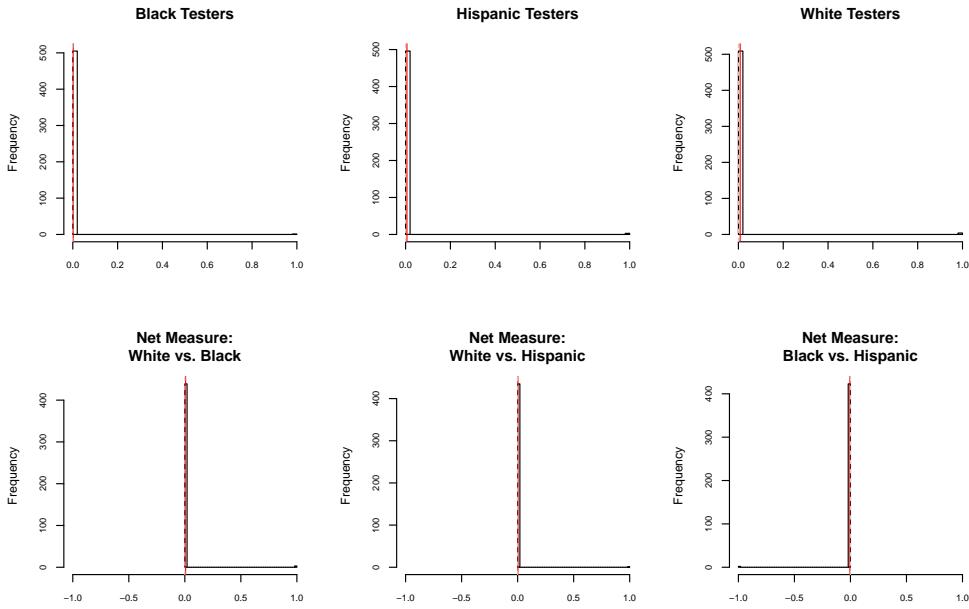


Figure D.22: Incidence of a Quoted Administrative or Processing Fee Requirement for the Listed Unit, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

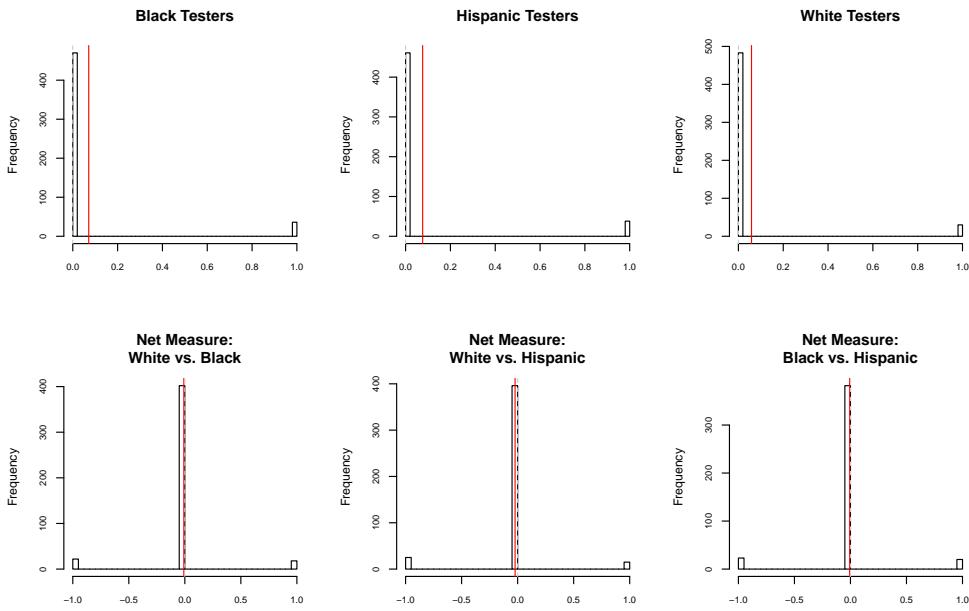


Figure D.23: Incidence of a Quoted Holding Fee Requirement for the Listed Unit, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

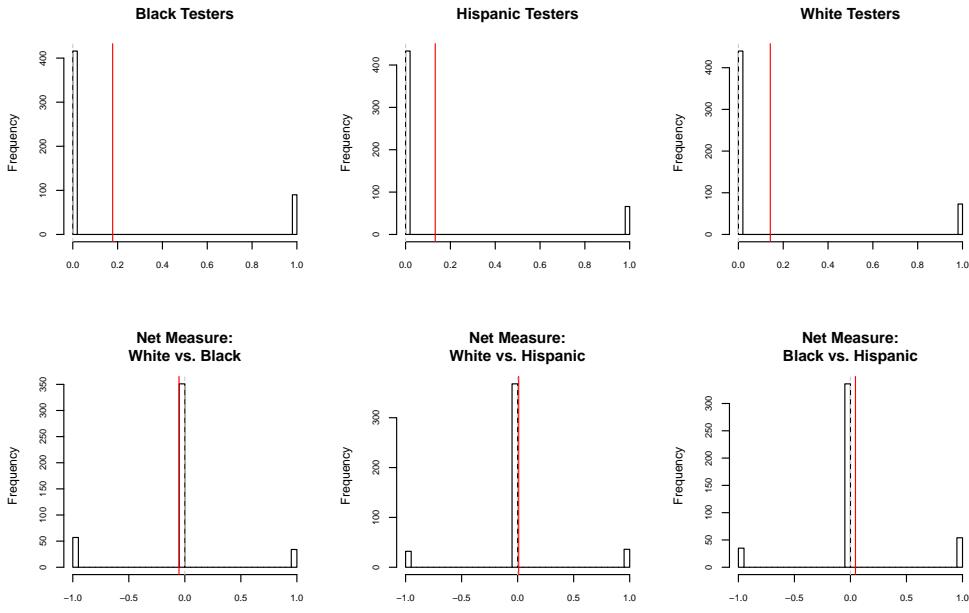


Figure D.24: Incidence of a Quoted Credit or Background Check Fee Requirement for the Listed Unit, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

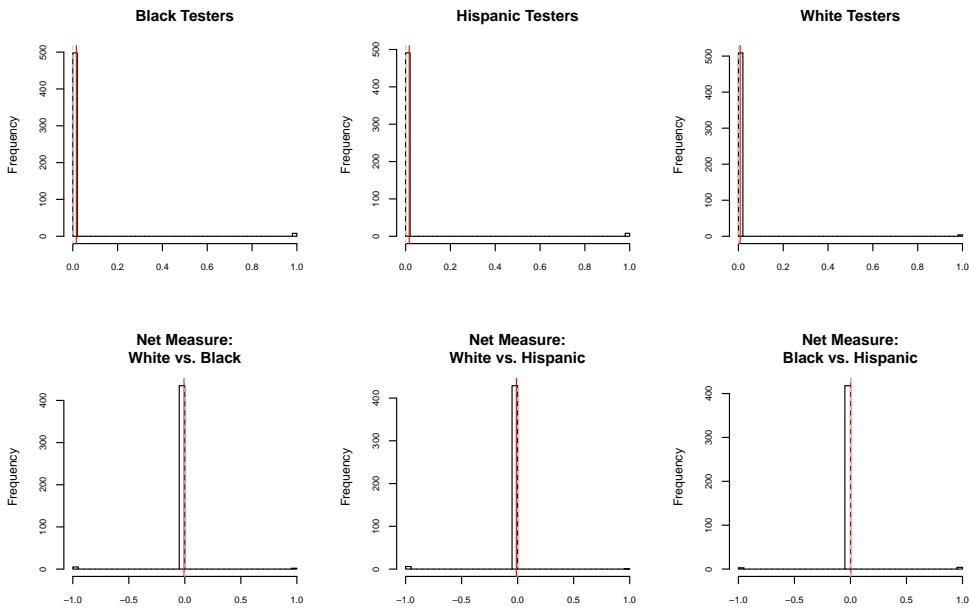


Figure D.25: Incidence of a Quoted Other Fees Requirement for the Listed Unit, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

D.1.4 Quoted Fee Amounts for Listed Unit, if Shown

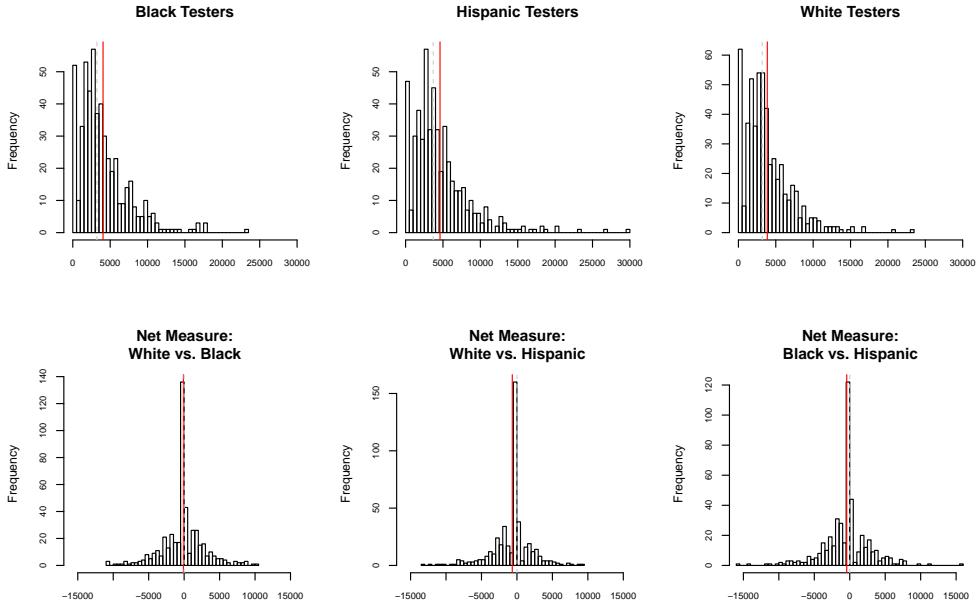


Figure D.26: Quoted Sum of All Fees for the Listed Unit, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

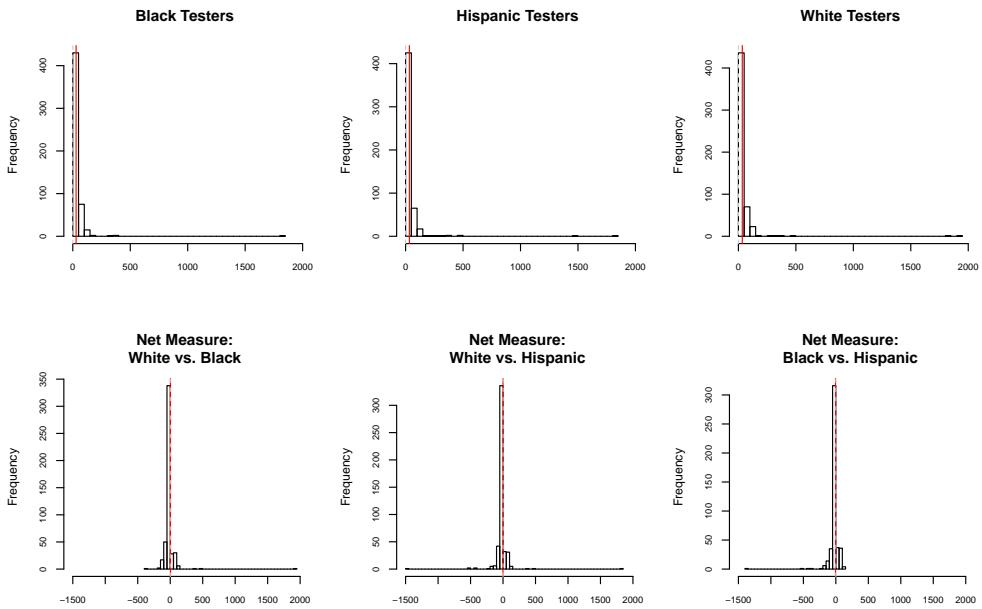


Figure D.27: Quoted Sum of Application Fees and Other Net Costs for the Listed Unit, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

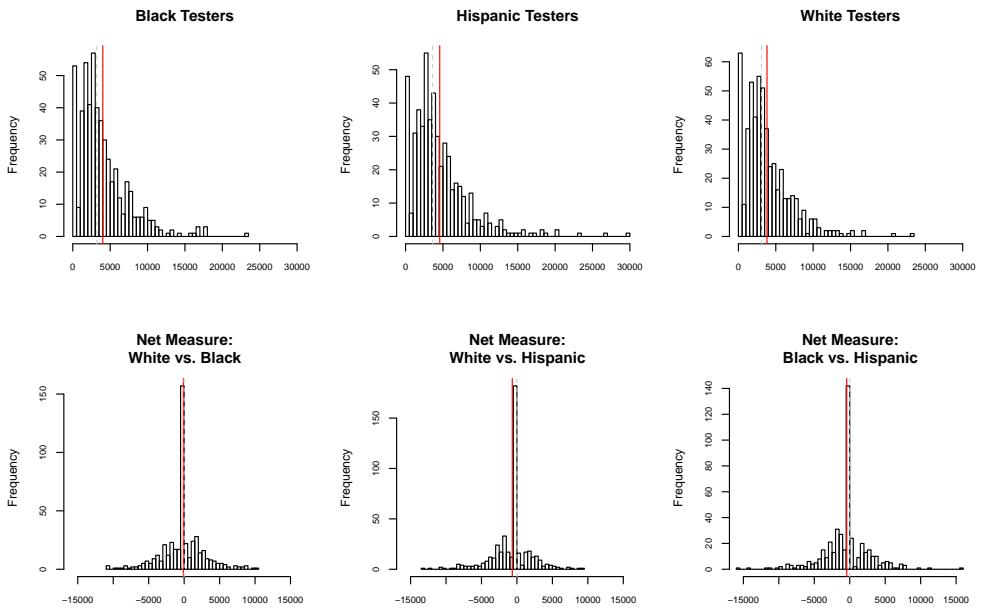


Figure D.28: Quoted Sum of Up-Front Rent, Security Deposit, Holding Fees, and Broker Fees for the Listed Unit, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

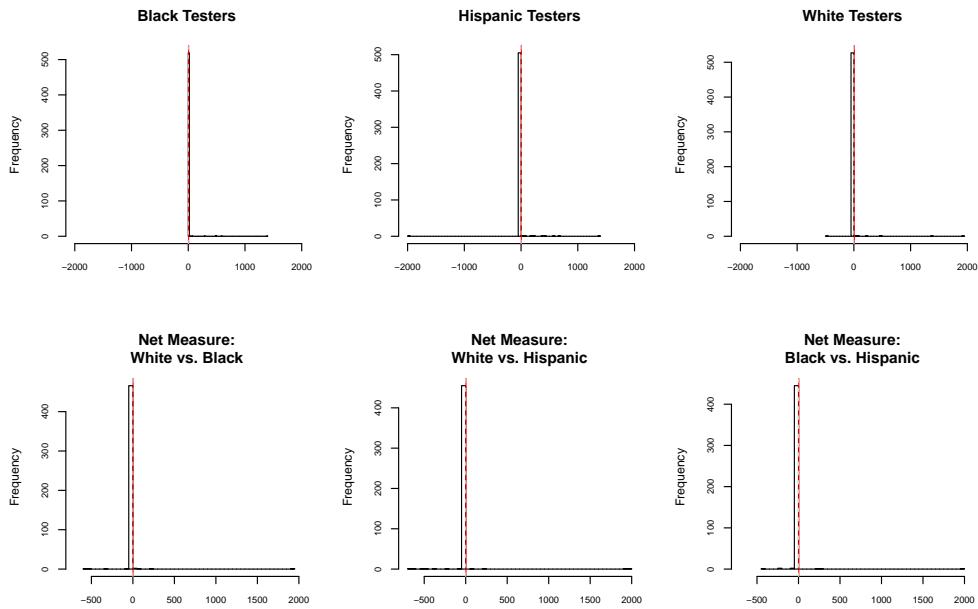


Figure D.29: Quoted Sum of Administrative Fees and Other Net Costs for the Listed Unit, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

D.1.5 Willingness to Negotiate for Listed Unit, if Shown

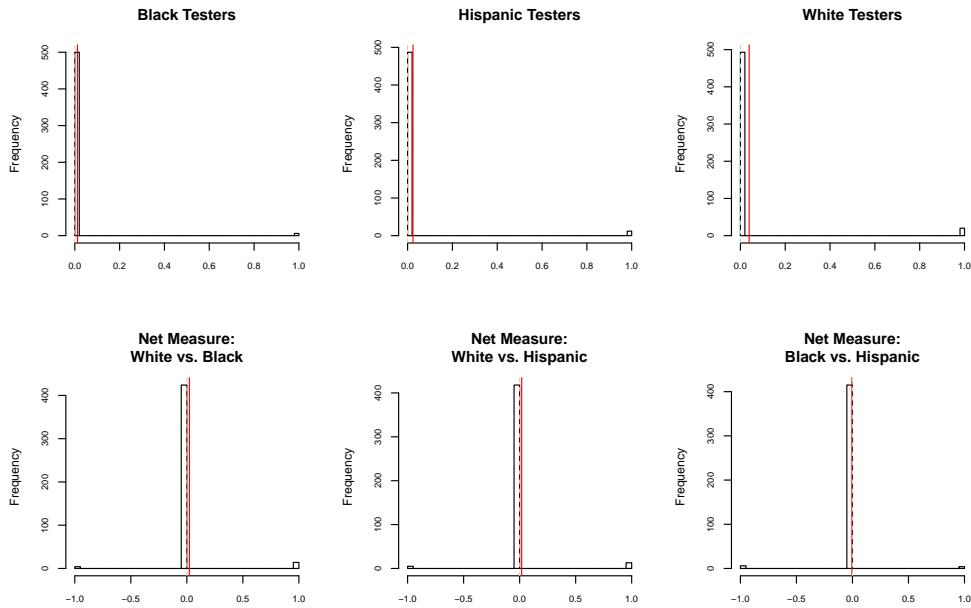


Figure D.30: Incidence of a Offer to Negotiate Fees for the Listed Unit, by Tester Race and by Group Differences. The first row presents the distribution by tester race, using valid case-tester-level reports. The second row presents the distribution by majority-minority group comparisons, using valid paired test data. The red line denotes the mean; the dashed grey line denotes the median.

D.2 Effects of Treatment Messaging on Net Discrimination: Unweighted Data

This section presents the raw unweighted data. We report mean levels of favorable treatment by tester race; differences in mean levels of favorable treatment across treatment groups, by tester race; differences in mean levels of favorable treatment between tester groups, by treatment (i.e., the mean net discrimination levels by treatment group); and the difference in net discrimination levels between treatment groups (i.e., estimates of the effects of treatment messaging on net discrimination levels). *Note: These estimates are presented only to provide a sense of the raw data, but should not be interpreted as the causal effects of sending different messages on net discrimination because they do not account for the randomization procedure.*

	I. Mean Levels, by Group			II. Differences in Means		
	Control	Monitoring	Punitive	Monitoring vs. Control	Punitive vs. Control	Punitive vs. Monitoring
Panel A. Percent Favorable						
White	0.993	0.989	1	-0.004 (0.651)	0.007 (0.158)	0.011 (0.158)
Black	0.996	0.977	0.995	-0.019 (0.106)	-0.001 (0.818)	0.018 (0.15)
Hispanic	0.996	0.989	1	-0.008 (0.373)	0.004 (0.318)	0.011 (0.158)
Panel B. Net Discrimination						
<i>(% Majority Favorable - % Minority Favorable)</i>						
White vs. Black	-0.004 (0.318)	0.011 (0.319)	0.005 (0.319)	0.015 (0.951)	0.009 (0.834)	-0.006 (0.486)
White vs. Hispanic	-0.004 (0.318)	0 (1)	0 (NA)	0.004 (0.757)	0.004 (0.687)	0 (0.82)
Black vs. Hispanic	0 (NA)	-0.011 (0.319)	-0.005 (0.319)	-0.011 (0.123)	-0.005 (0.278)	0.006 (0.576)
Sample Size	279	174	200	453	479	374

Table D.1: Unweighted Data: The Effects of Assignment to Messaging Conditions on Net Discrimination in Landlords Honoring Scheduled Appointments. Cells in the upper-left quadrant (quadrants denoted by double lines) contain estimates of the levels of favorable treatment toward each population (white, black, Hispanic testers). Cells in the ‘Control’ column in the bottom-left quadrant contain estimates of baseline net discrimination rates, defined as the share of favorable majority treatment minus the share of favorable minority treatment. The remaining cells in the bottom-left quadrant contain estimates of net discrimination rates in non-control treatment groups. Cells in the upper-right quadrant contain estimates of the treatment effects on favorable treatment rates for specific populations. Cells in the bottom-right quadrant contain the primary estimates of interest – estimates of the treatment effects on net discrimination rates. We show estimates without weighting for different probabilities of assignment to treatment to provide a sense of the raw data. For these quantities, we estimate p-values (shown in parentheses), which are the probability of obtaining an effect at least as large (in absolute value) as the one observed in the actual experiment for the monitoring-control and punitive-control (punitive-monitoring) comparisons. The *p*-value is reported as NA when the data in the two groups being compared are essentially constant.

	I. Mean Levels, by Group			II. Differences in Means		
	Control	Monitoring	Punitive	Monitoring vs. Control	Punitive vs. Control	Punitive vs. Monitoring
Panel A. Percent Favorable						
White	0.215	0.184	0.19	-0.031 (0.418)	-0.025 (0.5)	0.006 (0.881)
Black	0.168	0.144	0.12	-0.025 (0.478)	-0.048 (0.133)	-0.024 (0.502)
Hispanic	0.154	0.155	0.185	0.001 (0.976)	0.031 (0.378)	0.03 (0.444)
Panel B. Net Discrimination						
<i>(% Majority Favorable - % Minority Favorable)</i>						
White vs. Black	0.047 (0.107)	0.04 (0.264)	0.07 (0.026)	-0.006 (0.417)	0.023 (0.675)	0.03 (0.58)
White vs. Hispanic	0.061 (0.019)	0.029 (0.425)	0.005 (0.876)	-0.032 (0.271)	-0.056 (0.086)	-0.024 (0.575)
Black vs. Hispanic	0.014 (0.587)	-0.011 (0.733)	-0.065 (0.037)	-0.026 (0.346)	-0.079 (0.031)	-0.054 (0.246)
Sample Size	279	174	200	453	479	374

Table D.2: Unweighted Data: The Effects of Assignment to Messaging Conditions on Net Discrimination in Receiving a Post-Visit Callback. Cells in the upper-left quadrant (quadrants denoted by double lines) contain estimates of the levels of favorable treatment toward each population (white, black, Hispanic testers). Cells in the ‘Control’ column in the bottom-left quadrant contain estimates of baseline net discrimination rates, defined as the share of favorable majority treatment minus the share of favorable minority treatment. The remaining cells in the bottom-left quadrant contain estimates of net discrimination rates in non-control treatment groups. Cells in the upper-right quadrant contain estimates of the treatment effects on favorable treatment rates for specific populations. Cells in the bottom-right quadrant contain the primary estimates of interest – estimates of the treatment effects on net discrimination rates. We show estimates without weighting for different probabilities of assignment to treatment to provide a sense of the raw data. For these quantities, we estimate p-values (shown in parentheses), which are the probability of obtaining an effect at least as large (in absolute value) as the one observed in the actual experiment for the monitoring-control and punitive-control (punitive-monitoring) comparisons.

	I. Mean Levels, by Group			II. Differences in Means		
	Control	Monitoring	Punitive	Monitoring vs. Control	Punitive vs. Control	Punitive vs. Monitoring
Panel A. Percent Favorable						
White	0.118	0.08	0.105	-0.038 (0.183)	-0.013 (0.648)	0.025 (0.414)
Black	0.09	0.08	0.08	-0.009 (0.734)	-0.01 (0.709)	0 (0.987)
Hispanic	0.061	0.063	0.095	0.002 (0.922)	0.034 (0.178)	0.032 (0.254)
Panel B. Net Discrimination						
<i>(% Majority Favorable - % Minority Favorable)</i>						
White vs. Black	0.029 (0.239)	0 (1)	0.025 (0.319)	-0.029 (0.084)	-0.004 (0.239)	0.025 (0.51)
White vs. Hispanic	0.057 (0.011)	0.017 (0.514)	0.01 (0.706)	-0.04 (0.052)	-0.047 (0.012)	-0.007 (0.698)
Black vs. Hispanic	0.029 (0.17)	0.017 (0.44)	-0.015 (0.565)	-0.011 (0.413)	-0.044 (0.048)	-0.032 (0.251)
Sample Size	279	174	200	453	479	374

Table D.3: Unweighted Data: The Effects of Assignment to Messaging Conditions on Net Discrimination in Receiving a Post-Visit Offer for the Unit. Cells in the upper-left quadrant (quadrants denoted by double lines) contain estimates of the levels of favorable treatment toward each population (white, black, Hispanic testers). Cells in the ‘Control’ column in the bottom-left quadrant contain estimates of baseline net discrimination rates, defined as the share of favorable majority treatment minus the share of favorable minority treatment. The remaining cells in the bottom-left quadrant contain estimates of net discrimination rates in non-control treatment groups. Cells in the upper-right quadrant contain estimates of the treatment effects on favorable treatment rates for specific populations. Cells in the bottom-right quadrant contain the primary estimates of interest – estimates of the treatment effects on net discrimination rates. We show estimates without weighting for different probabilities of assignment to treatment to provide a sense of the raw data. For these quantities, we estimate p-values (shown in parentheses), which are the probability of obtaining an effect at least as large (in absolute value) as the one observed in the actual experiment for the monitoring-control and punitive-control (punitive-monitoring) comparisons.

D.3 Full Results: Bayesian Estimates of Average Treatment Effects of Message Content by Complier Type

This section presents the full results from the Bayesian data augmentation procedure used to estimate the average causal effects of message content, by complier type. Tables below summarize the estimated posterior distribution of average treatment effects of message content by complier type (Panel A), of outcomes by treatment received and complier type (Panel B), and the proportion of complier type (Panel C). The average effects, means, and proportions are reported in the column labeled “Posterior Mean.”

Parameter	Posterior			95% Credible Interval	Standard Deviation	\hat{R}	Effective N
	Mean	Median	Posterior				
A. Estimates of the Distribution of Treatment Effects by Complier Type							
Monitoring vs. Control for Partial Compliers	-0.023	-0.022	-0.207	0.151	0.091	1.001	5000
Monitoring vs. Control for Full Compliers	0.053	0.052	-0.014	0.132	0.037	1.001	5000
Punitive vs. Control for Full Compliers	0.006	0.006	-0.051	0.062	0.028	1.002	1800
Punitive vs. Monitoring for Full Compliers	-0.048	-0.046	-0.127	0.022	0.038	1.002	2300
B. Estimates of the Distribution of Outcomes by Treatment Received and Complier Type							
Control Outcome for Never-Takers	0.1	0.098	0.061	0.15	0.023	1.001	5000
Control Outcome for Partial Compliers	0.296	0.292	0.172	0.449	0.071	1.001	5000
Monitoring Outcome for Partial Compliers	0.273	0.27	0.164	0.401	0.062	1.002	2000
Control Outcome for Full Compliers	0.082	0.08	0.048	0.124	0.02	1.001	4300
Monitoring Outcome for Full Compliers	0.135	0.132	0.081	0.206	0.032	1.001	3900
Punitive Outcome for Full Compliers	0.087	0.086	0.052	0.132	0.021	1.002	2600
C. Estimates of Proportions of Complier Type							
Share of Never-Takers	0.3	0.299	0.242	0.361	0.03	1.003	1200
Share of Partial Compliers	0.077	0.075	0.046	0.115	0.018	1.003	1500
Share of Full Compliers	0.623	0.624	0.558	0.685	0.032	1.005	670

Table D.4: Estimates of Principal Causal Effects on Net Discrimination against Blacks (vs. Whites) in Landlords and Brokers Honoring Scheduled Appointments

Parameter	Posterior Mean	Posterior Median	95% Credible Interval	Standard Deviation	\hat{R}	Effective N	
A. Estimates of the Distribution of Treatment Effects by Complier Type							
Monitoring vs. Control for Partial Compliers	-0.022	-0.021	-0.205	0.154	0.091	1.001	5000
Monitoring vs. Control for Full Compliers	0.047	0.046	-0.021	0.121	0.036	1.002	1800
Punitive vs. Control for Full Compliers	0.005	0.005	-0.05	0.061	0.028	1.001	5000
Punitive vs. Monitoring for Full Compliers	-0.042	-0.041	-0.117	0.025	0.037	1.002	3000
B. Estimates of the Distribution of Outcomes by Treatment Received and Complier Type							
Control Outcome for Never-Takers	0.09	0.088	0.052	0.134	0.021	1.001	5000
Control Outcome for Partial Compliers	0.296	0.291	0.173	0.444	0.069	1.001	5000
Monitoring Outcome for Partial Compliers	0.274	0.271	0.161	0.404	0.062	1.001	5000
Control Outcome for Full Compliers	0.082	0.08	0.048	0.125	0.019	1.001	5000
Monitoring Outcome for Full Compliers	0.129	0.127	0.076	0.197	0.031	1.002	2300
Punitive Outcome for Full Compliers	0.087	0.086	0.051	0.13	0.02	1.001	4700
C. Estimates of Proportions of Complier Type							
Share of Never-Takers	0.3	0.3	0.241	0.361	0.03	1.002	2000
Share of Partial Compliers	0.076	0.075	0.046	0.113	0.017	1.003	1200
Share of Full Compliers	0.624	0.623	0.561	0.687	0.032	1.002	1800

Table D.5: Estimates of Principal Causal Effects on Net Discrimination against Blacks (vs. Whites) in Testers' Subjective Perceptions of Favorable Interactions during the Appointment

Parameter	Posterior Mean	Posterior Median	95% Credible Interval	Standard Deviation	\hat{R}	Effective N	
A. Estimates of the Distribution of Treatment Effects by Complier Type							
Monitoring vs. Control for Partial Compliers	0.026	0.027	-0.171	0.221	0.1	1.001	5000
Monitoring vs. Control for Full Compliers	0.043	0.043	-0.06	0.15	0.053	1.001	3300
Punitive vs. Control for Full Compliers	-0.029	-0.029	-0.112	0.053	0.042	1.003	1400
Punitive vs. Monitoring for Full Compliers	-0.072	-0.071	-0.175	0.028	0.052	1.001	5000
B. Estimates of the Distribution of Outcomes by Treatment Received and Complier Type							
Control Outcome for Never-Takers	0.218	0.217	0.156	0.283	0.033	1.001	5000
Control Outcome for Partial Compliers	0.355	0.351	0.218	0.516	0.076	1.001	5000
Monitoring Outcome for Partial Compliers	0.381	0.379	0.25	0.522	0.069	1.001	5000
Control Outcome for Full Compliers	0.2	0.198	0.143	0.262	0.03	1.002	2800
Monitoring Outcome for Full Compliers	0.243	0.241	0.164	0.336	0.043	1.001	3600
Punitive Outcome for Full Compliers	0.171	0.17	0.12	0.228	0.028	1.002	1600
C. Estimates of Proportions of Complier Type							
Share of Never-Takers	0.299	0.298	0.241	0.359	0.03	1.001	4900
Share of Partial Compliers	0.076	0.075	0.047	0.113	0.017	1.004	830
Share of Full Compliers	0.625	0.626	0.561	0.688	0.032	1.001	3900

Table D.6: Estimates of Principal Causal Effects on Net Discrimination against Blacks (vs. Whites) in Receiving Post-Visit Callbacks

Parameter	Posterior Mean	Posterior Median	95% Credible Interval	Standard Deviation	\hat{R}	Effective N	
A. Estimates of the Distribution of Treatment Effects by Complier Type							
Monitoring vs. Control for Partial Compliers	-0.011	-0.011	-0.204	0.176	0.098	1.001	5000
Monitoring vs. Control for Full Compliers	0.028	0.027	-0.06	0.123	0.048	1.001	5000
Punitive vs. Control for Full Compliers	-0.015	-0.015	-0.09	0.057	0.037	1.001	4000
Punitive vs. Monitoring for Full Compliers	-0.043	-0.043	-0.138	0.047	0.047	1.002	2500
B. Estimates of the Distribution of Outcomes by Treatment Received and Complier Type							
Control Outcome for Never-Takers	0.163	0.161	0.111	0.223	0.029	1.001	5000
Control Outcome for Partial Compliers	0.336	0.332	0.204	0.498	0.075	1.001	5000
Monitoring Outcome for Partial Compliers	0.324	0.322	0.204	0.459	0.066	1.001	5000
Control Outcome for Full Compliers	0.158	0.156	0.109	0.215	0.027	1.001	3300
Monitoring Outcome for Full Compliers	0.185	0.183	0.116	0.267	0.039	1.002	2400
Punitive Outcome for Full Compliers	0.142	0.141	0.096	0.196	0.026	1.001	5000
C. Estimates of Proportions of Complier Type							
Share of Never-Takers	0.298	0.298	0.241	0.358	0.03	1.003	1100
Share of Partial Compliers	0.077	0.076	0.046	0.115	0.018	1.003	1300
Share of Full Compliers	0.624	0.625	0.563	0.683	0.031	1.005	650

Table D.7: Estimates of Principal Causal Effects on Net Discrimination against Blacks (vs. Whites) in Receiving Post-Visit Offers

Parameter	Posterior Mean	Posterior Median	95% Credible Interval	Standard Deviation	\hat{R}	Effective N	
A. Estimates of the Distribution of Treatment Effects by Complier Type							
Monitoring vs. Control for Partial Compliers	-0.022	-0.021	-0.204	0.147	0.09	1.001	5000
Monitoring vs. Control for Full Compliers	0.052	0.051	-0.018	0.128	0.037	1.001	5000
Punitive vs. Control for Full Compliers	0.006	0.006	-0.047	0.062	0.028	1.001	3400
Punitive vs. Monitoring for Full Compliers	-0.046	-0.045	-0.124	0.024	0.037	1.001	5000
B. Estimates of the Distribution of Outcomes by Treatment Received and Complier Type							
Control Outcome for Never-Takers	0.09	0.088	0.054	0.136	0.021	1.001	5000
Control Outcome for Partial Compliers	0.296	0.291	0.17	0.454	0.071	1.001	5000
Monitoring Outcome for Partial Compliers	0.274	0.272	0.163	0.399	0.061	1.001	5000
Control Outcome for Full Compliers	0.082	0.08	0.048	0.124	0.019	1.001	5000
Monitoring Outcome for Full Compliers	0.134	0.132	0.08	0.204	0.031	1.001	5000
Punitive Outcome for Full Compliers	0.088	0.086	0.053	0.132	0.02	1.001	4200
C. Estimates of Proportions of Complier Type							
Share of Never-Takers	0.299	0.298	0.243	0.357	0.03	1.002	2000
Share of Partial Compliers	0.077	0.076	0.047	0.114	0.017	1.002	1900
Share of Full Compliers	0.624	0.624	0.561	0.685	0.032	1.004	950

Table D.8: Estimates of Principal Causal Effects on Net Discrimination against Hispanics (vs. Whites) in Landlords and Brokers Honoring Scheduled Appointments

Parameter	Posterior Mean	Posterior Median	95% Credible Interval	Standard Deviation	\hat{R}	Effective N	
A. Estimates of the Distribution of Treatment Effects by Complier Type							
Monitoring vs. Control for Partial Compliers	-0.025	-0.022	-0.209	0.147	0.09	1.001	5000
Monitoring vs. Control for Full Compliers	0.047	0.046	-0.021	0.121	0.036	1.001	4900
Punitive vs. Control for Full Compliers	0.006	0.006	-0.047	0.063	0.028	1.001	5000
Punitive vs. Monitoring for Full Compliers	-0.041	-0.039	-0.116	0.027	0.037	1.001	3300
B. Estimates of the Distribution of Outcomes by Treatment Received and Complier Type							
Control Outcome for Never-Takers	0.089	0.088	0.053	0.137	0.021	1.001	5000
Control Outcome for Partial Compliers	0.297	0.291	0.177	0.45	0.069	1.002	2300
Monitoring Outcome for Partial Compliers	0.272	0.269	0.161	0.401	0.062	1.002	2700
Control Outcome for Full Compliers	0.081	0.079	0.048	0.122	0.019	1.001	5000
Monitoring Outcome for Full Compliers	0.128	0.126	0.077	0.196	0.031	1.002	2500
Punitive Outcome for Full Compliers	0.088	0.086	0.051	0.133	0.021	1.001	5000
C. Estimates of Proportions of Complier Type							
Share of Never-Takers	0.298	0.298	0.239	0.357	0.03	1.002	1600
Share of Partial Compliers	0.077	0.076	0.047	0.115	0.018	1.004	910
Share of Full Compliers	0.625	0.625	0.565	0.687	0.031	1.004	910

Table D.9: Estimates of Principal Causal Effects on Net Discrimination against Hispanics (vs. Whites) in Testers' Subjective Perceptions of Favorable Interactions during the Appointment

Parameter	Posterior	Posterior	95% Credible Interval	Standard Deviation	\hat{R}	Effective N	
	Mean	Median					
A. Estimates of the Distribution of Treatment Effects by Complier Type							
Monitoring vs. Control for Partial Compliers	-0.006	-0.005	-0.205	0.187	0.1	1.002	3000
Monitoring vs. Control for Full Compliers	0.045	0.045	-0.059	0.153	0.054	1.001	5000
Punitive vs. Control for Full Compliers	-0.004	-0.004	-0.088	0.078	0.042	1.001	5000
Punitive vs. Monitoring for Full Compliers	-0.049	-0.048	-0.155	0.047	0.052	1.001	5000
B. Estimates of the Distribution of Outcomes by Treatment Received and Complier Type							
Control Outcome for Never-Takers	0.185	0.183	0.127	0.25	0.032	1.001	5000
Control Outcome for Partial Compliers	0.346	0.342	0.21	0.501	0.074	1.002	2700
Monitoring Outcome for Partial Compliers	0.34	0.337	0.213	0.477	0.068	1.002	2500
Control Outcome for Full Compliers	0.188	0.186	0.133	0.253	0.031	1.001	3600
Monitoring Outcome for Full Compliers	0.233	0.232	0.155	0.324	0.044	1.001	4200
Punitive Outcome for Full Compliers	0.184	0.183	0.13	0.242	0.029	1.001	5000
C. Estimates of Proportions of Complier Type							
Share of Never-Takers	0.299	0.298	0.242	0.357	0.029	1.006	560
Share of Partial Compliers	0.077	0.076	0.047	0.112	0.017	1.002	2800
Share of Full Compliers	0.624	0.625	0.562	0.685	0.031	1.005	610

Table D.10: Estimates of Principal Causal Effects on Net Discrimination against Hispanics (vs. Whites) in Receiving Post-Visit Callbacks.

Parameter	Posterior Mean	Posterior Median	95% Credible Interval	Standard Deviation	\hat{R}	Effective N	
A. Estimates of the Distribution of Treatment Effects by Complier Type							
Monitoring vs. Control for Partial Compliers	-0.045	-0.045	-0.233	0.142	0.096	1.001	4000
Monitoring vs. Control for Full Compliers	0.033	0.032	-0.058	0.13	0.048	1.001	5000
Punitive vs. Control for Full Compliers	0.013	0.013	-0.063	0.087	0.039	1.001	5000
Punitive vs. Monitoring for Full Compliers	-0.02	-0.019	-0.118	0.073	0.049	1.001	5000
B. Estimates of the Distribution of Outcomes by Treatment Received and Complier Type							
Control Outcome for Never-Takers	0.154	0.152	0.102	0.213	0.029	1.001	5000
Control Outcome for Partial Compliers	0.333	0.329	0.203	0.485	0.074	1.001	5000
Monitoring Outcome for Partial Compliers	0.288	0.284	0.175	0.423	0.063	1.002	2500
Control Outcome for Full Compliers	0.16	0.158	0.11	0.217	0.027	1.001	4500
Monitoring Outcome for Full Compliers	0.193	0.19	0.12	0.276	0.04	1.001	3800
Punitive Outcome for Full Compliers	0.173	0.172	0.123	0.228	0.027	1.001	5000
C. Estimates of Proportions of Complier Type							
Share of Never-Takers	0.3	0.3	0.242	0.359	0.03	1.002	2100
Share of Partial Compliers	0.078	0.077	0.047	0.116	0.017	1.002	2600
Share of Full Compliers	0.622	0.622	0.561	0.682	0.031	1.003	1300

Table D.11: Estimates of Principal Causal Effects on Net Discrimination against Hispanics (vs. Whites) in Receiving Post-Visit Offers.

Parameter	Posterior Mean	Posterior Median	95% Credible Interval	Standard Deviation	\hat{R}	Effective N	
A. Estimates of the Distribution of Treatment Effects by Complier Type							
Monitoring vs. Control for Partial Compliers	-0.022	-0.021	-0.208	0.154	0.091	1.001	5000
Monitoring vs. Control for Full Compliers	0.05	0.048	-0.016	0.127	0.037	1.001	5000
Punitive vs. Control for Full Compliers	0.006	0.005	-0.051	0.062	0.028	1.001	4000
Punitive vs. Monitoring for Full Compliers	-0.044	-0.043	-0.125	0.026	0.038	1.001	4500
B. Estimates of the Distribution of Outcomes by Treatment Received and Complier Type							
Control Outcome for Never-Takers	0.089	0.087	0.052	0.137	0.021	1.001	3900
Control Outcome for Partial Compliers	0.295	0.289	0.173	0.45	0.071	1.002	2200
Monitoring Outcome for Partial Compliers	0.273	0.269	0.163	0.404	0.062	1.002	2500
Control Outcome for Full Compliers	0.082	0.08	0.047	0.123	0.02	1.001	3100
Monitoring Outcome for Full Compliers	0.132	0.129	0.078	0.202	0.031	1.001	3600
Punitive Outcome for Full Compliers	0.087	0.086	0.051	0.132	0.02	1.001	5000
C. Estimates of Proportions of Complier Type							
Share of Never-Takers	0.299	0.299	0.241	0.36	0.03	1.002	3400
Share of Partial Compliers	0.077	0.075	0.046	0.113	0.017	1.003	1200
Share of Full Compliers	0.624	0.625	0.559	0.683	0.032	1.001	3000

Table D.12: Estimates of Principal Causal Effects on Net Discrimination against Hispanics (vs. Blacks) in Landlords and Brokers Honoring Scheduled Appointments.

Parameter	Posterior Mean	Posterior Median	95% Credible Interval	Standard Deviation	\hat{R}	Effective N	
A. Estimates of the Distribution of Treatment Effects by Complier Type							
Monitoring vs. Control for Partial Compliers	-0.023	-0.021	-0.2	0.146	0.089	1.001	5000
Monitoring vs. Control for Full Compliers	0.047	0.045	-0.019	0.122	0.036	1.001	5000
Punitive vs. Control for Full Compliers	0.006	0.006	-0.049	0.062	0.028	1.001	5000
Punitive vs. Monitoring for Full Compliers	-0.041	-0.04	-0.116	0.029	0.037	1.001	5000
B. Estimates of the Distribution of Outcomes by Treatment Received and Complier Type							
Control Outcome for Never-Takers	0.089	0.088	0.053	0.135	0.021	1.001	5000
Control Outcome for Partial Compliers	0.294	0.29	0.174	0.437	0.068	1.001	5000
Monitoring Outcome for Partial Compliers	0.271	0.268	0.164	0.402	0.061	1.001	5000
Control Outcome for Full Compliers	0.081	0.08	0.049	0.123	0.019	1.001	5000
Monitoring Outcome for Full Compliers	0.128	0.126	0.075	0.195	0.031	1.001	5000
Punitive Outcome for Full Compliers	0.087	0.086	0.051	0.132	0.021	1.001	3600
C. Estimates of Proportions of Complier Type							
Share of Never-Takers	0.299	0.299	0.243	0.357	0.029	1.005	720
Share of Partial Compliers	0.078	0.077	0.047	0.113	0.017	1.004	930
Share of Full Compliers	0.623	0.623	0.563	0.683	0.03	1.003	1400

Table D.13: Estimates of Principal Causal Effects on Net Discrimination against Hispanics (vs. Blacks) in Testers' Subjective Perceptions of Favorable Interactions during the Appointment.

Parameter	Posterior Mean	Posterior Median	95% Credible Interval	Standard Deviation	\hat{R}	Effective N	
A. Estimates of the Distribution of Treatment Effects by Complier Type							
Monitoring vs. Control for Partial Compliers	-0.027	-0.024	-0.219	0.159	0.097	1.001	4000
Monitoring vs. Control for Full Compliers	0.029	0.028	-0.065	0.125	0.049	1.001	5000
Punitive vs. Control for Full Compliers	-0.014	-0.015	-0.088	0.064	0.039	1.001	5000
Punitive vs. Monitoring for Full Compliers	-0.042	-0.042	-0.139	0.05	0.048	1.001	5000
B. Estimates of the Distribution of Outcomes by Treatment Received and Complier Type							
Control Outcome for Never-Takers	0.162	0.161	0.107	0.227	0.03	1.001	5000
Control Outcome for Partial Compliers	0.339	0.334	0.205	0.494	0.073	1.001	5000
Monitoring Outcome for Partial Compliers	0.312	0.308	0.19	0.448	0.066	1.001	3900
Control Outcome for Full Compliers	0.168	0.167	0.116	0.225	0.028	1.001	5000
Monitoring Outcome for Full Compliers	0.197	0.195	0.126	0.278	0.04	1.001	5000
Punitive Outcome for Full Compliers	0.154	0.153	0.105	0.21	0.027	1.001	5000
C. Estimates of Proportions of Complier Type							
Share of Never-Takers	0.3	0.299	0.245	0.36	0.029	1.006	560
Share of Partial Compliers	0.077	0.076	0.047	0.112	0.017	1.005	620
Share of Full Compliers	0.623	0.623	0.562	0.683	0.031	1.003	1100

Table D.14: Estimates of Principal Causal Effects on Net Discrimination against Hispanics (vs. Blacks) in Receiving Post-Visit Callbacks.

Parameter	Posterior	Posterior	95% Credible Interval	Standard Deviation	\hat{R}	Effective N	
	Mean	Median					
A. Estimates of the Distribution of Treatment Effects by Complier Type							
Monitoring vs. Control for Partial Compliers	-0.025	-0.023	-0.215	0.159	0.096	1.002	2300
Monitoring vs. Control for Full Compliers	0.007	0.006	-0.074	0.092	0.042	1.001	5000
Punitive vs. Control for Full Compliers	0.011	0.011	-0.066	0.085	0.038	1.002	2200
Punitive vs. Monitoring for Full Compliers	0.003	0.004	-0.084	0.087	0.043	1.001	5000
B. Estimates of the Distribution of Outcomes by Treatment Received and Complier Type							
Control Outcome for Never-Takers	0.145	0.144	0.096	0.206	0.028	1.001	5000
Control Outcome for Partial Compliers	0.328	0.323	0.198	0.485	0.074	1.002	2000
Monitoring Outcome for Partial Compliers	0.303	0.299	0.185	0.438	0.065	1.001	4300
Control Outcome for Full Compliers	0.146	0.145	0.097	0.201	0.027	1.002	1500
Monitoring Outcome for Full Compliers	0.154	0.151	0.096	0.227	0.034	1.001	5000
Punitive Outcome for Full Compliers	0.157	0.155	0.107	0.215	0.028	1.001	5000
C. Estimates of Proportions of Complier Type							
Share of Never-Takers	0.299	0.298	0.244	0.362	0.03	1.002	1800
Share of Partial Compliers	0.077	0.076	0.046	0.114	0.017	1.004	840
Share of Full Compliers	0.624	0.625	0.561	0.684	0.032	1.002	1900

Table D.15: Estimates of Principal Causal Effects on Net Discrimination against Hispanics (vs. Blacks) in Receiving Post-Visit Offers

D.4 Estimated Effects of Messaging by Putative Low versus High Employment Stability Signal

This section presents the full results from the models estimating the conditional effects of messaging by putative low (versus high) employment stability signal. Estimates presented in panel (I) correspond to a model specified with the control condition as the base (omitted) category. Estimates presented in panel (II) correspond to a model specified with the monitoring condition as the base (omitted) category.

Variable	Estimate	Std. Error	t value	Pr(> t)	Variable	Estimate	Std. Error	t value	Pr(> t)
A. Outcome: Net Discrimination in Receiving Callbacks									
(Intercept)	0.031	0.037	0.844	0.399	(Intercept)	0.038	0.058	0.649	0.517
Monitoring	0.007	0.069	0.099	0.921	Control	-0.007	0.069	-0.099	0.921
Punitive	0.064	0.062	1.025	0.306	Punitive	0.057	0.077	0.739	0.46
Low Empl Stability	0.033	0.051	0.656	0.512	Low Empl Stability	0.016	0.082	0.19	0.849
Monitoring:Low Empl Stability	-0.018	0.097	-0.183	0.855	Control:Low Empl Stability	0.018	0.097	0.183	0.855
Punitive:Low Empl Stability	-0.07	0.088	-0.795	0.427	Punitive:Low Empl Stability	-0.052	0.109	-0.478	0.633
B. Outcome: Net Discrimination in Receiving Offers									
(Intercept)	0.066	0.03	2.228	0.026	(Intercept)	-0.034	0.047	-0.72	0.472
Monitoring	-0.101	0.056	-1.796	0.073	Control	0.101	0.056	1.796	0.073
Punitive	-0.074	0.05	-1.467	0.143	Punitive	0.026	0.062	0.423	0.672
Low Empl Stability	-0.05	0.042	-1.197	0.232	Low Empl Stability	0.041	0.067	0.615	0.539
Monitoring:Low Empl Stability	0.091	0.079	1.152	0.25	Control:Low Empl Stability	-0.091	0.079	-1.152	0.25
Punitive:Low Empl Stability	0.112	0.072	1.564	0.118	Punitive:Low Empl Stability	0.021	0.089	0.237	0.812

Table D.16: Full Results: Estimated Conditional Effects of Messaging by Putative Low Employment Stability Signal: Net Discrimination against Blacks (versus Whites)

Variable	(I)				Variable	(II)			
	Estimate	Std. Error	t value	Pr(> t)		Estimate	Std. Error	t value	Pr(> t)
A. Outcome: Net Discrimination in Receiving Callbacks									
(Intercept)	0.049	0.035	1.418	0.157	(Intercept)	0.018	0.055	0.327	0.744
Monitoring	-0.031	0.065	-0.479	0.632	Control	0.031	0.065	0.479	0.632
Punitive	-0.024	0.059	-0.401	0.689	Punitive	0.008	0.073	0.105	0.916
Low Empl Stability	0.024	0.048	0.495	0.621	Low Empl Stability	0.046	0.078	0.593	0.553
Monitoring:Low Empl Stability	0.022	0.092	0.244	0.807	Control:Low Empl Stability	-0.022	0.092	-0.244	0.807
Punitive:Low Empl Stability	-0.052	0.083	-0.628	0.53	Punitive:Low Empl Stability	-0.075	0.103	-0.722	0.47
B. Outcome: Net Discrimination in Receiving Offers									
(Intercept)	0.085	0.029	2.943	0.003	(Intercept)	0.004	0.046	0.095	0.924
Monitoring	-0.081	0.055	-1.487	0.138	Control	0.081	0.055	1.487	0.138
Punitive	-0.103	0.049	-2.103	0.036	Punitive	-0.022	0.061	-0.368	0.713
Low Empl Stability	-0.03	0.04	-0.744	0.457	Low Empl Stability	0.008	0.065	0.12	0.904
Monitoring:Low Empl Stability	0.038	0.077	0.493	0.622	Control:Low Empl Stability	-0.038	0.077	-0.493	0.622
Punitive:Low Empl Stability	0.063	0.07	0.896	0.371	Punitive:Low Empl Stability	0.025	0.087	0.284	0.777

Table D.17: Full Results: Estimated Conditional Effects of Messaging by Putative Low Employment Stability Signal: Net Discrimination against Hispanics (versus Whites)

Variable	(I)				Variable	(II)			
	Estimate	Std. Error	t value	Pr(> t)		Estimate	Std. Error	t value	Pr(> t)
A. Outcome: Net Discrimination in Receiving Callbacks									
(Intercept)	0.018	0.034	0.53	0.596	(Intercept)	-0.02	0.055	-0.36	0.719
Monitoring	-0.038	0.065	-0.587	0.557	Control	0.038	0.065	0.587	0.557
Punitive	-0.087	0.058	-1.49	0.137	Punitive	-0.049	0.072	-0.678	0.498
Low Empl Stability	-0.01	0.048	-0.198	0.843	Low Empl Stability	0.031	0.078	0.394	0.693
Monitoring:Low Empl Stability	0.04	0.091	0.44	0.66	Control:Low Empl Stability	-0.04	0.091	-0.44	0.66
Punitive:Low Empl Stability	0.018	0.083	0.212	0.832	Punitive:Low Empl Stability	-0.023	0.103	-0.219	0.827
B. Outcome: Net Discrimination in Receiving Offers									
(Intercept)	0.019	0.027	0.704	0.482	(Intercept)	0.039	0.043	0.893	0.372
Monitoring	0.019	0.051	0.381	0.703	Control	-0.019	0.051	-0.381	0.703
Punitive	-0.029	0.046	-0.64	0.523	Punitive	-0.049	0.057	-0.858	0.391
Low Empl Stability	0.02	0.038	0.518	0.605	Low Empl Stability	-0.033	0.061	-0.546	0.585
Monitoring:Low Empl Stability	-0.053	0.072	-0.737	0.461	Control:Low Empl Stability	0.053	0.072	0.737	0.461
Punitive:Low Empl Stability	-0.049	0.065	-0.759	0.448	Punitive:Low Empl Stability	0.003	0.081	0.043	0.966

Table D.18: Full Results: Estimated Conditional Effects of Messaging by Putative Low Employment Stability Signal: Net Discrimination against Hispanics (versus Blacks)

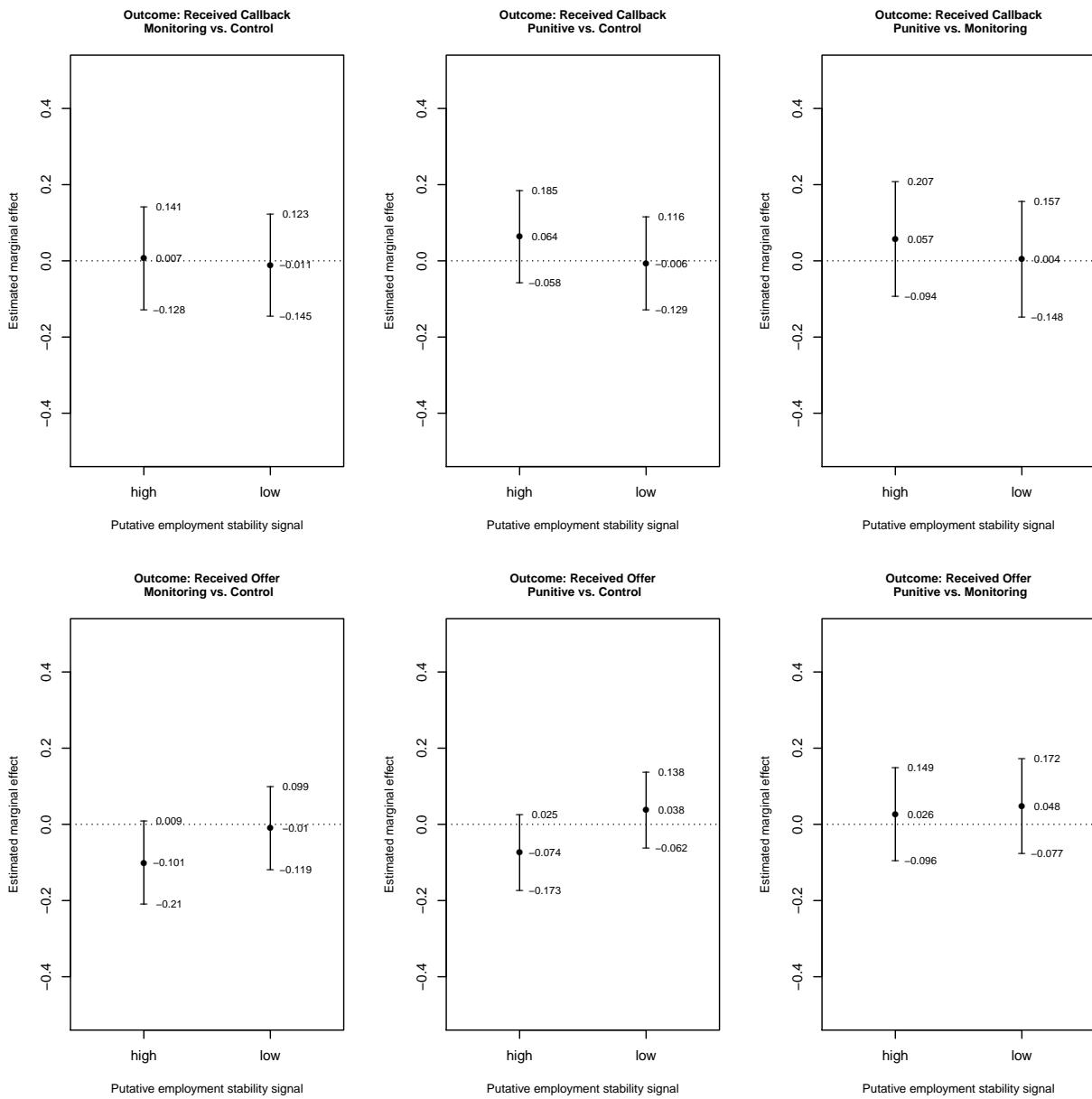


Figure D.31: Estimated Marginal Effects of Messaging Treatments by Putative Low Employment Stability Signal on Net Discrimination against Blacks (versus Whites)

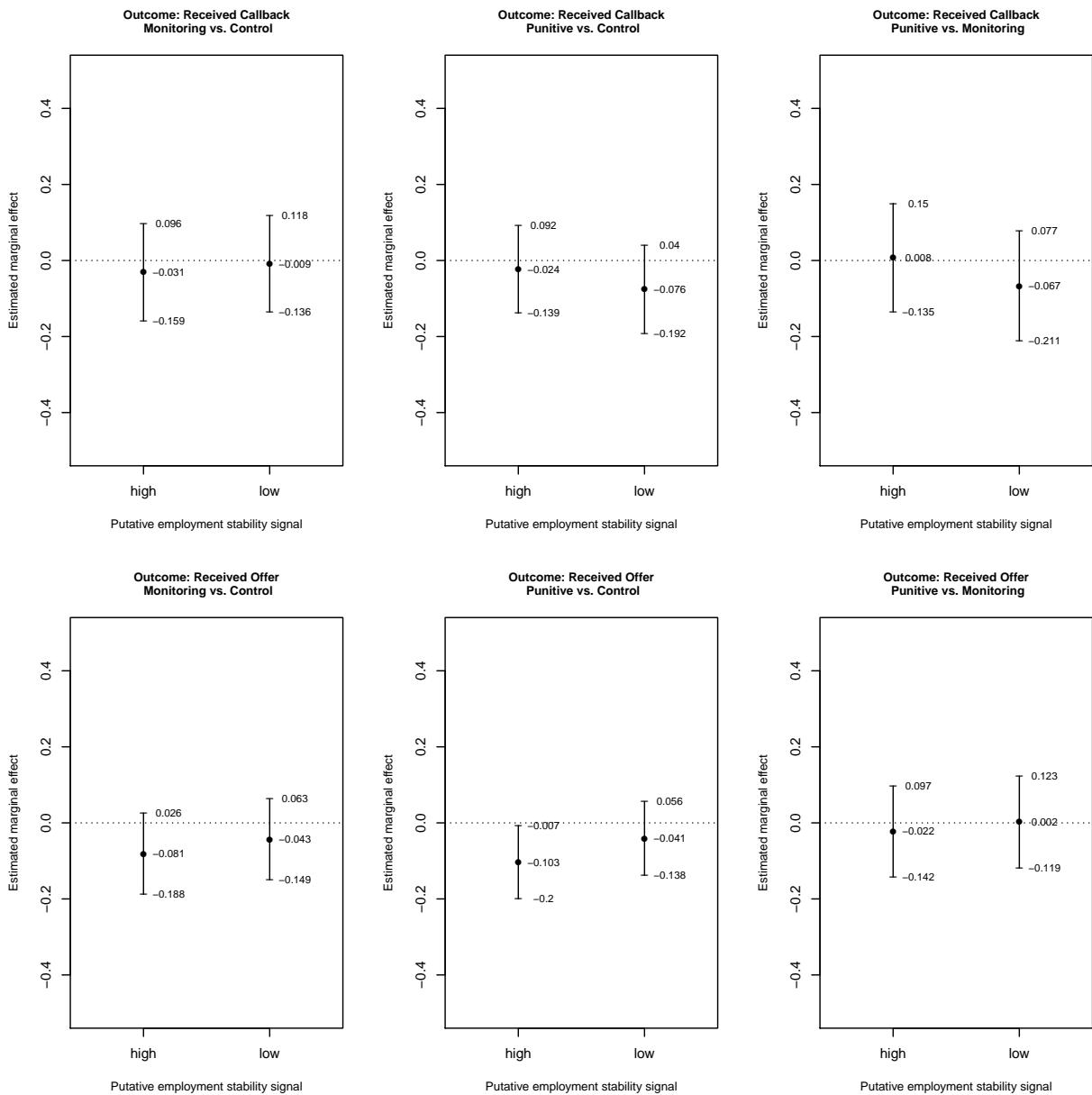


Figure D.32: Estimated Marginal Effects of Messaging Treatments by Putative Low Employment Stability Signal on Net Discrimination against Hispanics (versus Whites)

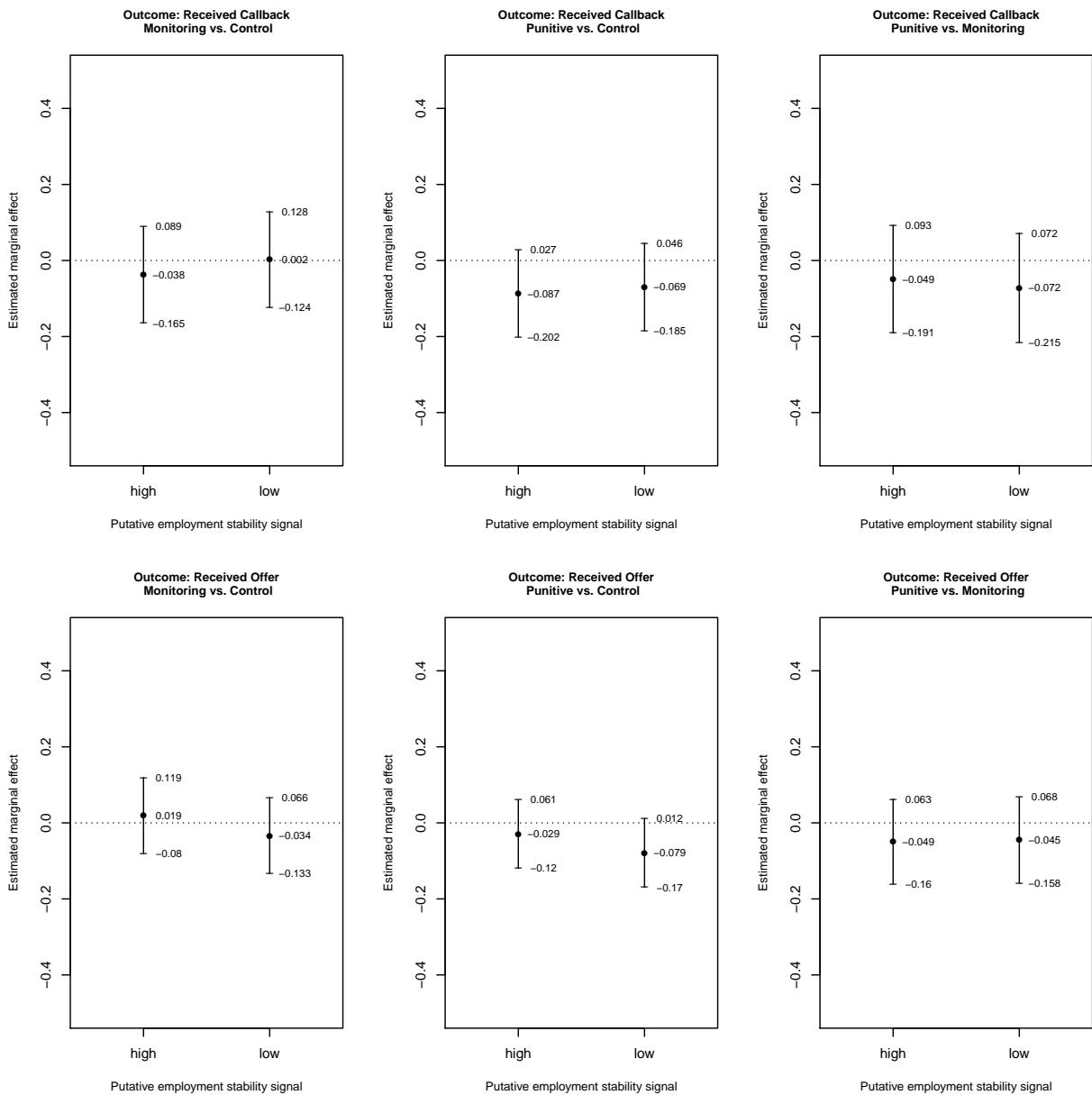


Figure D.33: Estimated Marginal Effects of Messaging Treatments by Putative Low Employment Stability Signal on Net Discrimination against Hispanics (versus Blacks)

Bibliography

- Kahan, Dan M. 2000. "Gentle nudges vs. hard shoves: Solving the sticky norms problem." *The University of Chicago Law Review* pp. 607–645.
- Mayo, Elton. 1949. *Hawthorne and the Western Electric Company*. Vol. 999 of *Spatial Choices and Processes* 999 ed. Holland: Elsevier Science Publishers 999 9, pp. 203–217.
- Mendelberg, Tali. 2001. *The race card: Campaign strategy, implicit messages, and the norm of equality*. Princeton University Press.
- Thaler, Richard H and Cass R Sunstein. 2008. *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.
- Turner, Margery A., Claudia Aranda, Diane K. Levy, Rob Pitingolo, Rob Santos and Doug Wissoker. 2013. Housing Discrimination Against Racial And Ethnic Minorities 2012. Technical report U.S. Department of Housing and Urban Development, Office of Policy Development and Research.