

Moving Beyond Measurement: Adapting Audit Studies to Test Bias-Reducing Interventions

Daniel M. Butler* and Charles Crabtree†

Abstract

This paper discusses how audit studies can be adapted to test the effectiveness of interventions aimed at reducing discrimination. We conducted an adapted audit experiment to test whether making officials aware of bias could reduce levels of racial bias. While the limitations of our design make it difficult to assess where information alone can reduce bias, our study makes two important contributions. First, we replicate prior studies by showing that white, local elected officials are less responsive to black constituents. That local officials exhibit biased behavior is particularly worrisome, as local government is often the level that most directly affects citizens' daily lives. Second, we provide several suggestions for future audit studies that draw from the strengths and weaknesses of our own design. We hope that they will help improve future work on identifying and reducing discrimination.

Keywords: audit studies, racial bias, local politics

This paper discusses how audit studies can be adapted to test the effectiveness of interventions aimed at reducing discrimination. While the approach we describe can be applied to studies of discrimination in many areas, we focus on racial discrimination because it affects numerous facets of life including—but not limited to—housing (Turner et al. 2002), employment (Bertrand and Mullainathan 2003; Gaddis 2014; Pager 2007), health (Burgess et al. 2007; Schulman et al. 1999), and civic life (Griffin and Newman 2008). Audit studies have become an increasingly popular tool for measuring discrimination across the social sciences (Vuolo et al. 2015). Most audit studies, however, simply focus on measuring the extent to which bias exists. We build on those works by showing how audit studies can be combined

The research reported in this paper was reviewed by the Pennsylvania State University Institutional Review Board (#STUDY00001114). Authors are listed alphabetically and contributed equally to study design, data collection, analysis, and manuscript preparation. All data and computer code necessary to replicate the results in this analysis will be made publicly available upon publication. We thank two anonymous reviewers for extremely helpful comments.

*Associate Professor, University of California, San Diego, 9500 Gilman Drive, #0521, La Jolla, CA 92093-0521, e-mail: daniel.butler@gmail.com

†PhD Candidate, University of Michigan, 5700 Haven Hall, 505 S. State, Ann Arbor, MI 48105, e-mail: crcrabtr@umich.edu

with randomly assigned interventions as a way of testing the effectiveness of those interventions.

We implemented an adapted audit study to test whether information can be used to reduce levels of racial discrimination that local officials exhibit toward constituents seeking answers to basic questions. We focus on an informational treatment because evidence suggests that many people do not realize that they are acting in a biased manner (Dovidio et al. 2002). Consequently, some have advocated using information to make people aware of their biases (Burgess et al. 2007; Devine and Monteith 1999; Pope et al. 2013; Rudman et al. 2001). For our study, we provided a random subset of the local officials in our sample with information about previous evidence of racial bias in how elected officials respond to citizens' requests. A few weeks later, we then conducted an audit study to measure levels of racial bias among all the officials in our sample (Adida et al. 2010; Butler and Broockman 2011; McClendon Forthcoming), both those who were sent the information and those who were not sent the information. We did this by sending an email to each public official in our sample, randomizing whether the request for help came from a putatively black or putatively white constituent. We then measured the level of racial bias by comparing the response rates to the white and black constituents.

Our audit study successfully replicates prior work in this area by showing that officials exhibit bias against black constituents. Indeed, the extent of bias is comparable to levels found by previous studies of officials at other levels of government. Because local officials often help constituents with the services that they use most regularly, discrimination at this level of government raises concerns about fairness in representation.

Our results regarding the effect of information on reducing bias are less conclusive. In our study, the information treatment group and the control group exhibited the same level of racial discrimination. While this is consistent with the possibility that information alone is insufficient to reduce bias, there are limitations with the design that may also be responsible for this null result. After presenting the details and results of our adapted audit study, we discuss those limitations and what may have been done differently.

We end our paper with a list of concrete recommendations for how to improve future audit studies. Our points draw on the lessons learned from both the limitations and strengths of our particular study. We provide these suggestions as a way to improve future work on identifying and reducing discrimination.

TESTING THE EFFECT OF INFORMATION

We use an audit study to test whether information can reduce racial bias. In our audit study, we sent elected, municipal officials (i.e., mayors and city councilors) in the U. S. requests for assistance from putative constituents, randomizing whether

the request came from someone with a distinctively black name or a distinctively white name.

We chose our sample to learn about how white, elected officials at the municipal level treat white and black constituents. We focus on local officials because they are responsible for the provision of many important public services and oversee the government employees with whom citizens most regularly interact. Because we could not identify the race of all city officials, we restricted our sample to the types of cities where, based on the racial make up of the city, the vast majority of elected officials are likely to be non-Latino, whites. This is important from a methodological perspective because if the sample represented a more racially diverse population, the in-group bias exhibited by officials from different racial and ethnic groups may have cancelled each other out (Butler and Broockman 2011).

We used the 2011 International City/County Management Association (ICMA) city survey, which includes information on the racial demographics of the elected officials, and data from the U.S. census to determine which cities to include in our sample. Based on these data, we restricted our sample to cities where 75% or more of the population is white and less than 15% of the population is Latino. On average, 96% of the officials serving in cities with these racial demographics are non-Latino, whites (see Figure 1 in the SI). We also restricted our sample to officials from cities with at least 50 black individuals and a total population of 3,000 or more.¹ We use these cut points to increase the plausibility that municipal officials would receive an email from a black constituent that they did not know. Our final sample included 11,801 city officials² from 2,160 cities from across the United States.

The email addresses for the elected city officials were collected by research assistants through web searches. Research assistants first searched for the website of each town or city taken from the census. If the research assistants were able to identify the city's website, they then collected the name and email address of the city's mayor and council members (or the equivalent).

To measure levels of differential treatment by race, we emailed all of the 11,801 officials in the sample, randomizing whether the email was sent from either a putatively black or a putatively white individual. We used information from previous studies and the U.S. census to identify common first and last names for black and white individuals (Butler 2014). In total, we used 76 different aliases: 35 black aliases and 41 white aliases (see Appendix E in the SI for list of aliases). For each alias, we created a separate Gmail account. When the emails were sent, the alias, which signaled the individual's race, was included both in the field identifying the sender and in the salutation in the body of the message. We carried out the study during the Fall of 2014.

¹The mean city in our sample has 23,949 residents, of whom 4.23% are black. Cities range in size from 3,011 to 583,776 people, and the black population within cities ranges from 50 to 42,188.

²This excludes the 338 officials with bad email addresses.

We carried out the study so that no two officials from the same city received a request from the same putative constituent. In particular, we first assigned officials to receive either the black or white constituent email and then block randomized, by city, aliases from the assigned racial treatment. Similarly, the request found in each email was randomly drawn from a list of simple requests adapted from the “frequently asked questions” sections of various city websites. We used 27 different requests, because all of the cities in the sample had 27 or fewer elected officials, and again block randomized so that no two officials from the same city received the same request (see Appendix F in the SI for the full list of requests used). We also randomized when we sent the emails so that they went out over a 5-day period. We spread the emails out to help ensure that no city in the study received all of the emails on the same day.

About two weeks earlier, we sent the elected officials assigned to the treatment group ($n = 4,004$) a pair of emails with information about prior research on racial bias exhibited by public officials.³ The information email we sent to the treatment group was ostensibly an invitation to take a survey about recent research on racial bias in how officials deal with constituent communications. We used this approach because it provided a rationale for sharing information about previous research, while minimizing the possibility that officials would assume that we were monitoring their behavior. The email states that research has shown “that office holders respond more often and provide better advice to individuals like them. White legislators, for instance, respond at higher rates to inquiries from white constituents than from black constituents.” We underlined this sentence and placed it in bold to draw attention to it. Within the main body of the email, we also provided links to (a) an NPR report on research showing racial bias among public officials and (b) a website we created to further highlight the findings from the racial bias literature (see <http://n.pr/1SRr6VA> for the NPR report and Appendix C in the SI for screenshots of the website). The sites that we linked to focused on the empirical finding that white legislators respond at higher rates to inquiries from white constituents than from black constituents, reinforcing the information provided in the email. Below the valediction of the email, we also included a one-paragraph summary of findings from the racial bias literature (see Appendix B in the SI for full email text).

Finally, we designed the study so that we could test both the direct effect of receiving the information on the recipient’s behavior, and also assess whether there was any spillover effect that changed the behavior of the other officials in the same city. We are able to test for this type of contagion effect because we used a multilevel design (Sinclair et al. 2012) where we first randomly assigned the cities into two groups: in one-third of the cities no officials received information; half of the officials in the remaining cities were randomly chosen to receive the information. After excluding the bad email addresses, we had three randomly chosen groups of

³The only difference between the waves was that the second email began with “Thanks to those who have already read this email and taken the survey. If you have not, please continue reading.”

officials: (a) 4,004 officials who received the information, (b) 3,963 officials who did not receive the information, but who serve with officials that did, and (c) 3,834 officials who did not receive the information and whose colleagues also did not receive the information. If the information had a direct effect on reducing bias, but no spillover effect, then the officials in group (a) should exhibit less bias than those in (b) and (c) and there should be no difference between (b) and (c). If there is a spillover effect, then the people in (a) and (b) should both exhibit less bias than officials in group (c). And if there is no effect at all, then all three groups should exhibit similar levels of bias.

RESULTS

Of the 11,801 contacted elected officials, 7,135 sent at least one email within two weeks of receiving our requests, for a response rate of approximately 60%.⁴ However, in line with previous research, not all putative constituents were treated equally. Elected municipal officials responded to emails from white constituents about 63% of the time (3,732/5,908) but only responded to emails from black constituents about 58% of the time (3,407/5,893). The five percentage point difference [3.6–7.1, 95% confident interval] between response rates is statistically significant⁵ and in line with other findings on racial bias among elected officials (Butler 2014; Costa 2016). A treatment effect of this size suggests that blacks and whites receive quite different treatment by local elected officials. All discrimination by elected officials harms the individuals discriminated against as well as the broader society, but discrimination by *local* elected officials can be particularly harmful. This is because the actions of local officials influence the lives of their constituents in fundamental ways: they administer many of the public services that constituents use, they oversee many of the street-level bureaucrats that constituents interact with, and they are often the first to be contacted by constituents who need assistance. It should then cause considerable concern that local officials treat constituents differently on the basis of race.

Unfortunately, providing information about racial bias to public officials did not appear to change their behavior. Table 1 gives the mean response rates by treatment, as well as the differences between how they treated black and white aliases. All three groups in our sample were about five percentage points more likely to respond to an email from a white alias than they were to respond to an email from a black alias. Because there does not seem to be evidence of a contagion effect, Table 2 directly compares the level of bias among those who were assigned to be made aware of the research and those who were not sent any information. Individuals assigned to receive the information treatment exhibited racial bias that was about 0.7 percentage points greater than those exhibited by the control group. Regardless

⁴We do not count autoresponses as replies.

⁵ $t = 5.96, p < 0.01$.

Table 1
Email Replies from Elected Municipal Officials

	Directly informed	Colleagues informed	No one in city informed
Response rate to black names	57.0% [N = 2,017]	58.6% [N = 1,949]	57.8% [N = 1,927]
Response rate to white names	62.8% [N = 1,987]	63.3% [N = 2,014]	63.4% [N = 1,907]
Racial bias	-5.8* (1.5)	-4.7* (1.6)	-5.6* (1.6)
Bias difference: Directly—no one (Std. error)		-0.2 (2.2)	
Bias difference: Colleagues—no one (Std. error)		0.9 (2.2)	

Note: Standard errors are shown in parentheses. * $p < 0.05$, two-tailed. The dependent variable, *Email Reply*, is coded as 1 if an elected municipal official sends a non-automated response to an information request within two weeks of receiving our email and 0 otherwise. While it appears that the directly treated group responded at a lower rate than the spillover or control groups, this difference is not statistically significant at any meaningful level.

Table 2
Email Replies from Elected Municipal Officials

	Directly informed	Not directly informed
Response rate to black names	57.0% [N = 2,017]	58.2% [N = 3,876]
Response rate to white names	62.8% [N = 1,987]	63.4% [N = 3,921]
Racial bias	-5.8* (1.5)	-5.1* (1.1)
Bias difference: Directly informed—not informed (Std. error)	-0.7 (1.9)	

Note: Standard errors are shown in parentheses. * $p < 0.05$, two-tailed. The dependent variable, *Email Reply*, is coded as 1 if an elected municipal official sends a non-automated response to an information request within two weeks of receiving our email and 0 otherwise.

of which way the control group is defined, the level of racial discrimination is comparable across the different groups. There are small differences in the level of bias (ranging from 4.7 to 5.8 percentage points), but the differences between these groups are statistically insignificant and point in the wrong direction: the largest difference was observed among those who were sent information. Being made aware of previous research on bias did not decrease the level of racial bias and did not appear to have much of a spillover effect.

LIMITATIONS

Our treatment, however, might have been too weak to cause a noticeable effect. First, not every official assigned to receive the information about previous results was actually exposed to that information. Some individuals simply did not open either of the two emails we sent. We know this because we sent the emails with

information on previous research through an email service that tracks whether recipients open their email. In our sample, 53% of the group assigned to receive the information treatment opened one or more of the emails with this treatment.

Second, even those who opened our treatment emails might not have read far enough to see the treatment. The key piece of our informational treatment appears in the third sentence of the email. Although we bolded and underlined the treatment text, we cannot know if people read past the first two sentences.

Third, even if officials did read our informational treatment, they might not have paid attention to it. The email referred to scholarly research and officials might be uninterested in academic research. Also, officials might not believe that the findings discussed in the email apply to them personally. While we cannot know for certain whether individuals carefully read our treatment emails, or thought that the content applied to them, we have anecdotal evidence that this might not be the case. We know that only 18% of the treatment group visited the survey link that was in the email, that only 15% of the treatment group completed the survey, and that only 4% of the treatment group clicked on the link to the research summary. Taken together, these statistics suggest that officials were not particularly interested in the information that we provided them.

The timing between our treatment and audit emails is another potential limiting factor. We sent the informational email to the treatment group a couple of weeks before conducting the audit because we wanted to assess whether our informational treatment had a meaningful long-term effect. Because our treatment was weak, it might have been unreasonable to expect that it had such a lasting effect. Perhaps, if we would have sent the audit emails shortly after our treatment emails, we would have found an effect.

Further, when conducting audit studies with elites' email addresses, researchers can rarely be certain that the individual they email is the one that reads and responds to their email. Elites often employ staffers to handle their correspondence. It is possible that one staffer read the informational, treatment email, and that another dealt the audit email. This might occur if staffers rotate email correspondence duties, there is turnover in staff, or staffers go on vacation.

SUGGESTIONS FOR FUTURE AUDIT STUDIES

Below is a list of suggestions for future audit studies. These suggestions are drawn from both what worked in our study and from what could have been improved. We discuss our suggestions in relation to our study to provide concrete examples of how these points can be implemented.

1. **Put the treatment at the forefront.** The information treatment may simply have been missed. If we were to do this again, we would place the treatment language

as close to the top of the email as possible. If it could not be placed in the first couple of sentences, we would put the message in color or highlight it.

2. **Increase the relevance of the treatment.** Individuals might not have thought that the information in the treatment applied to them. One could address this issue by increasing the perceived importance or influence of the sender. Our treatment may have been more powerful if the sender had been an individual or representative of a group from the same general geographic area, or from the local officials own constituency. It also might have been more powerful if it was sent from someone who could credibly sanction the officials if their future behavior was deemed inappropriate.
3. **Carefully consider how long to wait between the treatment intervention and the audit study.** Researchers are more likely to find treatment effects if the interval between treatment and the outcome is short, but in many cases, scholars might care more about the long-term effects of treatment. It is not always clear, unfortunately, how long treatment effects should last to be substantively meaningful. The proper length of time should be carefully considered at the design stage.
4. **Consider automatically coding the outcome measure.** We created our dependent variable by reading approximately 8,000 email responses, identifying the sender, and then coding whether they had replied to our email.⁶ As researchers who conduct email audit studies know, this can be a time-consuming procedure. If we were to do this again, we would use software to automatically process email responses and construct a response indicator. Our replication files contain an annotated R script and example data file that researchers can adapt for this purpose.
5. **Have a sufficiently large pool of email texts.** If officials think they are being studied, they are less likely to exhibit socially undesirable behavior (Findley and Nielson 2015). This can be a problem for audit studies that send the exact same text to all officials. This is because officials might share the communications they receive (or share staff who respond to emails). If this is the case, individuals might know that they are being studied and may simply respond differently to those communications. Having a sufficiently large number of different email texts helps minimize this threat to validity. Ideally, one would have enough different texts so that the officials who talk with each other do not receive the exact same email message. In our case, we had enough different email texts so that no one serving in the same city received the same message. In developing these questions, we looked for city pages that included answers to “frequently asked questions.” Those questions then helped us to write questions that would be relevant to our target population, thereby increasing the believability of the emails.
6. **Have a sufficiently large number of aliases and email accounts.** This will also minimize the possibility that officials respond to the emails differently than they normally would.

⁶Some officials sent more than one reply.

7. **Consider spillover at the design stage.** One of the advantages of studying public officials is that they have publicly observed social networks. Researchers can use the design proposed by Sinclair et al. (2012) to test for spillover in those networks.
8. **If the treatment designed to lower bias is delivered via email, track who opens the email.** Because we tracked who opened the emails, we were in a position to discuss what the bounds of the effect might be. If we were to do it over again, we would have also sent a placebo email (which included an invitation to participate in a survey on a different topic) to the control group and tracked who opened those emails. This would have given us a placebo design that would have increased the power of our analysis (Nickerson 2005; Gerber et al. 2010).

CONCLUSION

We conducted an adapted audit experiment to test whether making officials aware of bias could reduce levels of racial bias. Although limitations in our design make it difficult to assess whether information alone can reduce bias, our study makes two important contributions. First, we find that white, local, elected officials in the United States are less responsive to black constituents. The extent of this bias is in line with prior studies (Butler 2014; Butler & Broockman 2011; Costa 2016), suggesting that blacks face a similar degree of discrimination across governmental contexts. The fact that we find such bias among local government officials is worrying, however, as local government is often the level that most directly affects citizens' daily lives. Reducing this bias should be an important goal.

Second, we have described how audit studies can be adapted to help learn what measures help reduce bias. We draw these lessons from what worked in our experience and from what did not work. We offer these suggestions as a way to help improve future work on identifying and reducing discrimination.

SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit Cambridge Journals Online: <https://doi.org/10.1017/XPS.2017.11>.

REFERENCES

- Adida, Claire L., David D. Laitin, and Marie-Anne Valfort. 2010. "Identifying Barriers to Muslim Integration in France." *Proceedings of the National Academy of Sciences* 107(52): 384–90.

- Bertrand, Marianne and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal: A Field Experiment on Labor Market Discrimination." *American Economic Review* 94(4): 991–1013.
- Burgess, Diana, Michelle van Ryn, John Dovidio, and Somnath Saha. 2007. "Reducing Racial Bias Among Health Care Providers: Lessons from Social-Cognitive Psychology." *Journal of General Internal Medicine* 22(6): 882–7.
- Butler, Daniel M. 2014. *Representing the Advantaged: How Politicians Reinforce Inequality*. New York: Cambridge University Press.
- Butler, Daniel M. and David E. Broockman. 2011. "Do Politicians Racially Discriminate against Constituents? A Field Experiment on State Legislators." *American Journal of Political Science* 55(3): 463–77.
- Costa, Mia. 2016. "How Responsive are Political Elites? A Meta-Analysis of Experiments on Public Officials." Working Paper.
- Devine, Patricia G. and Margo J. Monteith. 1999. "Automaticity and Control in Stereotyping." In *Dual-Process Theories in Social Psychology*, eds. Chaiken and Y. Trope (pp. 339–360). New York: Guilford Press.
- Dovidio, John F., Kerry Kawakami and Samuel L. Gaertner. 2002 "Implicit and Explicit Prejudice and Interracial Interaction." *Journal of Personality and Social Psychology* 82(1): 62.
- Findley, Michael and Daniel Nielson. 2015. "Obligated to Deceive? Aliases, Confederates, and the Common Rule in International Field Experiments.? In *Ethics and Experiments: Problems and Solutions for Social Scientists and Policy Professionals*, ed. Scott Desposato. New York: Routledge.
- Gerber, Alan S., Donald P. Green, Edward H. Kaplan, and Holger L. Kern. 2010. "Baseline, Placebo, and Treatment: Efficient Estimation for Three-Group Experiments." *Political Analysis* 18(3): 297–315.
- Gaddis, S. Michael. 2014. "Discrimination in the Credential Society: An Audit Study of Race and College Selectivity in the Labor Market." *Social Forces* 93(4): 1451–79.
- Griffin, John D. and Brian Newman. 2008. *Minority Report: Evaluating Political Equality in America*. Chicago: University of Chicago Press.
- McClendon, Gwyneth. 2016. "Race and Responsiveness: A Field Experiment with South African Politicians." *Journal of Experimental Political Science* 3(1): 60–74.
- Nickerson, David W. 2005. "Scalable Protocols Offer Efficient Design for Field Experiments." *Political Analysis* 13(3): 233–52.
- Pager, Devah. 2007. "The Use of Field Experiments for Studies of Employment Discrimination: Contributions, Critiques, and Directions for the Future." *The Annals of the American Academy of Political and Social Science* 609(1): 104–33.
- Pope, Devin G., Joseph Price, and Justin Wolfers. 2013. "Awareness Reduces Racial Bias." *National Bureau of Economic Research* No. w19765.
- Rudman, Laurie A., Richard D. Ashmore, and Melvin L. Gary. 2001. "Unlearning" Automatic Biases: The Malleability of Implicit Prejudice and Stereotypes." *Journal of Personality and Social Psychology* 81(5): 856–68.
- Schulman, Kevin A., et al. 1999. "The Effect of Race and Sex on Physicians' Recommendations for Cardiac Catheterization." *New England Journal of Medicine* 340(8): 618–26.
- Sinclair, Betsy, Margaret McConnell, and Donald P. Green. (2012). "Detecting Spillover Effects: Design and Analysis of Multilevel Experiments." *American Journal of Political Science* 56(4): 1055–69.

- Turner, Margery A., Stephen Ross, George C. Galster, and John Yinger. 2002. *Discrimination in Metropolitan Housing Markets: National Results from Phase 1 of the Housing Discrimination Study (HDS)* (No. 2002-16).
- Vuolo, Mike, Christopher Uggen, and Sarah Lageson. 2015. "Statistical Power in Experimental Audit Studies: Cautions and Calculations for Matched Tests with Nominal Outcomes." *Sociological Research & Methods*. DOI: 10.1177/0049124115570066.