



Published in final edited form as:

J Econ Behav Organ. 2009 June 1; 70(3): 508–514. doi:10.1016/j.jebo.2008.02.009.

Computerizing Audit Studies

Joanna N. Lahey¹ and

Assistant Professor, The Bush School of Government, Texas A&M University

Ryan A. Beasley²

Assistant Professor, Department of Engineering Technology and Industrial Distribution, Texas A&M University

Abstract

This paper briefly discusses the history, benefits, and shortcomings of traditional audit field experiments to study market discrimination. Specifically it identifies template bias and experimenter bias as major concerns in the traditional audit method, and demonstrates through an empirical example that computerization of a resume or correspondence audit can efficiently increase sample size and greatly mitigate these concerns. Finally, it presents a useful meta-tool that future researchers can use to create their own resume audits.

I. Benefits of Audit Studies

Field experiments have gained popularity in recent years (Harrison and List 2004 and List 2006 provide reviews). There are several types of field experiments in the Harrison and List (2004) taxonomy, varying from “artefactual field experiments,” which are similar to laboratory experiments but are performed on real agents instead of on students, to “framed field experiments,” performed on real agents but in a controlled setting that is in a context familiar to these agents, to “natural field experiments,” in which the environment is less controlled and agents do not know they are being studied. This paper briefly discusses one type of “natural field experiment,” the audit study, explains difficulties and benefits of this type of experiment, and provides a new tool to mitigate some of those difficulties.

The audit study, used by urban and labor economists, is one of the earliest versions of the natural field experiment (Riach and Rich 2002 provide a review). Unlike many other field experiments (e.g., Ashraf et al. 2006, Karlan and List 2007, Lucking-Reily 1999, among many others), audit studies developed separately from the experimental economics tradition. The audit study was used to study discrimination as early as 1955 (Yinger 1995). Early studies sent pairs of trained “auditors,” matched in all respects except the variable of interest, usually race, to rent an apartment or to buy a house. Later this method was

The corresponding author can be reached at The Bush School of Government, TAMU 4220, Texas A&M University, College Station, TX 77843-4220 or at jlahey@nber.org, Fax (979) 845-4155, Phone (979) 458-3463.

¹This work was supported by funds from NSF Doctoral Dissertation Grant # 238 7480 and National Institute on Aging NBER Grant # T32-AG00186. We thank Elizabeth O. Ananat, M. Rose Barlow, Dora L. Costa, John A. List, Ren Mu, Lori Taylor, Christopher R. Wagner and two anonymous referees for helpful discussion and comments.

²The author can be reached at TAMU 3367, Texas A&M University, College Station, TX 77845-3367 or at beasley@entc.tamu.edu.

expanded to explore labor markets and other areas where discrimination could occur, such as flagging down a taxicab (e.g., Fix and Struyk 1993, Ridley et al. 1989).

Correspondence or resume audits, where resumes or letters are sent instead of matched pairs of people, have been popular in Europe (Jowell and Prescott-Clarke 1969, Rich and Riach 1991) and have been used recently in the United States (Bertrand and Mullainathan 2004, Lahey 2008). Although these types of audits can only measure the interviewing stage of the hiring process, they allow the experimenter much more control over the experimental variables. Equally importantly, they allow the experimenter to generate a large number of data points at a much smaller cost than does a traditional audit. Because resume audits allow for large sample sizes, large sample techniques can be used to analyze the data, providing more power and circumventing disagreements over which small sample technique is correct.

Experimental economists might initially feel distaste for the deception involved in this kind of field experiment. However, the impetus for lack of deception in the field of experimental economics is to keep subject pools pure, and to make sure behavior is not changed based on worries about deception. This worry is less of a problem with audit studies (or, indeed, any natural field experiment). Firms never know they are being tested and thus the experimental pool remains pure. Additionally, with academic audit studies, IRBs generally do not allow firms to be harmed in any way and insist upon conditions such that no individual firm can be identified or found to be discriminatory, so even a firm aware that such a study was going on would have little incentive to change its behavior. Finally, if firms did change their behavior because of fear of such audits, then that behavior would still be a true reflection of the current market.³ For this reason, large government audits are sometimes suggested as a way to change equilibrium housing or firm behavior (e.g., Charles 2005, Galster 1990).

Researchers may also worry about inconvenience to firms, and may worry that IRBs will not approve this type of study. Contrary to initial fears, in our experience and our knowledge of others' experience, IRBs generally approve resume audit studies without detailed review so long as no firm or hiring manager can be identified from the final data and the researcher promises not to use these data in a court of law. The reasoning is that the unit of study is the firm, not a human subject.⁴ Researchers may still worry about damage to the firm, based on concerns that evaluating false resumes may be time-consuming and costly. Human resource personnel have informed us, however, that the initial decision to pursue a candidate based on a resume generally takes much less than a minute. Additionally, in the job market, many applicants turn down jobs prior to interviews. Considerate researchers can promptly call firms that contact the applicant to inform them that the applicant has accepted a job elsewhere in order to minimize inconvenience to the firm.

Matched resume audit studies are an ingenious solution to the problem of the intensive resources required in creating hand-generated variation in resumes. Carefully thought-through designs have allowed researchers to posit that resumes are comparable and can

³A worry might be that most academic audits might occur in one city, Boston, for example, and results from that city are then extrapolated to the entire population. However, researchers generally use more than one city in any study and the choices for these cities vary widely.

⁴Further arguments for why IRBs are likely to accept an audit study can be found in Pager (2007).

appropriately be analysed using small sample techniques. However, due to advancements in computing power, this valuable classic technique, the audit study, can now be expanded in both scale and scope. This expansion will allow us to build on the results of existing audit studies and further our understanding of the fundamental issues identified in the audit literature. We introduce a computer program that aids researchers who wish to do correspondence audits. This program efficiently generates a large number of dissimilar resumes based on parameters set by the researcher and automates the resume creation and data-collection process. In this paper, we explain some of the problems that our program helps mitigate. We include an empirical example that demonstrates 1) the existence of these problems and 2) how our program can alleviate them. Finally, we detail how the program works.

II. Drawbacks of Audit Studies

Audit experiments (unlike field experiments in the style of List 2004) do not measure the actual level of market discrimination against a group.⁵ Typically there are demographic differences in population characteristics between the two groups being studied. Therefore, a matched pairs audit, in which each matched pair has identical characteristics by design, can never give the true average treatment effect in the overall market. Nevertheless, at their best, audits can help identify which individual characteristics best predict outcomes and which contribute to differential outcomes by group.

This capability is constrained by the limited number of tester pairs or resume templates typical of traditional audit studies. We term this problem “template bias.” In these traditional studies all items are correlated within each template or tester pair with the exception of the variable of interest. As a result, it is not clear whether the interaction effect is occurring between group status and one specific characteristic or some combination of those characteristics.⁶ Thus, not only do these studies lack the capacity to provide an accurate estimate of the extent of discrimination in the labor market, they also can only predict the outcomes and interaction effects of specific bundles of characteristics rather than of the characteristics themselves. To isolate the predictive effects of individual characteristics and their interactions with group status accurately, a large number of dissimilar resumes (or, less plausibly, testers) must be sent out.

Another potential problem with many audit studies is experimenter bias, a problem that is exacerbated when there are limited numbers of testing pairs or correspondence templates. With a standard pairs audit, there is some concern that, in a blind experiment, the minority

⁵For more discussion, see arguments in Heckman and Siegelman (1993) about heterogeneity across pairs.

⁶For a simple example, firms could choose to interview all candidates regardless of age if they had taken a computer class, and no candidates who had typos in their resumes. Discrimination could still exist if they preferred younger candidates to older in applicants with no typos and no computer class (or both typos and a computer class). However, this estimate would be mismeasured if all resumes had either computer classes or typos. Additionally, it would not be clear what the effects of classes or typos were if these items were correlated with other characteristics on the resume because of a limited number of templates. For example, if the resumes with computer classes were also from wealthy neighborhoods and the resumes with typos were from poor neighborhoods, then the experimenter would be unable to identify the fact that computer classes and typos, rather than neighborhoods, were driving firm decisions. One could argue that instead of being interesting in its own right, including any item that interacts with group status is poor experimental design and that pilot testing should eliminate any of these items. For those who wish to use a small “ideal” matched pairs resume template that does not contain these interactions, our program can be used in a pilot study to identify these effects so that they can be eliminated.

applicant may behave differently than the other applicant (for example, as a reaction to discriminatory treatment). If instead trained testers are used, then both testers may unconsciously influence the experiment (Heckman and Siegelman 1993). These concerns are minimized but do not disappear in a correspondence/resume audit as long as humans are responsible for matching templates to jobs, because the human may deviate from random assignment, especially when instructed to tweak resumes to better fit the position (as in Bertrand and Mullainathan 2004).⁷

Unlike a typical matched pairs audit using testers, a resume (or correspondence⁸) audit can be expanded cheaply, easily achieving a large sample size and corresponding statistical power (see Bertrand and Mullainathan 2004 for more detail). It has been standard in the resume audit literature to follow the person-audit literature by using a small number of “pairs” (in this case pairs of templates rather than people) when creating the resumes or cover letters being sent. With a large sample, however, there is no need to impose such a restriction, because there is no need to use small sample audit analysis. As long as a researcher first conducts an appropriate power test to estimate the minimum number of resumes to obtain the necessary power, then large sample characteristic properties apply. With these properties, standard econometric techniques such as OLS or Probit/Logit can be used rather than using traditional audit techniques, which were designed to address the problems associated with the necessarily small scale of person-audits. As part of the proper application of OLS or Probit/Logit techniques to large samples of matched pairs, the standard errors of the results should be clustered by pair. Clustering at the firm level should be used if multiple resumes are sent to the same firm.

Within a resume audit, the problems of “template bias” and experimenter bias can be greatly mitigated with the use of a computer-generated program that creates and matches resumes or other correspondence. Via this program, the researcher can provide a databank of items to randomize, such as employment history items, volunteer work, statements such as “I am flexible,” and many others. When randomly combined, every part of the resume becomes a potential control variable that can be interacted with the variable of interest, independently from other variables. Again, this technique will not measure the true extent of discrimination in a market, but it can show the existence of discrimination by group of interest, as well as showing which factors predict outcomes, either by themselves or in interaction with group status.⁹

Using the computer program, the experimenter chooses which job to apply for, and the program puts together the different aspects of each resume, including the indicator of group membership, reducing and perhaps even eliminating the threat or appearance of experimenter bias.¹⁰ A final benefit to computerization is that the program that creates the

⁷A well-designed traditional study could create and match the resumes before any “tweaking” was done. Our program greatly facilitates, even “idiot-proofs,” this step.

⁸Note, we will be referring to “resumes” and “firms” in the remainder of the paper, but the same arguments apply to any sort of paper or electronic correspondence and market being tested.

⁹It is true that the output is still only as good as the inputs; instead of having a limited number of templates, the researcher is responsible for inputting a database of items for the resume program to choose from. However, our methodology provides the experimenter with significantly more control over the ability to measure the effects of these inputs individually. We do not claim that an audit study can ever measure the true percentage of differential treatment in the labor market, but only that represented by the cross-section of correspondence tested.

resumes can also record the characteristics of the resumes, and even the date, automatically creating the database that will later be used for analysis. The experimenter need only record by hand the response that the resumes get (such as an offer of an interview) and any characteristics outside of the resume (such as firm location).

III. Empirical Examples

Classic Paired-Resume Audit Approach

To illustrate some of the problems with the traditional audit approach which relies on a limited number of templates, we have simulated a smaller matched pairs audit using a small subset of the data from a large randomized resume audit that investigated the effects of age on interview outcomes for entry-level jobs (Lahey 2008). We then show that the extent, and even existence, of discrimination varies based on the template chosen. Finally, we demonstrate how our methodology allows us to examine a much richer set of outcomes than solely existence of discrimination, including determinants of discrimination. Our smaller dataset matches 35-year-olds with 62-year-olds and sends resumes to firms in Boston, MA.¹¹ In addition, to imitate the use of a limited number of templates, we chose 3 items on the resumes—whether it contained a statement of flexibility, experience volunteering, and a work history longer than 5 years--which we will consider “observables.” We used these three binary characteristics to partition the resumes into 8 pseudo-templates,¹² ending up with a total of 464 observations.¹³

Using the entire set of 8 templates allows all possible combinations of the three variables to be sent out and tested, and thus allows for estimation procedures on the effects of each variable. However, standard person-audits and resume audits only send out a very limited number of tester pairs or templates,¹⁴ despite the fact that the testers and resumes possess a large number of diverse characteristics. Because firms may treat candidates differentially depending on these characteristics, estimates of overall discrimination can vary greatly based on which characteristics are chosen to represent the sample (Heckman and Siegelman 1993 provide further discussion). To approximate this limitation, we separated the 8 templates into two sets of 4. Then, as an illustration of the potential sensitivity of audit results to the particular templates chosen, we analyzed each set of 4 separately and compared our findings.

¹⁰If multiple resumes are sent to the same firm, a simple coin-flip, die toss, or random number generator can determine the order in which resumes are sent, further reducing experimenter bias.

¹¹This is a large difference in age, and was chosen from several age levels in the larger audit to maximize the results in this sample. A number of original methods of controlling for and separating the effects of experience from age were used in this study and are detailed in Lahey (2008).

¹²Other variables can be considered unobservables in this setting. These unobservables are not exactly analogous to unobservables in a paired testers audit, because in this study the unobservables are orthogonal to the observables and are orthogonal to each other by design. Relative to the estimation bias demonstrated in this example, therefore, bias in traditional audit studies may be exacerbated since items within templates are correlated. (Note, however, that the computer program presented in this paper allows the user to correlate items with some probability if the experimenter desires correlation between items.)

¹³Because the templates were created randomly in Lahey (2007), there are more observations of some templates than of others. In order to create an example with the same number of observations of each template in the two sets of four templates, we randomly dropped the necessary number of observations of each of the more common templates for each pair. Note that depending on your small-sample econometric persuasion, there are really only 6 or 11 observations with variation presented in Table 1. However, this size is consistent with many matched tester audits using small samples; the sign test and t-test still hold.

¹⁴The largest number of separate templates we have seen prior to Lahey (2008) is eight (Bertrand and Mullainathan 2004).

Table 1 provides estimates for differences in interview requests for these two sets of paired resumes for older and younger applicants. We use small sample audit analysis methods of the sign test (e.g., Heckman and Siegelman 1993) and paired-difference-of-means t-test (e.g., Yinger 1986) to estimate whether younger applicants are more likely to be called back for an interview than are older applicants. Results are strikingly different depending on which set of templates is used. Analysis of Set 1 suggests that younger applicants are statistically significantly more likely to be called than are older applicants with a t-test at the 3% level, and at the 6% level using the more restrictive sign test. Analysis of Set 2, on the other hand, identifies no statistical difference between older and younger applicants. Analysis of the entire sample using all eight templates shows younger applicants more likely to be called at the 5% level using the t-test and at the 10% level with the more restrictive sign test.

Updated Large-Scale Randomized Resume Audit Approach

As this simulation illustrates, using a limited number of templates (or testers) can strongly bias estimates of discrimination because of unmeasurable interactions between the resume characteristics and the characteristic of interest. Moreover, the bias cannot be signed—depending on the template, results may be biased towards either finding or rejecting discrimination. In addition, when resume characteristics are correlated with each other, for example because there is not a full set of permutations, a traditional audit study cannot be used to determine how these characteristics interact with the variable of interest.

By contrast, the benefits of using a computer generator to create hundreds of various resumes can be illustrated using a larger subset of the observations collected in Lahey 2008. Because computerization allows the experimenter to randomize characteristics across thousands of resumes, standard large sample analyses apply and there is room for interactions of characteristics with group status to be tested. In the following example, we test whether including the statement “I am flexible” on a resume has different effects on the interview probabilities of younger versus older job applicants. We conduct this test by running marginal probit regressions of the following form:

$$Interview = \beta_1 Older + X\beta_2 + \varepsilon \quad (1)$$

$$Interview = \beta_1 Older + \beta_2 Older * Flexible + \beta_3 Flexible + X\beta_4 + \varepsilon \quad (2)$$

where *Interview* is an indicator variable describing whether the applicant was asked to an interview, *Older* is an indicator of whether the applicant is over the age of 50, *Flexible* is an indicator describing whether the applicant’s resume says “I am flexible” or “I am willing to embrace change,” and *X* is a vector of resume and other characteristics.

The dataset analyzed in Table 2 contains paired resumes from several ages sent to firms in Boston, MA. Columns (1) and (2) use a subset of the data, specifically only the matched pairs of 35 year olds and 62 year olds, in order to provide a parallel with audit studies that use binary variables such as black/white or male/female. Column (1) provides the results of a marginal probit regression of a binary “older” variable (where 62 is older) on interview

responses; moving from age 35 to age 62 has a negative effect, -0.024 , on getting an interview, but that effect is not statistically significant at conventional levels. Column (2) shows the importance of adjusting the standard errors to address the non-independence of errors within firms; the 2.4% decrease in the probability of getting an interview becomes statistically significant at the 5% level with clustering of standard errors at the firm level.

However, unlike the case with a standard matched pairs audit, with this methodology we do not need to restrict our analysis to pairs of 35 and 62 year olds. In column (3), where we add all observations of intermediate ages to the universe and dichotomize older as age 50 or over, we find a decrease of 1.6% in the probability of an interview for those over 50 compared to those under 50 (significant at the 1% level). Alternatively, age could be examined as a continuous variable or entered as a set of separate age dummies (results on those regressions are available in Lahey 2008).

With the larger sample size created by including all ages, and with the full set of randomized characteristics on each resume recorded by the resume-creation program, controls can be added easily, as in column (4). In the case of this example, the signs on controls are as expected, and having had vocational training or having done volunteer work have positive effects on the probability of an interview that are marginally statistically significant. These results again emphasize the possibility that which templates are used may help determine the number of responses received.

The last column of Table 2 investigates whether there are interaction effects between age and other resume characteristics; it estimates equation (2) to test whether the effect of a control on the probability of interview varies by age. A significant coefficient on the interaction between group membership and a resume characteristic implies that templates that include that characteristic will tend to turn up more evidence of discrimination than will resumes that do not include that characteristic. As shown in column (5), putting down that one is flexible or “willing to embrace change” is shown to decrease the probability of interview for older workers relative to younger workers.¹⁵ Thus, a resume audit in which the template includes this statement will be more likely to conclude that there is age discrimination in interviewing than will an otherwise similar audit in which the template does not include this statement.

In addition, such investigations of interaction effects may be important for understanding how discrimination takes place. For example, the discovery that “I am flexible” hurts older workers may be particularly important to groups interested in combating age discrimination.

IV. The Program

To assist in creating a large number of resumes with a diverse set of experimenter-defined characteristics, we have created a new computer program. This program comes in two parts. The first is a web-based meta-program that will allow an experimenter to define and input the general characteristics of the resume or correspondence template.¹⁶ In this part, the

¹⁵One potential explanation for this counter-intuitive finding is that the AARP recommends that these statements be put on a resume, so having such cheap talk statements is another signal that the worker is older.

researcher decides on the outline of the resume and the probabilities that items will be represented in a resume. For an example, in a race resume audit, a researcher could determine at this stage that all resumes would contain the name of the applicant as a category (100% chance of being present) and 3 sets of work histories, each with a 75% chance of being present. The researcher could also determine the set of work histories to choose from, and, if using a matched pairs audit, set pairs of work histories such that each had equivalent but different characteristics (e.g., the same or similar job title at different companies).

The second part of the program, the “resume-generator,” is created from this web-based meta-program and uses the experimenter-defined template to generate and record resume information.¹⁷ Once the resume outline has been created via the web-program, the command-line resume generator can be run any number of times to generate resumes according to the outline and databank created in the first part of the program. Each time the second program is run, the experimenter can instruct it to generate any number of resumes, either matched or not matched.¹⁸ The resume-generator uses the probabilities assigned to each resume item in the outline to determine whether those items are represented when it is run. Along with each resume, the generator creates a space-delimited record of the random choices made in the creation of that resume, sufficient for exact re-creation of that resume. A companion program, filegather.exe, collects these data outputs into an Excel-readable dataset.

The program also contains more advanced features. For example, it also allows for parameters to be associated with each other; the researcher can specify that someone applying for a truck driving position also have a randomly chosen truck driving school and truck driving experience on the resume, but a waitress applicant would not. A repetition feature allows the researcher to designate that a set of resume items, such as job histories, be repeated a specified number of times, allowing the creation of a multi-item work history from a single set of work history items (and merging job experience listings when the same experience item is randomly selected multiples times in a row). Furthermore, when creating a matched pair of resumes, the program can be adjusted to make sure the matched resumes always include the same resume item as in the other resume, always include a different item, or include items independently. The program is provided with an example resume-generator demonstrating different possibilities.

¹⁶Available from the authors by request or at <http://www.nber.org/data/> (under “Other”). The web program creates .rtf files which can be opened from resume-randomizer-framemaster.exe, which then creates .doc resumes, .sav information files, and .txt tab delimited data. After all resumes in a session have been created, filegather.exe collects data information from .txt files into a tab delimited .dat file which can be opened in a spreadsheet program. To access it offsite, unzip resume-randomizer.zip and double-click on resume-randomizer-framemaster.html within the resume-randomizer folder (or open from Mozilla). If applicable, change any .ex files to .exe within the .zip file. The web program creates .rtf files which can be opened from resume-randomizer-framemaster.exe, which then creates .doc resumes, .sav information files, and .txt tab delimited data. After all resumes in a session have been created, filegather.exe collects data information from .txt files into a tab delimited .dat file which can be opened in a spreadsheet program.

¹⁷As in the earlier discussion, we discuss the use of the program for creating resumes, but the program is general enough to be used for other text randomization tasks such as creating cover letters.

¹⁸The program has the capacity to create triplets, quadruplets, etc. of resumes; we prefer single resumes for large sample analysis, but this preference is dependent on the study being done so we do not limit the scope of this program.

V. Conclusion

Randomized computer generation of resumes improves on previous resume audits by allowing the creation of many distinct templates at a low cost to researchers. This technological innovation allows variation in and measurement of the effects of many resume characteristics, producing thousands of distinct resumes and allowing analysis of the resulting data in a multivariate regression framework. By contrast, matched paired audits have a limited number of observations and templates, requiring controversial small sample econometric analysis, and making it difficult to test for interactions between the tested characteristic (e.g., race or gender) and other applicant or job attributes. The ability to easily create thousands of distinct resumes and to cheaply collect data on dozens of resume characteristics greatly expands the scale and scope of the correspondence audit technology.

Works Cited

- Ashraf N, Karlan D, Yin W. Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines. *Quarterly Journal of Economics*. 2006; 121(2):635–672.
- Bertrand M, Mullainathan S. Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review*. 2004; 94(4):991–1013.
- Charles, CZ. Can We Live Together? Racial Preferences and Neighborhood Outcomes. In: Briggs, editor. *The Geography of Opportunity*. Brookings Institution Press; 2005.
- Fix, M.; Galster, GC.; Struyk, RJ. An Overview of Auditing for Discrimination. In: Fix, Michael; Struyk, Raymond J., editors. *Clear and convincing evidence: Measurement of Discrimination in America*. Washington, D.C: Urban Institute Press; distributed by University Press of America Lanham Md; 1993. na
- Galster GC. Racial Steering in Urban Housing Markets: A Review of the Audit Evidence. *The Review of Black Political Economy*. 1990 Winter;:105–129.
- Harrison G, List JA. Field Experiments. *Journal of Economic Literature*. 2004; XLII:1013–1059.
- Heckman, JJ.; Siegelman, P. The Urban Institute Audit Studies: Their Methods and Findings: Response to Comments by John Yinger. In: Fix, M.; Struyk, RJ., editors. *Clear and Convincing Evidence: Measurement of Discrimination in America*. Washington, D.C: Urban Institute Press; distributed by University Press of America Lanham Md; 1993. na
- Heckman JJ. Detecting Discrimination. *Journal of Economic Perspectives*. 1998; 12(2):101–16.
- Jowell R, Prescott-Clarke P. Racial Discrimination and White-collar Workers in Britain. *Race & Class*. 1970; 11(4):397–417.
- Karlan D, List JA. Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment. *American Economic Review*. 2007; 97(5):1774–1793.
- Lahey JN. Age, Women and Hiring: An Experimental Study. *Journal of Human Resources*. 2008; 43(1):30–56.
- List JA. The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field. *Quarterly Journal of Economics*. 2004; 119(1):49–89.
- List JA. Field Experiments: A Bridge between Lab and Naturally Occurring Data. *Advances in Economic Analysis & Policy*. 2006; 6(2):Article 8.
- Pager D. The Use of Field Experiments for Employment Discrimination: Contributions, Critiques, and Directions for the Future. *The Annals for the American Academy*. 2007; 609:104–133.
- Riach, PA.; Rich, J. *Cambridge Journal of Economics*. Vol. 15. Oxford University Press; 1991. Testing for Racial Discrimination in the Labour Market; p. 239-56.
- Riach PA, Rich J. Field Experiments of Discrimination in the Market Place. *Economic Journal*. 2002; 112(483):F480–F518.

- Ridley, S.; Bayton, JA.; Hamilton Outtz, J. Taxi service in the District of Columbia: Is it influenced by patrons' race and destination?. Mimeograph, Washington Lawyers' Committee for Civil Rights under the Law; Washington, DC. 1989.
- Yinger J. Measuring Discrimination with Fair Housing Audits: Caught in the Act. *American Economic Review*. 1986; 76(December):881-93.
- Yinger, J. *Closed Doors, Opportunities Lost: The Continuing Costs of Housing Discrimination*. New York: Russell Sage Foundation; 1995.

Table 1

Problem: Traditional template analyses lead to inconsistent conclusions

	(1) Both	(2) Neither	(3) Only Younger	(4) Only Older	(5) sign test*	(6) t-test**
template 1	(1) 2.9%	(33) 94.3%	(1) 2.9%	(0) 0%		
template 2	(0) 0%	(43) 97.7%	(1) 2.3%	(0) 0%		
template 3	(0) 0%	(16) 100%	(0) 0%	(0) 0%		
template 4	(1) 4.8%	(18) 85.7%	(2) 9.5%	(0) 0%		
set 1	(2) 1.7%	(110) 94.8%	(4) 3.5%	(0) 0%	0.0625	-0.0345 p = 0.0225
template 5	(2) 5.7%	(33) 94.3%	(0) 0%	(0) 0%		
template 6	(1) 2.3%	(43) 97.7%	(0) 0%	(0) 0%		
template 7	(0) 0%	(16) 100%	(0) 0%	(0) 0%		
template 8	(0) 0%	(19) 90.5%	(1) 4.8%	(1) 4.8%		
set 2	(3) 2.6%	(111) 95.7%	(1) 0.9%	(1) 0.9%	0.7500	-0.0122 p = 0.5000
total	(5) 2.2%	(221) 95.3%	(5) 2.2%	(1) 0.4%	0.1094	-0.0172 p = 0.0513

Note: Columns (1) through (4) give percentages for pairs in each template where both older and younger, neither, only younger, and only older resumes received interview requests; figures in parentheses are the relevant number of audits. Set 1 includes templates 1–4, Set 2 includes templates 5–8. Templates are formed from the different possible combinations of the binary characteristics of flexibility, volunteering, 5+ year work history. Set 1 finds evidence of discrimination (marginal in the case of the sign-test), Set 2 does not find such evidence.

* p-value of one-tailed binomial that younger are more likely to be interviewed than older

** coefficient and p-value of a one-tailed t-test that younger are more likely to be interviewed than older

Table 2

Solution: Expand models using computerization and large sample techniques

	Interview Outcomes				
	(1)	(2)	(3)	(4)	(5)
older=50+	-0.024 (0.017)	-0.024 (0.011)*	-0.016 (0.006)**	-0.016 (0.005)**	-0.006 (0.006)
flexible* older					-0.017 (0.008)*
flexible				-0.001 (0.007)	0.009 (0.009)
gap in work history				-0.002 (0.007)	-0.002 (0.007)
vocational training				0.033 (0.018) ⁺	0.033 (0.018) ⁺
relevant computer class				0.010 (0.008)	0.010 (0.008)
volunteer work				0.013 (0.007) ⁺	0.013 (0.007) ⁺
sports				-0.005 (0.007)	-0.005 (0.007)
already has insurance				0.009 (0.007)	0.009 (0.007)
attendance award				0.005 (0.007)	0.005 (0.007)
Observations	490	490	4229	4229	4229
cluster?	no	yes	yes	yes	yes
controls?	no	no	no	yes	yes

Note: The universe for columns (1) and (2) is the set of paired resumes for 35 and 62 year olds. Columns (3)–(5) include resumes for applicants aged 35, 45, 50, 55, and 62. Additional controls include years in the labor force and years in the labor force squared, and occupational dummies for professional, education, health, manager, sales, craftsman, operative, service, and laborer dummies.

Standard errors in parentheses. Standard errors clustered on firm when noted.

⁺ significant at 10%;

* significant at 5%;

** significant at 1%.