Meta-analysis of field experiments shows no change in racial discrimination in hiring over time

Lincoln Quillian[a,b,1], Devah Pager[c,d], Ole Hexel[a,e], and Arnfinn H. Midtbøen[f]

[a]Department of Sociology, Northwestern University, Evanston, IL 60208; [b]Institute for Policy Research, Northwestern University, Evanston IL 60208; [c]Department of Sociology, Harvard University, Cambridge, MA 02138; [d]Kennedy School of Government, Harvard University, Cambridge MA 02138; [e]Sciences Po, Observatoire Sociologique du Changement (OSC), CNRS, 75007 Paris, France; and [f]Institute for Social Research, N-0208 Oslo, Norway

Edited by Douglas S. Massey, Princeton University, Princeton, NJ, and approved August 8, 2017 (received for review April 14, 2017)

This study investigates change over time in the level of hiring discrimination in US labor markets. We perform a meta-analysis of every available field experiment of hiring discrimination against African Americans or Latinos (n = 28). Together, these studies represent 55,842 applications submitted for 26,326 positions. We focus on trends since 1989 (n = 24 studies), when field experiments became more common and improved methodologically. Since 1989, whites receive on average 36% more callbacks than African Americans, and 24% more callbacks than Latinos. We observe no change in the level of hiring discrimination against African Americans over the past 25 years, although we find modest evidence of a decline in discrimination against Latinos. Accounting for applicant education, applicant gender, study method, occupational groups, and local labor market conditions does little to alter this result. Contrary to claims of declining discrimination in American society, our estimates suggest that levels of discrimination remain largely unchanged, at least at the point of hire.

discrimination | labor markets | field experiments | race | ethnicity

The American racial landscape has changed in fundamental ways since the Civil Rights Movement of the 1960s. During that time, sweeping legal and social reforms reduced the barriers facing African Americans in many important domains (1, 2). A rising African American middle class and a growing acceptance of the principles of inclusion led some to conclude that racial discrimination had declined to the point that it was no longer a primary determinant of life chances for African Americans and Latinos (2, 3).

Supporting this perspective, a variety of indicators pointed toward a reduction of discriminatory treatment. Surveys indicated that whites increasingly endorsed the principle of equal treatment regardless of race (4). Rates of high school graduation for whites and African Americans converged substantially, and the black–white test score gap declined (5, 6). Large companies increasingly recognized diversity as a goal and revamped their hiring to curtail practices that disadvantaged minority applicants (7). With the election of the country's first African-American president in 2008, many concluded that the country had finally moved beyond its troubled racial past (8).

Despite clear signs of racial progress, however, on several key dimensions racial inequality persists and has even increased. For example, racial gaps in unemployment have shown little change since 1980 (9, 10), and the black–white gap in labor force participation rates among young men widened during this time (11). Recently, the Black Lives Matter movement shone a spotlight on the ongoing struggles with racism and discrimination experienced by people of color in interactions with law enforcement. The election of Donald J. Trump as the 45th President of the United States with the support of antiimmigrant and white nationalist groups highlighted the persistence of racial resentment (12).

In light of persistent racial gaps in key social and economic indicators, some scholars have challenged prevailing assumptions about waning discrimination. Indeed, while expressions of explicit prejudice have declined precipitously over time, measures of stereotypes and implicit bias appear to have changed little over the past few decades (13–15). In this view, far from disappearing, racial bias has taken on new forms, becoming more contingent, subtle, and covert (15–18).

What can we reliably say about trends in discrimination over time? Has the role of race appreciably diminished across the board, or are there important domains in which little racial progress has been achieved? Answers to these questions are important for understanding the sources of persistent racial inequality.

In this study, we examine trends in racial and ethnic discrimination in American labor markets based on a meta-analysis of every available field experiment of hiring discrimination (with fieldwork dates through December 2015). Meta-analysis is a body of formal methods to synthesize data from a population of existing studies. Field experiments of hiring discrimination are experimental studies in which fictionalized matched candidates from different racial or ethnic groups apply for jobs. These studies include both resume audits, in which fictionalized resumes with distinct racial names are submitted online or by mail (e.g., ref. 19), and in-person audits, in which racially dissimilar but otherwise matched pairs of trained testers apply for jobs (e.g., ref. 20).

The field experimental method is a design with high causal (internal) validity because it benefits from aspects of experimental design. The experimenter carefully manages the application process, which provides control over many potential confounding variables. The exact basis of causal inference across the two main forms of field experiment, resume and in-person audits, is somewhat different. In the typical resume audit, clues indicating race (such as a racially identifiable name) are randomly assigned to otherwise similar resumes, allowing for treatment and control groups to be equated through randomization. In in-person audits, matched

Significance

Many scholars have argued that discrimination in American society has decreased over time, while others point to persisting race and ethnic gaps and subtle forms of prejudice. The question has remained unsettled due to the indirect methods often used to assess levels of discrimination. We assess trends in hiring discrimination against African Americans and Latinos over time by analyzing callback rates from all available field experiments of hiring, capitalizing on the direct measure of discrimination and strong causal validity of these studies. We find no change in the levels of discrimination against African Americans since 1989, although we do find some indication of declining discrimination against Latinos. The results document a striking persistence of racial discrimination in US labor markets.

Author contributions: L.Q. designed research; L.Q., O.H., and A.H.M. performed research; L.Q. and O.H. analyzed data; and L.Q., D.P., O.H., and A.H.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

See Commentary on page 10815.

[1]To whom correspondence should be addressed. Email: l-quillian@northwestern.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1706255114/-/DCSupplemental.

pairs of trained testers who differ on the basis of race but are otherwise similar apply for jobs; the between-race contrast is grounded in matching pairs of applicants to make them as similar as possible in all employment-relevant characteristics except race. Both resume and in-person audit methods provide a strong basis from which to draw conclusions about hiring discrimination, particularly relative to the nonexperimental methods widely used in the literature, including by all prior studies of discrimination trends over time (ref. 21 and *SI Appendix*, section 1).

We use meta-analytic techniques to investigate change in hiring discrimination over time based on all existing US field experimental studies of labor market discrimination. Our procedure follows three basic stages: First, we identified all existing studies, published or unpublished, that use a field experimental method and that provide contrasts in hiring-related outcomes between equally qualified candidates from different racial or ethnic groups. Second, we coded key characteristics of the studies into a database for our analysis based on a coding rubric. This produced 24 studies containing 30 estimates of discrimination against African Americans and Latinos since 1989, together representing 54,318 applications submitted for 25,517 positions. Finally, we performed a random-effects meta-regression to identify trends over time.

We assess discrimination for each study using the ratio of the proportion of applications that received "callbacks"—or invitations to interview—by white applicants relative to African-American or Latino applicants. We calculated the proportions based on counts of the number of callbacks received by each group (white/African American/Latino) within each study. This discrimination ratio measured at the study level is the outcome in our meta-regression. Other methods of calculating hiring disparities between groups produced substantively similar results (*SI Appendix*, section 8).

We analyze the relationship of discrimination ratios to years in which the data were gathered to provide an estimate of the trend in discrimination. Specifically, we regress the log of the discrimination ratio on year of survey, with controls for key characteristics of the studies, using meta-regression. Meta-regression is a procedure similar to standard regression, except covariates are measured at the level of the study rather than the level of the individual, and the outcome is an effect from the study of interest (in our case, the outcome is the estimate of discrimination against African Americans or Latinos). *Methods and Materials* discusses further methodological and modeling details.

Results

To explore trends over time, we estimate a series of meta-regressions. We take the natural log of the discrimination ratio (our outcome variable) to account for skew. In the simplest meta-regression models, the only covariate is the time trend. In later models, we include a more extensive set of predictors to control for other factors that might confound the time trend. To capture sources of variability not covered by the covariates, we use a random effects specification (22). Random effects incorporate a variance component capturing variation in outcomes across studies that are due to unobserved study-level factors (*Methods and Materials*).

Our core analysis focuses on studies that conducted their fieldwork from 1989 to 2015, allowing us to observe trends in discrimination over the past 25 years. For some supplementary analyses, we also add four field experiments conducted before 1989, although these studies use less standardized methodologies. On average, white applicants receive 36% more callbacks than equally qualified African Americans (95% confidence interval of 25–47% more), based on random-effects meta-analysis of data since 1989, representing a substantial degree of direct discrimination. White applicants receive on average 24% more callbacks than Latinos (95% confidence interval of 15–33% more). For more detailed results, see *SI Appendix*, section 2 and Figs. S1 and S2.

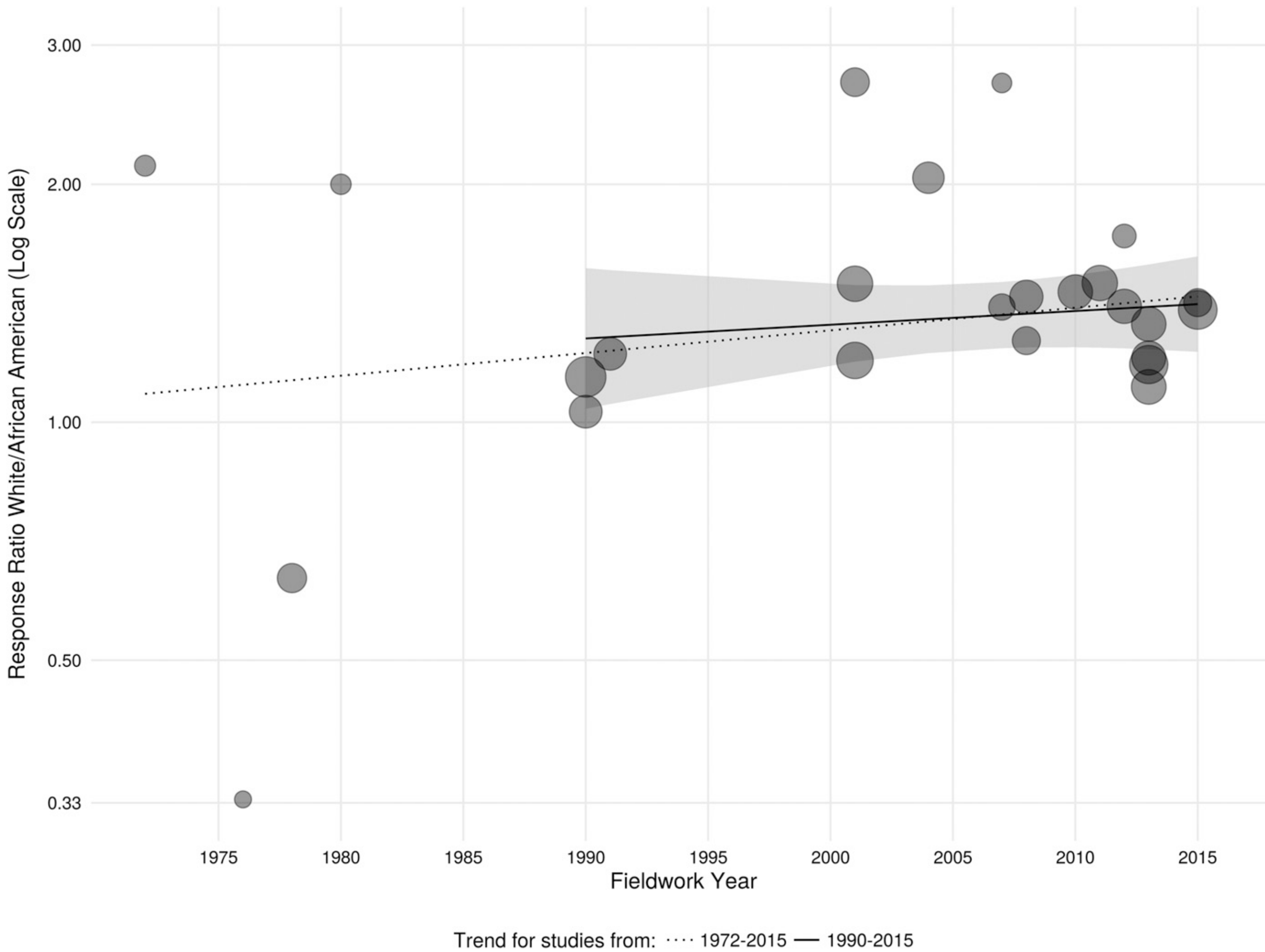
Do we find evidence of change over time in rates of hiring discrimination? With respect to African Americans, the answer is no. Fig. 1 plots estimates of discrimination by year, with linear trends of best fit and 95% confidence regions (detailed estimates are in *SI Appendix*, section 3 and Table S3; in Fig. 1, we exponentiate predictions to present predicted values as discrimination ratios rather than less interpretable log discrimination ratios). The solid line captures the trend since 1990. The dashed line extends this time trend back to 1972, adding four resume audits conducted from 1972 to 1980. The size of the symbol is proportional to the weight it is given in the meta-analysis. The line of best fit for studies since 1990 is close to flat, sloping slightly upward, suggesting no change in the rate of discrimination over the past 25 years. The longer time series includes studies that use a more heterogeneous set of procedures (*Methods and Materials*), but even here we see no clear change over time in the level of hiring discrimination against African Americans.

Is there sufficient power based on 21 studies to conclude that discrimination against African Americans did not decline? The confidence interval of the annual change provides a way to answer this question. The 95% confidence interval of the slope 1989–2015 is −0.007 to 0.015. [This is the confidence interval of the slope of "year" (our time trend variable) with the log discrimination ratio outcome. The regression is shown in *SI Appendix*, Table S3.] The lower end of this interval indicates a decline in the discrimination ratio of 0.7% per year. If we take this number as the smallest slope consistent with the data based on the confidence interval, this suggests only a slight decline in discrimination each year. We conclude that this evidence rules out all but a slow decline in discrimination—with the most likely estimate being the point estimate, which indicates no decline in discrimination at all.

Fig. 2 presents the trend for Latinos (as with Fig. 1, model predictions have been exponentiated to allow interpretation as discrimination ratios rather than log ratios). Here, we see the line slopes downward, indicating a possible decline in discrimination, although this trend is outside of conventional levels of significance ($P = 0.099$). The point estimate suggests a decline from whites receiving 30% more callbacks than Latinos in 1990 to 15% more callbacks in 2010 (1.30 vs. 1.15). Because of the small number of Latino field experiments ($n = 9$), there is high uncertainty in characterizing this trend. (Using the difference in proportions or the odds ratio as outcomes, rather than the discrimination ratio, results in downward slopes in discrimination against Latinos over time that are statistically significant at the $P < 0.05$ level; see *SI Appendix*, section 8 and Table S9. However, sensitivity checks that modified the outcome sample counts slightly result in nonsignificant year coefficients of the difference in proportion or odds ratio, see *SI Appendix*, section 4 and Table S5).

Is it possible that key aspects of study design changed over time, influencing our estimates of changes in discrimination? To consider this question, we estimate a meta-regression model of discrimination rates as a function of a time trend plus other study characteristics. We discuss only models for African Americans, because the number of studies with Latinos ($n = 9$) is too small to produce reasonable precise estimates in a meta-regression model with multiple covariates.

Fig. 3 graphs estimates of change over time when the outcome discrimination ratio is modified and when controls are added. Full coefficients of the models are shown in *SI Appendix*, Table S4, with additional discussion in *SI Appendix*, section 4. The coefficients can be interpreted as the one-year percentage change in the discrimination ratio. [Because the outcome is logged, and $\exp(b) \approx 1 + b$ for $b < 0.1$, coefficients with values less than about 0.1 can be multiplied by 100 to closely approximate percentage changes with a one-unit change in x.] The first coefficient graphed shows the annual percentage change in discrimination from 1990 to 2015, corresponding to solid line in Fig. 1. The second shows the annual change for the longer time period 1972–2015, corresponding to the dashed line in Fig. 1.



Note: Size of plotting symbols proportional to meta-regression weights. Shaded region gives 95% confidence interval.

Fig. 1. No reduction in hiring discrimination facing African Americans over time.

The next few models alter the dependent variable to see if this changes our results (using our base sample of 1990–2015). In one modification, we use "job offer" in place of callback as the outcome for studies for which the job offer outcome is available ($n = 3$), retaining callbacks as the outcome for studies in which the measure of job offer is not available. This makes the outcome variable less uniform across studies, although closer to the outcome of greatest substantive interest, getting a job. With this modification, the trend line for African Americans slants more downward, but is still close to zero (−0.008) and statistically nonsignificant. A second modification eliminates applicant profiles that included either a fictitious criminal background ($n = 7$) or a disability ($n = 1$). This limits the applicant profiles to those with more mainstream job backgrounds and credentials. The modified results show the trend line slanting slightly more upward, providing less evidence to support a downward trend than the results including a more heterogeneous set of applicant characteristics. A third modification uses only resume audit studies, discarding in-person audits. This results in an almost perfectly flat line (−0.002).

The next estimates are based on models that add controls for applicant attributes, region and area unemployment rates, and occupational categories to the baseline time-trend model. The "Applicant Attributes" model introduces covariates representing applicant characteristics (e.g., gender, education) and study design (resume or in-person audits). The "UE & Regions" model adds controls for the unemployment rate of the local metropolitan area and dummy variables for region. The "Occupations" model includes controls for occupational categories of blue collar, office-focused, and restaurant occupations. Finally we present the coefficient from a trimmed model in which only the predictors with the largest t ratios from prior models are included. In each case, we see coefficients for the time trend that are close to zero—ranging from an estimated increase of 0.1% per year (0.001) to an increase of 1.3% per year (0.013) —suggesting little change in the level of discrimination facing African Americans over time. Notably, then, we find evidence of stability, not change, in hiring discrimination against African Americans.

Few of the measured covariates in our analysis (*SI Appendix*, Table S4) demonstrate a clear relationship to patterns of discrimination. This likely is in part due to the relatively small overall sample of studies ($n = 21$ for African Americans since 1990), which limits our ability to detect statistical significance. However, even looking at the point estimates we find no large differences in magnitude across categories. This result is consistent with the findings within individual audit studies that suggest relative stability in measured discrimination across job types, applicant gender, and skill levels (e.g., refs. 19 and 20). (We also note that a meta-analysis designed to look specifically at effects of many of

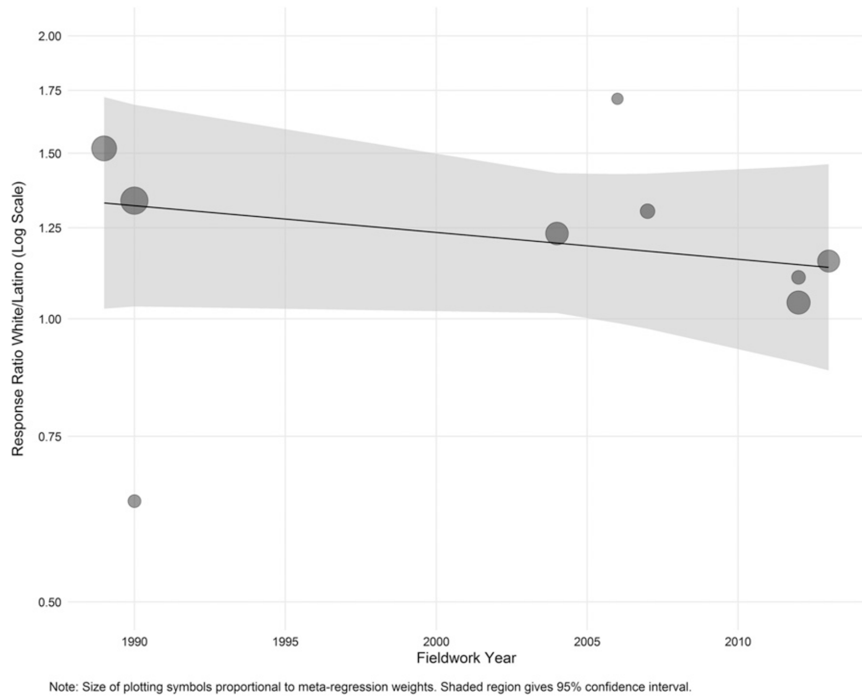


Fig. 2. Modest evidence of a reduction in hiring discrimination facing Latinos over time.

these covariates would use within-study variability—such as contrasting male and female auditors in the same study—which could provide more power to discern effects. Within-study variability cannot be applied to understand change over time since studies are generally conducted over the span of just a few months.) As a final check on the influence of covariates, we tested for time trends among our study-level and individual-level characteristics, finding no evidence of systematic change (*SI Appendix*, section 9 and Table S10). This suggests that covariates are unlikely to influence the observed time trend for discrimination among either the African-American or Latino samples.

In relation to our estimate of changes in discrimination over time, the inclusion of study-level and applicant-level characteristics has little impact. In all models, we see little evidence of a reduction in hiring discrimination against African Americans over time.

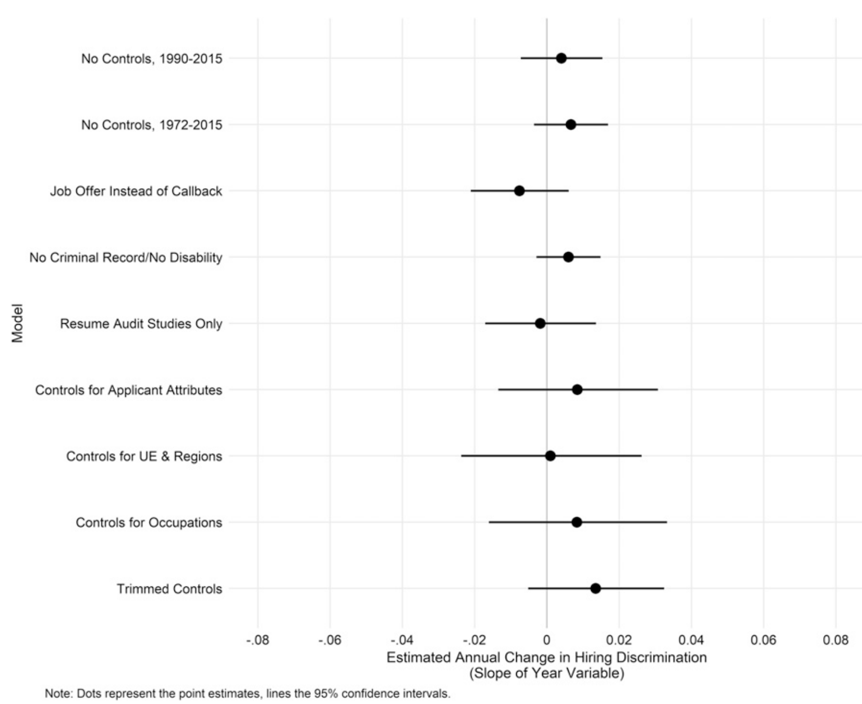


Fig. 3. Alternative estimates of change in hiring discrimination against African-Americans.

A potential concern of any meta-analysis is publication bias. In the present case, publication bias may entail studies that show no discrimination being less likely to be published and, thus, included in our study. We sought to address this issue by seeking out and including all nonpublished field experiments available (n = 11). Their inclusion did little to affect our estimates. Finally, in *SI Appendix*, section 5 and Table S7, we show that studies in which racial discrimination was the focus of the analysis (and for which there may be more pressure to demonstrate a positive effect) show no more discrimination than studies in which other characteristics were the main focus (with race included as an secondary or incidental covariate), further reducing concerns over publication bias for our results.

Discussion

Contrary to widespread assumptions about the declining significance of race, the magnitude and consistency of discrimination we observe over time is a sobering counterpoint. We note that our results do not address the possibility that hiring discrimination may have substantially dropped in the 1960s or early 1970s, during the civil rights era when many forms of direct discrimination were outlawed, as some evidence suggests (1). Further, we note that our results pertain only to discrimination at the point of hire, not at later points in the employment relationship such as in wage setting or termination decisions. Social psychological theories would predict hiring to be most vulnerable to the influence of racial bias, given that objective information is limited or unreliable (23–25). Likewise, from an accountability standpoint, discrimination is less easily detected, and therefore less costly to employers, at the point of hire (26). It may be the case, then, that more meaningful reductions in discrimination have taken place at other points in the employment relationship not measured here. What our results point to, however, is that at the initial point of entry—hiring decisions—African Americans remain substantially disadvantaged relative to equally qualified whites, and we see little indication of progress over time.

These findings lead us to temper our optimism regarding racial progress in the United States. At one time it was assumed that the gradual fade-out of prejudiced beliefs, through cohort replacement and cultural change, would drive a steady reduction in discriminatory treatment (27). At least in the case of hiring discrimination against African Americans, this expectation does not appear borne out.

We find some evidence of a decline in discrimination against Latinos since 1989. The small number of audit studies including Latinos limits our ability to include controls and the precision of our estimates—the decline is marginally significant statistically (P = 0.099). More evidence is needed to establish the trend in hiring discrimination against Latinos with greater certainty.

Our results point toward the need for strong enforcement of antidiscrimination legislation and provide a rationale for continuing compensatory policies like affirmative action to improve equality of opportunity. Discrimination continues, and we find little evidence in regards to African Americans that it is disappearing or even gradually diminishing. Instead, we find the persistence of discrimination at a distressingly uniform rate.

Materials and Methods

Our procedure follows three basic stages: first, to identify all existing field experiments of hiring discrimination; second, to develop a coding rubric and to code studies to produce a database of their results; and third, to perform a statistical meta-analysis to draw conclusions from the combined results. We discuss each of these steps in turn.

Identifying Relevant Studies. We aimed to include in our meta-analysis all existing studies, published or unpublished, that use a field experimental method and that provide contrasts in hiring-related outcomes between different race and ethnic groups in the United States. This includes both in-person audit studies and resume studies (or correspondence studies). We also required that contrasts of hiring outcomes between race or ethnic groups were made for groups that were on average equivalent in their labor market relevant characteristics, since otherwise discrimination estimates are confounded with the difference in nonracial characteristics.

We used three methods to identify relevant field experiments: searches in bibliographic databases, citation searches, and an email request to corresponding authors of field experiments of race-ethnic discrimination in labor markets and other experts on field experiments and discrimination.

We began with a bibliographic search. Our search covered the following bibliographic databases and working paper repositories: Thomson's Web of Science (Social Science Citation Index), ProQuest Sociological Abstracts, ProQuest Dissertations and Theses, Lexis Nexis, Google Scholar, and NBER working papers. We searched for some combination of "field experiment" or "audit study" or "correspondence study" and sometimes included the term "discrimination," with some variation depending on the search functions of the database. We also searched two French-language indexes, Cairn and Persée, and two international sources, IZA discussion papers, a German working paper archive, and ILO International Migration Papers.

Our second technique for identifying relevant studies relied on citation search. Working from the initial set of studies located through bibliographic search, we examined the bibliographies of all review articles and eligible field studies to find additional field experiments of hiring discrimination.

The last technique used was an email request of authors of existing field experiments of discrimination. From our list of audit studies identified by bibliographic and citation search, we compiled a list of email addresses of authors of existing field experiments of discrimination. To this we added the addresses of authors of literature review articles on field experiments. Our email request asked for citations or copies of field discrimination studies published, unpublished, or ongoing. We also asked that authors refer us to any other researchers who may have recent or ongoing field experiments.

The email requests were conducted in two phases. In the initial wave, 131 apparently valid email addresses were contacted. We received 56 responses. We also sent out a second wave of 68 e-mails which consisted of additional authors identified from the initial wave of surveys and some corrected email addresses. We received 19 responses to this second wave of email surveys.

Overall, our search located 34 studies that were US-based field experiments of hiring, included contrasts between white and nonwhite applicant profiles that were on-average equivalent in their labor-market relevant characteristics (e.g., education, experience level in the labor market). Six studies were excluded for various reasons, as explained in *SI Appendix*, section 6. Our remaining 28 studies yielded 24 estimates of discrimination against African Americans and 9 against Latinos relative to whites.

Coding and Selection of Analysis Period (1989–2015). We coded key characteristics of the studies into a database for our analysis. Coding was based on a coding rubric, which listed each potentially relevant characteristic of the research and included coding instructions. To develop the rubric, we initially read several studies and, based on this, developed an initial coding rubric of factors we thought might influence measured rates of discrimination. The initial rubric was reviewed and updated by all authors of this study for completeness. It was subsequently refined as coding progressed. Each study was coded independently by two raters, with disagreement resolved by the first author. See *SI Appendix*, section 7 for more discussion of coding procedures. A list of coded characteristics for the 1989–2015 studies are shown in the *SI Appendix*, Tables S1 and S2.

Studies have fieldwork periods range from 1972 to 2015 for African Americans and 1989 to 2015 for Latinos. For most analyses in this paper, we focus on the period 1989–2015. We focus on this period because the data are sparse before this period (only four studies before 1989) and because our reading of the early studies indicates key methodological differences among these early studies that may affect their results. Resume audits typically signal race by using race-typed names on resumes, but the pre-1989 studies either indicated race directly on the resume [McIntyre et al. (28) put "Race: BLACK" on the minority resumes and nothing about race on the "white" resumes] or attached photos to resumes (a procedure used by Newman; ref. 29). Excluding the early studies leaves us with 21 estimates of discrimination against African Americans and nine against Latinos from 24 studies (six studies include estimates of discrimination against both African Americans and Latinos).

The Meta-Analysis Model. A meta-analysis aggregates information from across studies to produce an estimate of an effect of interest (30). In this study, our basic measure of discrimination is the discrimination ratio. This is the ratio of the percentage of callbacks for interviews received by white applicants to the percentage of callbacks or interviews received by African Americans or Latinos. Formally, if c^w is the number of callbacks received by whites, and c^m is the number of callbacks received by African Americans or Latinos, and n^w is the number of applications submitted by white applicants, and n^m is the number of applications submitted by African American or Latino applicants, then the discrimination ratio is $(c^{w}/n^{w})/(c^{m}/n^{m})$. Ratios above 1 indicate whites received more positive responses than African Americans or Latinos, with the amount above 1 multiplied by 100 indicating the percentage higher callbacks for whites relative to the minority group. Because audit studies equate groups on their nonracial characteristics either through matching and assignment of characteristics (in-person audits) or through random assignment (most resume audits), no further within-study controls are required. *SI Appendix*, section 8 discusses potential alternative measures of discrimination using the difference in proportions and the odds ratio, and presents alternative results using these measures. Our basic result—no decline in discrimination against African Americans over time—holds using both of the alternative measures, whereas evidence of a decline in discrimination for Latinos appears somewhat stronger with the difference in proportions or the odds ratio.

The goal of a meta-analysis is to combine information across studies. This requires measuring the information each study contains about discrimination against a group. The information each study provides is inversely proportional to the square of the SE of the discrimination ratio. We calculate the SE of the ratio from counts reported in each study, accounting for audit pairs in the design when possible. In cases where information on paired outcomes is available from the study (counts of pairs in which both the white and the nonwhite tester receive a callback, white yes nonwhite no, white no nonwhite yes, neither get a callback), we calculated SEs of discrimination ratios accounting for the pairing (see *SI Appendix*, section 9 for details and formulas). For studies that are not paired between whites or nonwhites or where paired outcomes are not reported, we use formulas for the SE for unpaired groups. This formula will slightly overestimate the SE of the effect for studies that are paired but we treat as unpaired due to lack of information about the outcomes at the pair level, underweighting these studies a bit in computing the overall effect, and slightly inflating the overall cross-study SE.

Of course field experiments vary in their characteristics, such as the geographic area they cover, the exact job sectors covered, and details of their methodology. To account for this variability in understanding the time trend, we use two procedures. First, we include controls, discussed further below, for many study characteristics. Second, to capture sources of variability not covered by the covariates, we use a random effects specification (22). Random effects incorporate a variance component capturing variation in outcomes across studies that are due to unobserved study-level factors. Random effects are recommended whenever there is reason to believe that the effect in question is likely to vary as a function of design features of the study, rather than representing a single underlying effect that is constant over the whole population. This is surely the case in our analysis, as we expect that the level of racial discrimination may depend on the year of the study, the situation the study considers (e.g., the occupational categories), the skill level of the applicants, and so on. The random effect increases the SEs of estimates to correctly account for variabilities among studies in drawing inferences about overall trend.

More formally, random-effects meta-analysis allows the true effects of race on the callback rate in each situation estimated by each study, θ_i, to vary between studies by assuming that they have a normal distribution around a mean effect, θ. If y_i is the discrimination ratio in the ith study, then the meta-analysis model is as follows:

$$\ln\,(y_i) = \theta + u_i + e_i,\text{ where } u_i \sim N(0,\tau^2)\text{ and } e_i \sim N(0,\sigma_i^2).$$

Here, τ^2 is the between-study variance, estimated from between-study variance as part of the meta-analysis model, while σ_i^2 is the variance of the log response ratio in the ith study, estimated from study counts as described above. Following standard practice in the meta-analysis literature, we log the response ratio to reduce the asymmetry of the ratio.

Meta-regression allows that the rate of discrimination is a function of a vector of k characteristics of the studies and effects, x, plus (in the random effects specification) residual study-level heterogeneity (between study variance not explained by the covariates). The model assumes the study-level heterogeneity follows a normal distribution around the linear predictor:

$$\ln\,(y_i) = x_i\beta + u_i + e_i,\text{ where } u_i \sim N(0,\tau^2)\text{ and } e_i \sim N(0,\sigma_i^2),$$

where β is a $k \times 1$ vector of coefficients (including a constant), and x_i is a $1 \times k$ vector of covariate values in study i (including a 1 for a constant). Estimation is by restricted maximum likelihood. For details, see *SI Appendix*, section 9.

To explore trends over time, we include covariates for the year of fieldwork of the study. In the simplest models, the only covariate is this time trend. In later models, we include a more extensive set of predictors to control for other factors that might confound the time trend. These additional controls include resume audit vs. field audit as the study method, gender and education level of the fictitious applicants, occupations tested, unemployment rates at the field sites used for testing, criminal background of some fictitious applicants, and region of the country. For discussions of why these controls were selected, see *SI Appendix*, section 4 (see *SI Appendix*, Tables S1 and S2 for descriptive statistics on the controls; for a discussion of trends in covariates, see *SI Appendix*, section 10).

ACKNOWLEDGMENTS. We thank Anthony Heath, Fenella Fleischmann, Matthew Salganik, Frank Dobbin, András Tilcsik, Donald Green, David Neumark, Hedwig Lee, two anonymous PNAS reviewers, and the editor for comments; Larry Hedges for methodological advice; and Jim Cheng Chen and Joshua Aaron Klingenstein for excellent research assistance. We have received financial support for this project from the Russell Sage Foundation and the Institute for Policy Research at Northwestern University.

1. Donohue J-J, Heckman J (1991) Continuous versus episodic change: The impact of civil rights policy on the economic status of blacks. *J Econ Lit* 29:1603–1643.
2. Wilson W-J (1978) *The Declining Significance of Race: Blacks and Changing American Institutions* (Univ of Chicago Press, Chicago).
3. Heckman J (1998) Detecting discrimination. *J Econ Perspect* 12:101–116.
4. Schuman H, Steeh C, Bobo L, Krysan M, eds (1998) *Racial Attitudes in America: Trends and Interpretations* (Harvard Univ Press, Cambridge, MA), Rev ed.
5. Jencks C, Phillips M (2011) *The Black-White Test Score Gap* (Brookings Institution Press, Washington, DC), p 542.
6. Reardon S-F, Fahle E (2017) Education. State of the union: The poverty and inequality report (Stanford Center on Poverty and Inequality, Palo Alto, CA), special issue, *Pathways* Magazine.
7. Dobbin F (2009) *Inventing Equal Opportunity* (Princeton Univ Press, Princeton, NJ), p 321.
8. Tesler M, Sears D-O (2010) *Obama's Race: The 2008 Election and the Dream of a Post-Racial America* (Univ of Chicago Press, Chicago).
9. Austin A-A (2013) The unfinished march: An overview. (Economic Policy Institute). Avaliable at www.epi.org/publication/unfinished-march-overview/. Accessed May 25, 2016.
10. Cancio A-S, Evans D, Maume D-J (1996) Reconsidering the declining significance of race: Racial differences in early career wages. *Am Sociol Rev* 61:541–556.
11. Holzer H-J, Offner P, Sorensen E (2005) Declining employment among young black less-educated men: The role of incarceration and child support. *J Policy Anal Manage* 24:329–350.
12. Hajnal Z, Abrajano M (2016) Trump's all too familiar strategy and its future in the GOP. *Forum* 14:295–309.
13. Devine P-G, Elliot A-J (1995) Are racial stereotypes really fading? The Princeton trilogy revisited. *Pers Soc Psychol Bull* 21:1139–1150.
14. Bobo L-D, Charles C-Z, Krysan M, Simmons A-D (2012) The real record on racial attitudes. *Social Trends in American Life: Findings from the General Social Survey Since 1972*, ed Marsden P-V (Princeton Univ Press, Princeton), pp 38–83.
15. Dovidio J-F, Gaertner S-L (2010) Intergroup bias. *Handbook of Social Psychology*, eds Fiske S-T, Gilbert D-T, Lindzey G, Jongsma A-E (Wiley, Hoboken, NJ), 5th Ed.
16. Kinder D-R, Sears D-O (1981) Prejudice and politics: Symbolic racism versus racial threats to the good life. *J Pers Soc Psychol* 40:414–431.
17. McConahay J-B (1983) Modern racism and modern discrimination the effects of race, racial attitudes, and context on simulated hiring decisions. *Pers Soc Psychol Bull* 9: 551–558.
18. Bonilla-Silva E (2006) *Racism Without Racists: Color-Blind Racism and the Persistence of Racial Inequality in America* (Rowman & Littlefield, Lanham, MD), 2nd Ed.
19. Bertrand M, Mullainathan S (2004) Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *Am Econ Rev* 94:991–1013.
20. Pager D, Western B, Bonikowski B (2009) Discrimination in a low-wage labor market: A field experiment. *Am Sociol Rev* 74:777–799.
21. National Research Council (2004) *Measuring Racial Discrimination*, eds Blank RM, Dabady M, Citro CF (National Academies Press, Washington, DC).
22. Raudenbush S-W (2009) Analyzing effects sizes: Random effects coding. *The Handbook of Research Synthesis and Meta-Analysis*, eds Cooper HM, Hedges LV, Valentine JC (Russell Sage Foundation, New York), 2nd Ed, pp 295–316.
23. Oettinger G-S (1996) Statistical discrimination and the early career evolution of the black-white wage gap. *J Labor Econ* 14:52–78.
24. Fiske S (1998) Stereotyping, prejudice, and discrimination. *The Handbook of Social Psychology*, eds Gilbert D, Fiske S, Lindzey G (Wiley, New York), 4th Ed, pp 357–411.
25. Altonji J-G, Pierret C-R (2001) Employer learning and statistical discrimination. *Q J Econ* 116:313–350.
26. Wessel D (September 11, 2003) Fear of bias suits may be affecting hiring decisions. *Wall Street Journal*, section A, p 2.
27. Firebaugh G, Davis K-E (1988) Trends in antiblack prejudice, 1972–1984: Region and cohort effects. *Am J Sociol* 94:251–272.
28. McIntyre S, Moberg D-J, Posner B-Z (1980) Preferential treatment in preselection decisions according to sex and race. *Acad Manage J* 23:738–749.
29. Newman J-M (1978) Discrimination in recruitment: An empirical analysis. *Ind Labor Relat Rev* 32:15–23.
30. Borenstein M, Hedges L-V, Higgins J-P-T, Rothstein H-R (2009) *Introduction to Meta-Analysis* (John Wiley & Sons, Chichester, UK), p 421.

Supporting Information Appendix for:

The Persistence of Racial Discrimination: A Meta-Analysis of Field Experiments in Hiring

Lincoln Quillian, Devah Pager, Ole Hexel,
Arnfinn H. Midtbøen

correspondence to: l-quillian@northwestern.edu

Includes:

Supporting Text
Figs. S1 to S2
Tables S1 to S10

Supporting Information Appendix

1. Problems in Assessing Trends in Discrimination

Methods for measuring discrimination are notoriously flawed, a problem only compounded by attempts to make comparisons over time. The most common approach to studying trends in racial discrimination has been the residual method: based on a statistical model of an outcome, the residual race gap left unexplained after other factors are accounted for is attributed to discrimination. If this residual is reduced over time, discrimination is thought to have declined. This approach suffers from the significant weakness that its validity rests crucially on controlling for all other factors that influence the outcome and may vary between racial groups – a circumstance that is generally impossible to verify (1, 2). This problem is compounded by attempts to draw comparisons over time, as the source and importance of relevant controls (or unobservables) may shift over time.

Other approaches to assessing trends in discrimination have relied on survey and institutional reports. These methods have different but no less serious problems. One approach relies on self-reports of discrimination from targets (e.g. 3). The weakness of this method is that it cannot detect discrimination that targets are not aware of; and, conversely, targets may sometimes mistakenly attribute a poor outcome to racial discrimination when the outcome has a different basis. A second method uses the frequency of formal complaints of discrimination from targets or lawsuits alleging discrimination (4). This method too only captures discrimination that victims are aware of, and formal complaints or lawsuits are strongly influenced by institutional factors that discourage or encourage reporting or lawsuits (5).

A final method relies on interviews with potential perpetrators, but this approach faces obvious potential problems with underreporting of socially unacceptable conduct, and also cannot capture discrimination grounded in implicit or subtle attitudes that perpetrators may not be aware of. Again, we face the problem that these analyses require strong assumptions that cannot be tested, leading to unresolved questions of whether discrimination is really changing or whether apparent changes are artifacts in measurement (1, 6, 7).

In an effort to address the problems of measurement and potential omitted variable bias plaguing research in this area, researchers have increasingly turned to experimental methods (8). Field experiments, in particular, offer a powerful design for isolating the causal effect of discrimination within the context of real-world hiring decisions. Researchers conducting field experiments have a high degree of control over the experimental conditions (whether matched testers or racially-identifiable names randomly assigned to resumes), providing a strong basis from which to draw causal conclusions about hiring discrimination. Likewise, situating these experiments in the context of actual hiring decisions offers conclusions that readily generalize to real labor market dynamics.

That being said, field experiments of discrimination have also been subject to important critique. Field experiments typically focus on a single skill level (or a narrow range) for their tests of discrimination, thereby potentially missing variation in rates of discrimination facing those of greater of lesser skill levels and across a wider range of occupations. In-person audits can suffer from spurious effects due to the poor matching of test partners and/or experimenter effects due to the expectations of testers (9). Resume audits, which rely on racially-identifiable names for their key experimental treatment, may confound racial discrimination with effects driven by class, by relying on names that may signal both (10). Both types of audit studies typically rely on random samples of job listings for their tests of discrimination. Selective applications by job seekers and selective recruitment by employers may complicate the degree to which estimates of discrimination from audit studies map on to real-world experiences with discrimination (11, 12). These potential limitations are by now well known and have received extensive attention in the research literature. While by no means perfect, field experiments continue to be considered the most valid measure of hiring discrimination available (1). In the current study, we capitalize on the meta-analysis approach which draws information from across studies. We can therefore explicitly model design issues (such as resume-based versus in-person) and variability across research studies.

2. Estimates of Average Discrimination Levels

The first step of our analysis is to consider the overall levels of discrimination by group and the extent of heterogeneity across studies. Figure S1 shows a forest plot of the discrimination ratios of 24 studies in which the target group is African-Americans, contrasted to whites.

In Figure S1 an overall average discrimination ratio for 1990 to 2015 based on random-effects meta-analysis is shown as the final diamond on the table (see Methods and Materials and the SI Appendix section 9 for details of the meta-analysis model). The results indicate that on average whites receive 36% more positive responses to job applications than African-Americans. A 95% confidence interval for the effect is 25% to 47% more callbacks. This reinforces the conclusion of many in-person and resume audits regarding the persistence of hiring discrimination against African-Americans, and provides a broader overall estimate of the average prevalence of discrimination in hiring by combining information from 24 studies. If we use all data from 1972 to 2015, a total of 24 estimates from 24 studies, whites receive on average 34% more positive responses than African-Americans, with a 95% CI of 23% to 46% more.

What accounts for the variability in estimates across studies? The model estimates that 67.3% of variability (I-squared) reflects differences resulting from differences in study characteristics (e.g. year, applicant education level, in-person vs resume audit, etc.). The remaining 32.7% of variability between studies could be accounted for by random variation in outcomes of individual studies. A significance test strongly rejects the hypothesis that the between-study variability is zero ($p<.001$), supporting a random-effects specification.

Figure S2 provides a forest plot for studies estimating rates of discrimination against Latinos relative to whites. On average whites received 24% more positive responses than Latinos, with a 95% confidence interval of 15% to 33% more.

3. Discrimination Trends

Table S3 presents estimates of the random-effect meta-regression used to create the trend line estimates and weights in Figures 1 and 2. In the model the discrimination ratio is logged. The coefficient of year in table S3 may be interpreted as percentage change in the discrimination ratio each year. For instance the coefficient of .004 for 1990-2015 indicates a trend upward of

.4% in the discrimination ratio per year. Both lines slope upward (coefficients of year in table S3 of greater than zero).

The rightmost numeric column of table S3 shows the coefficient that is the basis for the line of best fit in figure 2. Here the coefficient is less than zero, indicating the downward slope in discrimination against Latinos shown in figure 2.

4. Models of Discrimination Trends with Controls and Sensitivity Analyses

We then performed sensitivity analyses of changes in the dependent variable and built a model with additional controls to account for characteristics of studies that may confound the time trend. Because there are too few studies for Hispanics to support controls, we only perform the multivariate analyses for African-Americans. Model estimates are shown in table S4.

Coefficients of the model are shown in table S4. The coefficients for year are shown with confidence interval lines in figure 3. Models 2 to 4 alter the dependent variable (models 2 and 3) or sample (model 4). The results are robust to this change: we never have a statistically significant downward slope, and in all but one modification the point estimate is still an upward trend.

Models 5 to 8 add controls for characteristics of applicants and studies. These are added to control for characteristics of studies that could confound the trend over time and might influence discrimination levels. We include as controls common covariates that have been used to model discrimination rates in field experimental studies (e.g. 13, 14) and studies suggested by theoretical accounts of factors that might influence discrimination (e.g. 15, 6, 7). This includes study method, gender of the applicants, education level of applicants, presence of a fictional criminal background, metropolitan unemployment rates, region, and occupational sectors. Past theoretical literature has suggested reasons each of these may influence discrimination rates; for instance when unemployment rates are high, employers may be more likely to indulge discriminatory tastes given the wide range of applicants they have to choose from. However, we find that few of these controls are statistically significant predictors of discrimination.

Model 8 trims the model by dropping variables that consistently had t-ratios below 1.5. None of the controls has much effect on the time trend in discrimination, which remains insignificant and usually slightly upward in direction for white vs. African-American (indicated by the positive coefficient).

Alterations of the dependent variables for the Latino models are shown in table S5. These results show a bit less evidence for downward trend than for the model in table S3. The number of studies with Latino respondents is too low to allow for additional covariates.

In table S6 we use an alternative procedure to compute discrimination trend estimates: a model that pools African-American and Latino effect sizes. We use African-American as the base trend and allow an interaction of Latino by year. Coefficients of controls are constrained to be similar for the two groups.

Six of the studies in the analysis provide estimates of discrimination against both African-Americans and Latinos, which are represented as separate effects. We account for the clustering of the estimates within study by using robust standard errors allowing for correlated effects within study and a small sample adjustment (procedures discussed in 16). Model estimates are based on assumed correlation of rho=.8. Sensitivity analysis showed only slight changes in estimates with different values of rho.

The first two models show results including the linear trend, and dummies for whether the effect size is for the Latino target group (with the black target group as the reference). The coefficient for the Latino effect suggests less discrimination against Latinos, although this difference is only statistically significant (at $p<.1$) in the initial two models. Like the models estimated only for African-Americans, the pooled estimates provide no evidence of downward trend: all point estimates except one show a positive trend, and the one point estimate with negative trend is almost flat (model 4, year coefficient=-.004).

5. Publication Bias

Is it possible that recent studies finding little or no discrimination are more likely to remain unfinished or unpublished, thus excluded from our analysis and biasing our estimates? This problem, known as publication bias, represents a possible threat to the validity of any meta-analysis (17).

We find some evidence of publication bias in the analysis presented in table S7. Studies published in academic journals show somewhat higher discrimination ratios than those published elsewhere (reports, working papers, etc.), a result that is marginally statistically significant ($p<.1$). Because we were aware of this potential issue, we went to great lengths to include unpublished studies and reports. By contacting all known authors of audit studies and other researchers who specialize in the study of discrimination, we attempted to learn of any past, recent, or ongoing study that would not show up in our bibliographic search. Indeed, 12 of the 31 discrimination estimates in our study come from sources that were unpublished when initially included in our analysis. We believe, then, that any existing publication bias is unlikely to have affected our estimates drawing from both published and unpublished sources.

While this reduces the danger from selection into formal publication in academic journals or books, if studies that do not find discrimination are less likely to be written up at all—as working papers, reports, journal articles, etc.—there still may be an important form of publication bias even when unpublished studies are included ("write-up bias" may be a more accurate term for this).

As a more formal investigation into the possibility of publication or write-up bias (both of which create the problem that only studies finding race differences are likely to end up producing a report that can be included in our analysis), we included a study-level predictor that is diagnostic these problems. Specifically, we created a dummy variable to indicate whether the study design implied a primary focus on race or some other attribute. For instance, some studies focused on effects of particular educational qualifications, labor market histories, or a criminal record on receiving a positive response from an employer. While these studies also included racially diverse applicant profiles, allowing for an estimate of the effects of race on hiring outcomes, the primary emphasis of the study was on something other than racial discrimination. Note that in many of these studies, the variable of interest is included as a within-pairs contrast while secondary variables (i.e., race) are included as a between-pairs contrast. Given the way these priorities are expressed in the design of the study, the question of whether or not a study has race as its primary focus is not simply a matter of post-hoc interpretation. If write-up bias or publication bias is a serious problem, then we should find that race-focused studies tend to find more racial discrimination than not-race-focused studies, since a significant finding on race should be more important for write-up and publication in the case of studies with race as their primary emphasis. We prefer this test to tests in the meta-analysis literature based on symmetric distributions of effects, because these tests require the assumption that other confounding variables are not related to study size or effect sizes (see 17, section 23.3.1).

Results are shown in table S7. We find on average effect sizes in not-race-focused studies are somewhat larger than effect sizes of studies focused on race (not a statistically significant difference), as indicated by the coefficient above one, opposite the direction that publication or write-up bias would predict. According to this test, publication or write-up bias is unlikely to have produced inflated discrimination estimates.

6. Excluded studies.

Overall our search located 34 studies that were U.S.-based field experiments of hiring and included contrasts between white and non-white applicant profiles who were on-average equivalent in their labor-market relevant characteristics (e.g. education, experience level in the labor market, etc.). Two studies were excluded because it was not clear if employers were the ones making decisions producing discrepant outcomes because applications were conducted through an employment agency. One study contrasted whites to Arab-Americans; we excluded this study since it was the only study with this target group. One study did not report counts and the authors declined our request for counts of outcome by target groups. Two studies were excluded because they used mixed non-white groups and did not break out results separately for African-American and Latino applicants. All other studies focused on whites contrasted to African-Americans or Latinos (or both). Our remaining 28 studies yielded 24 estimates of discrimination against African-Americans and 9 against Latinos relative to whites. For most analyses in this paper, we exclude studies before 1989, which leaves us with 21 estimates of discrimination against African-Americans and 9 against Latinos from 24 studies (six studies include estimates of discrimination against both African-Americans and Latinos).

7. Coding

To ensure reliability, each study was coded independently by two raters. The first rating was completed by the third author under direction of the first author of this paper. The second rating was performed by two undergraduate students who were hired to conduct a second coding using the rubric. We then reconciled the results of the two codings, performing further investigation to find the correct answer on coding decisions in cases of disagreement. The variables coded were factual in nature (e.g. year of publication, counts of positive and negative responses for the white and non-white group, etc.); the main sources of disagreement in coding were difficulty in understanding the text or procedures of a particular study, or occasional judgment calls about what “fit” on a particular category. For instance, such judgments include decisions about whether working in a warehouse stockroom counts as “blue collar” employment (we did code it as blue collar), or whether an employer’s response that they would keep an applicant’s resume on file and might eventually request an interview constitutes a callback (we did not count this as a callback). In cases of disagreement or high uncertainty in the reconciliation process, the first author examined the study and broke the tie by assigning a code.

The coding involved two levels of information: study level and effect level. Study level characteristics are constant for the entire study, such as year of publication and type of publication outlet (academic journal, government report, etc.). Effect estimates refer to estimates of discrimination against a non-white group, with the number of effect sizes for a study depending on the number of target groups the study includes. A study that contrasts both African-Americans and Latinos to whites would produce two effect sizes.

We coded effects that measure discrimination based on counts of hiring outcomes by racial or ethnic group. Most studies included this information in the write-up. When the study did not include counts of outcomes in their research report, we requested counts from the authors, which we received for all studies with black or Latino target groups. We used all white and minority applicant profiles in computing discrimination ratios, except for cases in which the groups were non-equivalent in their labor market characteristics (most often where minorities were given somewhat stronger background qualifications than whites). For instance, one study included contrasts between white applicants with criminal records and black and Latino applicants without criminal records (19), and we do not include these audits because of the non-equivalence of the characteristics of white and minority testers (this study also included some audits between equivalent groups and is thus not excluded entirely). In our baseline analysis we include applicant profiles with characteristics such as a criminal background or a disability as long as this condition was equally present between white and minority testers. We perform some sensitivity analysis to illustrate changes in results when perhaps atypical groups like the disabled or those with a criminal background are eliminated (see SI Appendix section 4 and table S4).

We excluded some audits that were part of the New York Audit study reported in (14, 18) because they were based on between-pair comparisons, when within-pair comparisons focused on race were available from the same audit study. Adding in these audits has no effect on our results.

8. The Discrimination Ratio contrasted to the Difference in Proportions or the Log Odds Ratio

We use the ratio of the proportion of callbacks received by white applicants to the ratio received by nonwhite applicants to measure discrimination. Two other candidate measures that could be used instead are the difference in proportions or the odds ratio (19).

The difference in proportions is a measure widely used in the correspondence and audit literature. In our context this measure is the difference between the percentage of callbacks received by whites and the percentage of callbacks received by minorities ($\frac{c^w}{n^w} - \frac{c^m}{n^m}$). This is a logical measure for single studies, but we think it is less well-suited for a meta-analytic context in which base rates vary over studies. When the base rate of callbacks gets relatively close to 0% (many studies have low callback rates), then the floor places a limit on the size of the difference in proportions. When base rates differ over studies, the floor places varying limits across studies.

This problem is especially clear in paired field experiments. In paired studies one or more majority and minority applicants apply for the same job. In most field experiments the most common outcome is that neither the majority nor the minority auditor gets a callback for an interview. This "neither" outcome provides no information about discrimination. But the frequency of this outcome sets an upper limit on the size of a difference ratio of discrimination. For instance, if in 80% of audits neither applicant gets a callback, then the maximum difference in the proportion of callbacks between groups is .2. By contrast, ratio measures are not automatically limited by low base-rates of callbacks.

The result of using a difference measure in the presence of unequal base rates is greater heterogeneity in the outcome across studies in the meta-analysis, because the difference measure absorbs base rate variability. This is a problem that has been recognized in methodological discussions in the meta-analysis literature as an undesirable feature of the difference in proportions as an outcome measure (see section 13.2.1 in 19).

We also find the ratio measure is preferable because the base rate of callbacks must implicitly be invoked to understand the implications of differences in callback rates for racial disparities in hiring. If whites receive 8% callbacks and African-Americans receive 4% callbacks, then whites receive 200% as many callbacks per application submitted, an outcome strongly favoring whites in hiring. By contrast, if whites receive 44% callbacks and African-Americans receive 40% callbacks, whites receive 10% more callbacks per application submitted, a much smaller advantage in getting a job. Using the difference measure in both cases the racial disparity is measured to be the same at 4%.

Because the difference in proportions is a widely used measure in the field experimental literature, we conducted a sensitivity analysis using this measure. Results using the difference in proportion measure for African-Americans are shown in table S8, column 1. We use the applicant attributes model, which we take as our most basic model. The point estimate for the time trend in this model is upward, and the coefficient is small and not statistically significant – similar to what we find with the ratio measure.

Result for the difference measure for Latinos are shown in column 2 of table S8. There is a statistically significant downward trend with this measure.

A second candidate measure is the natural log of the odds ratio. The odds ratio measure takes the ratio in the odds of the outcome between groups in place of the ratio in the proportions. Formally, if c^w is the number of callbacks received by whites, and c^m is the number of callbacks received by blacks or Latinos, and n^w is the number of applications submitted by white applicants, and n^m is the number of applications submitted by black or Latino applicants, then the odds ratio is $\frac{\frac{c^w}{n^w - c^w}}{\frac{c^m}{n^m - c^m}}$. The odds ratio is logged to make its distribution more symmetric before regression analysis. This measure has good statistical properties, but we prefer the ratio of proportions to the odds ratio because of the greater interpretability of the ratio of proportions. Ratios of proportions are more intuitive than ratios of odds. Further, many field experimental studies also use ratio of proportion as their basic measure of discrimination (e.g. 14), but we know of none that use odds ratios for basic group comparison of outcomes.

Results using the log odds ratio as the outcome for white vs. African-Americans are shown in table S9 column 1. Again, the slope of the year variable is slightly upward and very close to zero. The main result is unchanged from the log ratio of proportions. Results for Latinos are shown in table S9 column 2.

For Latinos, we find a statistically significant downward trend with both the difference in proportions and the log odds ratio. This strengthens the case for decline in discrimination against Latino job seekers across the years we consider. However, our inability to include controls due to the small number of studies including Latinos weakens our ability to draw inferences about trends in anti-Latino discrimination. And modifications to the dependent variable discussed in section 4 and shown in table S5 result in non-significant year coefficients for Latinos.

9. Weighting and Model Estimation

Meta-analysis requires estimating the sampling variability of the discrimination estimate for each study. To estimate this we use standard formulas for variability of a ratio due to sampling error and the counts of outcomes from each studies. For studies that are unpaired or do not report paired outcome, the variance of the log risk ratio in the ith study is estimated by (see 20, formula 5.3):

$$\sigma_i^2 = \mathrm{Var}(\ln(y_i)) = 1/c^w_i - 1/n^w_i + 1/c^m_i - 1/n^m_i$$

Where c^w is the number of callbacks received by whites, c^m is the number of callbacks received by blacks or Latinos, and n^w is the number of applications submitted by white applicants, and n^m is the number of applications submitted by black or Latino applicants

For studies that use a paired design – with one minority and one white applicant applying for each job – and report paired outcomes, we use an alternative formula to account for the pairing. If p^a are the number of pairs in which both majority and minority testers receive a callback, p^b are the number of pairs in which the majority tester received a callback but not the minority, p^c are the number of pairs in which the minority tester received a callback but not the majority, and p^d are the number of pairs in which neither tester received a callback, then the variance of the log odds ratio in the ith study with paired data is (see 21):

$$\sigma_i^2 = \mathrm{Var}(\ln(y_i)) = \frac{p_i^a + p_i^d}{(p_i^a + p_i^b)(p_i^a + p_i^c)}$$

The meta-regression model is:

$$\ln(y_i) = x_i\beta + u_i + e_i, \text{ where } u_i \sim N(0, \tau^2) \text{ and } e_i \sim N(0, \sigma_i^2)$$

where β is a $k \times 1$ vector of coefficients (including a constant), and x_i is a $1 \times k$ vector of covariate values in study i (including a 1 for a constant).

The random effect variance (τ^2) is estimated with the parameters as part of the meta-analysis model. Effectively, the random effect is estimated by the extent of residual variation in the outcome that cannot be accounted for by sampling variability for each study (σ_i^2) or the covariates. Estimation is by restricted maximum likelihood, which has the advantage of giving nearly fully efficient estimates and nearly unbiased estimation of variance components in smaller samples. For further details of estimation, see 22.

In practice we estimated the basic meta-analyses and meta-regressions in Stata using the "metan" suite of commands (23). The Knapp-Hartung modification, which we employ in all tables except table S6 uses the t-distribution rather than the normal distribution for inference. The t distribution has been shown to provide more accurate tests and confidence intervals in simulation studies (24).

For the pooled model shown in table S6, we use the "robumeta" command in Stata written by Tipton and co-authors (16). This include robust estimation methods to adjust for correlated effect sizes in cases for which the African-Americans and Latinos effect sizes are from the same studies.

A bibliographic list of studies in the meta-analysis, data, and code are available from the discrimination meta-analysis project website at: http://sites.northwestern.edu/dmap

10. Trends in Covariates

Altonji, Elder, and Taber (25) argue that confounding by observable covariates can provide evidence about confounding by unobservable covariates, viewing the observable covariates as a random sample on a larger set of unobserved covariates. This suggests that, under assumptions they outline, we can test for potential bias in unobservables by examining the correlation of observed variables with year of survey, our primary variable.

Table S10 shows an OLS regression of year of fieldwork on all of the covariates. An F-test regressing the year of study on all of the control variables has a p-value of .1938, suggesting no clear evidence of trend in the observed covariates over years. Under the logic that Altonji et al. outline, this suggests unobservables also are unlikely to be correlated with year in a strong enough way to create substantial bias in estimates of year.

Looking at bivariate relationships to year and models with smaller sets of covariates (not shown), two variables are sometimes statistically significant in predicting survey year: audit design (in-person vs. resume audit) and the inclusion of applicants with a criminal background. These are not surprising in light of trends in job application practices and trends in incarceration: resume audits have become more popular as the rise of the internet has made online application procedures increasingly common; including testers with fake criminal records has become more popular because of the growth of incarceration and concern about its consequences.

We include sensitivity analyses in our base results excluding in-person audits (resume audit studies only) and excluding applicant profiles with fake criminal backgrounds. These results are shown in table S4, and the slope estimates shown in figure 3 in the main text. Excluding auditors with criminal records and limiting our study to only resume audit studies both produce no trend in discrimination ratios for African-Americans. We conclude that there are no trends in the covariates that appear potentially problematic for our results.

References

1. National Research Council (2004) Measuring racial discrimination. eds Blank R-M, Dabady M, Citro C-F (National Academies Press, Washington, DC).
2. Charles K-K, Guryan J (2011) Studying discrimination: Fundamental challenges and recent progress. *Annu Rev Econom* 3:479-511.
3. Feagin J-R, Sikes M-P (1994) *Living with racism: The black middle-class experience* (Beacon Press, Boston).
4. Roscigno V-J (2007) *The face of discrimination: How race and gender impact work and home lives* (Rowman & Littlefield Publishers, Lanham).
5. Nielsen L-B, Nelson R-L (2005) Scaling the pyramid: A sociolegal model of employment discrimination litigation. *Handbook of Employment Discrimination Research*, eds Nielsen L-B, Nelson R-L), pp 3-34.

6. Quillian L (2006) New approaches to understanding racial prejudice and discrimination. *Annu Rev Sociol* 32:299-328.
7. Pager D, Shepherd H (2008) The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets. *Annu Rev Sociol* 34:181-209.
8. Pager D (2007) The use of field experiments for studies of employment discrimination: Contributions, critiques, and directions for the future. *Ann Am Acad Pol Soc Sci* 609:104-133.
9. Heckman J-J (1998) Detecting discrimination. *J Econ Perspect* 12(2):101-116.
10. Gaddis, S-M (in press) How black are Lakisha and Jamal? Racial perceptions from names used in correspondence audit studies. *Soc. Sci.*
11. Fix M, Struyk R-J eds (1993) *Clear and convincing evidence: Measurement of discrimination in America* (Urban Institute Press, Washington DC).
12. Pager D, Pedulla D-S (2015) Race, self-selection, and the job search process. *Am J Sociol* 120(4):1005-1054.
13. Bertrand M, Mullainathan S (2004) Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. Am Econ Rev 94(4):991-1013.
14. Pager D, Bonikowski B, Western B (2009) Discrimination in a low-wage labor market: A field experiment. Am Sociol Rev 74(5):777-799.
15. OECD (2013) Discrimination against immigrants – measurement, incidence and policy instruments. *International migration outlook 2013*, (OECD Publishing), pp 191-230. Retrieved April 1, 2017. (http://www.oecd-ilibrary.org/social-issues-migration-health/international-migration-outlook-2013/discrimination-against-immigrants-measurement-incidence-and-policy-instruments_migr_outlook-2013-7-en).
16. Tipton E (2015) Small sample adjustments for robust variance estimation with meta-regression. *Psychol Methods* 20(3):375-393.
17. Sutton A-J (2009) Publication bias. *The handbook of research synthesis and meta-analysis*, eds Cooper H-M, Hedges L-V, Valentine J-C (Russell Sage Foundation, New York), 2nd Ed.
18. Pager D, Western B, Sugie N (2009) Sequencing disadvantage: Barriers to employment facing young black and white men with criminal records. *Ann Am Acad Pol Soc Sci* 623(1):195-213.
19. Fleiss, JL, Berlin JA (2009) Effect sizes for dichotomous data. The handbook of research synthesis and meta-analysis, eds Cooper H-M, Hedges L-V, Valentine J-C (Russell Sage Foundation, New York), 2nd Ed.
20. Borenstein M, Hedges L-V, Higgins J-P-T, Rothstein H-R (2009) *Introduction to meta-analysis* (John Wiley & Sons, Chichester, U.K) p 421.
21. Zou G-Y (2007) One relative risk versus two odds ratios: Implications for meta-analyses involving paired and unpaired binary data. *Clinical Trials* 4(1):25-31.
22. Raudenbush S-W (2009) Analyzing effects sizes: Random effects coding. *The handbook of research synthesis and meta-analysis*, eds Cooper H-M, Hedges L-V, Valentine J-C (Russell Sage Foundation, New York), 2nd ed, pp 295-316.
23. Bradburn M, Deeks J, Altman D (1998) Metan – an alternative meta-analysis command. *Stata Technical Bulletin* 44: 4-15.
24. Knapp G, Hartung J (2003) Improved tests for a random-effects meta-regression with a single covariate. *Stat med* 22: 2693-2710.
25. Altonji, J, Elder T, Taber C (2005) Selection on observed and unobserved variables: assessing the effectiveness of catholic schools. *J Political Economy* 113 (1): 151-184.

Figure S1: Discrimination Ratios African-Americans

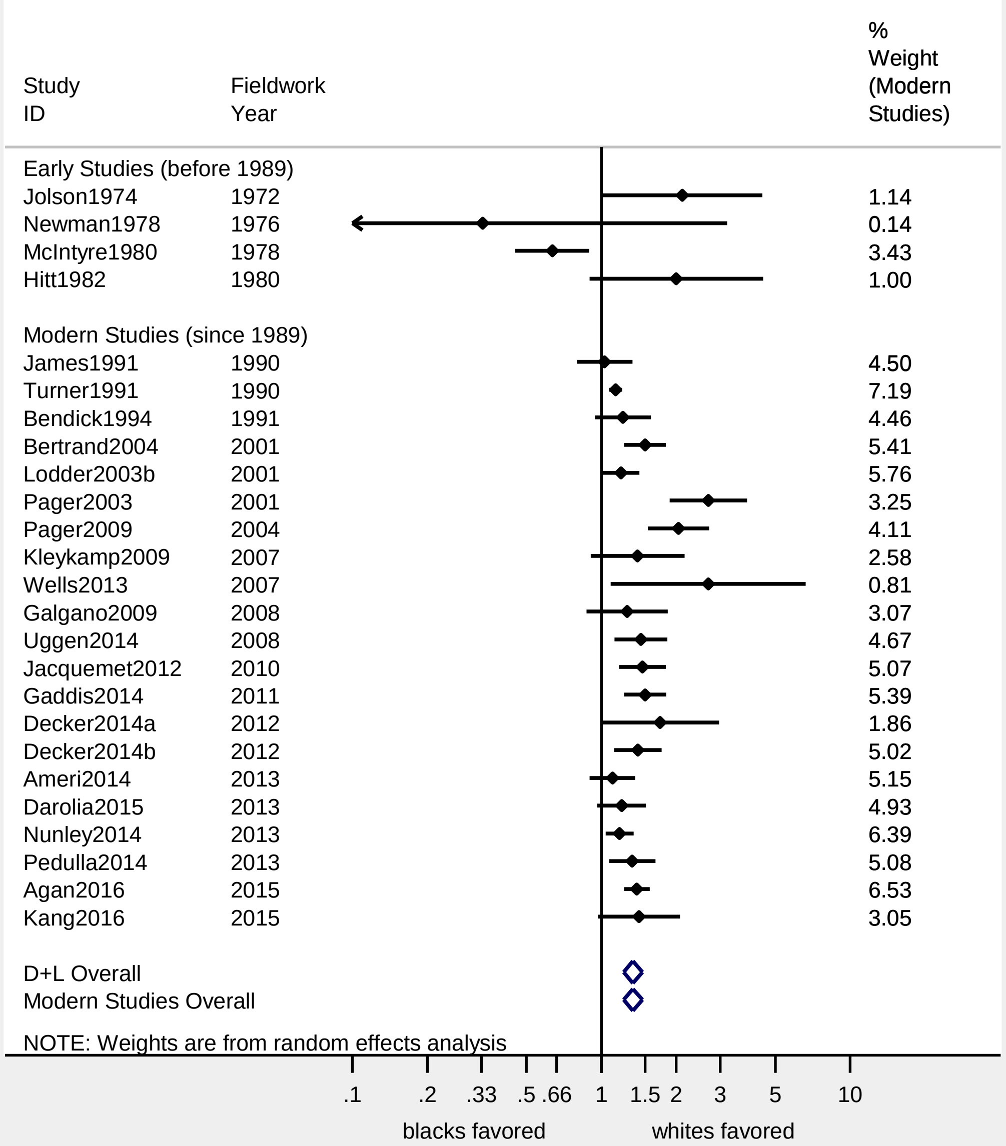


Figure S2: Discrimination Ratios, Latinos

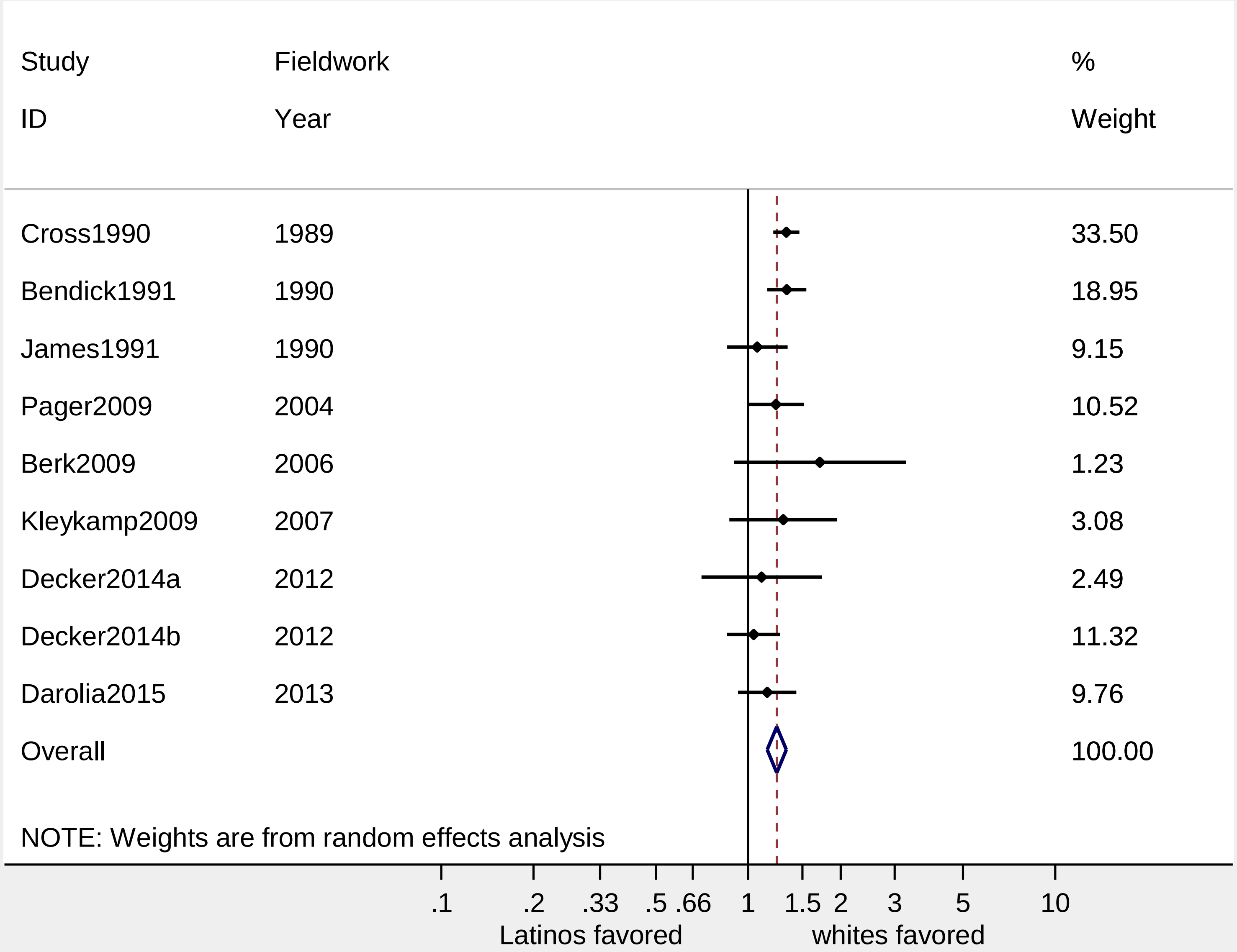


Table S1: Characteristics of Field Experiments of Discrimination Since 1989 (Categorical variables)[a]

	Count	Percentage
Study Method		
In-person audit	12	40%
Resume audit	18	60%
Target Group		
Black/African-American	21	70%
Hispanic/Latino	9	30%
Race is Not a Primary Focus of the Publication		
Yes	14	45%
Publication Type		
Dissertation or MA thesis	3	10%
Journal article	18	60%
Report	7	23%
Working paper	2	7%
Paired design with mixed-race pairs? [b]		
Yes	7	23%
No	23	66%
Sample Frame (1-Yes; More than one per study possible)		
Newpaper Ads	17	46%
Online ads or job bank	19	51%
Other (e.g., industry lists, employment agencies)	2	5%
Gender of Applicants [c]		
Female	3	10%
Male	14	47%
Both	13	43%
Criminal Record		
Yes (some auditors in study have fictitious criminal records)	9	29%
Region [d]		
Northeast	6	19%
Midwest	6	19%
South	3	10%
West	7	23%
Occupational Categories (more than one per study possible) [e]		
Includes blue collar	14	47%
Includes office focus	24	80%
Includes restaurant jobs	20	67%

[a] Tabulated at effect level, excluding four studies with fieldwork dates before 1989

[b] Studies with =>2 applications/job from different racial/ethnic groups that provide counts at the pair level

[c] Two studies did not clearly state gender of testers. These were coded to both.

[d] Eight studies (not included in this tabulation) use national or multi-regional samples.

[e] One study lacking clear occupatonal information was coded as not including restaurant jobs

Table S2: Characteristics of Field Experiments of Discrimination Since 1989 (Continuous variables)

	Mean[a]	Std. Dev.
Year of Fieldwork [b]	2005.3	8.7
Jobs Applied For		
In-person Audits	185.7	146.6
Resume Audits	1422.3	1636.8
Applications Submitted		
In-person Audits	440.4	244.2
Resume Audits	3061.2	2331.7
Positive Response Rates of Callbacks		
White Auditors/Resumes	25.1%	18.6%
Non-white Auditors/Resumes	18.7%	15.6%
Response Ratio (white/minority)	1.42	0.41
Unemployment Rate [c]	6.6%	1.5%
Education in Years [d]	13.6	1.7

[a] Unweighted means at the effect level, n=30

[b] For studies that do not indicate year of fieldwork (n=7), we coded fieldwork year to be year of publication minus 2 years for published articles, and minus 1 year for reports and working papers.

[c] Average unemployment rate of metropolitan areas in the study. Rates are averaged across months or years of the fieldwork

[d] Nineteen effect sizes are based on applicants with a single education level, the other eleven included applicants with varying levels of education. We use a continuous education measure coded high school degree=12, some college, no degree =13, associate's degree=14, four year college degree=16, graduate degree=17.

Table S3: Random-Effects Meta-Regressions of Log Discrimination Ratios on Year

Outcome: Log of Ratio of Callback Proportions White/Minority

	African-Americans		Latinos
Predictor Variable	Since 1989	All Years	Since 1989
Fieldwork Year	0.004	0.007	-0.006 +
	(0.005)	(0.005)	(0.003)
Tau-squared	0.021	0.033	0.000
N (effects/studies)	21	25	9

Notes: Models are estimated with a constant, but it is not shown.

Standard errors are shown in parentheses. Significance tests employed the Knapp-Hartung modification.

Tau-squared is the estimated variation between studies (attributable to study differences).

** = p<.01; * = p<.05, + = p< .1; two-tailed tests

Table S4: Random-Effects Meta-Regressions of Log Discrimination Ratios, African-American, Studies After 1989

Outcome: Log of Ratio of Callback Proportions White/African-American

		Modified Outcomes			Base Outcome with Controls			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Predictor Variable	Base Model, No Controls	Job Offer in Place of Interview / Callback	No Criminal Record or Disability	Resume Audit Studies Only	Applicant Attributes	Metro. & Region	Occupations	Trimmed
Fieldwork Year	0.004	-0.008	0.006	-0.002	0.008	0.001	0.008	0.013
	(0.005)	(0.007)	(0.004)	(0.007)	(0.010)	(0.011)	(0.011)	(0.009)
Study Method is In-Person Audit					0.315	0.228	0.361	0.279
(1=In-person audit, ref.=resume audit)					(0.182)	(0.273)	(0.212)	(0.164)
Applicants Male Only (1=yes, ref.= male & female)					-0.007	-0.166	-0.073	-0.139
					(0.113)	(0.139)	(0.152)	(0.126)
Applicants Female Only (1=yes, ref.= male & female)					0.100	-0.311	0.151	
					(0.155)	(0.457)	(0.173)	
Applicant Education in Grades Completed					0.047	0.060	0.063	
					(0.047)	(0.067)	(0.063)	
Some Applicants have (Falsified) Criminal Records (1=yes)					0.206	0.109	0.158	-0.035
					(0.166)	(0.245)	(0.206)	(0.149)
Unemployment Rate of Metropolitan Area(s), Percentage						-0.006		
						(0.052)		

Table S4, Continued

	(1) No Controls	(3) Job Offer in Place of Interview / Callback	(4) No Criminal Record or Disability	(4) Resume Audit Studies Only	(5) Applicant Attributes	(6) Metro. & Region	(7) Occupations	(8) Trimmed
Region = Midwest (1=yes, vs. reference of multi-region)						0.405		0.147
						(0.367)		(0.113)
Region = Northeast (1=yes)						0.372 +		0.155
						(0.198)		(0.168)
Region = South (1=yes)						-0.005		
						(0.190)		
Region = West (1=yes)						0.023		
						(0.195)		
Includes blue collar occupations (1=yes)							0.229	0.117
							(0.185)	(0.158)
Includes occupations with an office focus (1=yes)							-0.009	
							(0.171)	
Includes restaurant occupations (1=yes)							-0.073	
							(0.188)	
Tau-squared	0.021	0.021	0.012	0.004	0.023	0.023	0.029	0.021
Number of Studies	21	21	21	13	21	21	21	21

Notes: Standard errors are shown in parentheses. Significance tests employ the Knapp-Hartung modification.
Models are estimated with a constant, but constant is not shown.
Tau-squared is the estimated variation between studies (attributable to study differences).
** = p<.01; * = p<.05, + = p< .1; two-tailed tests

Table S5: Random-Effects Meta-Regressions of Log Discrimination Ratios, Latinos, Modified Outcomes

Outcome: Log of Ratio of Callback Proportions White/Latino

Predictor Variable	Latinos: Job Offer in Place of Interview / Callback	Latinos: No Criminal Record or Disability	Latinos: Resume Audit Studies Only
Fieldwork Year	-0.009	-0.006	-0.008
	(0.005)	(0.004)	(0.005)
Tau-squared	0.000	0.000	0.000
N (studies)	9	9	5

Notes: Standard errors are shown in parentheses. Significance tests employ the Knapp-Hartung modification.
Models are estimated with a constant, but constant is not shown.
Tau-squared is the estimated variation between studies (attributable to study differences).
None of the coefficeints are significant at p<.1, two-tailed tests.

Table S6: Random-Effects Meta-Regressions of Log Discrimination Ratios, African-Americans and Latino Effect Sizes Pooled

Outcome: Log of Ratio of Callback Proportions White/Minority

Predictor Variable	(1) No Controls	(2)	(3) Applicant Attributes	(4) Metro. & Region	(5) Occupations
	Since 1989	All Years	Since 1989	Since 1989	Since 1989
Fieldwork Year (Year 2015 = 0)	0.004	0.007	0.003	-0.004	0.006
	(0.005)	(0.006)	(0.010)	(0.008)	(0.011)
White/Hispanic	-0.193 +	-0.209 +	-0.200	-0.198	-0.233
(1=yes, vs. white/black)	(0.079)	(0.090)	(0.131)	(0.115)	(0.136)
White/Hispanic * Fieldwork Year	-0.009	-0.011	-0.010	-0.009	-0.009
	(0.005)	(0.007)	(0.008)	(0.007)	(0.010)
Study Method is Audit			0.1248	0.055	0.198
(1=Audit, ref.=correspondence)			0.1109	(0.108)	(0.126)
Applicants Male Only (1=yes, ref.= male & female)			-0.033	-0.147	-0.082
			(0.077)	(0.102)	(0.137)
Applicants Female Only (1=yes, ref.= male & female)			0.006	-0.491	0.075
			(0.087)	(0.318)	(0.050)
Applicant Education in Grades Completed			0.024	0.047	0.030
			(0.041)	(0.033)	(0.032)
Some Applicants have (Falsified) Criminal Records (1=yes)			0.179	0.053	0.114
			(0.194)	(0.118)	(0.243)
Unemployment Rate of Metropolitan Area(s), Percentage				0.001	
				(0.026)	

Table S6 Continued:

	(1)	(2)	(3)	(4)	(5)
	No Controls		Applicant Attributes	Metro. & Region	Occupations
	Since 1989	All Years	Since 1989	Since 1989	Since 1989
Region = Midwest (1=yes, vs. reference of multi-region)				0.534	
				(0.302)	
Region = Northeast (1=yes)				0.355 *	
				(0.133)	
Region = South (1=yes)				0.011	
				(0.194)	
Region = West (1=yes)				0.047	
				(0.109)	
Includes blue collar occupations (1=yes)					0.200
					(0.108)
Includes occupations with an office focus (1=yes)					0.041
					(0.088)
Includes restaurant occupations (1=yes)					-0.061
					(0.162)
Tau-squared	0.019	0.014	0.020	0.014	0.024
Number of Effects	30	34	30	30	30
Number of Studies	24	28	24	24	24

Notes: Standard errors are shown in parentheses. Significance tests employ the Knapp-Hartung modification.
Models are estimated with a constant, but constant is not shown.
Correlated effects model, robust standard errors with small-sample adjustment (see Tipton 2015).
Tau-squared is the estimated residual variation between studies.
** = p<.01; * = p<.05, + = p< .1; two-tailed tests

Table S7: Random-Effects Meta-Regression of Log Discrimination Ratios, African-Americans, Publication Predictors

Outcome: Log of Callback Ratio White/African-American

Predictor Variable	Coefficient (SE)
Fieldwork Year	-0.001
	(0.009)
Authors=Advocacy Groups (1=yes, vs. academic authors)	-0.130
	(0.211)
Authors=Government (1=yes, vs. academic)	0.237
	(0.357)
Publication Type is Journal (1=yes)	0.092
	(0.122)
Race is not primary focus (1=yes)	0.051
	(0.112)
Tau-squared	0.025
Number of Studies	21 (all since 1989)

Notes: Standard errors are shown in parentheses. Significance tests employ the Knapp-Hartung modification.
Models are estimated with a constant, but constant is not shown.
Tau-squared is the estimated variation between studies (attributable to study differences).
** = p<.01; * = p<.05, + = p< .1; two-tailed tests

Table S8: Random-Effects Meta-Regressions of Discrimination, Difference in Proportions Outcome

Outcome: Difference in Callback Proportion, White to Minority

Predictor Variable	(1) African-Americans		(2) Latinos	
Fieldwork Year	0.001		-0.005	**
	(0.001)		(0.001)	
Study Method is In-Person Audit	0.095	**		
(1=In-person audit, ref.=resume audit)	(0.027)			
Applicants Male Only (1=yes, ref.= male & female)	0.006			
	(0.014)			
Applicants Female Only (1=yes, ref.= male & female)	0.040			
	(0.023)			
Applicant Education in Grades Completed	0.004			
	(0.006)			
Some Applicants have (Falsified) Criminal Records (1=yes)	0.018			
	(0.023)			
Tau-squared	0.0003		0.0001	
Number of Studies	21		9	

Notes: Outcome is the difference in the proportion of callbacks received by white applicants minus the proportion received by minority applicants.

Standard errors are shown in parentheses. Significance tests employ the Knapp-Hartung modification.

Models are estimated with a constant, but constant is not shown.

Tau-squared is the estimated variation between studies (attributable to study differences).

** = p<.01; * = p<.05, + = p< .1; two-tailed tests

Table S9: Random-Effects Meta-Regressions of Discrimination, Log Odds Ratio Outcome

Outcome: Log of Odds Ratio of Callback, White to Minority

	(1)		(2)	
Predictor Variable	African-Americans		Latinos	
Fieldwork Year	0.005		-0.021	**
	(0.011)		(0.005)	
Study Method is In-Person Audit	0.420	+		
(1=In-person audit, ref.=resume audit)	(0.202)			
Applicants Male Only (1=yes, ref.= male & female)	-0.006			
	(0.121)			
Applicants Female Only (1=yes, ref.= male & female)	0.136			
	(0.174)			
Applicant Education in Grades Completed	0.058			
	(0.051)			
Some Applicants have (Falsified) Criminal Records (1=yes)	0.265			
	(0.182)			
Tau-squared	0.0002		0.0024	
Number of Studies	21		9	

Note: Standard errors are shown in parentheses. Significance tests employ the Knapp-Hartung modification.
Models are estimated with a constant, but constant is not shown.
Tau-squared is the estimated variation between studies (attributable to study differences).
** = p<.01; * = p<.05, + = p< .1; two-tailed tests

Table S10: OLS Regressions of Fieldwork Year on Other Covariates, African Americans, Studies After 1989

Outcome: Year of Fieldwork of Study

Predictor Variable	Coef. (SE)
Study Method is In-Person Audit	-4.589
(1=In-person audit, ref.=resume audit)	(10.521)
Applicants Male Only (1=yes, ref.= male & female)	6.248
	(5.821)
Applicants Female Only (1=yes, ref.= male & female)	0.093
	(21.193)
Applicant Education in Grades Completed	0.805
	(2.675)
Some Applicants have (Falsified) Criminal Records (1=yes)	18.072
	(10.972)
Unemployment Rate of Metropolitan Area(s), Percentage	0.397
	(2.298)
Region = Midwest (1=yes, vs. reference of multi-region)	-1.097
	(15.232)
Region = Northeast (1=yes)	11.393
	(8.431)
Region = South (1=yes)	8.290
	(10.238)
Region = West (1=yes)	6.168
	(8.740)
Includes blue collar occupations (1=yes)	-4.680
	(9.490)
Includes occupations with an office focus (1=yes)	5.570
	(7.028)
Includes restaurant occupations (1=yes)	-14.537
	(13.880)
F-statistic (13, 7 df)	1.930
P-value for F-statistic	0.194
Number of Studies	21

Notes: Standard errors are shown in parentheses.

Models are estimated with a constant, but constant is not shown.

None of the coefficients produce significance tests with p-values < .1