

Trends and Patterns in Racial Discrimination in Hiring in America, 1974-2015

Paper for Presentation at the Meetings of the Population Association of America

Seattle, Washington, March-April 2016

Lincoln Quillian, Northwestern University

Ole Hexel, Northwestern University

Address correspondence to: Lincoln Quillian, Department of Sociology, Northwestern University, 1810 Chicago Avenue, Evanston, IL 60208, e-mail: l-quillian@northwestern.edu. This paper is a part of a larger project involving joint work with Devah Pager, Harvard; Arnfinn H. Midtbøen, Oslo Institute for Social Research; Anthony Heath, Oxford; and Fenella Fleischmann, Utrecht University Netherlands. We have benefitted from methodological conversations and advice of Larry Hedges, Northwestern University. We have received financial support from the Russell Sage Foundation and the Institute for Policy Research at Northwestern University.

Trends and Patterns in Racial Discrimination in Hiring in America, 1974-2015

This study investigates the trend in hiring discrimination in American Labor Markets. We perform a meta-analysis of every available field experiment of hiring discrimination (n=27), containing 32 estimates of discrimination against non-white groups, conducted between 1974 and 2015. Field experiments of hiring discrimination are experimental or quasi-experimental studies in which fictionalized candidates from different race or ethnic groups apply for jobs. Whites receive on average 44% more positive responses than African-Americans and 18% more than Latinos. The ratio of positive responses white to African-American has remained stable across the range of years we observe (from 1974 to 2015 for African-Americans, from 1989 to 2015 for Latinos). Levels of racial and ethnic discrimination are fairly uniform across categories of occupation, education, and gender. 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.

Introduction

Has labor market discrimination declined? This question has been central to discussions of race and racial inequality in America in recent decades. Most white Americans, and many scholars, think of racial discrimination as a product of America's open racism of the pre-civil rights era that has declined significantly as time from that era has increased. Others think that discrimination continues unabated, and claims of progress miss the shift of discrimination to new and subtler forms. The extent of continuing discrimination remains a central issue in public discussions about the importance of active policy efforts to eliminate discrimination as well as compensatory policies to reduce racial and ethnic inequality like affirmative action.

In this study, we examine trends and major patterns of discrimination against race and ethnic groups in American labor markets based on a comprehensive meta-analysis of every available field experiment of discrimination. Field experiments of hiring discrimination are experimental or quasi-experimental studies in which fictionalized candidates from different race or ethnic group apply for jobs. Field experiments come in two major types: correspondence studies, in which fictionalized resumes are submitted, and audit studies, in which ethnically dissimilar but otherwise matched pairs of trained testers apply in person for jobs. Meta-analysis is a secondary statistical analysis of the results of existing studies, which has most commonly been used in medical research to summarize results of clinical trials.

Field experiments capitalize on experimental and quasi-experimental designs to measure discrimination, making them by far the most valid measure to directly assess the prevalence of discrimination on an outcome. Our major finding is that there has been almost no change in levels of hiring discrimination during the period 1974 to 2015. The ratio of positive responses of white to black testers has remained stable with whites receiving on average 44% more callbacks than blacks and 18%

more callbacks than Latinos. The results also suggest that discrimination appears surprisingly uniform across levels of education and major occupational groups. These findings call into question the idea of a “declining significance of race”, of which surveys on racial attitudes in the US imply, and point to the need for continuing and strengthened efforts of anti-discrimination.

BACKGROUND

Discrimination in labor markets has long been recognized as an important social problem. Discrimination directly contributes to persistent gaps in employment between majority and minority racial and ethnic groups and increases racial inequality. Discrimination has been shown to have negative effects on mental and physical health (Pascoe and Smart Richman 2009). In addition, discrimination can have wide-ranging diffuse negative effects. Economic arguments suggest employment discrimination wastes talent and reduces productivity by lowering the quality of matches between applicants and jobs (Becker 1971). Discrimination violates norms of fairness and undermines confidence in legitimate means of mobility such as schooling, perhaps resulting in reduced efforts to succeed in the legitimate labor market in favor of crime or other illicit ways to make a living.

While the negative consequences of discrimination in labor markets are widely recognized, the prevalence and importance of discrimination remains a topic of extensive debate. Some scholars, perhaps most notably W.J.Wilson, have argued that current discrimination has declined from the pre-civil rights era to the point that it is no longer a major impediment to black socio-economic achievement (e.g. Wilson 2012). Supporting this view, surveys indicate that white support for openly racist beliefs

regarding the biological inferiority of non-whites and the need to keep races separate has declined substantially since 1970 (Schuman, Steeh, Bobo, and Krysan 1997; Krysan 2011).

Correspondingly, residential segregation has slowly declined and interracial marriages have become more common. Further suggesting change, in the labor market, laws that made discrimination based on race illegal gained a legal apparatus to enforce these policies through the Equal Employment Opportunity Commission (EEOC) and the right to sue for civil damages resulting from discrimination through private lawsuits. In short, several influential theoretical accounts and many facts suggest that labor market discrimination may have declined across this period. These perspectives suggest that discrimination is no longer a first order cause of racial disparities in employment, with scholarly attention turning to education, skills, and other pre-market factors as the major sources of persistent racial disparities (Neal and Johnson, 1996).

By contrast, other scholars describe discrimination as hardly lessened with the end of legalized discrimination, but only becoming more subtle or taking on new forms (e.g. Feagin and Sikes 1994; Bobo, Kluegel, and Smith 1997; Sidanius and Pratto 2001). These scholars point toward the fact that while whites have increasingly rejected biological inferiority as a source of racial inequality, racial attributions have shifted toward the insinuation that blacks do not work hard enough or are subject to a dysfunctional culture. Correspondingly, studies of stereotypes and implicit prejudice find far less change than explicit prejudice measures would suggest (Devine & Elliot, 1995; XX). Finally, the black unemployment rate remains stubbornly high at about twice the white rate, a relationship that has not changed much since the early 1970s. Scholars emphasizing these facts suggest that discrimination in the

labor market has seen little change, despite some obvious improvements in racial inequality in other areas of life.

These debates persist, often without clearly progress toward a resolution, in significant part because of the lack of clear and reliable measures over time. Perhaps the most common approach to studying racial discrimination has been the residual method: based on a statistical model of an outcome, the residual between majority and minority group members left unexplained after other factors are accounted for is attributed to discrimination. This 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 second method relies on self-reports of incidents of discrimination from targets, often gathered on surveys. 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 third method uses the frequency of formal complaints of discrimination from targets or lawsuits alleging discrimination. 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. A final method are interviews with potential perpetrators, but these face 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 (National Research Council 2004).

In contrast to these basic methods to measure discrimination, a more valid approach is the field experimental method (Pager and Shepherd 2008). Field experiments of hiring discrimination are experimental or quasi-experimental studies in which fictionalized candidates from different race or ethnic group apply for jobs. These include both correspondence studies, in which fictionalized resumes are submitted (e.g., Bertrand and Mullainathan 2004), and audit studies, in which ethnically dissimilar but otherwise matched pairs of trained testers apply for jobs (e.g. Pager, Bonikowski, and Western

2009). We think it useful to include both audit and correspondence studies in our analysis, first simply since both provide valid information about rates of discrimination, and second because they tend to cover somewhat different segments of the labor market, because low-skill jobs often require in-person applications at place of employment, while correspondence studies focus on somewhat higher-skill jobs that are based on paper or online applications as the initial step.

The field experimental method is a design with high causal (internal) validity because it benefits from aspects of experimental design. The experimenter controls 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, correspondence and audit studies, is somewhat different. In correspondence studies, clues indicating race (such as an ethnically typed name) are randomly assigned to otherwise identical resumes, allowing for treatment and control groups to be equated through randomization. In audit studies, matched pairs of trained testers who are ethnically different but 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. Some audit studies track applicants all the way through the hiring process, but more often field experiments record whether their applicants receive an initial indication of further interest, such as a callback for an interview.

Because field experiments are conducted over a short period, they have not been used to examine changes in levels of discrimination over time. But field experiments have a long history, with early U.S. field experiments on labor market discrimination in the 1970s (Riach and Rich 2002). In this study, we capitalize on the history of field experiments to examine change over time by re-analyzing results from the extensive history of field experiments on discrimination. We need to combine the information from all available studies, while not ignoring the differences between studies. There is a body of a well-developed methods to do this in the literature on meta-analysis.

PROCEDURE

Our goal is to use existing field experiments of discrimination in the United States to assess trends in discrimination. We focused on field experiments of employment discrimination, which have the most extensive history. We follow steps that have been developed in the meta-analysis literature (for authoritative discussions, see Higgins and Green 2008; Cooper, Hedges, and Valentine 2009). In performing this analysis we followed the basic stages of a meta-analysis: first to identify all existing field experiments of discrimination relevant for understanding trends in discrimination in labor markets, second developing a rubric and coding the studies to produce a database of their results for later analysis, and third performing the statistical meta-analysis to draw conclusions from the combined study results. We discuss each of these steps in turn.

1. Identifying Relevant Studies

We aimed to include all existing studies, published or unpublished, that use a field experimental method, either in person audit or correspondence testing, and that provide contrasts in hiring-related outcomes between different race and ethnic groups in the United States. We also required the contrasts were made between on-average equivalent racial or ethnic groups in their labor-market relevant characteristics. We attempted to identify all relevant studies through three methods: searches in bibliographic databases, citation searches, and by fielding an e-mail survey of every known author of a field experiment of race-ethnic discrimination in labor markets.

We began with a bibliographic search looking for audit and correspondence studies. 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.¹ 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.²

Our second technique of search was citation search. Working from the initial set of studies located through bibliographic search, we examined the bibliographies of all review articles and eligible audit studies to find further field experiments of hiring discrimination.

The last technique employed was a survey of authors of existing field experiments of discrimination. From our list of found audit and correspondence studies based on the methods above, we compiled a list of e-mail addresses of all known authors of existing field experiments of discrimination. We sent out an e-mail survey to all authors of existing audit studies we identified asking 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 not yet.

The survey was conducted in two phases. In the initial survey 131 apparently valid e-mail addresses were mailed to. We received 56 responses. We also sent out a second wave of 68 e-mails which consisted of corrected e-mail addresses for authors who had changed jobs or e-mail addresses from our initial information. We received 19 responses to this second wave.

Overall our search revealed 27 studies that were field experiments of hiring and included contrasts between white and non-white groups who were on-average equivalent in their labor-market relevant characteristics (e.g. education, experience level in the labor market, etc.). We excluded two

¹ A list of exact search terms used for each database are available from the authors on request.

² These were included as part of an international study this project is part of.

studies where it was not clear if employers were the ones making decisions producing discrepant outcomes.³ Some of these studies included contrasts between more than one target group and whites (e.g. blacks and Hispanics) producing multiple effect sizes.

By searching a dissertation database, web-based database that includes many working and unpublished studies on the web (e.g, Google Scholar, NBER working papers), and by asking experts and study investigators for unpublished results, we hoped to identify many studies that were never published in academic journals. Of 32 usable effect sizes we coded, 11 were not published in academic journals. This is useful in helping to combat the potential problem of publication bias, in which studies with statistically significant evidence of discrimination are more likely to be published (Higgins and Green 2011). We also employed several analytic techniques to assess the possibility of publication bias, discussed in the results section below.

2. Coding

Coding was based on a systematic rubric identifying key study characteristics and coding instructions. We initially read several studies selected from the larger sample, and based on this developed an initial coding rubric of factors we thought might influence rates of discrimination. The initial rubric was reviewed and updated by all authors of this study for completeness. It was subsequently refined through the coding process.

To ensure reliability, each study was coded independently by two raters. The first rating was completed by graduate student Ole Hexel in consultation with Lincoln Quillian. The second rating was done by two undergraduate students who were hired to perform a second coding using the rubric. We

³ Two studies assessed outcomes of auditors who applied through an employment agency (Nunes and Seligman 1999; Thanasombat and Trasviña 2005). It was unclear if the employment agency or employers discriminated in producing the outcome.

then reconciled the results of the two codings, performing further investigating 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.);⁴ 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.⁵ 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 nature of the published outlet of the study (academic journal, government report, etc.). Studies include one or more analyzed effects. For the analysis in this paper, a study can have more than one effect size if it assesses discrimination against multiple target groups, such as African-Americans and Hispanics. We only coded effect sizes for which the white and minority racial groups were given on-average equivalently strong labor market characteristics, omitting a few effect estimates where the groups were non-equivalent in their labor market characteristics.⁶

A list of coded characteristics that we use in our meta-analysis are shown in table 1A (categorical variables) and table 1B (continuous variables). We have overall 32 effect sizes based on 27 distinct studies (5 studies included effect contrasting whites with African-Americans and Hispanics). We

⁴ We did not ask the coders to make subjective evaluations as part of the coding, such as evaluative ratings of study quality.

⁵ 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 employers’ 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).

⁶ For instance, Pager, Bonikowski, and Western (2009) include contrasts between white applicants with criminal records and black and Latino applicants without criminal records. We exclude this effect size because the criminal record for white applicants but not black or Latino applicants means this contrast cannot be used to estimate average discrimination across comparable testers of different races.

have 18 correspondence studies (done by mail, web, e-mail, or phone) and 14 in-person audits. In the audit studies, on average 542 submissions were made for 233 positions. In the correspondence studies, on average 2312 submissions were made for 514 positions. Overall our results combine information from 7,616 in-person job applications made for 3,262 positions and 41,616 resumes submitted for 9,252 positions.

3. The meta-analysis model

Our basic outcome, the “effect” measure in meta-analysis terminology, is the ratio of positive responses to job applications by members of the majority group to the minority group (this is a measure commonly used in epidemiology and medical studies known as the “relative risk”). We calculated this ratio based on counts of number of positive responses to the majority group and minority group from each study. Ratios above 1 indicate the majority received more positive responses than the minority, with the amount above 1 multiplied by 100 indicating the percentage higher callbacks for the majority group relative to the minority group. Greater numbers above indicate higher discrimination against non-whites. Because studies equate the groups on their non-racial characteristics either through matching and assignment of characteristics (typical for audits) or through random assignment (common in correspondence studies), no further controls are required for these contrasts.

Two other candidate measures that could be used instead of the ratio of responses are the difference in proportions of positive responses and the odds ratio. We prefer the ratio of responses to the difference in proportions because it is less sensitive to the base rate of the outcome (see Bornstein, Hedges, Higgins, Rothstein 2009, chapter 5). For the risk difference, high baserate studies dominate low baserate studies. We prefer the ratio of positive responses to the odds ratio because relative risk is more easily interpretable.

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, which is summarized through the standard error of the effect. In cases where information on paired outcomes is available from the study (e.g. counts of pairs in which both the white and the non-white tester receive a callback, white yes non-white no, white no non-white yes, neither get a callback), we calculated standard errors accounting for the pairing using standard error formulas from Zhou (2007). For studies that are not paired between whites or non-whites or where paired outcomes are not reported, we use formulas for the standard error for unpaired groups (see Bornstein, Hedges, Higgins, Rothstein 2009, chapter 5).⁷

The primary goal of the meta-analysis is to estimate rates of discrimination against target groups over time accounting for methodological features of studies. By “target group”, we mean the race/ethnic target group that is the target of discrimination, as contrasted to non-Hispanic whites.

To capture sources of variability not covered by the covariates, we employ a random effects specification for the meta-analysis model (Raudenbush 2009). The random effects specification incorporates in the error structure a variance component capturing variation in outcomes across studies because of factors that vary across studies but are not controlled. Random effects is recommended whenever there is reason to believe that the effect estimated by the various studies in a meta-analysis should vary due to design features of studies, rather than representing a single underlying parameter that is constant over the whole population. This is surely the case in our analysis, since we expect discrimination against a target group may depend on the year of the study, the situation the study considers (e.g. the occupational categories), the (fake) credentials of the testers, etc. We employ a random effects meta-regression model to incorporate covariates in predicting the (log) risk ratio.

⁷ This formula will slightly over-estimate the standard error 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 standard error.

More formally, random-effects meta-analysis allows the true effects of race on 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 observed outcome (the log of the positive response percentage for whites divided by the positive response percentage for the non-white target group) in the i th study, then the meta-analysis model is:

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

Here τ^2 is the between-study variance, estimated from the data, while σ_i^2 is the variance of the log response ratio in the i th study.

Meta-regression allows that the rate of discrimination is a function of a vector of characteristics of the studies and effects, \mathbf{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:

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + u_i + e_i, \text{ where } u_i \sim N(0, \tau^2) \text{ and } e_i \sim N(0, \sigma_i^2)$$

where $\boldsymbol{\beta}$ is a $k \times 1$ vector of coefficients (including a constant), and \mathbf{x}_i is a $1 \times k$ vector of covariate values in study i (including a 1 if a constant is fit). In this study the covariates include many study characteristics shown in table 1.

RESULTS

We begin the results of our meta-analysis by examining the overall levels of discrimination by group and the extent of heterogeneity across studies. The basic study-level outcome is the response ratio of whites to the minority group. Ratios of 1 indicate equal number of responses, suggesting no

discrimination. Ratios above 1 indicate discrimination against minorities while those lower than 1 suggest “reverse” discrimination against whites.

Figure 1 shows a forest plot of the results of 21 studies in which the target group is African-Americans, contrasted to whites. There are two main types of outcome: callback/interview or job offer. Callback or interview reflects whether the applicant received a callback to be interviewed or for further consideration. In figure 1, for each study, information about a callback or interview is shown depending on the term used in the study, but the difference between these terms is not meaningful. The “further consideration” for a callback could in some cases represent a request that the applicant fill out an application form or provide other more detailed information. All correspondence studies and many audit studies use callbacks or interviews as their main outcome: 26 of the 32 effect sizes in our analysis have callbacks or interviews as the outcome.

Our meta-analysis includes six studies that pursued job applications all the way to the actual job offer. For these studies, we use the ratio of the percentage of white applications that received a job offer to the percentage of minority applications that received a job offer as the outcome. Many of the studies that went all the way to the job offer also have data for an interview or callback outcome, indicating if the applicant got to the stage of an interview. We use job offer as the basic outcome (and not interview) as the outcome for these studies since this is the outcome of greatest importance. Contrasting the job offer and interview outcomes suggests somewhat higher rates of discrimination (higher white to minority response ratios) for the job offer outcome than for the interview outcome. However, the difference between the callback/interview and job offer response ratios is not statistically significant (at $p < .1$; appendix table added later).

The forest plot shows confidence intervals for measures of discrimination for each study. Finally an overall effect based on the random-effects meta-analysis model is shown at the bottom. The results

suggest that on average whites receive 44% more positive responses to job applications than African-Americans. A 95% confidence interval for the effect is 27% to 64% more callbacks. This confirms the finding of many audit and correspondence studies of hiring discrimination against African-Americans, and provides a broader overall estimate of the average prevalence of discrimination in hiring by combining information from 21 studies.

The I-squared statistic shown on the forest plot is the estimated share of the variation between effects sizes due to variation among studies. The model estimates that 74% of variability reflects differences resulting from differences in the situation (e.g. year and occupation) and procedures of studies. The remaining 25.7% of variability between studies could be accounted for by random variations in study outcomes (resulting from random variation in audit outcomes for example). A significance test strongly rejects the hypothesis that the between-study variability is zero ($p < .001$).

Table 2 provides a forest plot for white against Hispanics or Latinos. It also shows results for one audit (the only one we know of that met our inclusion criterion) that tested Arab-Americans against whites,⁸ and results for two studies that used aggregated non-white groups.⁹ We include the two combined groups because they provide evidence of how whites contrast to African-Americans and Latinos overall, even if we cannot disaggregate among these two groups from their results.

The results for Latinos suggest discrimination in hiring relative to whites, but less than that targeted at African-Americans. On average whites received 18% more positive responses than Latinos,

⁸ Ghumman and Ryan (2013) conducted an audit focused on wearing a headscarf. They had both white and Arab auditors, but did not report results separately for these two groups, and when contacted, the authors declined to provide the data we needed to calculate effects.

⁹ For the Deming et al. (2014) study, the non-white group is composed of African-Americans and Latinos. For Bendick et al. (2010), an in-person audit study, the testers are a mix of non-white persons, including some who are mixed-race and Asian. In a personal communication, Bendick described the non-white group as a group of clearly non-white black and brown people for whom it was sometime difficult to identify single groups as their racial or ethnic background.

with a 95% CI of 0% to 39% (1.00 to 1.39) more. Statistical test of the hypothesis that the ratio is 1 gives a p-value of .052, just outside of significant at the 5% range (two-tailed).

The results for Arab-Americans are obviously limited by the availability of only a single correspondence study, but it is interesting to note that this one study finds a high level of discrimination (white to Arab-American positive response ratio of 2.8). Still this outcome is consistent with the possibility of a wide range of levels of hiring discrimination (from 7% more responses to 736% more).

Trends in Discrimination Over Time

The studies available performed their fieldwork from 1974 to 2014 for African-Americans and from 1989 to 2014 for Latinos. Do we find evidence of change over time in rates of hiring discrimination?

The short answer is no. Figure 3 shows a graph of study positive response ratios against year, with lines for the linear trends for African-Americans (solid) and Hispanics (dashed). The size of the symbol is proportional to the weight it is given in the meta-analysis. Larger symbols indicate the study provides more information about discrimination in hiring.¹⁰

Both lines are close to flat, with lines sloping slightly upward for African-Americans, and slightly downward for Hispanics. Table 2 shows the exponentiated coefficients of the slopes of the lines; these can be interpreted as the multiplicative change in the positive response ratio white/minority per year. The coefficient suggests a multiplicative increase in the ratio of .004 by year for African-Americans and a

¹⁰ Using notation from the method section, the weight is $1/(\tau^2 + \sigma_i^2)$. These are the random-effect meta-analysis weights. That is, studies are weighted inversely to the estimated variance based on the sum of the between-study variation (a constant number for all studies in an analysis) and the variation based on the sample size and proportion of successes from the individual study.

multiplicative decline of .008 (coefficient of .992) for Latinos. Neither are significant at any conventional level, with t-ratios below one.

A second model for whites and Latinos uses dummy variables to represent decades instead of a linear year trend. This different approach to trends again provides no significant evidence of change over time for either group.

A third model for African-Americans only takes studies with fieldwork from 1989 or later. This limits the range of studies to the same as those available for Latinos (since we know of no field experiments of hiring discrimination against Latinos before 1989). We also perform this analysis because our reading of the early studies (pre-1985) shows they have some differences in methods from later studies. Later correspondence studies signal race by using race-typed names on resumes, but the pre-1985 studies either indicate race directly on the resume (McIntyre 1980 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 1978). These early studies also all used resumes with college education or better. Using only the post-1989 studies, the slope of change for African-Americans switches from very slight positive to very slight negative, but never becomes close to statistically significant.

Notably, then, these results suggest very little change in rates of hiring discrimination in American labor markets over the period studied. We find evidence of stability, not change, in hiring discrimination. To the extent we find any evidence, however weak, of change over time, we find it for Latinos contrasted to whites, where there is a slight downward trend, but not a statistically significant one.

A Fuller Model of factors influencing Discrimination

We turn now to introducing study characteristics that may be related to disparities in hiring rates between whites and non-whites. In table 3, we pool together effect sizes for African-Americans and Hispanics (also adding the two effect sizes from studies that combined these groups). Like in table 2, we use a random-effects meta-regression with the logged response ratio white to minority as the outcome. Coefficients are exponentiated to give multiplicative changes in the ratio with a one unit change in the independent variable.

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 for Latinos, although this difference is not significant at conventional levels in the first model, and only marginally significant using the post-1989 studies (significant at $p < .1$ two-tailed). Figures 4 and 5 show the linear trends together with confidence interval regions based on the regression, whites and African-Americans in figure 4, and whites and Latinos in figure 5.

We then add controls for whether the study used male or female testers or resumes (contrasted to using both), whether the testers had a college education or more, and whether the method of the study was an in-person audit (vs. correspondence). None of these variables are statistically significant in predicting the rate of discrimination. The non-significance of these predictors may be surprising, but it is somewhat consistent with Bertrand and Mullinathan's (2004) finding from their correspondence study of relatively uniform discrimination across job characteristics.

None of these controls changes the conclusion that the disparity in rates of positive response between whites and minorities shows almost no trend over time. The trend in the data gets slightly flatter after controls are added.

Publication Bias

A possible threat to validity of our meta-analysis is the possibility of publication bias: that studies with statistically insignificant results might not get published, with the result that the literature is “biased” toward finding significant discrimination. If publication bias increases or decreases over time, this could also influence estimates of the trend.

We used several procedures to check for publication bias. First, we included several study-level predictors as determinants of discrimination, including one that is diagnostic of publication bias. These are shown in table 4. Of particular interest is the characteristic of “race is not primary focus”, which is a dummy variable indicating whether the publication was primarily focused on race effects or focused on some other effect (even though it had racially diverse testers, producing a tester effect). For instance, some studies focused on effects of types of educational qualification on receiving a positive response from an employer or criminal background effects on job offers.

If publication bias is a serious problem, then we should find that race-focused studies tend to find more discrimination than not-race-focused studies, since a positive finding on race is probably more important for publication of the race-focused studies. Our results suggest this is not the case, and in fact on average not-race-focused studies find somewhat more discrimination in our sample.

As a second procedure, we employed several tests suggested in the meta-analysis literature for publication bias. These tests generally rely on the principle that if there is publication bias, and the underlying (“true”) distribution of effects about the center is symmetric, then we should observe an asymmetric distribution of observed effects among published studies, with disproportionately many “missing” studies in the non-significant portion of the symmetric distribution. We used two well-known publication bias tests: the Eggers test and the Duvall and Tweedie’s trim-and-fill procedure (see Sutton 2009 for descriptions). The Eggers test did not find statistically significant evidence of publication bias

($p > .9$). The trim-and-fill procedure suggested two effects were missing due to publication bias, with the results that the positive response ratio for African-Americans would be decreased from 1.44 to 1.40, a very small change of little significance.

Overall these results suggest that publication bias is not a serious problem for our results.

SUMMARY

Labor market discrimination has long been a major focus of social science attention, but due to limitations in methods for measuring discrimination we have lacked direct information on trends. Studies of trends in discrimination over time have relied on indirect methods to assess its prevalence. We capitalize on the extensive history of audit and correspondence studies to assess the frequency of discrimination grounded in the superior validity of field experiments.

We find striking evidence of continuity rather than change in levels of hiring discrimination against blacks and Latinos. We find almost no change for African-Americans in positive response rates to applications relative to whites from 1974 to 2014. The evidence is especially strong after 1989, because audit studies become more frequent after this time, and these studies employ some improvements in methodology relative to earlier studies. For Latinos there are no studies before 1989, but the data suggests only a small (and not statistically significant) drop in discrimination in the twenty-five years from 1990 to 2015.

We find that discrimination in employment is significantly higher against African-Americans than Latinos, but that both groups are disadvantaged in hiring relative to whites. On average whites receive about 44% more positive responses than whites, compared to 18% more than Latinos.

We find little evidence of major variation in discrimination levels depending on educational level of the testers, gender of testers, or type of occupation tested. Our evidence suggests relatively uniform levels of hiring discrimination,

We note as well that our results do not address the possibility that hiring discrimination might have substantially dropped in the 1960s or early 1970s, during the civil rights era when many forms of direct discrimination were outlawed. We know of no audit or correspondence studies from these eras. Some evidence suggests discrimination declined notably during the 1960s as a direct result of federal policy actions (Heckman and Payner 1989). Further, we note that our results only pertain to hiring discrimination, not discrimination after the point of hire, or discrimination in wage setting. Some evidence suggests that antidiscrimination enforcement has increased the costs to employers for wage discrimination or wrongful termination. Because discrimination is extremely difficult to document at the point of hire, by contrast, enforcement has had less impact at this stage (Donahue & Heckman, 1991; Wessell, 2003). 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 – blacks and Latinos remain substantially disadvantaged relative to equally qualified whites, and we see little indication of progress over time.

DISCUSSION

We believe that these results suggest the need for rethinking of the factors that drive hiring discrimination and the level of discrimination in American society. Despite the fact that on some measures of attitudes there have been major changes in white views on race and ethnicity, and despite evidence of declining discrimination in some respects (e.g. declining residential segregation, more interracial unions), we find no evidence of change in hiring discrimination against African-Americans or

Latinos using the most direct measure of discrimination available. These results suggest that hiring discrimination may be driven by subtler processes than overt racist attitudes, for which surveys have shown substantial declines over time (see Quillian 2006). The idea that gradually fading beliefs from before the civil rights movements, and the replacement of cohorts raised before the civil rights movement, should drive a gradual reduction in discrimination does not appear borne out in the case of hiring discrimination.

Our results regarding the persistence of discrimination are consistent with the fact of very stable differentials between whites and blacks in unemployment from 1970s to the present. The black unemployment rate has long been about twice the white rate, and this ratio has seen little change over the period from 1974 to the present covered by this study.

We think the results also point toward the need for strong enforcement of anti-discrimination legislation and provide a rationale for continuing compensatory policies like affirmative action to help improve equality of opportunity. Discrimination continues, and we find little evidence that it is gradually diminishing to nothing. Instead, we find the persistence of discrimination at a distressingly uniform and unchanged rate.

Recent efforts to reduce racial inequality have focused on improving schools in poor, non-white neighborhoods, and reforms of the criminal justice system. While clearly needed, our results suggest the continuing importance of a more basic factor: direct discrimination in hiring decisions by employers. Aggressive enforcement of anti-discrimination laws focused on hiring, including possibly tester methods to detect discrimination, are needed to help address persistent racial inequality in the labor market.

References

- Arrow, K.J., 1973. The theory of discrimination., in: Ashenfelter, O., Rees, A. (Eds.), *Discrimination in Labor Markets*. Princeton University Press, Princeton, N.J.
- Becker, Gary S. 1971. *The Economics of Discrimination*. 2nd edition. Chicago: University Of Chicago Press.
- Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *The American Economic Review* 94 (4): 991–1013.
- Bobo Lawrence, Kluegel JR, Smith R. 1997. Laissez-faire racism: the crystalization of a kinder, gentler, antiblack ideology. In *Racial Attitudes in the 1990s: Continuity and Change*, ed. SA Tuch and JK Martin, 15-42. Westport, Conn: Praeger.
- Cooper, Harris, Larry V. Hedges, and Jeff C. Valentine. 2009. *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation.
- Devine, P. G., and A. J. Elliot. 1995. "Are racial stereotypes really fading? The Princeton trilogy revisited." *Personality and Social Psychology Bulletin* 21:1139-1150.
- Donahue, John J. III and James Heckman. 1991. "Continuous Versus Episodic Change: The Effect of Federal Civil Rights Policy on the Economic Status of Blacks," *Journal of Economic Literature* 29: 1603-43.
- Feagin, Joe R., and Melvin P. Sikes. 1994. *Living with Racism: The Black Middle-Class Experience*. Beacon Press.
- Ghumman, S., and A. M. Ryan. 2013. "Not Welcome Here: Discrimination towards Women Who Wear the Muslim Headscarf." *Human Relations* 66(5):671–98.
- Heckman, James and Brook Payner. 1989. "Determining the Impact of Federal Antidiscrimination Policy on the Economic Status of Blacks: A Study of South Carolina." *American Economic Review* 79(1): 138-177.
- Higgins, Julian P.T. and Sally Green. 2008. *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester: Wiley.
- Krysan, Maria. 2011. "Data Update to Racial Attitudes in America." An update and website to complement *Racial Attitudes in America: Trends and Interpretations*, Revised Edition, Howard Schuman, Charlotte Steeh, Lawrence Bobo, and Maria Krysan, 1997, Harvard University Press.
<http://www.igpa.uillinois.edu/programs/racial-attitudes/>.
- National Research Council. 2004. *Measuring Racial Discrimination*. Washington, DC: National Academies Press.
- Neal, Derek and William Johnson. 1996. "The Role of Premarket Factors in Black-White Wage Differences," *Journal of Political Economy* 104(5): 869-95

- Orwin, R.G., Vevea, J.L., 2009. Evaluating Coding Decisions, in: Cooper, H.M., Hedges, L.V., Valentine, J.C. (Eds.), *The Handbook of Research Synthesis and Meta-Analysis*. Russell Sage Foundation, New York.
- Pager, Devah, and Hana Shepherd. 2008. "The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets." *Annual Review of Sociology* 34(1):181–209.
- Pager, D., B. Bonikowski, and B. Western. 2009. "Discrimination in a Low-Wage Labor Market: A Field Experiment." *American Sociological Review* 74 (5): 777–99.
- Pascoe, Elizabeth A., and Laura Smart Richman. 2009. "Perceived Discrimination and Health: A Meta-Analytic Review." *Psychological Bulletin* 135 (4): 531–54.
- Quillian, Lincoln. 2006. "New Approaches to Understanding Racial Prejudice and Discrimination." *Annual Review of Sociology* 32: 299-328.
- Riach, P. A., and J. Rich. 2002. "Field Experiments of Discrimination in the Market Place." *The Economic Journal* 112(483):F480–F518.
- Raudenbush, S., 2009. Analyzing Effects Sizes: Random Effects Coding, in: Cooper, H.M., Hedges, L.V., Valentine, J.C. (Eds.), *The Handbook of Research Synthesis and Meta-Analysis*. Russell Sage Foundation, New York.
- Schuman, Howard, Charlotte Steeh, Lawrence Bobo, and Maria Krysan. 1998. *Racial Attitudes in America: Trends and Interpretations*. Cambridge, Mass.; London: Harvard University Press.
- Sidanius, Jim, and Felicia Pratto. 2001. *Social Dominance: An Intergroup Theory of Social Hierarchy and Oppression*. Cambridge University Press.
- Sutton, A.J., 2009. Publication bias, in: Cooper, H.M., Hedges, L.V., Valentine, J.C. (Eds.), *The Handbook of Research Synthesis and Meta-Analysis*. Russell Sage Foundation, New York.
- David Wessell, David. 2003. "Fear of Bias Suits May be Affecting Hiring Decisions." *The Wall Street Journal*, 9/11/03.
- Wilson, D.B., 2009. Systematic Coding, in: Cooper, H.M., Hedges, L.V., Valentine, J.C. (Eds.), *The Handbook of Research Synthesis and Meta-Analysis*. Russell Sage Foundation, New York.
- Wilson, William Julius. 2012. *The Declining Significance of Race: Blacks and Changing American Institutions*. Third Edition. Chicago: University Of Chicago Press.
- Zhou, Guang Yong. 2007. "One Relative Risk Versus Two Odds Ratios: Implications for Meta-Analyses Involving Paired and Unpaired Binary Data." *Clinical Trials* 4: 25-31.

Table 1A: Characteristics of Audit and Correspondence Studies of Discrimination Against Non-White Groups

<u>Study Method</u>	Count	Percentage of Total (N=32)
In-person audit	14	45%
Correspondence study (by mail or online)	18	58%
<u>Target Group</u>		
Black/African-American	21	68%
Hispanic/Latino	8	26%
"Tester of Color" (aggregated)	2	6%
Arab-American	1	3%
<u>Outcome (Closest to Hire Available)</u>		
Callback or Interview	26	84%
Job Offer	6	19%
<u>Race is a Primary Focus of the Publication</u>		
No	13	42%
Yes	19	61%
<u>Publication Type</u>		
Dissertation	1	3%
Journal article	21	68%
Report	8	26%
Working paper	2	6%
<u>Paired design with mixed-race pairs?</u>		
Yes	10	32%
No	22	71%
<u>Sample Frame (more than one per study possible)</u>		
Newspaper Ads	19	53%
Online job bank or online ads	8	22%
Other	9	25%
<u>Gender of Testers</u>		
Female	4	13%
Male	14	45%
Both	12	39%
Not Stated	2	6%
<u>Education Level of Testers/Resumes</u>		
High School Degree	11	35%
Some College	1	3%
College Degree	6	19%
Graduate Degree	2	6%
Mixed (multiple levels)	10	32%
Not clearly stated	2	6%

Table 1B: Means and Standard Deviations of Characteristics of U.S. Audit and Correspondence Studies

<u>Jobs Applied For</u>	<u>Mean</u>	<u>Std. Dev.</u>
Audit Studies (N=14)	233.3	128.5
Correspondence Studies (N=18)	514.4	337.8
<u>Applications Submitted</u>		
Audit Studies (N=14)	542.9	295.8
Correspondence Studies (N=18)	2312.4	2683.4
<u>Positive Response Rates, Whites</u>	<u>Mean</u>	<u>Std. Dev.</u>
White, Outcome=Callback	18.2%	11.6%
Non-white, Outcome=Callback	13.1%	7.9%
White, Outcome=Job Offer	35.2%	24.6%
Non-White, Outcome=Job Offer	26.8%	22.5%
Positive Response Ratio (white to minority)	ADD	ADD

Note: means shown are unweighted means at the effect level, n=32

Table 2: Random-Effects Meta-Regression of Response Ratios (White to Minority),
 Separate Models for African-Americans and Hispanics

	<u>White/African-American</u>			<u>White/Hispanic</u>	
	<u>All Years</u>	<u>All Years</u>	<u>After 1989</u>	<u>All Years</u>	<u>All Years</u>
Fieldwork Year (four digit year)	1.004 (0.480)		0.995 (-0.520)	0.992 (-0.750)	
<u>Decade Dummy Variables</u>					
Year = Before 1989 (vs. years 2000-2009)		0.698 (-1.210)		1.133 (0.390)	
Year = 1990 to 1999 (vs. years 2000-2009)		1.071 (0.260)		0.778 (-0.910)	
Year = 2010 to 2015 (vs. years 2000-2009)		0.911 (-0.440)		0.683 (-1.380)	
I-squared	75.4%	74.9%	72.5%	58.6%	60.3%
N	21	21	17	8	8

Notes: Figure shown is the exponentiated coefficients (exp[coef.]).

Outcome is the log response ratio white to minority.

T-ratios are shown in parentheses. Models are estimated with a constant, but constant it not shown.

I-squared is the estimated share of variation between studies (attributable to study differences).

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

Table 3: Random-Effects Meta-Regression of Response Ratios (White to Minority), African-Americans and Hispanics

	White/Minority (African-American or Hispanic)			
	<u>All Years</u>	<u>After 1989</u>	<u>All Years</u>	<u>After 1989</u>
Fieldwork Year (four digit year)	1.001 (0.130)	0.993 (-0.960)	1.000 (0.060)	0.995 (-0.480)
White/Hispanic (1=yes, vs. white/black)	0.793 (-1.550)	0.767 + (-2.010)	0.724 + (-1.860)	0.722 * (-2.110)
Combined non-white (1=yes, vs. white/black)	0.739 (-1.190)	0.757 (-1.310)	0.651 (-1.460)	0.733 (-1.180)
Testers Male Only (1=yes, ref. = male & female)			0.952 (-0.230)	1.128 (0.570)
Tester Female Only (1=yes, ref.= male & female)			0.802 (-0.900)	0.859 (-0.710)
Tester Education College Plus (1=Yes, ref.= other)			0.808 (-0.960)	0.937 (-0.300)
Study Method (1=Audit, ref.=correspondence)			1.046 (0.230)	0.936 (-0.360)
I-squared	73.6%	68.6%	73.2%	69.2%
N	31	27	31	27

Notes: Figure shown is the exponentiated coefficients (exp[coef.]).

Outcome is the log response ratio white to minority.

T-ratios are shown in parentheses. Models are estimated with a constant, but constant it not shown.

I-squared is the estimated share of variation between studies (attributable to study differences).

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

Table 4: Random-Effects Meta-Regression of Response Ratios White to Minority, Study Type Predictors

Outcome: Response Ratios White to Minority (African-American or Hispanic)

	<u>All Years</u>
Fieldwork Year (four digit year)	1.002 (0.280)
Hispanic (1=yes)	0.763 (-1.630)
Combined non-white (1=yes)	0.693 (-1.400)
Authors=Advocacy Groups (1=yes, vs. academic authors)	1.323 (1.420)
Authors=Government (1=yes, vs. academic)	1.167 (0.560)
Race is not primary focus (1=yes)	1.221 (1.110)
Publication type is journal (1=yes)	1.312 (1.660)
I-squared	71.49%
N	31

Notes: Figure shown is the exponentiated coefficients (exp[coef.]).

Outcome is the log response ratio white to minority.

T-ratios are shown in parentheses. Models are estimated with a constant, but constant it not shown.

I-squared is the estimated share of variation between studies (attributable to study differences).

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

Fig 1: Response Ratios Whites to African-Americans

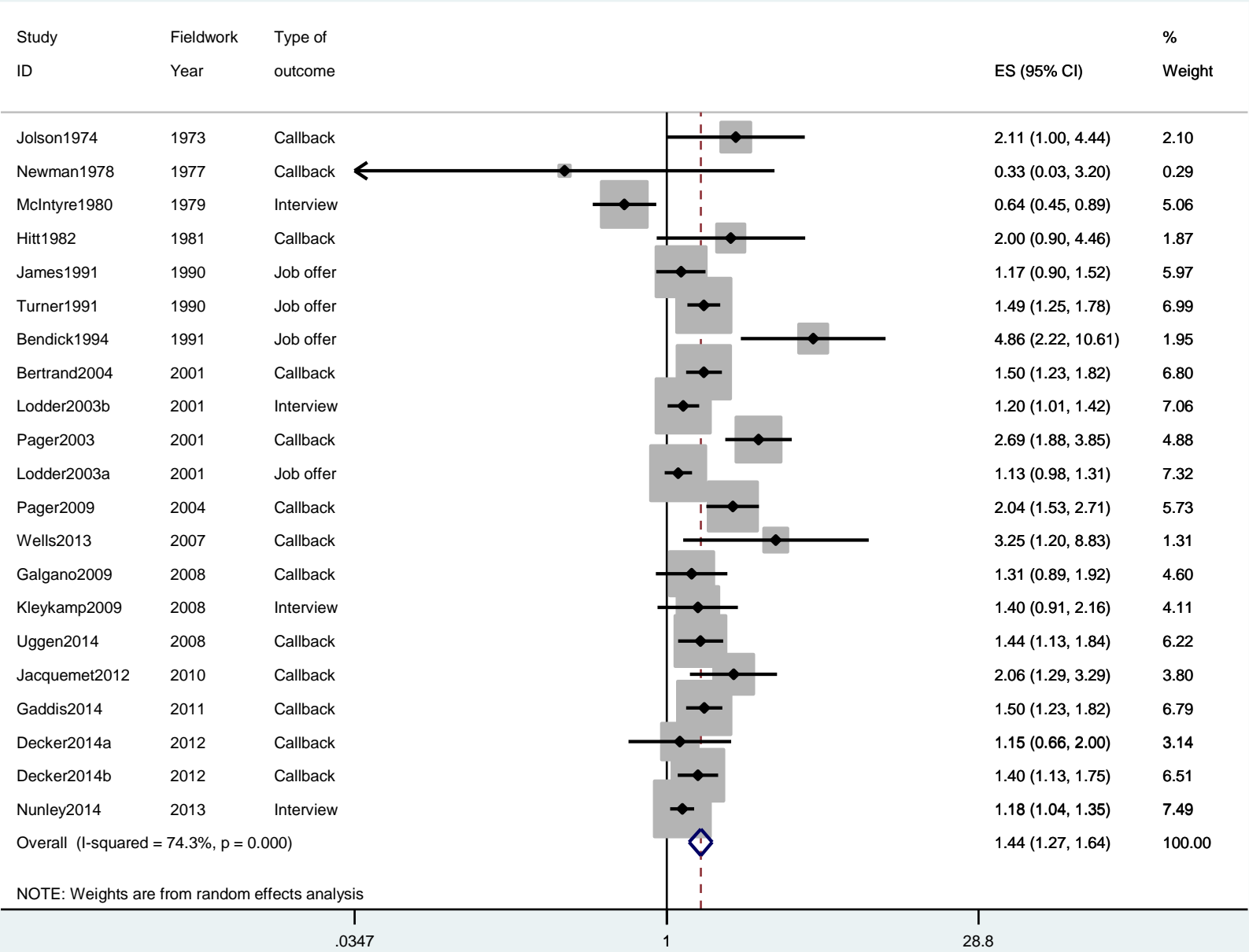


Fig 2: RRs Whites to Hispanics, Arab-Americans, Other

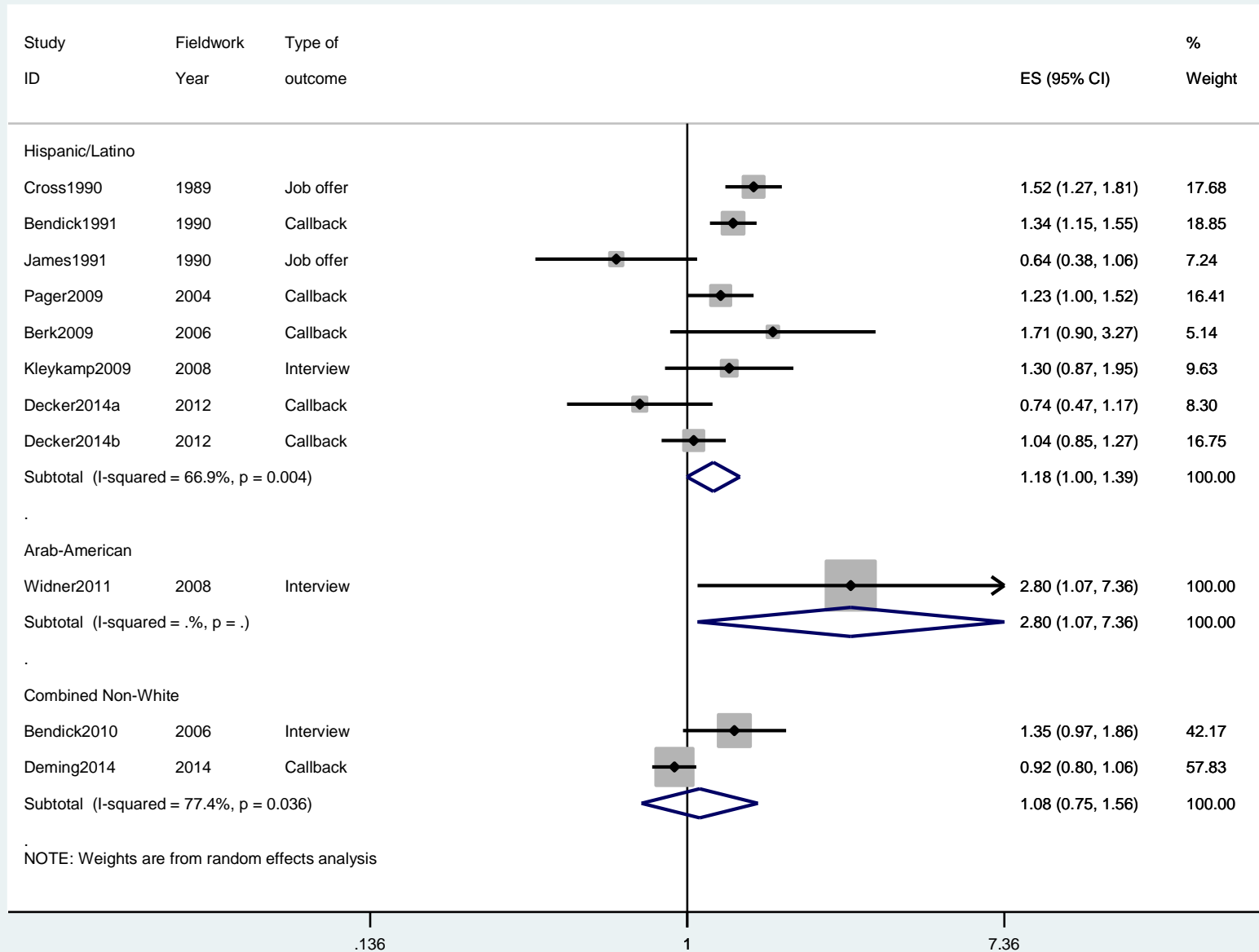


Fig. 3: Trends in RRs, White/AA and White/Hispanic

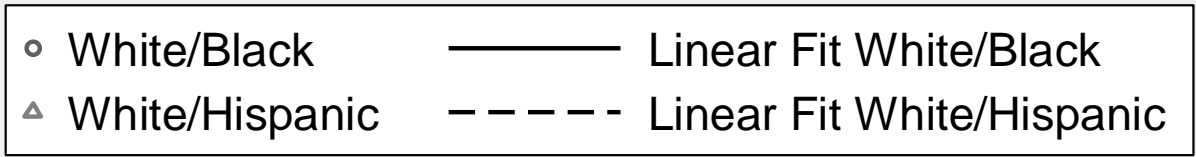
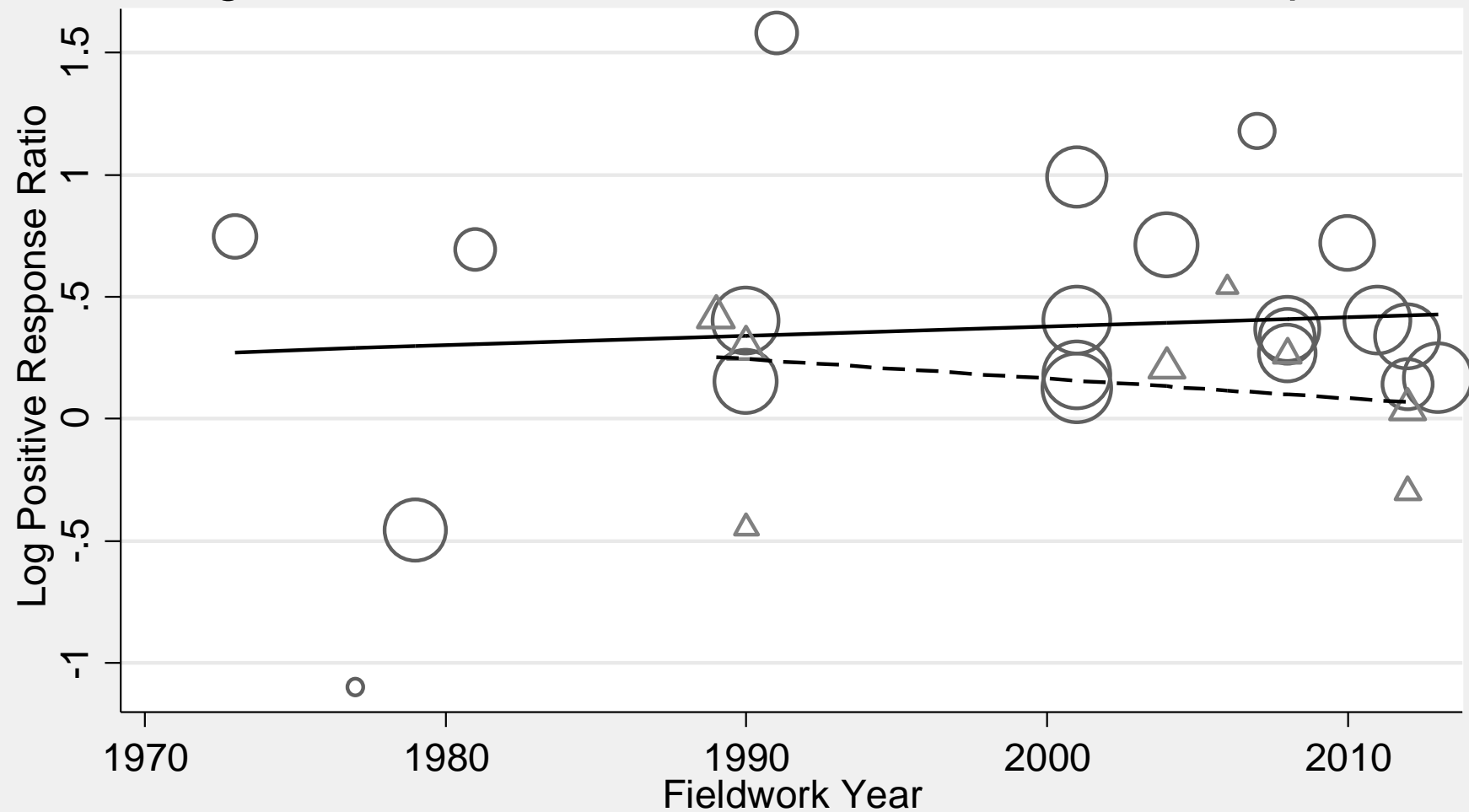
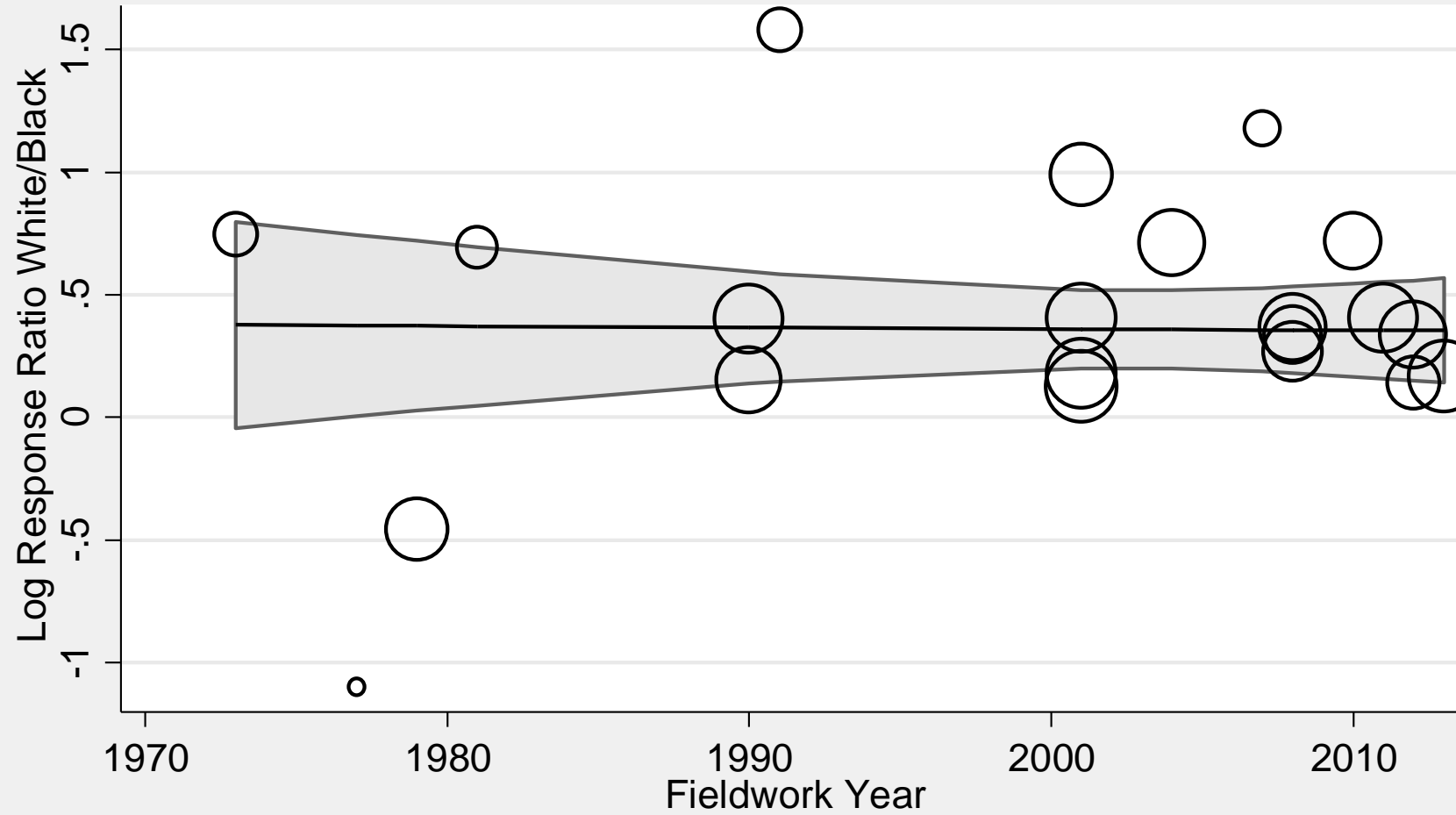


Fig 4: RR White/AA Trend with CI



95% CI Upper/95% CI Lower Linear prediction
◦ Observed Effect Sizes

Fig. 5: RR White/Hispanic Trend with CI

