Northwestern | INSTITUTE FOR POLICY RESEARCH

IPR Working Paper Series

WP-22-25

Trends in Hiring Discrimination Against Racial and Ethnic Minorities in Six Western Countries

Lincoln Quillian
Northwestern University and IPR

John J. Lee
Northwestern University

Version: August 8, 2022

DRAFT

Please do not quote or distribute without permission.

Abstract

The researchers examine trends in hiring discrimination against racial and ethnic minority groups in six European and North American countries: Canada, France, Germany, Great Britain, the Netherlands, and the United States. Their sample is composed of all available discrimination estimates from 90 field experimental studies of hiring discrimination, encompassing more than 170,000 applications for jobs. The years covered vary by country, ranging from 1970–2017 for Great Britain to 1994–2017 for Germany. Across countries, minority groups, and controlling for study attributes, the results indicate that levels of discrimination in callbacks for interviews have remained either unchanged or slightly increased. There are three notable exceptions. First, hiring discrimination against minorities with ethnic origins in the Middle East and North Africa (MENA) increased during the 2000s relative to the 1990s. Second, the researchers find that discrimination in France declined, although from very high to “merely” high levels. Third, they find evidence that discrimination in the Netherlands has increased over time. Contrary to the idea that discrimination will tend to decline in Western countries, they find that discrimination has not fallen in five of the six Western countries they examine over the last few decades.

Significance Statement

Trends in discrimination are critical to evaluate the extent to which Western societies are (or are not) achieving the fundamental goal of ensuring fair and equal treatment regardless of race and ethnicity. The researchers examine trends in hiring discrimination based on 90 field experiments of hiring over time in six Western countries, providing the first national estimates of discrimination trends in four of these countries. They examine trends by country and minority group and incorporate more extensive controls for study characteristics that could confound time trends than previous studies. Only in France do they find evidence that discrimination declined. Further efforts are needed in Western societies to reduce persistent biases against racial and ethnic minorities in the labor market.

The authors received helpful comments from audiences at the American Sociological Association Annual Meeting 2019 in New York, the Institute for Policy Research and the Department of Statistics at Northwestern University, the Centre for the Experimental-Philosophical Study of Discrimination at Aarhus University, Denmark, and the IMISCOE 2022 Conference in Oslo, Norway.

Introduction

From 1940 to 1960, much changed regarding race and racism in Western Europe and North America. The defeat of Nazi Germany delegitimized many forms of biological racist thinking (Fredrickson 2002). The American Civil Rights Movement and decolonization movements highlighted the contradiction between the claims of the West to represent freedom while minorities at home and abroad in colonial empires continued to face repression. As these movements succeeded, open racial discrimination and support for white preference were increasingly viewed as illegitimate (Blinder, Ford, and Ivarsflaten 2013). Codifying these changes was a new body of international law, beginning with the Universal Declaration of Human Rights, which declared equal treatment and freedom from discrimination as universal human rights. In the 1950s and 1960s, international bodies like the International Labor Organization and UNESCO passed resolutions calling for the elimination of discrimination based on race and ethnicity (Simon 2021). Racial discrimination in employment and housing was made illegal in the U.S. in the 1960s and in Western European countries in the late 1960s and 1970s.

Despite the establishment of a normative and legal framework against discrimination in the West, the goal of fair and equal treatment—without regard to race or ethnicity—has not been achieved. There remain substantial gaps in important economic and social outcomes between white and non-white populations in Western countries, with some of the largest and most critical gaps in the labor market (Heath and Cheung 2007). Discrimination plays a role in creating and exacerbating these disparities: persistent discrimination in labor markets in North America and Europe has clearly been demonstrated by field experiments in which investigators use testers or submit applications by mail or over the internet for jobs with clues indicating the race or ethnicity of applicants (Zschirnt and Ruedin 2016, Quillian et al. 2019; Thijssen et al. 2021).

Given the persistence of discrimination in Western countries, a critical question is whether these societies have made significant progress toward reducing it. Popular beliefs and academic discussions often assume a trajectory of declining racial discrimination and inequality over time (Seamster and Ray 2018; Kraus, Rucker, and Richeson 2017). Trends in discrimination are essential to evaluate these

assumptions and the success of policy efforts to curtail discrimination such as strengthened anti-discrimination legislation.

Scholars have suggested at least four factors that may have contributed to reduced discrimination since the mid-1970s. First is the "internationalization" of anti-discrimination and minority group rights. During World War II and the Cold War, Western countries attempted to claim superiority based on their democratic and inclusive systems, which were contradicted by the persistence of open discrimination against minority groups (Fredrickson 2002, Klinker and Smith 2002). International organizations and minority social movements subsequently adopted and spread a discourse declaring minority group rights as fundamental human rights (Skrentny 2004, Kymlicka 2009). Countries increasingly faced international scrutiny for their treatment of minority groups, with beneficial effects for minority groups in many contexts.

A second factor is changes in attitudes. Survey data shows sharp declines after 1970 in biological racist beliefs and support for the “right” of whites to discriminate, corresponding to a rising norm against discrimination. This change is especially well-documented in the United States (Schuman, Steeh, Bobo, and Krysan 1997), but evidence from national datasets in Europe indicates changes in attitudes similar to those in the U.S. (Pettigrew and Meetens 1995) with some variation across countries (Rubin et al. 2014).

A third factor is the strengthening of anti-discrimination laws, especially in Europe. In the U.S., tort-based discrimination law has been strengthened to increase chances for plaintiffs, most notably in 1991 (Corbett 2014). In Europe, the legal framework of anti-discrimination has been systematized and extended through reforms mandated by the European Union. In 2000, the European Union adopted a racial equality directive that required E.U. states to ban racial discrimination and to adopt certain standards in the enforcement of anti-discrimination laws. European states have gradually incorporated these anti-discrimination provisions into national law (Jacob-Owens 2019).

Finally, corporations and other large bureaucracies have increasingly adopted policies to increase their workforces' racial and ethnic diversity. These efforts include hiring chief diversity officers, diversity training, mentoring programs for minorities and women seeking promotion, recruitment from diverse hiring venues, and monitoring their workforces' racial/ethnic composition (Dobbin 2011). Large American companies and bureaucracies have most readily adopted these policies, but they are also

present in European and Canadian companies (for a discussion of corporate diversity policies in European government and corporate bureaucracies, see ENAR 2009, p. 24). While some of these measures are weak lip-service, evidence suggests that some measures adopted by diversity bureaucracies have positively affected the hiring and promotion of non-whites and women (Kalev, Dobbin, and Kelly 2006; Dobbin, Schrage, and Kalev 2015).

However, other social and political developments suggest a less optimistic picture of discrimination trends. In response to racial attitude changes, skeptics have pointed out that support for active policies to reduce racial inequality and beliefs in negative stereotypes has shown relatively minor change over time. A number of "new racism" accounts have argued that the apparent decline of racism from attitude surveys misses the fact that racism largely took on a more subtle and covert form (e.g., McConahay 1986; Bobo, Kluegel, and Smith 1997; Dovidio, Gaertner, and Pearson 2016; Sears and Henry 2004; Bonilla-Silva 2017; Pettigrew and Meertens 1995 pp. 57-58 and Dovidio, Gaertner, and Pearson 1996, p. 271 discuss European and Canadian studies in this tradition).

A growing backlash against population diversity and the increased status of minorities has been evident in Western countries. The 1980s saw the rise of a new far-right in Europe that politicized debates about immigration, emphasized the difficulty of incorporating new migrants and cast non-white immigrants as a threat to national identities. Far-right parties have won increasingly large shares of the vote in European elections since 1980 (Golder 2016). Correspondingly, European surveys show increased hostility to immigrants from 1985 to 1995 (Semyonov, Raijman, and Gorodzeisky 2006). In the U.S., the election of Donald Trump in 2016 can be seen as a capturing of the Republican Party by far-right elements (Bonikowski, Yuval, and Bock 2021). Recent studies suggest political events like these can affect hiring discrimination. Gorzig and Rho (2022) show that Donald Trump's election increased employment discrimination against Somali immigrants in the Minneapolis area, especially in occupations involving customer contact (see also Goldstein and Hamilton 2022 and Di Stasio and Heath 2021).

A final reason to believe discrimination may have increased, especially against groups perceived as Muslim, is international conflicts between Islamic extremists and Western countries. Negative perceptions and hate crimes against Muslims increased following terror attacks by Islamic extremists,

most notably the attacks of September 11, 2001 and subsequent attacks in Europe by Al Qaeda and the Islamic State, and the wars in Afghanistan and Iraq carried out by the United States and allies (Singh 2002; Branton et al. 2011; Legewie 2013; Jungkunz, Helbling Schwemmer 2018).

While anti-Muslim discrimination may be motivated by religious prejudice, it is grounded in racialized markers such as phenotype, name, and dress rather than direct indicators of religion. Immigrants from Muslim-majority countries who are not Muslim are often subject to anti-Muslim discrimination because they are perceived to be likely Muslims (Naber 2008). Because racial and ethnic markers drive it, many scholars argue that Islamophobia is best regarded as a type of racism, even though Islam is not a race (Allen 2010).

Studies of Trends in Hiring Discrimination: We examine trends in the results of field experiments on discrimination over time. In field experimental studies, investigators make fictitious job applications. The applications include signals of the race or ethnicity of the applicant. Differential responses to applicants with distinct race-ethnic signals are used to estimate discrimination. Most commonly, the application is made by mail or over the internet, and race or ethnicity is signaled by a racially or ethnically identifiable name (Gaddis 2017, 2018). Some studies use face-to-face applications by persons hired to portray job seekers (e.g., Bovenkerk, Gras, and Ramsoedh 1995; Bendick, Rodriguez, and Jayaraman 2010). In either case, the control the experimenter exercises over the procedure ensures that members of the majority and minority racial groups are given resumes of on-average equivalent strength in terms of their non-racial qualifications, making discrimination the likely explanation of systematic differences in outcomes between groups, and giving these studies higher internal validity than other approaches to measuring discrimination (National Research Council 2004, Pager and Shepherd 2008).

Four previous articles examine trends in discrimination over time using field experimental data. Zschirt and Ruedin (2016, Table 4) contrasted rates of hiring discrimination from a sample of 26 correspondence experiments from European Union countries before and after 2000, finding an increase in discrimination after 2000. Their early analysis was important, but they did not examine country-specific or group-specific trends and did not account for changes in study characteristics over time. Quillian et al.

(2017) examined trends in hiring discrimination in the U.S., finding no change in discrimination against African-Americans since 1985 and some evidence of a decline in discrimination against Latinos. Heath and Di Stasio (2019) examined trends in discrimination against minority groups in Great Britain, finding no changes in discrimination overall or against specific minority groups over time. Finally, a recent working paper by Lippens, Vermeiren, and Baert (2021) finds evidence of a decline in hiring discrimination against racial-ethnic minority groups in Europe in correspondence studies in English between 2005 and 2020. They do not examine country or group-specific trends or include study controls (their primary focus is a contrast over several bases of discrimination beyond race-ethnicity).

Our paper adds to this literature in four ways. First, we assess temporal trends using more data, allowing better estimates of trends and allowing us to produce the first national trend estimates of hiring discrimination for Canada, France, Germany, and the Netherlands. Second, we also measure trends for specific minority groups rather than treating all ethnic and racial minorities as one category, which masks potentially significant variability in discrimination trends. Third, we consider country-specific trends because evidence suggests that country is an important structuring factor of discrimination lost in samples aggregated over a broad region like Western Europe (Quillian et al. 2019, Thijssen et al. 2021). Moreover, countries vary greatly in the time range of data available. Fourth, we control for study characteristics that are potential confounders of the time trend, such as whether the study is conducted via postal mail or over the internet. As Ross (2017) points out, this is a potentially critical issue in drawing valid conclusions about discrimination trends from trends in the outcomes of field experiments.

Data and Approach: We conduct an analysis that combines all available field experimental estimates of discrimination against race-ethnic minority groups from six countries: Canada, France, Germany, Great Britain, the Netherlands, and the United States. We chose these six countries because they have at least six field experiments of race-ethnic discrimination conducted more than ten years apart, a minimum for trend analysis.

Table 1 shows the number of effects - estimates of discrimination against a minority group - and the number of studies in each country and for each minority group. There are more effects than studies

Table 1. Number of Effects, Studies, and Applications by Country and Minority Groups

Country	Effects	Studies	Applications	Year Range of Data
Canada	14	7	22735	1985-2011
France	23	20	44586	1976-2018
Germany	8	6	9447	1994-2017
Great Britain	30	12	9089	1969-2017
Netherlands	25	15	10727	1976-2017
US	40	30	77495	1985-2019
Total	140	90	174079	

Minority Group	Effects	Studies	Notes
African/Black	57	49	Excludes North African
Middle-Eastern/N. African	41	37	Includes Turkish
Latin Am./Hispanic	12	12	
Asian	30	19	Includes East Asian and South Asian
Total	140		

Notes: For minority group, a study can be present in more than one racial/ethnic category. Effects are discrimination estimates against minority groups. Some studies have estimates for multiple minority groups.
A few US. Studies before 1985 excluded for problems in how they signaled race.

because many studies include estimates of discrimination against multiple groups, such as a study that examines discrimination against Black and Latino applicants. The range of years available varies significantly over countries and groups: in some countries like Great Britain, we can have discrimination estimates as far back as 1970, whereas for others like Germany the first field experiments of discrimination begin in the early 1990s. We excluded studies with fieldwork conducted in 2020 or 2021 because COVID may have influenced discrimination in hiring. There are too few discrimination studies yet available with fieldwork conducted during COVID to characterize this period separately.

To allow an analysis of trends by minority group, we coded specific minority groups into four racial/ethnic categories based on the region of minority group origin: African/Black, Middle Eastern/North African, Latin American/Hispanic, and Asian (see Supporting Information Table S8 for a list of the specific minority groups). We use these categories because outsiders recognize them in relatively quick interactions such as hiring reviews. By contrast, ethnic distinctions within these categories are often incorrectly perceived by outsiders (Abdelkader 2017). We only examine non-European racial-ethnic groups because there are too few discrimination field experiments including white or European ethnic minority groups to allow a trend analysis.

Results

We begin with a pooled analysis and then move to the group and country-specific analyses. Figure 1 shows a scatterplot of all the data in our study by year. Each dot indicates a discrimination estimate against a minority group. The discrimination ratio is on the y-axis, and the year of the fieldwork is on the x-axis. The discrimination ratio can be interpreted as the expected number of applications a minority applicant must submit to expect the same number of callbacks as a majority applicant (see Materials and Methods). Larger numbers indicate more discrimination; for instance, a ratio of 1.5 means the minority applicant must submit 50% more applications to receive the same number of callbacks as a white native applicant. The size of each dot is proportional to the weight given the observation in the meta-analysis (the random-effect weights).

We overlay a meta-regression trend line on the dots in Figure 1. The line is flat: The meta-regression slope of discrimination on year is 0.0001, indicating an estimated increase in the discrimination

ratio of .01% per year, or almost no change over time. On average, hiring discrimination against minority groups in our sample has not changed.

This trend estimate lacks controls for study characteristics. Table 2 lists controls we introduce to account for potentially confounding variables, including dummies for country and minority group, occupational categories, applicant gender, applicant education, immigrant status, source of jobs online/offline, study in-person or correspondence, local unemployment rate, and share of the local population made up of immigrants (on area predictors see Thijssen 2021).

Figure 2 shows estimates of the linear trend in discrimination and 95% confidence interval ranges for various meta-regression models using the pooled data. Some models include additional controls and the last two use different subsets of data. The meta-regressions are a subgroup correlated effects model outlined in Pustejovsky and Tipton (2022, pp. 429-430), allowing for country-specific residual variability but estimating a single average year slope over countries. The model estimates are shown including Tau estimates in Supporting Information Table S1. For model details, see Materials and Methods. Table S1 shows significant heterogeneity in discrimination over studies not accounted for by model controls. The variability also differs considerably by country.

In Figure 2, the first model has no controls and corresponds to the top line. The point estimate of the linear trend remains small and statistically insignificantly different from zero when we add dummy variables for minority group and year in model 2. The third model adds controls for the basic model study characteristics listed in Table 2. The estimated slope lines change slightly, becoming positive and suggesting an upward shift in discrimination. However, the line is still relatively flat and non-significant, with an estimated slope of 0.37% per year in the base model. The fourth and fifth models add additional controls for immigrant minorities' nationality and place of education as contextual characteristics. Again, these still produce flat lines with slopes near zero.

Finally, the two models at the bottom (models 6 and 7) alter the sample used to estimate the trend, first by dropping studies that only include one occupation, then by only using resume audits. Again, the linear trend estimates remain nearly flat. As discussed in Materials and Methods, tests for non-linearity found no significant evidence of non-linearity. Beginning the time series in the mid-1980s produces similar estimates; see Supporting Information Table S2.

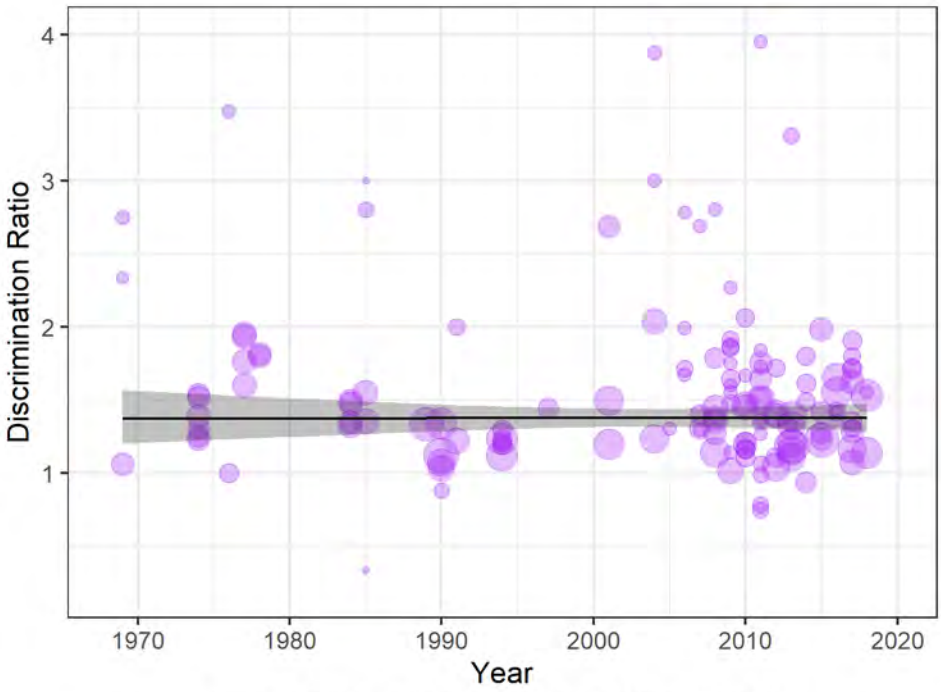
Table 2. Control Variables Included in Meta-Analysis Models of Discrimination Trends

Variables	**Measurement / Categories**
Country	Indicators: USA (ref.), Canada, England, France, Germany, Netherlands
Minority Group	Indicators: African/Black (ref.), Latin Am./Hispanic, Asian, MENA
Controls added in the "Basic Model"	
Study Method	Indicators: Resume Audit (ref.), In-Person Audit
Tester Gender	Indicators: Male(ref.), Female, Mixed
Applicant Education	Indicators: HS or Less, Some College or Post-HS Vocational Degree(ref.), College or More, Missing
Occupations Included (All that apply)	Indicators: Blue Collar Jobs, Jobs with Customer Contact, Office Jobs
Foreign Birth	Indicators: Includes Foreign Born Minority Persons, Domestic born(ref.)
Source of Jobs Online	Indicators: Online Source, Offline Source, Both Online and Offline(ref.)
Additional Controls in Some Models in Table 2	
Highest Degree from Foreign School?	Indicators: Yes, No(ref.), Mixed
Foreign Nationality (Citizenship)	Indicators: Yes, No(ref.), Mixed
Local Unemployment Rate	Percentage
Foreign Born in Region	Percentage

For descriptive statistics for controls see Supporting information Table S6

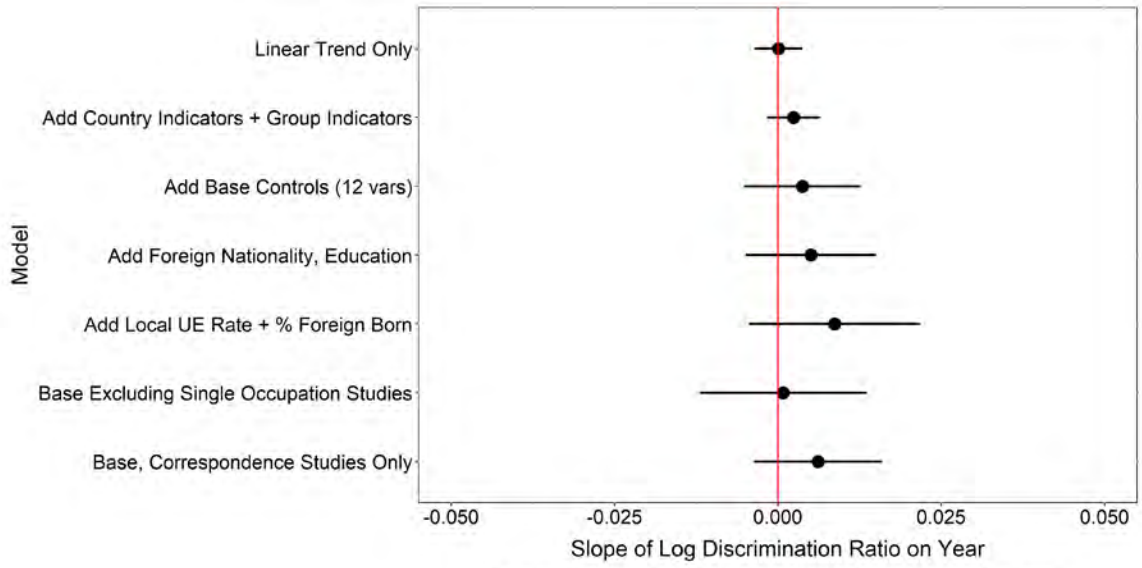
"Indicators" variables are represented by dummy variables, with reference category indicated by "ref."

Figure 1. Discrimination Ratios over Time, Pooled



Notes: 2 points outside the range. Line from Model 1, Table 2.
Shaded area is 95% Confidence Region. Size of symbol proportional to meta-analysis weight.

Figure 2. Annual Trend in Hiring Discrimination, Pooled Models



Notes: Dots represent the point estimates, lines the 95% confidence intervals.
Model shown in Table 2.

Trends by Minority Group

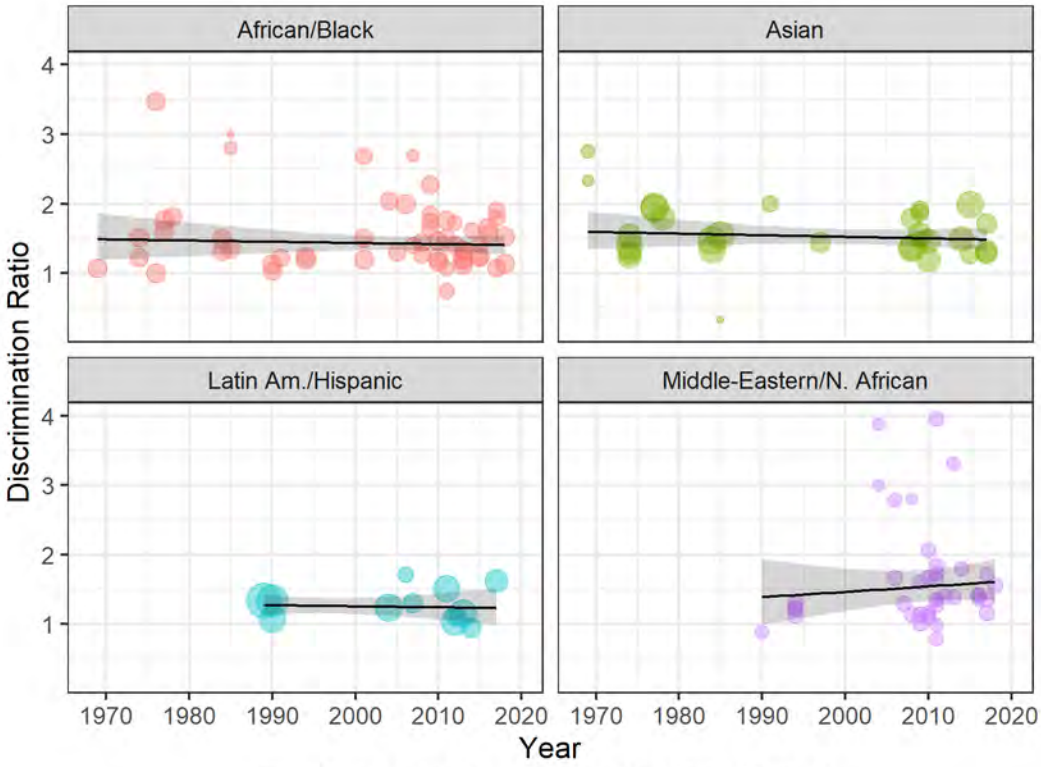
We now turn to estimates of trends specific to the minority group's region of origin. In Figure 3, we overlay a meta-regression line on scatterplots of discrimination ratios over time by group. Figure 4 contains dot-and-line plots of estimates of minority group-specific linear trends from meta-regression models. We show results for three sets of models. The first model shows slope estimates of a separate meta-regression for each minority group with no controls. The second model shows results from a separate meta-regression for Black and MENA minorities with "base" controls (listed in Table 2). There is insufficient data to estimate this model for Latino and Asian minorities. The third model estimates a model pooling all countries and groups together with base controls allowing country-specific residual variance and group-specific slopes but constraining slopes of the controls to be the same across countries. (Model estimates are in Supporting Information Table S3.)

Similar to the pooled analysis, trends by minority group are approximately flat for three of the four groups. The one exception is for MENA minorities, where we find evidence of an upward trend in discrimination, but the slope is not statistically significantly different from zero. However, an investigation of the scatterplots (Figure 3) suggested a possible non-linear pattern for the MENA group: studies in the 1990s appear to find less discrimination than those after 2000. (No MENA studies are available before 1990.)

Figure 5 shows estimates for MENA minorities modeled with a series of dummy variables representing decades. We divide the data into the years 2000 or earlier, 2001 to 2010, and after 2010. Years 2000 or earlier is the reference category. (Model estimates are shown in Supporting Information Table S4.)

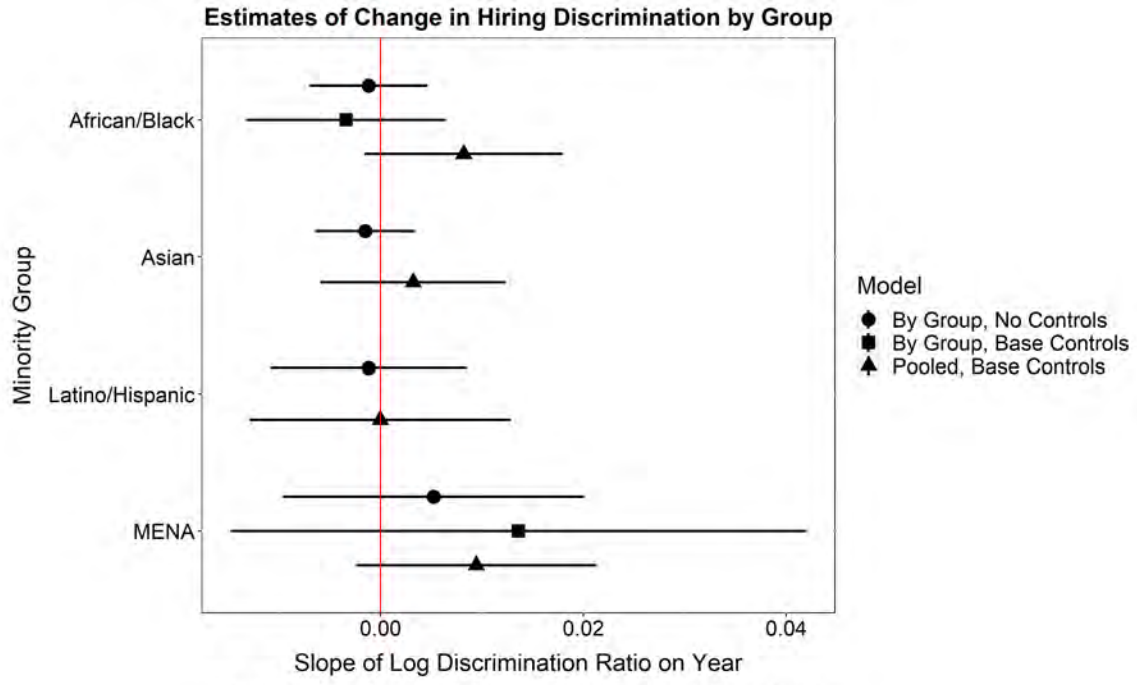
The model estimates show a statistically significant increase in discrimination against MENA groups from 2001-2010 and after 2010 compared to the 1990s. This trend persists in Model 3 when we include a complete set of controls for country and study characteristics. The change from the 1990s is substantively significant: it is an increase in the discrimination ratio of 40% (exp(.3351)) in the baseline model and more when we add base model controls. These results are consistent with the idea that attacks carried out by Islamic extremists during the 2000s in Europe and the U.S. (including September

Figure 3. Discrimination Ratios over Time by Group



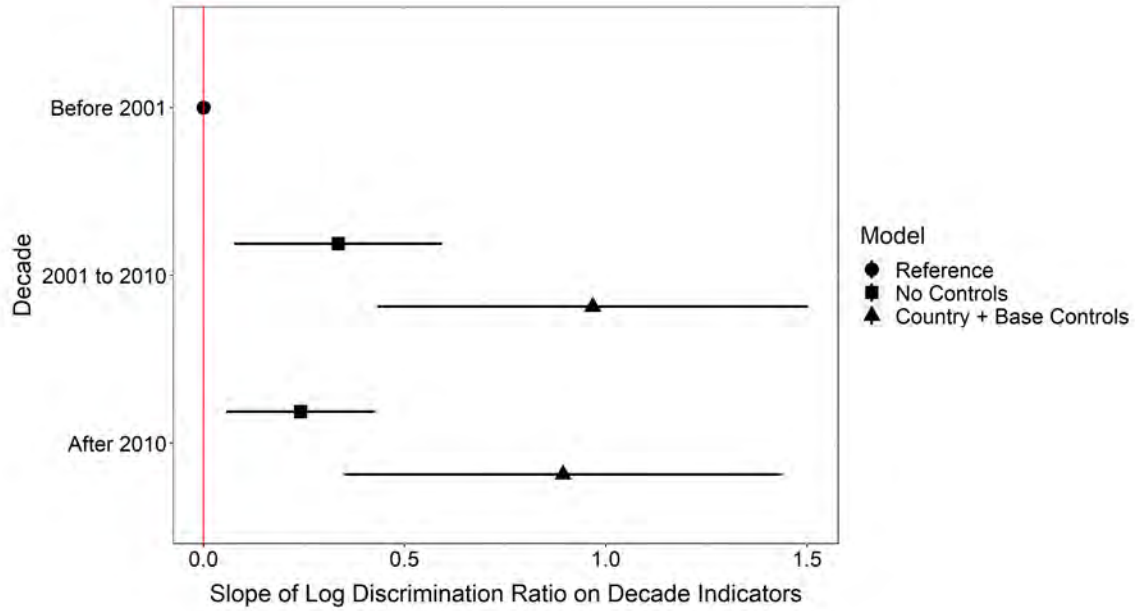
Notes: 2 points outside the range. Lines from Model 1, Table S1.
Shaded area is 95% Confidence Region. Size of symbol proportional to meta-analysis weight.

Figure 4. Trends in Hiring Discrimination by Group



Note: Dots represent the point estimates, lines the 95% confidence intervals.

Figure 5: MENA Hiring Discrimination by Decade



Notes: Dots represent the point estimates, lines the 95% confidence intervals. Model estimates shown in Table S2 models 1-3.

11, 2001) and the wars in Afghanistan and Iraq provoked an increase in hiring discrimination against MENA minorities in Western countries.

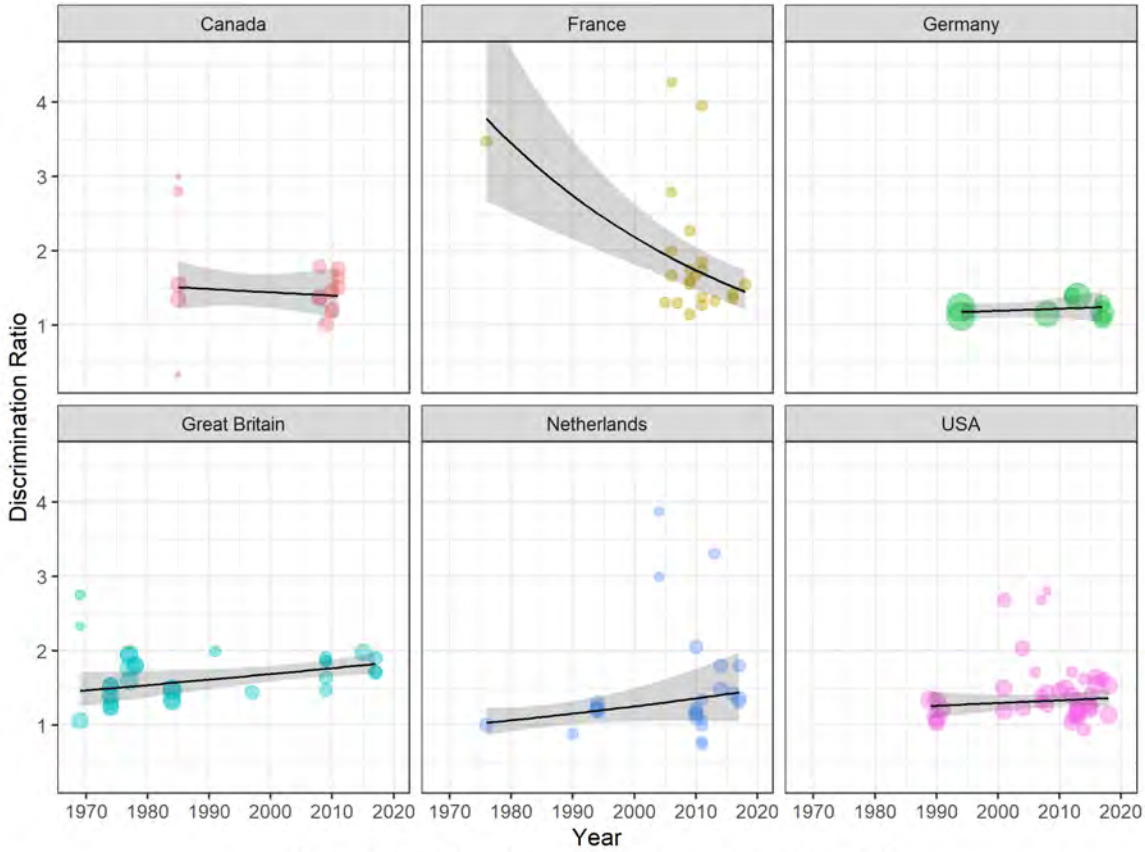
Discrimination against MENA minorities may result from discrimination based on the perception that applicants are Muslim. Religion is not signaled directly in most field experiments, because job candidates do not typically indicate their religion in job applications. To further explore this increase in discrimination and its relation to Muslim origins, we examined trends in discrimination against minorities with roots in Muslim-majority countries (shown in the right panel of Supporting Information Table S4). Muslim-majority countries include ethnic minorities from persons in the MENA region and minorities from some Muslim-majority countries in South Asia (like Pakistan) and sub-Saharan Africa (like Senegal). These results show similar patterns of change in the point estimates for minority groups from Muslim-majority countries of origin as we saw for MENA minorities. Hiring discrimination against minority groups of MENA origin and Muslim-majority origin countries saw an increase after 2000, with stability thereafter.

Trends by Country

Scatterplots of discrimination ratios over year by country are shown in Figure 6. Figure 7 shows dot-and-line graphs of the slopes (trends) by country from meta-regressions. (Corresponding model estimates of trends in discrimination are shown in Supporting Info Table S5.) The figures demonstrate that the trends over time vary by country: there is little change in Canada, Germany, Great Britain, and the U.S.; some increase in the Netherlands; and a decrease in France. Most countries do not have enough studies to estimate country-specific slopes for controls. To allow us to add the base controls in national models, we constrain the slope of the control variables to be the same across countries. Adding controls does not significantly change the estimates of the by-country slopes.

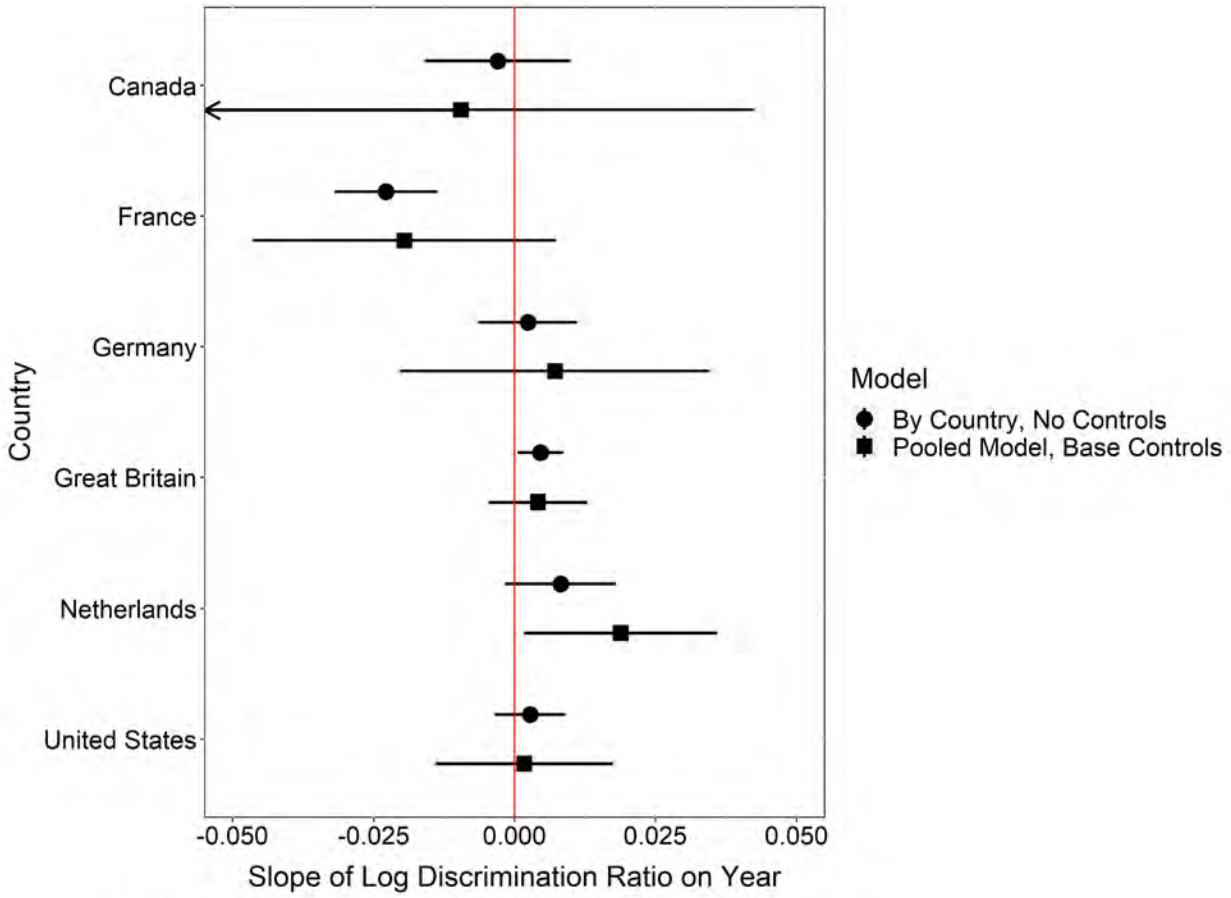
Examination of scatterplots for France in Figure 6 shows one point well outside the year range of the other studies: one study from the 1970s in France examining discrimination against migrants from sub-Saharan Africa. This study is the only one before 1985 in France; it found very high levels of discrimination. This outlier is a plausible influential point, but removing it only slightly alters the slope of

Figure 6: Discrimination Ratios over Time by Country



Notes: 1 point outside the range. Lines from Table S3 Column 1.
Shaded area is 95% confidence region. Size of symbol proportional to meta-analysis weight.

Figure 7: Trends in Hiring Discrimination by Country



Notes: Dots represent the point estimates, lines the 95% confidence intervals. Model estimates shown in Table S2 models 1-3.

the trend line, making it a bit steeper in models 1 and 2 and a bit less steep in model 3 (shown in Supporting Information Table S5).[1]

Trends by Country and Group Combined

Finally, we consider results for country and minority group combinations. We calculate trends for country-minority group combinations with at least five studies. Results are shown in Supporting Information, Figure S1 (scatterplots), Figure S2 (dot-and-line graphs of slopes), and Table S6 (model estimates).

Discrimination trends against different minority groups within a country usually follow the same direction. For most country-group combinations, there is a pattern of relative stability in discrimination over time. However, there is evidence of declining discrimination against minorities in France and increasing discrimination in the Netherlands. In short, the combined country-group results are highly consistent with results in the separate country and group analyses.

Discussion

Hiring discrimination against non-white minorities has not declined significantly over the last 20-40 years in five of the six countries we examine. In pooled analyses and four of the six countries—Canada, Germany, Great Britain, and the United States—we find relative stability over time in the level of hiring discrimination against racial and ethnic minority groups.

There are three exceptions to stability over time. The first is an increase in hiring discrimination against minority groups with ethnic origins in Middle Eastern and North-African countries and Muslim-majority countries after 2000. The most plausible explanations for this sudden shift include the terrorist attacks carried out by Al-Qaeda in the early 2000s (notably, the attacks of September 11, 2001 and bombings in Madrid in 2004 and London in 2005), the subsequent Western response of military interventions in the "War on Terror," and the increasing politicization of immigration from Muslim-majority countries. Our results suggest the backlash normalized biases against MENA minorities, leading to more hiring discrimination against this group beginning in the 2000s. We also find this pattern of increased

[1] However, it causes the standard error of the year coefficient for France to increase so the decline for France is no longer statistically significant at p<.1.

discrimination in the early 2000s to hold if we examine minority groups with origins in Muslim-majority countries rather than MENA minorities. This result adds to the literature suggesting political effects on hiring discrimination.

The second exception to the pattern of stability is a decline in hiring discrimination in France. However, it is worth noting that in relative terms, this decline was from a very high level relative to other countries to "merely" high levels. This decline meant native white French applicants "only" received 50% to 70% more callbacks than similarly qualified Black and MENA applicants after 2010. The decline in hiring discrimination in France suggests a convergence in France to a level of hiring discrimination similar to other Western European countries like Great Britain.

The third exception is we find some evidence of an increase in discrimination in the Netherlands. The upward trend in the Netherlands is a slightly sharper version of a trend we find in several countries, enough to make it a statistically significant increase in some analyses, rather than a unique trend for the Netherlands.

The lack of decline in discrimination in five of the six countries we examine, despite new anti-discrimination legislation, adoption of hiring practices aiming to increase diversity in many large corporations, and some evidence of attitudinal changes, is disturbing. It suggests that hiring discrimination results from enduring stereotypes, prejudices, or racist ideologies. Perhaps legal and social changes have been offset by a political backlash against immigration and ethnic minorities. Our results contradict the belief, widespread among the public and some academics (Kraus, Rucker, and Richeson 2017; Seamster and Ray 2018) that there are elements of Western societies or cultures that will "naturally" produce gradual reductions in discrimination over time. Yet persistent labor market discrimination is not inevitable, as demonstrated by our results for France and evidence of reductions in labor market discrimination in the U.S. in the 1960s (Heckman and Payer 1989, Wilson 2011).

Our study has several limitations. The data we use to estimate time trends has many more studies from more recent years (especially after 2000) than earlier years (except in Great Britain). For Germany and Canada, our trends analysis is based on fewer than ten studies, and for Germany our time series of studies only begins in 1994. Our studies mostly use broad samples of entry-level jobs available in public sources; ethnic job submarkets and positions that primarily recruit through social networks are

mostly not included in our sample. Our results cover minority groups included in field experiments of discrimination, generally groups that are larger and thought to be more likely to be victims of discrimination. Finally, declines in hiring discrimination against racial and ethnic minorities before our data begin are possible. Discrimination declined in the U.S. during the 1960s and early 1970s with anti-discrimination legislation (Heckman and Payner 1989; Tomaskovic-Devey et al. 2006). Changes for other countries before the beginning of our time series are possible as well. Finally, we excluded studies that conducted fieldwork after 2019 because we believe COVID may have affected hiring discrimination. Few field experiments of discrimination conducted during COVID are available, not enough to characterize COVID in contrast to earlier periods.

Without confronting discrimination, persistent gaps in employment and earnings in Western nations between minority groups and white majority populations—so-called "ethnic penalties" (Heath and Cheung 2007)—are likely to persist. Discrimination impedes the economic incorporation of minority groups and immigrants in ways likely to contribute to further white backlash. Because of a lack of data on race and ethnicity in Europe, relatively little is known about how racial and ethnic inequality has evolved in European societies. Still, our results provide no reason to support the hope that racial and ethnic inequalities have decreased. An important question for future research is how hiring procedures and anti-discrimination legislation can be revamped to yield sustained declines in hiring discrimination.

Materials and Methods

Study Search and Coding: We used three methods to identify relevant field experiments: searches in bibliographic databases, citation searches, and an e-mail request to corresponding authors of field experiments on racial-ethnic discrimination in labor markets and other experts on field experiments and discrimination. Details of the procedures are in Supporting Information Appendix A.

We coded effects that measure discrimination based on counts of applications and callbacks by racial or ethnic group. Most studies included these counts in their research report. When the study report did not have counts of callback outcomes, we requested counts from the authors and excluded the study if we did not receive the counts. We also coded other study characteristics such as the occupations

covered, the level of education of the applicants, and the gender of the applicants. Supporting Information Table S7 shows descriptive statistics for these characteristics.

Outcome: The Discrimination Ratio: Our outcome is the ratio of a majority group's callback rate to a minority group's callback rate. A callback is a request for more information or an invitation to interview and is the primary outcome used in the large majority of field experiments of hiring. Suppose c^w is the number of callbacks received by white natives, c^m is the number of callbacks received by a minority racial or ethnic group, n^w is the number of applications submitted by white native applicants, and n^m is the number of applications submitted by minority applicants. The discrimination ratio is $(c^w/n^w)/(c^m/n^m)$. We calculated this ratio based on the results reported in the studies or provided by the study authors. Ratios above 1 indicate that the majority group received more positive responses than the minority group, with the amount above one multiplied by 100 indicating the relative scale of the advantage enjoyed by native white applicants. The discrimination ratio may be interpreted as the number of applications a minority applicant must submit to have an equal chance of receiving a callback as an otherwise similar native white applicant. With a discrimination ratio of 2.0, for example, a minority candidate has to send out two applications for every application submitted by the white tester to expect to receive the same number of positive indications of employer interest.

Meta-Analysis Model: To examine time trends, we model the discrimination ratio as a function of fieldwork year and other study characteristics using meta-regression (Borenstein et al. 2009, chapter 20). The field experiments in our analyses have some similar design features that make their results relatively comparable, such as using the "callback" outcome and selecting jobs from public listings.

We use two procedures to deal with potential non-comparability across studies. First, as discussed above, we code many characteristics of field experiments and control for them in the model, allowing us to account for potential changes in attributes of studies over time that may confound the time trend. Second, we use a model with a random component at the study level, as is common in meta-regression studies (Raudenbush 2009). The random effects specification incorporates a variance component capturing unexplained variation in the outcome across studies. This increases standard errors to reflect additional uncertainty from unaccounted-for study-level characteristics.

We use meta-regression to model the discrimination ratio, y, as a function of a vector of characteristics of the studies and effects, **x**, plus residual study-level heterogeneity (between-study variance in the outcome not explained by the covariates). We use two related meta-analysis models.

We estimate standard random-effect meta-regression models for specific minority groups, countries, or country-group combinations. In these cases the model is:

$$ln(y_{ij}) = \alpha t_i + x_{ij}\boldsymbol{\beta} + u_i + e_{ij}, where\ u_i \sim N(0, \tau^2)\ and\ e_{ij} \sim N(0, \sigma_{ij}^2)$$

where t is the year the study's fieldwork was conducted, α is the slope of change over years, **β** is a $k \times 1$ vector of coefficients (including a constant), and x_{ij} is a $1 \times k$ vector of control covariate values in study i and effect size j (k is the number of covariates including a 1 for a constant). Following standard practice in the meta-analysis literature, we log the discrimination ratio to reduce the asymmetry of the ratio. Residual between-study variance is τ^2, estimated as part of the meta-analysis model. The logged discrimination ratio variance for the jth effect size in the ith study, σ_{ij}^2, reflects sampling variation for estimating each discrimination ratio for each study, and is calculated from counts of applications and callbacks (formulas are in Supporting Information Appendix C). The model is estimated by constrained maximum likelihood.

This model has a single time-trend for the model (α) and a single variability parameter estimated. Because of the much smaller sample sizes of model estimates for individual countries or minority groups, in models estimated separate by country or minority group we employ fewer controls.

We can employ more controls by pooling, assuming that the slopes of the controls are the same over countries. In models that pool data from multiple countries, we use a meta-regression with a random effects structure of subgroup correlated effects from Pustejovsky and Tipton (2022). We estimate a single meta-regression with country-specific year slopes and country-specific residual variance(τ_c^2).

The subgroup correlated meta-analysis model is:

$$ln(y_{ij}) = \alpha_c t_i + x_{ij}\boldsymbol{\beta} + u_i + e_{ij}, where\ u_i \sim N(0, \tau_c^2)\ and\ e_{ij} \sim N(0, \sigma_{ij}^2)$$

Where parameters are as above, except α_c is the slope of change over years in country c, and residual between-study variance for country c is τ_c^2, estimated as part of the meta-analysis model. We find significant differences in the slope of time for distinct countries and also highly different residual variability of the discrimination ratio by country, so this model corresponded to the data better than a standard meta-analysis model fit to a pooled dataset. We constrain the slopes of the controls to be the same over countries, allowing us to include controls in countries with too few cases to estimate slopes separately by country, which are most countries in our analysis. We estimate the models using the "metafor" package in the R statistical language (Viechtbauer 2010) with procedures from Pustejovsky and Tipton (2022) to estimate the subgroup correlated effects models.

Standard Errors: To account for the correlation of effect estimates from the same study, we use robust standard errors clustered at the study level. The standard errors are calculated with the "robust" command in "metafor" and the "clubSandwich" module (see Pustejovsky and Tipton 2018).

Significance Testing and Multiple Comparisons: We show significance test results at significance levels (α) of .1, .05, and .01. In tables with tests for multiple countries and/or groups, we also show results of the Benjamini-Hochberg (B.H.) adjustment for multiple comparisons with a false discovery rate (FDR) of .2. We discuss results when they are significant at the relatively high levels of α=.1 and FDR=.2 because we believe the problems caused by false negative in our analysis (concluding there is no trend when in fact there is a trend) are no less significant than are false positives; in contrast for causal-effects analysis the problems of false positives often exceed false negatives. We calculate confidence intervals and hypothesis tests with the "clubsandwich" procedure in R employing small-sample adjustments using the t-distribution (see Pustejovsky and Tipton 2018).

Non-linearity: In some models, we also replace the single linear year trend with dummy variables representing decade categories to capture potential non-linear trends in discrimination. We also estimated models with a squared term (Supporting Information Table S9 shows results from the pooled single-trend model). Except for changes in discrimination for MENA minorities, we find no clear evidence of non-linearity.

Odds Ratio Outcome: We estimated basic models using odds ratios of a callback in place of the risk ratio of a callback. This produced similar general conclusions to the risk ratios. See Supporting Information Table S10.

Publication Bias: Publication bias is examined in Supporting Information Appendix D. Some tests, but not all, suggest publication bias. Importantly, none of the tests indicated that the extent of publication bias changed over time or showed significant changes to trend estimates after adjusting for publication bias.

References

Abdelkader, Engy. 2017. "A Comparative Analysis of European Islamophobia: France, UK, Germany, Netherlands and Sweden." *Journal of Islamic and Near Eastern Law* 16(1): 29–63.

Allen, Chris. 2010. "What Is Islamophobia?" Pp. 123–38 in Islamophobia. New York, NY: Routledge.

Bendick, Marc, Rekha Eanni Rodriguez, and Sarumathi Jayaraman. 2010. "Employment Discrimination in Upscale Restaurants: Evidence from Matched Pair Testing." The Social Science Journal 47(4):802–818.

Blinder, Scott, Robert Ford, and Elisabeth Ivarsflaten. 2013. "The Better Angels of Our Nature: How the Antiprejudice Norm Affects Policy and Party Preferences in Great Britain and Germany." *American Journal of Political Science* 57(4):841–57. doi: 10.1111/ajps.12030.

Bobo, Lawrence, James R. Kluegel, and Ryan A. Smith. 1997. "Laissez-Faire Racism: The Crystallization of a Kinder, Gentler, Antiblack Ideology." in Racial Attitudes in the 1990s: Continuity and Change, edited by S. A. Tuch and J. K. Martin. Westport, CT: Praeger Publishers.

Bonikowski, Bart, Yuval Feinstein, and Sean Bock. 2021. "The Partisan Sorting of 'America': How Nationalist Cleavages Shaped the 2016 U.S. Presidential Election." *American Journal of Sociology* 127(2):492–561. doi: 10.1086/717103.

Bonilla-Silva, Eduardo. 2017. *Racism without Racists: Color-Blind Racism and the Persistence of Racial Inequality in America*. 5th ed. Lanham, MD: Rowman & Littlefield Publishers.

Borenstein, Michael, Larry V. Hedges, Julian P. T. Higgins, and Hannah R. Rothstein. 2009. *Introduction to Meta-Analysis*. West Sussex, UK: John Wiley & Sons.

Bovenkerk, Frank, Mitzi J. I. Gras, and Dhurender Ramsoedh. 1995. *Discrimination against Migrant Workers and Ethnic Minorities in Access to Employment in the Netherlands*. Geneva: International Labour Office, Employment Department. Retrieved February 20, 2014 (http://www.ilo.org/migrant/publications/working-papers/WCMS_201033/lang--en/index.htm).

Branton, Regina, Erin C. Cassese, Bradford S. Jones, and Chad Westerland. 2011. "All Along the Watchtower: Acculturation Fear, Anti-Latino Affect, and Immigration." *The Journal of Politics* 73(3):664–79.

Corbett, William. 2014. "What Is Troubling About the Tortification of Employment Discrimination Law?" *Ohio State Law Journal* 75:1028–77.

Di Stasio, Valentina, and Anthony Francis Heath. 2021. "London Calls? Discrimination of European Job Seekers in the Aftermath of the Brexit Referendum." *Frontiers in Sociology* 6:737857. doi: 10.3389/fsoc.2021.737857.

Dobbin, Frank. 2011. *Inventing Equal Opportunity*. Princeton, NJ: Princeton University Press.

Dobbin, Frank, Daniel Schrage, and Alexandra Kalev. 2015. "Rage against the Iron Cage: The Varied Effects of Bureaucratic Personnel Reforms on Diversity." *American Sociological Review* 80(5):1014–44.

Dovidio, John F., Samuel L. Gaertner, and Adam R. Pearson. 2016. "Aversive Racism and Contemporary Bias." Chapter. In The Cambridge Handbook of the Psychology of Prejudice, edited by Chris G. Sibley and Fiona Kate Barlow, 267–94. Cambridge Handbooks in Psychology. Cambridge: Cambridge University Press doi:10.1017/9781316161579.012.

ENAR (European Network Against Racism). 2017. *Shadow Report: Racism and Discrimination in Employment in Europe 2013-2017*. http://www.enar-eu.org

Fredrickson, George. 2002. *Racism: A Short History*. Princeton, NJ: Princeton University Press.

Gaddis, S. Michael. 2017. "How Black Are Lakisha and Jamal? Racial Perceptions from Names Used in Correspondence Audit Studies." Sociological Science 4: 469–89. https://doi.org/10.15195/v4.a19.

Gaddis, S. Michael. 2018. "An Introduction to Audit Studies in the Social Sciences." Pp. 3-44 in Audit Studies: Behind the Scenes with Theory, Method, and Nuance, Ed. S. Michael Gaddis. Spring, Cham, Switzerland.

Golder, Matt. 2016. "Far Right Parties in Europe." *Annual Review of Political Science* 19:477–97.

Goldstein, Adam, and Tod Hamilton. 2022. “Dog Whistles and Work Hours: The Political Activation of Labor Market Discrimination.” *Sociological Science* 9:40–74. doi: 10.15195/v9.a3.
Gorzig, Marina Mileo, and Deborah Rho. 2022. “The Effect of the 2016 United States Presidential Election on Employment Discrimination.” *Journal of Population Economics* 35(1):45–88. doi: 10.1007/s00148-021-00837-2.
Heath, Anthony F., and Sin Yi Cheung. *Unequal Chances: Ethnic Minorities in Western Labour Markets*. British Academy, 2007.
Heath, Anthony F., and Valentina Di Stasio. 2019. “Racial Discrimination in Britain, 1969–2017: A Meta-Analysis of Field Experiments on Racial Discrimination in the British Labour Market.” The British Journal of Sociology 70(5):1774–98. doi: 10.1111/1468-4446.12676.
Heckman, James J., and Brook S. 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):138–77.
Jacob-Owens, Timothy. 2019. 25 Years of the European Commission against Racism and Intolerance: Impact in the Field of Equality. Council of Europe. Retrieved February 16, 2021 from https://rm.coe.int/25-years-ecri-impact-in-the-field-of-equality/1680972f3e
Jungkunz, Sebastian, Marc Helbling, and Carsten Schwemmer. 2019. “Xenophobia before and after the Paris 2015 Attacks: Evidence from a Natural Experiment.” Ethnicities 19(2):271–91.
Kalev, Alexandra, Frank Dobbin, and Erin Kelly. 2006. “Best Practices or Best Guesses? Assessing the Efficacy of Corporate Affirmative Action and Diversity Policies.” *American Sociological Review* 71(4):589–617. doi: 10.1177/000312240607100404.
Klinkner, Philip A., and Rogers M. Smith. 2002. *The Unsteady March: The Rise and Decline of Racial Equality in America*. 1st edition. Chicago, Ill.: University of Chicago Press.
Kraus, Michael W., Julian M. Rucker, and Jennifer A. Richeson. 2017. “Americans Misperceive Racial Economic Equality.” *Proceedings of the National Academy of Sciences* 114(39):10324–31. doi: 10.1073/pnas.1707719114.
Kymlicka, Will. 2009. *Multicultural Odysseys: Navigating the New International Politics of Diversity*. Oxford ; New York: Oxford University Press.
Legewie, Joscha. 2013. “Terrorist Events and Attitudes toward Immigrants: A Natural Experiment.” American Journal of Sociology 118(5):1199–1245.
Lippens, Louis, Siel Vermeiren, and Stijn Baert. 2021. “The State of Hiring Discrimination: A Meta-Analysis of (Almost) All Recent Correspondence Experiments.” IZA Discussion Paper No. 14966.
McConahay, J. B. 1986. Modern racism, ambivalence, and the Modern Racism Scale. In J. F. Dovidio & S. L. Gaertner (Eds.), Prejudice, discrimination, and racism (pp. 91–125). Academic Press.
Naber, Nadine. 2008. “Look, Mohammed the Terrorist Is Coming!: Cultural Racism, Nation-Based Racism, and the Intersectionality of Oppressions after 9/11.” in Race and Arab Americans before and after 9/11: From invisible citizens to visible subjects, edited by A. Jamal and N. Naber. Syracuse, NY: Syracuse University Press.
National Research Council. 2004. *Measuring Racial Discrimination*. Washington, DC: National Academies Press.
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. doi: 10.1146/annurev.soc.33.040406.131740.
Pettigrew, T. F., and R. W. Meertens. 1995. “Subtle and Blatant Prejudice in Western Europe.” *European Journal of Social Psychology* 25(1):57–75. doi: 10.1002/ejsp.2420250106.
Pustejovsky, James E., and Elizabeth Tipton. 2018. “Small-Sample Methods for Cluster-Robust Variance Estimation and Hypothesis Testing in Fixed Effects Models.” *Journal of Business & Economic Statistics* 36(4):672–83. doi: 10.1080/07350015.2016.1247004.
Pustejovsky, James E., and Elizabeth Tipton. 2022. “Meta-Analysis with Robust Variance Estimation: Expanding the Range of Working Models.” *Prevention Science* 23(3):425–38. doi: 10.1007/s11121-021-01246-3.

Quillian, Lincoln, Anthony Heath, Devah Pager, Arnfinn Midtbøen, Fenella Fleischmann, Ole Hexel. 2019. "Do Some Countries Discriminate More than Others? Evidence from 97 Field Experiments of Racial Discrimination in Hiring." Sociological Science 6: 467-496.

Quillian, Lincoln, Devah Pager, Ole Hexel, and Arnfinn H. Midtbøen. 2017. "Meta-Analysis of Field Experiments Shows No Change in Racial Discrimination in Hiring over Time." *Proceedings of the National Academy of Sciences* 114(41):10870–10875.

Raudenbush, Stephen, 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.

Ross, Stephen L. 2017. "Measuring Trends in Discrimination with Field Experiment Data." *Proceedings of the National Academy of Sciences* 114 (41): 10815–17. https://doi.org/10.1073/pnas.1715243114.

Rubin, Jennifer, Jirka Taylor, Alexandra Pollitt, Joachim Krapels, and Mafalda Pardal. 2014. Intolerance in Western Europe: Analysis of Trends and Associated Factors. Santa Monica, CA: RAND Corporation.

Schuman, Howard, Charlotte Steeh, Lawrence D. Bobo, and Maria Krysan. 1997. *Racial Attitudes in America: Trends and Interpretations*. Cambridge, MA: Harvard University Press.

Seamster, Louise, and Victor Ray. 2018. "Against Teleology in the Study of Race: Toward the Abolition of the Progress Paradigm." Sociological Theory 36(4):315–42. doi: 10.1177/0735275118813614.

Sears, David, and P.J. Henry. 2004. "Over thirty years later: A contemporary look at symbolic racism." In M. P. Zanna (Ed.), Advances in experimental social psychology (Vol. 37, pp. 95–150). San Diego, CA: Academic Press.

Semyonov, Moshe, Rebeca Raijman, and Anastasia Gorodzeisky. 2006. "The Rise of Anti-Foreigner Sentiment in European Societies, 1988-2000." *American Sociological Review 71(3):426–49.*

Singh, Amardeep. 2002. "'We are Not the Enemy': Hate Crimes against Arabs, Muslims, and those Perceived to be Arab or Muslim after September 11." Human Rights Watch Report, November.

Simon, Patrick. 2021. "Discrimination: Studying the Racialized Structure of Disadvantage." Chapter 5 in Contested Concepts in Migration Studies, edited by Dirk Jacobs, Riva Kastoryano, Ricard Zapata Barrero, New York: Routledge.

Skrentny, John D. 2004. *The Minority Rights Revolution*. Cambridge, Mass.: Belknap Press: An Imprint of Harvard University Press.

Thijssen, Lex. 2021. "Regional Ethnic Concentrations and Ethnic Discrimination in Hiring: Field Experimental Evidence from the Netherlands." *Journal of Ethnic and Migration Studies* 1–20. doi: 10.1080/1369183X.2021.1905505.

Thijssen, Lex, Frank van Tubergen, Marcel Coenders, Robert Hellpap, and Suzanne Jak. 2021. "Discrimination of Black and Muslim Minority Groups in Western Societies: Evidence From a Meta-Analysis of Field Experiments." *International Migration Review* 019791832110450. doi: 10.1177/01979183211045044.

Tomaskovic-Devey, Donald, Catherine Zimmer, Kevin Stainback, Corre Robinson, Tiffany Taylor, and Tricia McTague. 2006. "Documenting Desegregation: Segregation in American Workplaces by Race, Ethnicity, and Sex, 1966–2003." *American Sociological Review* 71(4): 565–88. https://doi.org/10.1177/000312240607100403.

Viechtbauer, W. 2010. "Conducting meta-analyses in R with the metafor package." Journal of Statistical Software 36(3): 1–48.

Wilson, William Julius. 2011. "The Declining Significance of Race: Revisited & Revised." *Daedalus* 140(2):55–69. doi: 10.1162/DAED_a_00077.

Zschirnt, Eva, and Didier Ruedin. 2016. "Ethnic Discrimination in Hiring Decisions: A Meta-Analysis of Correspondence Tests 1990–2015." Journal of Ethnic and Migration Studies 42(7):1115–34.

Supporting Information for

Trends in Hiring Discrimination against Racial and Ethnic Minorities in Six Western Countries

Lincoln Quillian* and John J. Lee

Lincoln Quillian
E-mail: l-quillian@northwestern.edu

Includes:

Supporting text: Appendices A to D
Figures S1 to S3
Tables S1 to S11
SI References

Supporting Information Text

Some parts of the procedures described in the appendices below reprise text from Quillian et al. (2017) and Quillian et al. (2019)

Appendix A. Study Search Methods

Our search for studies included the methods of bibliographic search, citation search, and an e-mail request of authors of field experiments.

Our bibliographic 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. To improve our coverage of non-English publications, we also searched two French-language indexes, Cairn.info and Persée; two international sources: IZA discussion papers, a German working paper archive; and ILO International Migration Papers. Finally, we conducted a search with Italian, Spanish, Portuguese and Dutch translations of the search terms and other terms frequently used in these languages to describe field experiments in hiring discrimination in Google Scholar. The search was first performed in March 2014 and repeated in August and September 2014 and in November 2015. We conducted searches in Italian, Spanish, Portuguese, and Dutch in November 2015 and February 2016.

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 audit studies to find additional field experiments of hiring discrimination.

The last technique employed was an e-mail request to authors of existing field experiments of discrimination. From our list of audit studies identified by bibliographic and citation searches, we compiled a list of e-mail addresses of authors of existing field experiments of discrimination. We added the e-mail addresses of several well-known experts on field experiments, notably authors of literature review articles on field experiments. Our e-mail request asked for citations or copies of experimental field studies of discrimination that were published,

unpublished, or ongoing. We also asked that the authors refer us to other researchers who may have recent or ongoing field experiments.

We conducted the e-mail requests in two phases. In the initial wave 131 apparently valid e-mail 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 e-mail addresses. We received 19 responses to this second wave of e-mail surveys.

Overall, our search located more than 100 studies that 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.) and who otherwise met our inclusion criterion.[1] Some of these studies included contrasts between more than one target group and whites (e.g. blacks and Hispanics), producing multiple estimates of discrimination against non-whites.

Finally, since this procedure was originally undertaken, we have added new studies to our sample based on a refreshed bibliographic and citation search in 2021 and reference searches of new studies located through bibliographic search.

Appendix B. Adjustment to Discrimination Ratios in Some Multi-Stage Studies

A few studies in our sample follow a multi-stage design in measuring discrimination. This was a study design used by some studies commissioned by the International Labor Organization. In these studies, applicants applied for advertised jobs in pairs, and the applicants first called employers by phone to inquire if a job was still available.

The complication we run into is the following: In five studies, if one applicant was told the job was available and the other was not, no application was submitted by *either* tester. The last aspect of this design – that when one applicant received a positive response and the other did not, the applicant who could have then submitted a resume did not – requires some adjustment. We want to capture callback rates for all minority and majority applicants from the point of initial application. We know that respondents who were told "no job is available" did not receive a callback. For situations where both applicants were told the job is still available or both were told it is not available, this is straightforward: we include these counts in calculating the rates of callbacks. However, when one pair member was told the job is available and the other was not, we do not know how often the pair member who was told the job was available would have received a callback if they had applied. We need to estimate this to get complete callback outcomes from the point of application.

To estimate callback rates in these studies, we assume that the member of the pair who received the invitation to interview but did not submit a resume (because their partner was told the job was no longer available) was as likely to get a callback if they had submitted a resume as applicants of the same race/ethnic group in the same study for which an application was submitted.

Appendix C. Variances Estimation for Effect Sizes

For studies that are unpaired or do not report paired outcomes, the variance of the logged discrimination ratio for the mth minority group in the ith study for callbacks is estimated by:

$$\sigma_{ij}^2 = Var\left(\ln\left(y_{ij}\right)\right) = \frac{1}{c_{ij}^w} - \frac{1}{n_{ij}^w} + \frac{1}{c_{ij}^m} - \frac{1}{n_{ij}^m}$$

[1] We excluded some studies where it was unclear if employers were making decisions producing discrepant outcomes because applications were made through an employment agency. We excluded a few other studies because they lacked basic information on counts of outcomes by target group and the authors could not be located or declined to provide these data when contacted.

This is Bornstein, Hedges, Higgins, and Rothstein's (2009) formula 5.3. 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 from Zhou (2007). If p^a is the number of pairs in which both majority and minority testers receive a callback, p^b is the number of pairs in which the majority tester received a callback but not the minority, p^c is the number of pairs in which the minority tester received a callback but not the majority, and p^d is the number of pairs in which neither tester received a callback, then the variance of the logged discrimination ratio for the jth minority group in the ith study with paired data is:

$$\sigma_{im}^2 = Var\left(\ln\left(y_{ij}\right)\right) = \frac{p_{ij}^b + p_{ij}^c}{\left(p_{ij}^a + p_{ij}^b\right)\left(p_{ij}^a + p_{ij}^c\right)}$$

Appendix D. Publication Bias

Publication bias results when studies that fail to find statistically significant effects are less likely to be published. If studies that find no statistically significant discrimination are less likely to be published, this will lead to overestimating discrimination in meta-analysis. See Borenstein et al. (2009), chapter 30.

Our primary interest is change over time. Publication bias which is constant over time will not affect our estimates in principle; instead, it may just produce a constant upward bias in discrimination estimates, leading to discrimination estimates that are too large but without biasing the trend. However, it could be that publication bias increased or decreased over time because of changes in the social or academic context.

We tried two methods to examine publication bias. First, we examined publication bias based on funnel plots and the trim-and-fill method (Duval and Tweedie 2000) with the pooled data broken into three periods: before 2000, 2001-2010, and 2011 and later. Funnel plots are shown in Supporting Information Figure S3. The trim-and-fill analysis found no evidence of publication bias in any period – funnel plots are fairly symmetric and no points were found to be "missing" in any of the three periods based on the trim-and-fill analysis (estimated with "metafor" with default options).

Second, we used the PEESE procedure. PEESE readily accommodates moderators such as our year variables and country and group controls (Stanley and Doucouliagos 2012). This method uses a weighted regression with the variance of the effect as a predictor to capture publication bias.

Estimates of PEESE models are shown in Supporting Information Table S11. Models 3-4 allow for interaction of study year and variance, as suggested by Stanley and Doucouliagos (2012). We find evidence of publication bias in some (but not all) models. Importantly, we never find that the slope of year is significantly changed by adjusting for publication bias, which is of primary significance for our analysis. Not do we find a significant interaction of variability and the slope of year in models including this interaction.

We note, however, that some simulation studies have found that the PEESE method works well when there is little residual variability but does not work well under "realistic" conditions of heterogeneity in population effects (see Alinaghi and Reed 2018, Hong 2019).

In sum, neither the trim-and-fill by decade nor the PEESE finds evidence that publication bias is likely to affect the trend over time.

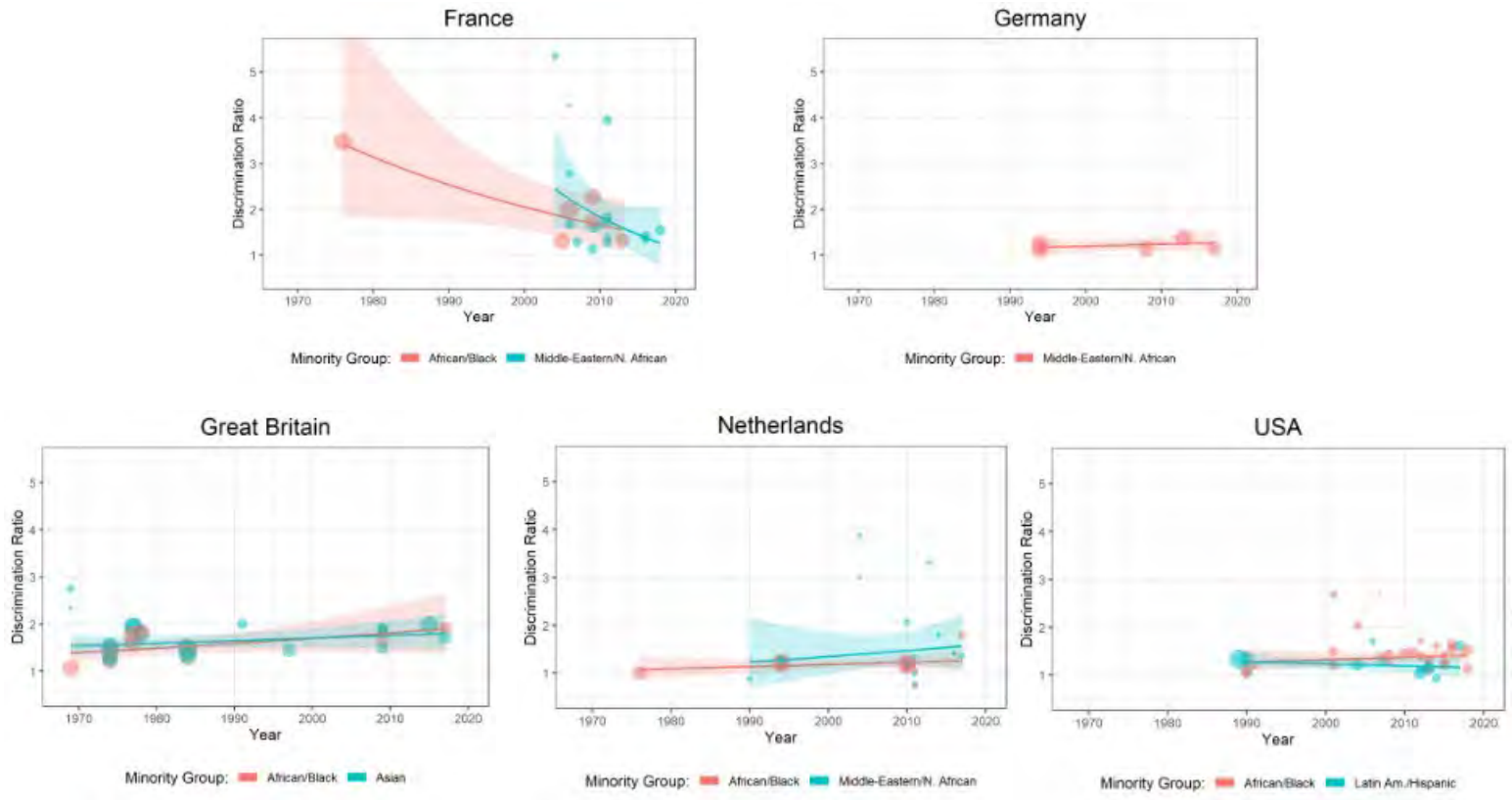


Fig. S1. Discrimination Ratios over Time by Country and Group

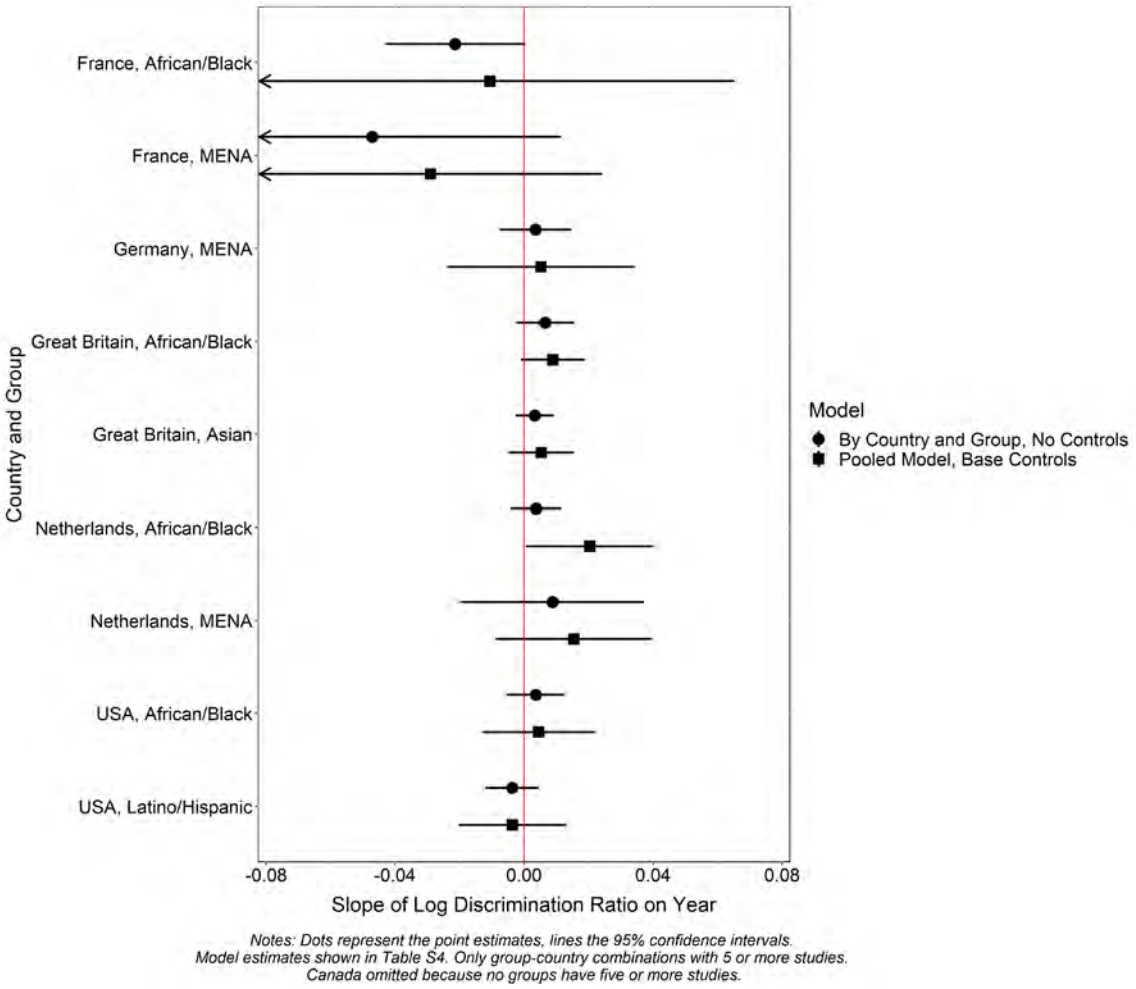
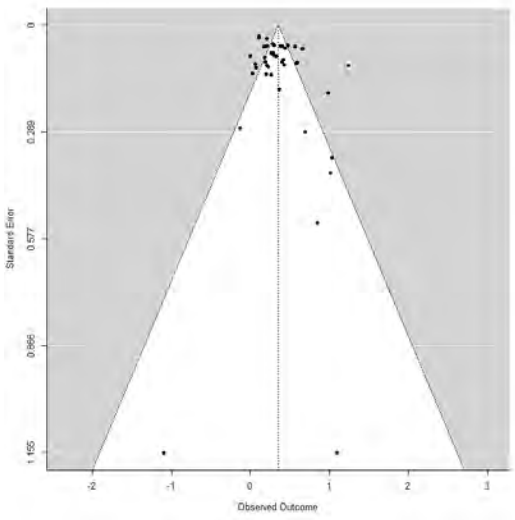
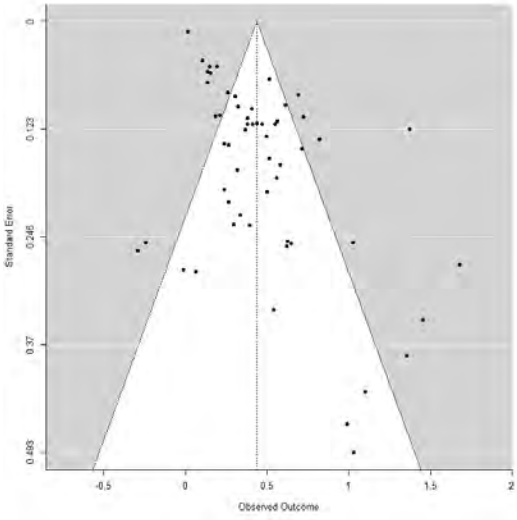


Fig. S2. Trends in Hiring Discrimination by Country and Group

Funnel plot, year 2000 and before:



Funnel Plot 2001 to 2010:



Funnel Plot 2011 and later:

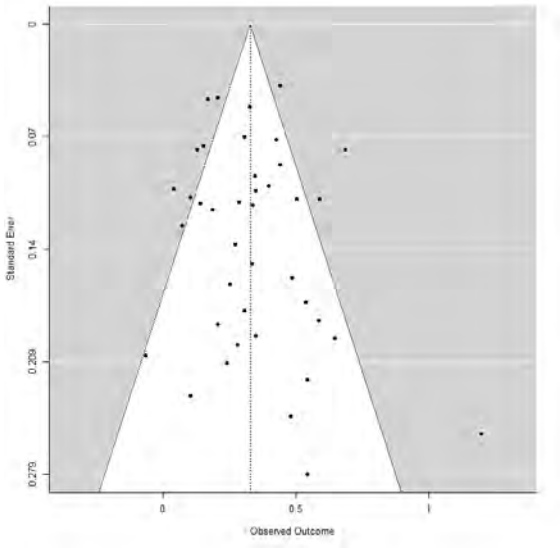


Fig. S3. Funnel Plots by Year Categories

Table S1: Pooled Meta-Regression Estimates of Log Discrimination Ratio on Country, Target Group, and Controls

Variable	Linear Trend Only (1)	Country + Group (2)	Base Controls (3)	Add Foreign Birth, Ed, Nationality (4)	Add Local UE Rate + % Foreign-Born (5)	Excluding single-occupation studies, base controls (6)	Base Controls, Resume Audits Only (7)
Year of Fieldwork	0.0001	0.0024	0.0037	0.0051	0.0087	0.0008	0.0061
(Four Digit Year)	(0.0018)	(0.0019)	(0.0043)	(0.0048)	(0.0062)	(0.0059)	(0.0047)
Dummy variable for country (6 countries)	No	Yes	Yes	Yes	Yes	Yes	Yes
Dummy variables for minority group (4 minority groups)	No	Yes	Yes	Yes	Yes	Yes	Yes
Basic Study Controls (see Table 1) (12 variables)	No	No	Yes	Yes	Yes	Yes	Yes
Controls for Foreign Educ. and Nationality (4 variables)	No	No	No	Yes	Yes	No	No
Controls for Unemployment and % Foreign Born ir	No	No	No	No	Yes	No	No
N effects / N studies	140/90	140/90	140/90	140/90	113/72	119/72	118 / 75
Tau-squared, USA	0.014	0.017	0.009	0.009	0.011	0.010	0.009
Tau-squared, Canada	0.031	0.042	0.140	0.153	0.199	0.126	0.066
Tau-squared, France	0.224	0.159	0.157	0.129	0.157	0.085	0.172
Tau-squared, Germany	0.019	0.004	0.020	0.024	0.002	0.017	0.030
Tau-squared, Great Britain	0.036	0.013	0.000	0.000	0.002	0.001	0.000
Tau-squared, Netherlands	0.072	0.080	0.145	0.143	0.058	0.178	0.165

Notes: +=p<.1; * = p<.05; ** = p<.01; *** = p<.001. Two-tailed tests. Standard error in parentheses. All models use the "subgroup correlated effects" model of Pustejovsky and Tipton (2022). Robust standard errors clustered at the study level.
Base controls are variables shown in table 1: study method dummy, tester gender (2 dummies), applicant education (3 dummies), applicant occupation (3 dummies), immigrant status, job source online (2 dummies).

Table S2: Pooled Meta-Regression Estimates of Log Discrimination Ratio on Country, Target Group, and Controls, Post-1984 Studies Only

Variable	Linear Trend Only (1)	Country + Group (2)	Base Controls (3)	Add Foreign Birth, Ed, Nationality (4)	Add Local UE Rate + % Foreign-Born (5)	Excluding single-occupation studies, base controls (6)	Base Controls, Resume Audits Only (7)
Year of Fieldwork	0.0032	0.0019	-0.0008	-0.0009	0.0099	0.0027	-0.0032
(Four Digit Year)	(0.0021)	(0.0025)	(0.0083)	(0.0086)	(0.0095)	(0.0073)	(0.0086)
Dummy variable for country (6 countries)	No	Yes	Yes	Yes	Yes	Yes	Yes
Dummy variables for minority group (4 minority groups)	No	Yes	Yes	Yes	Yes	Yes	Yes
Basic Study Controls (see Table 1) (12 variables)	No	No	Yes	Yes	Yes	Yes	Yes
Controls for Foreign Educ. and Nationality (4 variables)	No	No	No	Yes	Yes	No	No
Controls for Unemployment and % Foreign Born ir	No	No	No	No	Yes	No	No
N effects / N studies	119 / 81	119 / 81	119 / 81	119 / 81	94 / 65	100 / 64	100 / 67
Tau-squared, USA	0.011	0.017	0.009	0.019	0.014	0.016	0.006
Tau-squared, Canada	0.036	0.042	0.141	0.129	0.191	0.068	0.086
Tau-squared, France	0.185	0.131	0.139	0.132	0.160	0.000	0.140
Tau-squared, Germany	0.012	0.005	0.015	0.003	0.000	0.008	0.027
Tau-squared, Great Britain	0.058	0.000	0.000	0.000	0.000	0.000	0.000
Tau-squared, Netherlands	0.059	0.084	0.162	0.107	0.081	0.204	0.181

Notes: +=p<.1; * = p<.05; ** = p<.01; *** = p<.001. Two-tailed tests. Standard error in parentheses. All models use the "subgroup correlated effects" model of Pustejovsky and Tipton (2022). Robust standard errors clustered at the study level.

Basic study controls are variables shown in table 1: study method dummy, tester gender (2 dummies), applicant education (3 dummies), applicant occupation (3 dummies), immigrant status, job source online (2 dummies).

Table S3: Meta-Regression Estimates of the Trend in the Log Discrimination Ratio by Minority Group

	Models by Group			Pooled Model with Group-Specific Slopes	
Linear Trends by Group	Linear Trend Only (1)	Add Country Controls (2)	Add Base Controls (3)	Base Controls (4)	N effects / N studies
African/Black, Year of Fieldwork	-0.0012	0.0023	-0.0034	0.0082 +	57 / 51
(Four Digit Year)	(0.0029)	(0.0044)	(0.0048)	(0.0047)	
MENA, Year of Fieldwork	0.0052	-0.0037	0.0136	0.0095	41 / 39
(Four Digit Year)	(0.0073)	(0.0143)	(0.0136)	(0.0054)	
Latin/Hispanic, Year of Fieldwork	-0.0012	-0.0040	Insufficient	0.0000	12 / 12
(Four Digit Year)	(0.0043)	(0.0041)	Data	(0.0061)	
Asian, Year of Fieldwork	-0.0015	0.0008	Insufficient	0.0032	30 / 21
(Four Digit Year)	(0.0024)	(0.0027)	Data	(0.0043)	
Controls					
Dummy variables for Country	No	Yes	Yes	Yes	
Dummy for In-Person vs. Resume Aud	No	Yes	Yes	Yes	
Basic Study Controls	No	No	Yes	Yes	

Notes: Columns 1-3 based on a separate random-effects meta-regression estimated for each minority group. Column 4 based on a single model with country-specific year trends, base model controls, and country-specific Tau parameters. Tau parameters not shown.

+=p<.1; * = p<.05; ** = p<.01; *** = p<.001. Two-tailed tests. Standard error in parentheses. Robust standard errors clustered at the study level.

With the Bonjamini-Hochberg adjustment for multiple comparisons with four tests (see Materials & Methods) no tests are significant.

The base controls are twelve variables shown in Table 1.

Table S4: Meta-Regression Estimates of Decade Changes in Log Discrimination Ratio, MENA and Muslim-Origin Minority Groups

	Subsample: MENA			Subsample: Muslim-Origin Country		
Variable	Period Dummies Only (1)	+ Country (2)	+ Base Controls (3)	Period Dummies Only (4)	Country + Group (5)	+ Base Controls (6)
Period Dummies (Reference = 2000 or earlier)						
Year 2001 to 2010	0.3351 *	0.0510	0.9672 **	0.1927	0.0399	0.7783 **
	(0.1072)	(0.1113)	(0.2345)	(0.1125)	(0.0919)	(0.1395)
Year After 2010	0.2417 *	0.1146	0.8936 **	0.1333	0.1306 +	1.0677 ***
	(0.0773)	(0.0761)	(0.2361)	(0.0767)	(0.0579)	(0.1690)
Dummy variable for country (6 countries)	No	Yes	Yes	No	Yes	Yes
Dummy variables for minority group (4 minority groups)	No	Yes	Yes	No	Yes	Yes
Basic Study Controls (9 variables)	No	No	Yes	No	No	Yes
N effects / N studies	41 / 39	41 / 39	41 / 39	57 / 51	57 / 51	57 / 51
Tau-squared, USA	0.053	0.022	0.011	0.000	0.000	0.000
Tau-squared, Canada	0.126	0.107	0.628	0.076	0.064	0.328
Tau-squared, France	0.145	0.161	0.000	0.134	0.155	0.014
Tau-squared, Germany	0.040	0.005	0.033	0.042	0.003	0.010
Tau-squared, Great Britain	0.000	0.022	0.011	0.045	0.020	0.000
Tau-squared, Netherlands	0.085	0.091	0.083	0.082	0.091	0.138

Notes: +=p<.1; * = p<.05; ** = p<.01; *** = p<.001. Two-tailed tests. Standard error in parentheses. Robust standard errors clustered at the study level.

Model 1 to 3 include only Middle-East North Africa (MENA) minority groups. Model 4 to 6 include all ethnic minorities with origins in Muslim-majority countries.

Basic study controls are twelve variables shown in Table 1.

Table S5: Meta-Regression Estimates of Linear Trend in Log Discrimination Ratio by Country

Variable	Linear Trend Only, Models by Country (1)		Group Controls, Models by Country (2)		Pooled Model, Country-Specific Slopes, Base Controls (3)		N effects / N studies
Linear Trend by Country (Slope of Year of Fieldwork Variable)							
Canada	-0.0030		0.0007		-0.0095		14 / 7
	(0.0050)		(0.0046)		(0.0170)		
France	-0.0228	***, †	-0.0260	**, †	-0.0195	+, ns	23 / 20
	(0.0044)		(0.0069)		(0.0055)		
Germany	0.0023		0.0035		0.0072		8 /6
	(0.0032)		(0.0052)		(0.0119)		
Great Britain	0.0046	*, †	0.0044	+, ns	0.0042		30 / 12
	(0.0018)		(0.0021)		(0.0042)		
Netherlands	0.0081	+, ns	0.0063		0.0188	*, ns	25 / 15
	(0.0045)		(0.0059)		(0.0067)		
USA	0.0028		0.0016		0.0017		40 / 30
	(0.0031)		(0.0039)		(0.0075)		
Trend Dropping Early Study for France							
France post-1985	-0.0344		-0.0401		-0.0197		22 / 19
	(0.0217)		(0.0237)		(0.0240)		
Model Controls							
Dummy variables for minority group	No		Yes		Yes		
Dummy for In-Person vs. Resume Audit	No		Yes		Yes		
Base Controls	No		No		Yes		

Notes: Models 1 and 2 based on separate random-effects meta-regression for each country. Model 1 only includes a linear year predictor, model 2 adds controls for minority group. Model 3 is a single pooled model with country-specific slopes, base controls, and country-specific tau parameters.

+=p<.1; * = p<.05; ** = p<.01; *** = p<.001. Two-tailed tests, no multiple-comparison adjustment. Standard error in parentheses. Robust standard errors clustered at the study level.

† = Benjamini-Hochberg multiple comparison adjusted test significant with 6 tests, ns = not significant after adjustment.

Table S6: Meta-Regression Estimates of Linear Trend in Log Discrimination Ratio by Country and Group

	(1)		(2)		
Country and Group	Trend from Meta-Regressions by Country and Group		Country-Group Specific Trend, Pooled Model, Base Controls		N effects / N studies
USA, African/Black	0.0036		0.0045		27 / 27
	(0.0044)		(0.0084)		
USA, Latin Am./ Hispanic	-0.0037		-0.0036		11 / 11
	(0.0036)		(0.0079)		
France, African/Black	-0.0213	+	-0.0105		6 / 6
	(0.0078)		(0.0116)		
France, MENA	-0.0471		-0.0290		16 / 16
	(0.0273)		(0.0232)		
Germany, MENA	0.0035		0.0053		6 / 6
	(0.0040)		(0.0125)		
Great Britain, African/Black	0.0066		0.0090	+	11 / 11
	(0.0040)		(0.0045)		
Great Britain, Asian	0.0033		0.0053		18 / 18
	(0.0028)		(0.0048)		
Netherlands, African/Black	0.0038		0.0204	*	8 / 8
	(0.0032)		(0.0076)		
Netherlands, MENA	0.0088		0.0155		15 / 15
	(0.0132)		(0.0109)		
<u>Trend Dropping Early Study for France</u>					
France, African/Black, Post-1985	-0.0123		0.0236		5 / 5
	(0.0056)		(0.0628)		

Notes: +=p<.1; * = p<.05; ** = p<.01; *** = p<.001. Two-tailed tests. Standard error in parentheses. Robust standard errors clustered at the study level.

In model 1, each year coefficient is estimated in a separate meta-regression with year of fieldwork as the only predictor. Model 2 is subgroup correlated effect model including base controls.

With the Bonjamini-Hochberg adjustment for multiple comparisons with 10 tests, (see Materials & Methods) no tests are significant.

Table S7: Descriptive Statistics, Predictor Variables

Base Model Variables

Annual Trend	Mean	Std. Dev.	N (effects / studies)
Year of Fieldwork	2002.6	14.0	140 / 88

Study Method	Effects	Studies
Resume Audit / Correspondence	118	73
In-person Audit	22	15

Tester Gender	Effects	Studies
Testers Male Only	43	37
Testers Female Only	9	9
Testers Both Male and Female	84	43

Applicant Education (most common level)	Effects	Studies
High School or Less	55	37
Some College or Post-HS Vocational Degree	34	23
College or More	30	20
Education information missing	21	8

Occupation Controls (all that apply)	Effects	Studies
Includes Blue Collar Jobs (1=yes)	69	39
Includes Jobs with Customer Contact (1=yes)	100	62
Includes Jobs with an Office Focus (1=yes)	104	62

Job Source	Effects	Studies
Online	72	44
Offline	50	32
Both Online and Offline	18	12

Minority Group Includes Foreign-Born Persons?	Effects	Studies
Native-Born Only	97	70
Includes Foreign-Born	43	20

Additional Controls (used in some models in Table 2)

Minority Nationality	Effects	Studies
Minority Applicants Citizens of Host Nation	134	84
Minority Applicants Not Citizens of Host Nation	4	3
Minority Applicants Mix of Citizens/Noncitizens	2	2

Minority Highest Education Credential Foreign?	Effects	Studies
Domestic Highest Education Credential	134	84
Foreign Highest Education Credential	4	3
Highest Credential Mix of Foreign/Domestic	2	2

Contextual Controls	Mean	Std. Dev.	N (effects / studies)
Unemployment Rate of Local City/Region	7.0%	2.5%	115 / 74
Percentage Immigrants in Local City/Region	13.3%	10.9%	115 / 73

Notes: Effects are distinct estimates of discrimination against minority groups. Some studies include discriminaton estimates against multiple minority groups.

Table S8: Specific Minority Groups in Field Experiments by Country

Country	Minority Groups with Effect Sizes (Study Term)
Canada	African, Arab, Black, Chinese,Greek,Indian, Indo-Pakistani, Latino,Middle Eastern, West Indian
France	African, Antillean, Asian, Franco-North African, Moroccan, North African, Senegalese, Subsaharan African, Vietnamese
Germany	MENA, Turkish, Southeast Asian, Sub-Saharan African
Great Britain	African, Asian (South Asian), Black African, Black Caribbean, Chinese, Cypriot, Greek, Indian, Pakistani, Pakistani/Bangladeshi, West Indian
Netherlands	Antillean, Arab, Black Surinamer, Hindustani, Moroccan, Spanish, Surinamese, Turkish
US	African American, Arab American, Asian, Black, Hispanic, Latino, Somali
Total	

Table S9: Pooled Meta-Regression of Log Discrimination Ratio on Year, Non-Linearity Analysis

Variable	Period Dummies Only 1	Country + Group 2	Base Controls 3	Squared Year Predictor 4
Period Dummy Variables (Reference=Before 1991)				
Year 1991 to 2000 (1=yes)	0.0218	-0.0089	-0.0436	
	(0.0656)	(0.0810)	(0.0797)	
Year 2001 to 2010 (1=yes)	0.1765 *	0.1316	0.2233	
	(0.0677)	(0.0823)	(0.1273)	
Year After 2010 (1=yes)	0.1014 +	0.0731	0.2232	
	(0.0530)	(0.0731)	(0.1402)	
Year as Continuous Variable with Squared Term				
Year (Year 2000=0)				0.2233
				(0.1273)
Year Squared (Year 2000=0)				0.2232
				(0.1402)
Dummy variable for country (6 countries)	No	Yes	Yes	Yes
Dummy variables for minority group (4 minority groups)	No	Yes	Yes	Yes
Basic Study Controls (12 variables)	No	No	Yes	Yes
N effects / N studies	140/90	140/90	140/90	140/90
Tau-squared, USA	0.0077	0.0158	0.0097	0.0079
Tau-squared, Canada	0.0700	0.0596	0.1387	0.1380
Tau-squared, France	0.1624	0.1588	0.1541	0.1674
Tau-squared, Germany	0.0100	0.0096	0.0160	0.0228
Tau-squared, Great Britain	0.0122	0.0131	0.0000	0.0000
Tau-squared, Netherlands	0.0611	0.0727	0.1219	0.1488

Notes: +=p<.1; * = p<.05; ** = p<.01; *** = p<.001. Two-tailed tests. Standard error in parentheses. Standard errors clustered at the study level.
Basic study controls are shown in table 1.

Table S10: Random-Effects Meta-Regression Models with Odds-Ratio Outcome, Pooled Models with Base Controls

Variables in Model	Single Trend	Country-Specific Trends		Group-Specific Linear Trends		Muslim-Origin Country Trends		
Linear Trend Coef(s)	-0.0089	Canada	-0.0383 +	African/Black	-0.0034	Year 2000 or Earlier	(ref.)	
(Slope of Year)	(0.0077)		(0.0110)		(0.0082)			
		France	-0.0304 +	MENA	-0.0012	Year 2001 to 2011	1.1066	**
			(0.0077)		(0.0101)		(0.3047)	
		Germany	0.0053	Latin/Hispanic	-0.0150	Year 2012 or Later	1.0605	**
			(0.0070)		(0.0093)		(0.2976)	
		Great Britain	-0.0125	Asian, Year	-0.0108			
			(0.0080)		(0.0083)			
		Netherlands	0.0180 *					
			(0.0063)					
		USA	-0.0052					
			(0.0049)					
Control Variables								
Dummy variables for country (6 countries)	Yes		Yes		Yes		Yes	
Dummy variables for minority group (4 minority groups)	Yes		Yes		Yes		Yes	
Basic Study Controls (12 variables, see table 1)	Yes		Yes		Yes		Yes	
Variance Structure								
Country-Specific Tau's	Yes		Yes		Yes		Yes	
N effects / N studies	140 / 90		140 / 90		140 / 90		41 / 39	

Notes: +=p<.1; * = p<.05; ** = p<.01; *** = p<.001. Two-tailed tests. Standard error in parentheses. Standard errors clustered at the study level. The models are pooled subgroup correlated effects models.
Basic study controls are variables shown in table 1: study method dummy, tester gender (2 dummies), applicant education (3 dummies), applicant

Table S11: Publication Bias (PEESE) Estimates of Pooled Models

Variable	Period Dummies Only (1)	+ Base Controls (2)	Period Dummies Only (3)	+ Base Controls (4)
Year of Fieldwork (Four Digit Year)	-0.0047	0.0070	-0.0066	0.0055
	(0.0041)	(0.0040)	(0.0047)	(0.0043)
Estimated Variance of Effect	4.3055	2.2296 **	5.7208	2.9894 *
	(1.8687)	(0.6456)	(2.4717)	(0.9439)
Year * Variance			0.2704	0.1107
			(0.1480)	(0.0618)
Dummy variable for country (6 countries)	No	Yes	No	Yes
Dummy variables for minority group (4 minority groups)	No	Yes	No	Yes
Basic Study Controls (9 variables)	No	Yes	No	Yes
N effects / N studies	140 / 90	140 / 90	140 / 90	140 / 90

Notes: +=p<.1; * = p<.05; ** = p<.01; *** = p<.001. Two-tailed tests. Standard error in parentheses. Robust standard errors clustered at the study level.

Basic study controls are dummies for year and minority group plus twelve variables shown in Table 1.

SI References

Alinaghi, Nazila, W. Robert Reed. 2018. "Meta-analysis and publication bias: How well does the FAT-PET-PEESE procedure work?" Research Synthesis Methods 9: 285-311. DOI: 10.1002/jrsm.1298

Borenstein, Michael, Larry V. Hedges, Julian P. T. Higgins, and Hannah R. Rothstein. 2009. Introduction to Meta-Analysis. West Sussex, UK: John Wiley & Sons.

Duval S, Tweedie R. 2000. A nonparametric “trim and fill” method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association* 95:89–98.

Hong, Sanghyun. 2019. “Meta-analysis and publication bias: How well does the FAT-PET-PEESE procedure work? A replication study of Alinaghi & Reed (Research Synthesis Methods,2018).” International Journal for Re-Views in Empirical Economics, Vol 3(2019-4). DOI: 10.18718/81781.13.

Quillian, Lincoln, Anthony Heath, Devah Pager, Arnfinn Midtbøen, Fenella Fleischmann, Ole Hexel. 2019. "Do Some Countries Discriminate More than Others? Evidence from 97 Field Experiments of Racial Discrimination in Hiring." *Sociological Science* 6: 467-496.

Quillian, Lincoln, Devah Pager, Ole Hexel, and Arnfinn H. Midtbøen. 2017. “Meta-Analysis of Field Experiments Shows No Change in Racial Discrimination in Hiring over Time.” *Proceedings of the National Academy of Sciences* 114(41):10870–10875.

Stanley, T. D., and Hristos Doucouliagos. 2012. *Meta-Regression Analysis in Economics and Business*. 1st edition. New York: Routledge.

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.