

Closed doors everywhere? A meta-analysis of field experiments on ethnic discrimination in rental housing markets

Katrin Auspurg^a, Andreas Schneck^a and Thomas Hinz^b

^aDepartment of Sociology, LMU Munich, Munich, Germany; ^bDepartment of History and Sociology, University of Konstanz, Konstanz, Germany

ABSTRACT

Discrimination is long seen as a meaningful factor for ethnic inequalities on rental housing markets. Yet empirically, the extent of discrimination is still debatable. For the first time, this article provides a quantitative meta-analysis of field experiments (in person audits and correspondence tests) that were run over the last four decades in the United States, Canada and Europe ($N = 71$). Special focus is given to a possible inflation of effect sizes by publication bias; to time trends; and to evidence for statistical discrimination. Taken together, nearly all experiments document the occurrence of ethnic discrimination. Effect sizes are inflated by publication bias, but there is still substantial evidence left once the bias is removed. The analysis reveals a consistent decline in the extent of discrimination over time, from moderate levels of discrimination in the 1970s and 1980s, up to only small but still statistically significant levels in the 1990s and 2000s. A significant part of the discriminatory behaviour can be attributed to missing information about the social status of applicants, which supports theories on statistical discrimination. It is discussed how future research could move our knowledge on the underlying mechanisms forward.

ARTICLE HISTORY

Received 9 March 2018
Accepted 4 June 2018


KEYWORDS

Ethnic discrimination; rental housing market; meta-analysis; field experiments; statistical discrimination

Introduction

Discrimination in housing markets has long been seen as a key problem for the social integration of ethnic or racial minorities (see e.g. Pager and Shepherd 2008; Yinger 1998). Housing discrimination raises the costs for housing, which harms minorities' wealth accumulation. Furthermore, it has been shown to fuel residential segregation, which is associated with poor health outcomes and educational and employment disadvantages for ethnic minorities (Galster 1992, 1996; Pager and Shepherd 2008; Turner et al. 2002), and also known to strengthen group conflicts, stereotypes or prejudicial attitudes (Galster 1992; Quillian 2014).

CONTACT Katrin Auspurg  katrin.auspurg@lmu.de

 Supplemental data for this article can be accessed at <https://doi.org/10.1080/1369183X.2018.1489223>.

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

While these and other negative consequences are well known, we still have too little understanding of the prevalence and conditions (Riach and Rich 2002). Did the occurrence of ethnic discrimination decline during the last decades? Is there still a substantial discrimination left once a possible publication bias is removed, that might exist in form of an overrepresentation of strong findings in literature? Does the amount of discrimination vary with the amount of information applicants disclose on their social status? It is the purpose of this paper to address these and other research questions in regard to the huge number of field experiments on ethnic discrimination in rental housing markets. To summarise the large international literature, we use a quantitative meta-analytical methodology for the first time. Our meta-analysis summarises 71 experimental trials (audit and correspondence studies) that were published in 46 studies on ethnic discrimination in rental housing markets. It includes evidence from experiments run in 10 different Western countries during 42 years (1973–2015), covering a large number of different ethnicities.

Going back to the 1970s, a large number of field experiments evidenced ethnic discrimination on rental housing markets. However, the reported effect sizes vary extremely, and consequently there is an ongoing debate whether discrimination is much worth worrying about (for this debate see e.g. Galster 1990; Heckman 1998; Thernstrom and Thernstrom 1997; Yinger 1998; Zhao, Ondrich, and Yinger 2006). Resolving this issue at the current stage of research seems also impossible as far as it is completely unknown to what extent some of the effect sizes reported in literature might have been exaggerated by a publication bias. This bias exists when in particular strong findings (in line with scholars' research hypotheses) are more likely published than other findings (Dickersin 2006). Given that part of the field experiments were conducted by advocacy groups with the main purpose of documenting discrimination, the risk of such publication bias seems high. This even more so, as several studies recently suggested that also much of the scientific experimental research can be exacerbated by a publication bias (see e.g. Open Science Collaboration 2015). We are, however, so far not aware of any study testing for publication bias in the field of housing discrimination.

Aside from a lack of research on the overall level of discrimination, few studies have systematically analysed if there were changes over time (the few exceptional studies all focus on the U.S., and all used the same database: Galster 1992; Massey 2005; Pager and Shepherd 2008; Turner et al. 2013). In particular, studies for the last two decades are lacking. It is therefore unclear to what extent intensified anti-foreign sentiments in reaction to terrorist events (Legewie 2013; Ross and Galster 2005) or increasing inflows of immigrants in European countries might have increased discrimination during the last decades; and it is also unresolved if there is still a downward trend in discrimination as it was observed in prior periods in the U.S.

At the current stage of research, it is also difficult to conclude on the kind of discrimination. Some studies found evidence for 'statistical discrimination' in that sense that in particular minority applicants who disclose only few (positive) information on their social status are discriminated (e.g. Bosch, Carnero, and Farré 2010). But the evidence is not consistent (see e.g. Ahmed, Andersson, and Hammarstedt 2010), and it is therefore still debatable to what extent animus or statistical discrimination might underlie the findings.

In our meta-analysis, special focus will be given on all three issues: The amount of discrimination besides a possible publication bias; changes in the level of discrimination over time; and evidence for statistical discrimination. There exist some narrative reviews on the

field-experimental literature (e.g. Bertrand and Duflo 2017; Galster 1990; Oh and Yinger 2015; Pager and Shepherd 2008; Rich 2014). However, by virtue of the huge number of available experiments with a large heterogeneity in designs, any narrative report will have difficulties to condense the existing evidence in a satisfactory way (for similar arguments on the labour market: Weichselbaumer and Winter-Ebmer 2005). Our quantitative meta-analysis allows to combine the different experiments within one coherent framework, and also to quantify how the level of discrimination varies with study characteristics, such as tested ethnicities or points in time.

Background and literature

There is rich literature for nearly all Western countries that ethnic minorities are disadvantaged on rental housing markets. Ethnic minorities have shown to live in relatively smaller apartments, to pay higher rents per square metre, and also to live more likely in segregated poor-quality neighbourhoods (for literature on Europe see e.g. Drever and Clark 2002; Harrison, Law, and Philipps 2005; for the U.S. e.g. Cort 2011; Krivo 1995).

But, of course, there are other reasons for these inequalities beyond discrimination.¹ Part of the inequalities might reflect preferences (for ethnic segregation) or differences in resources that constrain individuals' choices (such as household income). In the 'residual approach' all unexplained inequalities in observational data are seen as evidence for discrimination. There is, however, always a risk that these remaining differences are simply caused by further unmeasured disparities in individuals' preferences or resources (Galster 1992).

The most compelling evidence for discrimination is therefore probably provided by field experiments. In prior decades, researchers mainly used in-person audits, where test persons with different ethnicities (but otherwise mostly similar characteristics) applied to the same housing vacancies.² With apartments becoming increasingly advertised on the internet, audit studies have been more and more replaced by correspondence studies, where standardised, written (e-mail) applications are sent out with typically only the names of applicants signalling their ethnicity. The vast majority of these studies detected discrimination: Ethnic minority test persons were found to be less likely invited to showings than majority group testers (for overviews: Galster 1990; Pager and Shepherd 2008; Riach and Rich 2002; Rich 2014). However, sometimes no meaningful ethnicity effects were found, and in general, there is a large variance in effect sizes. In the following, we shortly discuss three possible reasons for cross-study variance: publication bias; variation over time; and varying amounts of information on applicants' social status.

Are effect sizes exacerbated by publication bias?

A publication bias (also known as the 'file drawer effect') exists when statistically significant or 'spectacularly' strong findings are more likely submitted or published than other findings (Dickersin 2006). Many field experiments are based on only few observations, and in particular these small-scale studies are prone to publication bias. This is because results in small studies are in general attached with a large sampling error, and accordingly their findings show a huge variance in effect sizes (which is reflected in large standard errors (SEs) and confidence intervals). Due to this large random variation, some of these studies will find extraordinary large effects just by chance. If only these large findings

end up being published, the literature will overstate the evidence for discrimination. Recent literature suggests that much experimental work is suffering from such bias (see for this ‘replication crisis’ e.g. Open Science Collaboration 2015). However, to best of our knowledge, there was so far not any concern on this bias in the literature on housing discrimination. This is probably for the fact that the bias cannot be spotted on the ground of individual studies alone (Auspurg and Hinz 2011).

Have there been changes over time?

Few scholars have systematically analysed if there were changes over time. Exceptions exist with few studies on the U.S. that were all based on the large-scale testing studies of the U.S. Department of Housing and Urban Development (HUD) with data from 1977, 1989, 2000 and 2012. These studies found a steady decline of discrimination of Black compared to White Americans since the 1990s, with in 2012 nearly all discrimination being gone (Massey 2005; Pager and Shepherd 2008; Turner et al. 2013). If there was already a decline in the 1980s is more debatable (Galster 1992), and there were also some exceptions to the decline later on: for instance, discrimination of Hispanics in rental housing markets was not found to decline (Ross and Turner 2005).

For Europe, similar research on time trends is lacking. In many European countries legislations explicitly prohibiting housing discrimination came into force in the second half of the 2000 decade (while in the U.S. they have been implemented already in the 1960s; Harrison, Law, and Philipps 2005; Oh and Yinger 2015, 36). Yet besides tighter anti-discrimination laws, both in the U.S. and in Europe, discrimination might have even intensified during the last two decades. Plausible reasons exist with increasing anti-foreigner sentiments, whether in reaction to terrorist events (such as 9/11) in the U.S. or an increasing inflow of immigrants in European countries (Disha, Cavendish, and King 2011; Legewie 2013; McLaren 2003).³ Besides one single comparison of the 2012 and 2000 HUD testing, we are not aware of any longitudinal research covering the last two decades that is going beyond a pure narrative survey (for the latter: Rich 2014).

Animus or statistical discrimination?

Some studies have sought to use field-experimental techniques not only to document discrimination, but also to explore underlying mechanisms, seeking especially to distinguish between taste-based (animus) and statistical discrimination (Rich 2014). Theories on taste-based discrimination assume that gatekeepers discriminate because contact with a disliked minority group causes them a psychological disutility (e.g. Becker 1957). In contrast to this, in the perspective of statistical discrimination, discriminatory practices are motivated by attempts to increase economic profits (e.g. Arrow 1973; Phelps 1972). Actors are supposed to solve problems of missing information. It is for instance difficult to predict the extent to which possible renters will provide regular rental payments. Landlords and agencies might therefore use the ethnicity of apartment seekers as a proxy for these unknown characteristics, since individuals of many ethnic minorities are (statistically) known to have a lower or more unstable household income than typical members of the majority population.

Meanwhile, there were some experiments that tried to separate both strands of theories by testing if providing more (positive) information on applicants' social background lowers discrimination. Empirical findings are mixed, with some, but not all studies finding evidence for this assumption supporting statistical discrimination (Rich 2014). Again, the huge diversity in used research designs (such as tested ethnicities or countries) makes it difficult to come up with a clear conclusion.

Data and methods

We conducted a meta-analysis on field experiments (in-person audits and correspondence studies) on ethnic discrimination in the housing market. We only focused on the rental market, since there are only very few field experiments on real estate markets in Europe, and screening procedures likely differ across both markets.⁴

Data collection

We tried to locate all field experiments that were available up to December 2015, irrespective of whether they were published or not. Studies qualified for our study if they reported all relevant statistics: First, the effect sizes (i.e. discrimination rates or regression coefficients), and, second, their SEs (or statistics that could be used to calculate them). Relevant studies were identified by an extensive keyword search on several literature databases (Google Scholar, Web of Knowledge, Sociological Abstracts, Wiso).⁵ We also checked the reference lists of the literature reviews for further experiments.

We coded the abovementioned statistics, along with a large number of characteristics of the studies (for an overview of key variables see Table A1 in the Online Supplement: <http://dx.doi.org/10.6084/m9.figshare.5966671>). The collection of data was undertaken by three research assistants and supervised by a principal investigator to double-check all coding. When a study compared more than one minority to the majority population (e.g. both Arabs and Blacks were compared to White Americans), we treated the results on each minority as a different trial (i.e. in the above example, we would code two different trials).⁶ Raw studies often also reported several estimates for one minority group, for example, when they split results across different cities or estimated different regression models. In that case we collected all different estimates.

Using these inclusion criteria, we were able to trace 710 estimates from 71 trials reported in 46 studies done between 1973 and 2015.⁷ In-person audit studies contributed 56% of all trials (with these being overrepresented in the first two periods; see again Table A1 in the Online Supplement). Around one-third of experimental trials tested Black, Arabs/Muslim applicants or other ethnicities. Other ethnicities were mostly Hispanics, East-Europeans or Asians. Due to the low number of cases and these groups showing similar discrimination levels we combined them within one category.⁸ Around half of the experiments were done in the U.S. or Europe. To estimate a possible time trend, three different periods were defined that all included about the same number of trials, and also covered at least about one decade (1970–1990; 1991–2007; and 2008–2014). Eleven per cent of the trials varied the amount of (positive) information on socio-demographic characteristics of applicants to test for statistical discrimination. We collected for these trials information on the discrimination level for both, the low- and the high-

information condition (for details: see our Online Supplement). There were only two (nine) studies that focused on private (commercial) suppliers only, and studies that included both market segments frequently did not provide discrimination rates separately for the two different types of suppliers.⁹ Thus, we throughout report results pooling private and commercial landlords.

Meta-analytical procedure

In meta-analysis, the estimates of different studies (here the observed levels of discrimination) are pooled to estimate a mean overall effect, and to explain effect size variations across studies (for a general introduction: Borenstein et al. 2009). All outcome measures reported in the raw studies have to be converted to one joint metric. Following recent recommendations (Aloe 2014; Stanley and Doucouliagos 2012), we converted all effect sizes to partial correlations (r_{pc}). Due to their standardised range $[-1; 1]$, these correlations allow a straightforward interpretation. Negative coefficients would indicate that the minority group was disadvantaged compared to the majority group, while positive coefficients would point to a discrimination against the majority group. The more closely the absolute value to the maximum of one, the higher the level of discrimination.

To allow additional interpretations, some effects were converted to response differences that measure the differences in the proportion of (positive) responses to the majority versus minority group (Borenstein et al. 2009, 48). To calculate these differences, the response rate for the majority group had to be fixed to one specific value. We always imputed the mean of the response rates observed in raw experiments that reported this information. This mean response rate was 40.1%.

For all raw experiments also SEs for the partial correlations were calculated (for details see again the Online Supplement). This allows to give more informative studies with a higher amount of measurement precision a relatively higher weight in the meta-analysis. In addition, SEs can be used to identify publication bias. If there is publication bias, effect sizes tend to be larger in studies with low measurement precision (high SEs).

To explore such bias, we begin with visual tools. In *funnel plots* the partial correlations are plotted against their measurement precision (inverse SEs). To bypass the subjective interpretation of such graphs (Tang and Liu 2000), also linear regression analyses were employed (Egger et al. 1997; Stanley 2008). We selected the linear weighted-least-square (WLS) regression approach, as this method was shown in simulation studies to be superior to alternative regression models (Stanley 2008; Stanley and Doucouliagos 2015; Stanley and Jarrell 1989). At the first stage of this two-step regression approach, the outcome (in our case: the partial correlations coefficient) is regressed on its' SE to spot possible publication bias (that would exist with a positive significant association). In case there is evidence for a publication bias, and at the same time also for a true discrimination effect (which would be indicated by a significant regression intercept), in a second step it is tried to explain substantial variance in effect sizes by including predictor variables (such as the tested ethnicity or time periods). At this second stage, possible publication bias is now removed by controlling for the variance (squared SE) of partial correlations, as this was proved in simulation studies to be the most reliable bias correction in meta-regressions (formulas are provided in the Online Supplement).¹⁰

In cases where researchers reported several estimates for single trials, we calculated mean estimates that probably provide the best summary measure for the observed level of discrimination.¹¹ But we also did robustness tests based on all raw estimates. In addition, the within-trial variance in effect sizes was used to explore statistical discrimination for trials that varied the amount of information on applicants. We therefore used multilevel regressions with fixed effects for the level of experimental trials. These fixed effects regressions (for a general introduction e.g. Brüderl and Ludwig 2015) allowed us to explore exactly what we are interested in: Does a higher amount of (positive) information on applicants' social status – all else of the experimental trial being equal – lead to a lower level of discrimination?

Often, several estimates in our meta-analysis originated from the same study or trial. In that case, the observations are no longer independent from each other. We adjusted for this violation of assumptions underlying standard regression analyses by estimating cluster-robust SEs (Rogers 1994). We decided to cluster on the level of studies when using only one (mean) estimate per experimental trial, and otherwise on the level of experimental trials.

Results

Overall discrimination level and publication bias

Nearly all experimental trials ($N = 69$; 97%) included in our meta-analysis resulted in a mean negative partial correlation, meaning that they evidenced discrimination of the minority group. Only two trials led to a small, positive (but not significant) effect of a minority status on receiving a (positive) response when applying for a vacancy.¹² The other extreme exists with a strong negative correlation ($-.446$), found in a small audit study in France with Black testers in 1977. When summarising the evidence of all 71 experimental trials (always using one mean estimate per trial), the mean effect size amounted to a partial correlation of $-.148$ (SE: $.013$), pointing to an overall low, but significant level of discrimination. When weighting studies by their measurement precision (inverse of the variance), the overall mean reduces to $-.097$ (SE: $.015$). This is a first indication of publication bias: More precise (larger) studies report lower levels of discrimination.

Further evidence of such bias is provided by the funnel plot in [Figure 1](#). In this figure, the effect sizes are plotted against the inverse of their SE. Experimental trials with lower measurement precision (lower values of $1/SE$) are placed at the bottom and show – as one would expect – a higher variation in effect sizes. This leads to the typical shape of an inverted funnel. In the absence of publication bias, the funnel should be symmetrically arranged around the mean effect size, which is depicted in [Figure 1](#) by the dotted line (showing the weighted mean: $-.097$). However, the funnel is markedly skewed to the left, as also proved by a statistically significant correlation of effect sizes with their measurement precision ($r = .353$; $p = .003$).

The measurement precision of studies might be confounded with study years, since researchers conducted very large experimental trials mostly only during the last recent years. We therefore also employed multivariate regression techniques. In Model 1 in [Table 1](#) the partial correlations are regressed on the study year and on their SEs to test

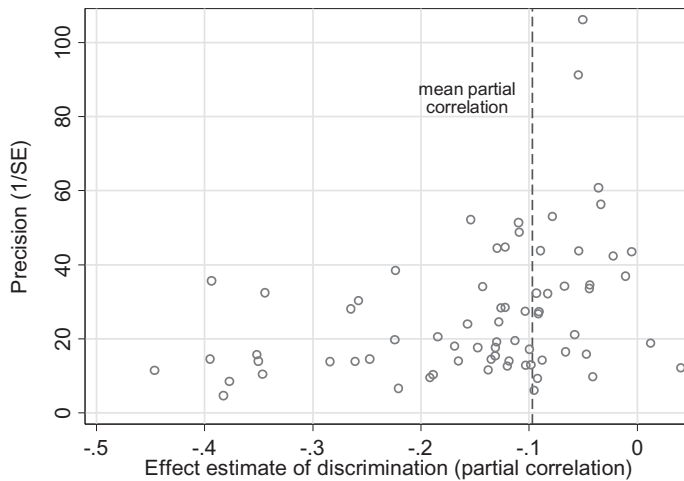


Figure 1. Funnel plot of discrimination rates against their measurement precision.

Notes: In this figure the effect size of discrimination (partial correlation) is plotted against the inverse of its standard error. More (less) precise studies are at the top (bottom) of the figure. The dashed line shows the weighted mean (−.097) of the partial correlation. The figure is significantly skewed to the left: $r = .353$; $p = .003$.

for publication bias. The statistically significant negative coefficient found for the SEs supports the interpretation of the funnel plot: Particularly small and imprecise studies provide strong evidence of discrimination. But also the intercept is significant, indicating that there is still a genuine effect besides the bias. As indicated in the method section, for those constellations it was shown that an even more consistent estimate of the genuine effect is obtained by a regression that controls for publication bias by including the variance (i.e. SE squared) instead of the SE. Results of this approach are shown in Model 2. The genuine mean effect for the earliest period (the reference category) is indicated by the

Table 1. Linear WLS regressions of partial correlations to test and correct for publication bias (regression coefficients; in parentheses: cluster-robust standard errors).

	Model 1 WLS testing for bias	Model 2 WLS with bias correction
SE ^a	−1.223* (0.508)	
SE ^a squared		−7.135 (3.623)
Year (Ref.: 1970–1990) ^b		
1991–2007	0.060 (0.040)	0.071 (0.042)
2008–2014	0.106*** (0.024)	0.122*** (0.019)
Intercept	−0.152*** (0.029)	−0.190*** (0.017)
N Trials	71	71
R ²	0.317	0.283

Notes: The regressions are weighted by the inverse variance of the partial correlation.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

^aStandard error of partial correlation.

^bYear of experiment (if reported, otherwise imputed by average publication delay).

intercept, and amounts to a statistically highly significant partial correlation of $-.190$ ($p < .001$). The discrimination level estimated for the last period (2008–2014) is substantially smaller, but still reaches statistical significance ($r_{pc} = .067$, $p < .001$).¹³

For illustrative purpose, we converted these measures to the response rate differences. By assuming a response rate of 40.1% for a majority applicant, in the latest experimental period (2008–2014) a minority applicant on average would have a 6.6 percentage points (ppts) lower response probability than its majority counterpart. In contrast, the difference was much more pronounced in the previous years, with 11.7 ppts (1991–2007) and even 18.9 ppts in 1970–1990. This strong response difference in the period 1970–1990 corresponds to response rates for the minority applicant of only 21.2%.

In sum, these first analyses suggest that evidence for discrimination is exacerbated by publication bias, but when removing the bias, a significant discrimination level remains. Although these regressions already point to time trends, more reliable estimates are provided in the next section where we added further controls.

Table 2. Linear WLS regressions of partial correlations to explain substantial variation across studies (regression coefficients; in parentheses: cluster-robust standard errors).

	Model 1	Model 2	Model 3a only North America	Model 3b only Europe	Model 4
Ethnicity (Ref.: Blacks)					
Arabs/Muslims		−0.046* (0.020)	−0.065 (0.044)	−0.025 (0.016)	−0.068*** (0.019)
Others ^a		0.037 (0.019)	0.035 (0.029)	0.045** (0.015)	0.025 (0.016)
Country (Ref.: U.S.)					
Canada	0.103 (0.057)				0.112* (0.050)
Germany	−0.095* (0.044)				−0.040 (0.043)
Great Britain	−0.009 (0.038)				−0.006 (0.032)
South Europe ^b	0.005 (0.031)				0.025 (0.024)
North Europe ^c	0.007 (0.039)				0.067 (0.036)
Year (Ref.: 1970–1990) ^d					
1991–2007	0.098** (0.032)	0.084** (0.025)	0.060 (0.035)	0.142* (0.060)	0.079** (0.027)
2008–2014	0.238*** (0.047)	0.210*** (0.036)	0.205*** (0.050)	0.199** (0.057)	0.215*** (0.041)
SE ^e squared	−6.908** (2.449)	−6.522** (2.034)	−2.152 (2.900)	−5.142 (5.761)	−6.696** (2.318)
Intercept	−0.340*** (0.068)	−0.313*** (0.051)	−0.289** (0.080)	−0.317*** (0.062)	−0.301*** (0.059)
N Trials	71	71	40	31	71
R ²	0.606	0.635	0.700	0.566	0.705

Notes: The regressions are weighted by the inverse variance of the partial correlation.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

^aMostly Hispanics, East-Europeans and Asians.

^bItaly, France, Spain, Greece.

^cSweden, Norway.

^dYear of experiment (if reported, otherwise imputed by average publication delay).

^eStandard error of partial correlation. Further controls: in person audit study (yes/no); within-experiment with both minority and majority person applying to the same apartment (yes/no); article published in journal (yes/no).

Time trends, effects of countries and ethnicities

Table 2 shows the results when additionally controlling for the ethnicities of testers and the countries where the experiments were run.¹⁴ All models in addition control for the experimental method (e.g. audit versus correspondence test; see the table notes for the controls). Due to strong collinearities, we show models that each include only the country or ethnicity, and in addition a full model (Model 4). To see if effects of ethnicity vary across regions, we also provide separate analyses for North America (Canada and the U.S., Model 3a) and Europe (Model 3b).

What can be learned from these regressions? First, in all models there is a time trend of declining discrimination (see the positive coefficients observed for later periods). Particularly the last period (2008–2014) shows a remarkable drop in the level of discrimination. In Europe, a statistically significant decline is also observed for the mid-period (1991–2007; see Model 3b). This strong, steady and early decline is remarkable, given that in many European countries anti-discrimination directives became effective not before the late 2000s. Also notable, both in the U.S. and Europe the drop in discrimination seemed to be in particular strong for the last period (2008–2014), despite anti-immigration attitudes were observed to increase in many Western countries at the same period (see e.g. Ogan et al. 2014).¹⁵

Second, there was some variation across ethnicities. Arabs/Muslims were found to be slightly more targeted by discrimination than Blacks, which represent the reference group. ‘Others’ (mainly Hispanics, Eastern Europeans or Asians) were throughout least affected by discrimination (see the positive effect for this group; with it, however, only reaching statistical significance in Model 3b for Europe).

Third, there were also some differences across countries. Germany ($N = 9$ trials) showed the strongest level of discrimination. This effect lost statistical significance once ethnicity was controlled in Model 4 (most experiments in Germany tested Muslims signalled by Turkish names). In Canada, the level of discrimination was the lowest, and this result also holds when controlling for ethnicity. Note, however, that there were so far no more than five trials in Canada (embedded in three studies covering Blacks, Muslims and other ethnicities). Additional research is highly warranted to see if these results are reliable, and if yes, what are the specific institutions or context conditions that lower discrimination in Canada. What is also noteworthy: All in all, results for European countries and the U.S. tended to be very similar. This can be seen even more clearly from Figure A2 in the Online Supplement, where we plotted the findings for North American and European countries against each other. Confidence intervals for both regions remarkably overlap.

Finally, also in the most recent period, there was a statistically significant level of discrimination left in both North America ($r_{pc} = -.08$; $p < .1$) and Europe ($r_{pc} = -.12$; $p < .01$). Converting the estimates of the full model (Model 4 in Table 2) to the difference in response rates (again imputing a 40.1% response rate for the majority group) gave the following results: The response rate to a minority applicant was in the last period 2008–2014 on average 8.4 ppts lower than that to a majority applicant. Given these estimates, in this period a minority rental apartment seeker would need on average nearly six applications [6.3] to get two replies, while for majority applicants there would be on average five applications needed.¹⁶ For the group mostly targeted by discrimination, Arab/Muslim

applicants, the response rate difference was 15.3 ppts. In Canada, the country with the lowest discrimination rate overall, in the last period the difference in response rates nearly vanished (difference of 2.6 ppts).

Evidence for statistical discrimination

To assess the evidence for statistical discrimination, we switch to regressions with fixed effects for the trial level, where the outcome represents all estimates per trial (and not only one mean estimate). This allows us to observe if – all characteristics of single trials held constant – disclosing more information on applicants significantly lowers the level of discrimination. In addition, fixed effects regressions (with fixed effects for studies) also provide higher confidence on the causal effects of ethnicities, as they allow to level out possible confounders in form of stable study characteristics.¹⁷

Table 3 shows the results. In Model 1, the analyses are restricted to the 14 studies ($N = 305$ estimates, $N = 35$ trials) that tested several minority groups. Again, there is a statistically negative effect for Arabs/Muslims. In particular, this finding based on a fixed effects approach ensures a great confidence that Arabs/Muslims are even slightly more strongly targeted by discrimination than Blacks.

Model 2 provides fixed effects regressions for the impact of information (using only the 9 trials that varied the amount of information; $N = 212$ estimates). One can clearly see that providing more information on applicants substantially reduces the level of discrimination (by about one-third). Converted to response rate differences, the effect sizes are as follows: In the low-information condition the response rate difference is 9.6 ppts (again imputing a 40.1% response rate for the majority group). Providing more information reduces this difference to 6.6 ppts. However, also in high-information conditions,

Table 3. Linear fixed effect regressions of partial correlations using all estimates per study or trial (regression coefficients; in parentheses: cluster-robust standard errors).

	Model 1 studies varying ethnicity	Model 3 trials varying information
Ethnicity (Ref.: Blacks)		
Arabs/Muslims	-0.030** (0.010)	
Others ^a	0.018* (0.008)	
High-information condition (yes)		0.030*** (0.007)
SE ^b squared	-1.643 (4.521)	-16.286 (8.730)
Intercept	-0.085*** (0.007)	-0.098*** (0.006)
N Estimates	305	212
N Studies	14	6
N trials	35	9
R^2	0.432	0.278

Notes: The regressions are weighted by the inverse variance of the partial correlation.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

^aMostly Hispanics, East-Europeans and Asians.

^bStandard error of partial correlation.

some discrimination still remains (overall: estimated mean partial correlation of $r_{pc} = -.067$; $p < .001$).

To sum up: There is strong evidence for statistical discrimination, but this evidence is so far grounded on only few trials that tested this mechanism. Although in high-information conditions discrimination was still present, which could, but does not have to be firm evidence for animus discrimination. We will get back on this in our discussion.

Robustness checks

In our main analyses, we used only one mean value for trials that reported several estimates. A threat to the validity of our findings exists when this mean value does not represent a good estimate for the discrimination evidenced in these trials. To ensure that our results are robust, we repeated all analyses with (1) using all estimates per trial (and clustering SEs on the study or trial level); and (2) by means of bootstrap regressions with 1000 replications, always sampling only one estimate per trial. In regressions with all estimates, studies were weighted by their measurement precision (inverse of variance, as used in our main analyses), and in further robustness checks additionally by the number of estimates per trial (see our Online Supplement). Our main results were found to be remarkably stable, with in particular the bootstrap specification providing very similar results.¹⁸

We also tested the robustness in regard to alternative specifications of time periods. In a metric specification, the year of study was found to be highly collinear with the used experimental method (audit or correspondence study; $r = -.589$). We therefore tested this metric specification only in sparse models without further controls. Main results (publication bias; true effect besides publication bias; decline of discrimination over time) were found to be stable. We also tried alternative categorical specifications, for example, grouping the study years to four or five categories based on quantiles. In terms of avoiding collinearities with other predictor variables (such as countries), the three-level specification presented in the main analyses turned out to be superior. But note that all main results even hold when using the alternative year specifications.

Finally, we also checked if some experiments might have been too obtrusive. With increasing numbers of experiments being reported (in newspapers and magazines), landlords might have become more and more suspicious of being targeted by those experiments, which might have led them to exert discrimination at a later stage of the selection process. In our case this would mean that we underestimated the overall level of discrimination, while at the same time we overestimated its decline over time. To assess such possible bias, we compared two different experimental designs: in some studies, only one inquiry per housing ad was sent (between-design), while in most studies matched pairs of inquiries with applications from both majority and minority test persons were sent (within-design). In particular in the later variant, the experimental manipulation might have been detected by landlords (Weichselbaumer 2015). In our meta-analytical data, however, overall the discrimination rates measured by within- and between-experiments did not differ significantly from each other after controlling for other experimental conditions such as countries.¹⁹

Summary

We summarise the results in five points. First, the raw experiments clearly document the widespread occurrence of discrimination in all researched countries so far, with only a very small number of experiments failing to provide any evidence of discrimination against ethnic minority groups.

Second, there was a statistically significant negative association between effect sizes and study precision, which can probably be best explained by a publication bias in form of predominantly strong positive results (randomly found in small studies) getting published. However, also when controlling for this bias by means of meta-analytic tools, a meaningful level of discrimination was left.

Third, the experiments showed a strong and steady decline of discrimination during the last four decades. While there were moderate levels of discrimination in the 1970s and 1980s (mean partial correlation of $-.19$), the level of discrimination observed for the period 1991–2007 had already dropped by about one-third, and for the last period (2008–2014) there was a further sharp decline of about half (resulting in a mean partial correlation of $-.07$).

Fourth, there were some differences across minority groups and countries. Arab/Muslim applicants were consistently found to be slightly more affected by discrimination than Blacks, with other ethnicities (that mainly consisted of Hispanics, Eastern Europeans and Asians) suffering from the lowest level of discrimination. Regarding countries, only Canada differed statistically significantly from other countries. The level of discrimination detected there was relatively low, but only based on few observations (five trials, testing different ethnicities). All in all, variations across ethnicities and countries tended to be small.

Finally, analysing the subgroup of the nine trials that varied the amount of (positive) information on applicants' socio-economic status revealed evidence for statistical discrimination: Providing more information on applicants reduced the level of discrimination on average by about one-third.

Discussion

Summarising the evidence of 71 field experiments that were run during the last four decades in European and North American countries, our meta-analysis all in all suggests that studying ethnic discrimination is still worth worrying about: in all decades and in all 10 Western countries under study a consistent level of ethnic discrimination on rental housing markets was found. Since the places where individuals live, and the quality of their housing are associated with numerous dimensions of social inequality, studying the conditions that fuel (or prevent) discrimination seems still important. In this last section, we discuss how results of our meta-analyses point to particularly fruitful directions for future research.

First, experiments varying the amount of information on applicants' social status suggested that a substantial part of discrimination is statistical. Interpreting the remaining part of discrimination is more difficult. It might be evidence for other forms of discrimination (animus discrimination), or it might be that even in high-information conditions still information for the transaction is missing. At the current stage of research conclusions

are difficult. More experiments that focus on underlying mechanisms, for example, by not only varying the amount of information, but also considering institutional differences (such as housing regulations that allow more or less security deposits to cover possible rental debts) would be highly warranted (for a reasoning on the effects of institutional regulations see e.g. Wasmer 2005).

Future research might also focus more explicitly on different market segments. Private and commercial providers probably differ in their incentives to discriminate. For instance, only private landlords decide on a long-term relationship with their possible tenants, and for private landlords (who typically have other sources of income besides rental payments) economic profits might be also less important compared to agencies who compete on a commercial market. Both aspects suggest that private landlords have more incentives and possibilities to follow their tastes for discrimination. So far, there are however not enough studies to allow for any reliable conclusions.

Further research that rests on a broader spectrum of well-chosen treatment variables might also help to understand why some groups are in particular targeted by discrimination. Our meta-analysis revealed that these are groups that are characterised by a relatively high salience of their minority status (Blacks), and/or relative large cultural distance to the majority population (Arabs/Muslims). This pattern speaks to animus discrimination, but it might also relate to statistical discrimination. Blacks and Arabs/Muslims are affected by strong resentments and ethnic prejudices in Western countries (see e.g. Ogan et al. 2014). At the same time, for this and other reasons, they are also especially disadvantaged on the labour market, which could stimulate statistical discrimination. In particular, for the joint group of Arabs/Muslims, our data did not allow for the level of fine-grained analyses that are needed to disentangle theoretical mechanisms. The broad category certainly captures diverse mechanisms such as discrimination motivated by an assumed lower economic background, a supposed lack of social integration regarding socio-political (democratic) attitudes, and/or resentments against the typical religious or cultural identity of test persons with this migration background. The same problem already applies to the raw studies since experiments on the housing market throughout only used names that signalled both at once, a Muslim and Arabic background. Further studies with innovative designs that allow for a better separation of group identities would be highly warranted (e.g. by including subtle information on the religious identity; for an example on the labour market see Valfort 2018).

In regard to country differences, it is remarkable that only Canada stands out with a somewhat lower level of discrimination. All in all, results were found to be surprisingly similar across countries, with the downward trend in discrimination also holding for geographical regions that recently experienced strong immigration or terrorist threats. One might explore in future research if the downward trend is due to effective interventions (such as tighter anti-discrimination legislations), or due to attitudes not necessarily translating into discriminating behaviour. So far there is a lack of studies that link experimental evidence with evidence on gatekeepers' attitudes (for rare examples on the labour market: Carlsson and Rooth 2012; Rooth 2010).

Regarding countries, it is also notable that Germany shows similar (or even slightly higher) levels of discrimination than other countries observed in our meta-analysis. In a recent meta-analysis on the labour market, Zschirnt and Ruedin (2016) found the extent of discrimination to be fairly *smaller* in German-speaking countries, which they

interpreted as evidence for statistical discrimination. The more detailed and standardised application procedures on the labour market (including e.g. extensive reference letters from former employers) might have lowered the level of discrimination in German-speaking countries. In contrast, applications on the German housing market are less standardised and typically kept very short.²⁰ Taken together, the results for both markets provide even further evidence for statistical discrimination: Only the more standardised, extensive application procedure in the German labour market might keep discrimination there at a low level. To allow for firmer conclusions, additional cross-country studies or those that compare different markets are highly warranted.

With only 71 different experiments included in our meta-analysis, it was hardly possible to explore interactions between different moderator variables. This is also because there is typically a strong association of countries and ethnic minorities. In line with the representations of minority groups in the different countries, experiments in North America typically focus on Blacks, while researchers in Europe are more likely concerned about Arabic or Turkish minorities. Providing a higher overlap in experimental settings, and harmonising methodological details of studies, would allow researchers to come up with more firm conclusions on cross-country differences.

Another suggestion for the future agenda would be to better prevent publication bias. If only the most dramatic findings get published, the overall picture is too pessimistic, and one risks misplacing effort in a problem that is overstated.²¹ To prevent such bias in future, one might think about pre-registrations of all field-experimental studies, and also about basing approvals of experiments by ethics committees on reasonable study sizes and researchers' obligation to publish *all* findings, never minding if they are positive or negative, statistically significant or not (for similar suggestions, see e.g. Franco, Malhotra, and Simonovits 2014).

One also has to bear in mind that the field experiments in our meta-analysis could only observe the very first step of the screening process – the invitation to a showing. There might be additional discrimination going on at later stages. This, however, rather points to a still meaningful preoccupation with discrimination to fight housing inequalities: although showing small effect sizes, several unequal treatments might add up to an overall substantial level of discrimination.

We therefore conclude by encouraging research that allows for a better understanding of the reasons underlying discrimination: How, and under which conditions is it possible that discrimination occurs (for similar suggestions: Neumark 2016; Zschirnt and Ruedin 2016). To move our knowledge in that regard forward, one might go beyond just detecting some kind of discrimination, for example, by using further experimental variations that more precisely relate to theories, or by combining the great strength of field experiments for uncovering causal relationships with more in-depth information on the institutional settings of studied markets, and also by explicitly starting cross-country research.

Notes

1. Discrimination refers to an 'unequal treatment of persons or groups on the basis of their race or ethnicity' (Pager and Shepherd 2008, 182). In the housing market, ethnic discrimination exists if some individuals are denied access to housing (in some locations), or if they have to pay relatively higher prices to get the same housing quality as other individuals, with this being true only because of their race or ethnicity.

2. For instance, at the beginning of the 1970s, audit-like studies were conducted in New York City to check for evidence of racial discrimination in the renting out of apartments by the Trump Management Company (<https://vault.fbi.gov/trump-management-company> last visited 7 March 2018). In later years, similar audits were done for scientific purpose.
3. For the U.S. there is also some reasoning but no empirical analysis if institutional changes in recent decades, such as budget constraints for organisations trying to enforce fair housing acts, might have thwarted the former downward trend in discrimination (Ross and Galster 2005).
4. In the rental market, applicants have to be screened for a long-term relationship, while on the real estate market the transaction typically includes only a few interactions between former and future owners.
5. We used combinations of the following keywords: ‘field experiment’ [or ‘audit’, ‘correspondence’], ‘ethnic discrimination’ and ‘housing market’.
6. We separated trials along the following four ethnicities: Blacks, Arabs/Muslims, Hispanics, and ‘Others’. Ethnic groups pooled in the ‘Others’ category showed too low numbers of cases to allow for reliable estimates when kept separately.
7. The data and all replication files are available at: <https://doi.org/10.6084/m9.figshare.5966671>.
8. For the same reason we pooled Arabs (in raw studies often also referred to as ‘North-Africans’) and Muslims within one joint category. There is a discussion whether Muslims in Western Europe face similar barriers than Blacks in the United States (Foner 2015).
9. A possible reason is that it is often difficult to identify the kind of supplier with the information provided in housing ads only. In particular, small commercial suppliers are often indiscernible from private landlords.
10. The employed WLS-approach was also found to be superior in terms of bias correction than classical fixed or random effects meta-regressions as they are frequently used in psychology (Stanley and Doucouliagos 2015).
11. Similarly, we calculated the mean standard errors in cases where there were several estimates. An alternative would have been to pick only the estimate considered by the authors of raw studies as being the ‘best’ one (for such a procedure see e.g. Doucouliagos and Stanley 2009). In many studies, however, the authors do not flag estimates in such a way, and this procedure might also lead to an overestimation of discrimination when authors only highlight their strongest findings.
12. The two exceptional trials were done in 2007 in Canada with Jewish testers (Hogan and Berry 2011), and in 2009 in France with North-African testers (Bonnet et al. 2016).
13. This partial correlation is the sum of the intercept and effect for the last period in Model 2. The p -value was estimated by a regression model with modified reference category (results not shown).
14. Some countries had to be pooled to clusters to have at least five observations in each category.
15. The trend of falling discrimination was also found to hold for Arabs/Muslims (there was no significant interaction between this ethnicity and time periods; results not shown, but available on request). Note, however, that there were so far only few observations for this group.
16. Calculated as follows: $2/[0.401-0.086]$, respectively, $2/0.401$.
17. For this reason, it no longer makes sense (or even is possible) to include controls for countries or other factors that are stable within single trials or studies. For details, see again our Online Supplement.
18. There was only one notable difference across the many different model specifications: When simply using all estimates (which gives studies reporting more estimates a, respectively, higher weight), there was no longer a statistically significant effect for the last time period (2008–2014). However, also this model specification pointed to a steady downward trend in discrimination.
19. This was true although the within-experiments reported throughout slightly higher discrimination rates. In our regression tables, we therefore decided to report the more conservative estimates for between-experiments, which represent our reference category.

20. In the first round of an application process, rental apartment seekers often simply ask for a showing, revealing only contact information. This is at least the default on many internet platforms for housing ads in Germany.
21. Our first explorative analyses on this issue showed that in particular small audit studies (that were often conducted by advocacy groups) are prone to such bias (analyses not shown, but available on request).

Acknowledgement

Johannes Laufer, Helene Rickert and Laila Schmitt helped us in compiling our database. For helpful suggestions, the authors thank two anonymous reviewers, participants of the RC33 workshop at the ISA Forum of Sociology 2016, participants of the Rational Choice Sociology Seminar at Venice International University in 2016 and the MAER-Net colloquium in 2017. This paper was written while Katrin Auspurg was a visiting researcher at the Berlin Social Science Center (WZB).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by a small non-profit foundation in Germany, the Wolfgang and Anita Bürkle foundation.

References

- Ahmed, Ali M., Lina Andersson, and Mats Hammarstedt. 2010. "Can Discrimination in the Housing Market be Reduced by Increasing the Information About the Applicants?" *Land Economics* 86 (1): 79–90. doi:10.3368/le.86.1.79.
- Aloe, Ariel M. 2014. "An Empirical Investigation of Partial Effect Sizes in Meta-analysis of Correlational Data." *The Journal of General Psychology* 141 (1): 47–64. doi:10.1080/00221309.2013.853021.
- Arrow, Kenneth J. 1973. "The Theory of Discrimination." In *Discrimination in Labor Markets*, edited by Orley Ashenfelter and Albert Rees, 3–33. Princeton, NJ: University Press.
- Auspurg, Katrin, and Thomas Hinz. 2011. "What Fuels Publication Bias?" *Jahrbücher für Nationalökonomie und Statistik* 231 (5–6): 636–660. doi:10.1515/jbnst-2011-5-607.
- Becker, Gary S. 1957. *The Economics of Discrimination*. 2nd ed. Chicago, IL: University of Chicago Press.
- Bertrand, Marianne, and Esther Duflo. 2017. "Field Experiments on Discrimination." In *Handbook of Economic Field Experiments*, edited by Abhijit Vinayak Banerjee and Esther Duflo, 309–393. Amsterdam: North-Holland.
- Bonnet, Francois, Etienne Lalé, Mirna Safi, and Etienne Wasmer. 2016. "Better Residential Than Ethnic Discrimination! Reconciling Audit and Interview Findings in the Parisian Housing Market." *Urban Studies* 53 (13): 2815–2833. doi:10.1177/0042098015596107.
- Borenstein, Michael, Larry V. Hedges, Julian P. T. Higgins, and Hannah R. Rothstein. 2009. *Introduction to Meta-analysis*. Chichester: John Wiley & Sons.
- Bosch, Mariano, Angeles Carnero, and Lidia Farré. 2010. "Information and Discrimination in the Rental Housing Market: Evidence from a Field Experiment." *Regional Science and Urban Economics* 40 (1): 11–19. doi:10.1016/j.regsciurbeco.2009.11.001.
- Brüderl, Josef, and Volker Ludwig. 2015. "Fixed-effects Panel Regression." In *The Sage Handbook of Regression Analysis and Causal Inference*, edited by Henning Best and Christof Wolf, 327–358. Los Angeles, CA: Sage.

- Carlsson, Magnus, and Dan-Olof Rooth. 2012. "Revealing Taste-based Discrimination in Hiring: A Correspondence Testing Experiment with Geographic Variation." *Applied Economics Letters* 19 (18): 1861–1864. doi:10.1080/13504851.2012.667537.
- Cort, David A. 2011. "Reexamining the Ethnic Hierarchy of Locational Attainment: Evidence from Los Angeles." *Social Science Research* 40 (6): 1521–1533. doi:10.1016/j.ssresearch.2011.07.003.
- Dickersin, Kay. 2006. "Publication Bias: Recognizing the Problem, Understanding its Origins and Scope, and Preventing Harm." In *Publication Bias in Meta-analysis*, edited by Hannah R. Rothstein, Alexander J. Sutton, Michael Borenstein, and Kay Dickersin, 9–33. Chichester: John Wiley & Sons.
- Disha, Ilir, James C. Cavendish, and Ryan D. King. 2011. "Historical Events and Spaces of Hate: Hate Crimes against Arabs and Muslims in Post-9/11 America." *Social Problems* 58 (1): 21–46. doi:10.1525/sp.2011.58.1.21.
- Doucouliaqos, Hristos, and Tom D. Stanley. 2009. "Publication Selection Bias in Minimum-wage Research? A Meta-regression Analysis." *British Journal of Industrial Relations* 47 (2): 406–428. doi:10.1111/j.1467-8543.2009.00723.x.
- Drever, Anita I., and William A.V. Clark. 2002. "Gaining Access to Housing in Germany: The Foreign-minority Experience." *Urban Studies* 39 (13): 2439–2453. doi:10.1080/0042098022000027040.
- Egger, Matthias, George D. Smith, Martin Schneider, and Christoph Minder. 1997. "Bias in Meta-analysis Detected by a Simple, Graphical Test." *British Medical Journal* 315 (7109): 629–634. doi:10.1136/bmj.315.7109.629.
- Foner, Nancy. 2015. "Is Islam in Western Europe Like Race in the United States?" *Sociological Forum* 30 (4): 885–899. doi:10.1111/sof.12199.
- Franco, Annie, Neil Malhotra, and Gabor Simonovits. 2014. "Publication Bias in the Social Sciences: Unlocking the File Drawer." *Science* 345 (6203): 1502–1505. doi:10.1126/science.1255484.
- Galster, George C. 1990. "Racial Discrimination in Housing Markets during the 1980s: A Review of the Audit Evidence." *Journal of Planning Education and Research* 9: 165–175. doi:10.1177/0739456X9000900302.
- Galster, George C. 1992. "Research on Discrimination in Housing and Mortgage Markets: Assessment and Future Directions." *Housing Policy Debate* 3 (2): 637–683. doi:10.1080/10511482.1992.9521105.
- Galster, George C. 1996. "Future Directions in Mortgage Discrimination Research and Enforcement." In *Mortgage Lending, Racial Discrimination, and Federal Policy*, edited by John M. Goering and Ron Wienk, 697–716. Washington, DC: The Urban Institute Press.
- Harrison, Malcolm, Ian Law, and Deborah Philipps. 2005. *Migrants, Minorities and Housing: Exclusion, Discrimination and Anti-discrimination in 15 Member States of the European Union*. Vienna: European Monitoring Centre on Racism and Xenophobia (EUMC).
- Heckman, James J. 1998. "Detecting Discrimination." *Journal of Economic Perspectives* 12 (2): 101–116. doi:10.1257/jep.12.2.101.
- Hogan, Bernie, and Brent Berry. 2011. "Racial and Ethnic Biases in Rental Housing: An Audit Study of Online Apartment Listings." *City and Community* 10 (4): 351–372. doi:10.1111/j.1540-6040.2011.01376.x.
- Krivo, Lauren J. 1995. "Immigrant Characteristics and Hispanic-Anglo Housing Inequality." *Demography* 32 (4): 599–615. doi:10.2307/2061677.
- Legewie, Joscha. 2013. "Terrorist Events and Attitudes toward Immigrants: A Natural Experiment." *American Journal of Sociology* 118 (5): 1199–1245. doi:10.1086/669605.
- Massey, Douglas S. 2005. "Racial Discrimination in Housing: A Moving Target." *Social Problems* 52 (2): 148–151. doi:10.1525/sp.2005.52.2.148.
- McLaren, Lauren M. 2003. "Anti-immigrant Prejudice in Europe: Contact, Threat Perception, and Preferences for the Exclusion of Migrants." *Social Forces* 81 (3): 909–936. doi:10.1353/sof.2003.0038.
- Neumark, David. 2016. *Experimental Research on Labor Market Discrimination*, NBER Working Paper Series No. 22022. Cambridge, MA: National Bureau of Economic Research.

- Ogan, Christine, Lars Willnat, Rosemary Pennington, and Manaf Bashir. 2014. "The Rise of Anti-Muslim Prejudice: Media and Islamophobia in Europe and the United States." *International Communication Gazette* 76 (1): 27–46. doi:10.1177/1748048513504048.
- Oh, Sun Jung, and John Yinger. 2015. "What Have We Learned from Paired Testing in Housing Markets?" *Cityscape* 17 (3): 15–60. doi:10.1142/9789813206670_0030.
- Open Science Collaboration. 2015. "Estimating the Reproducibility of Psychological Science." *Science* 349 (6251). doi:10.1126/science.aac4716.
- Pager, Devah, and Hana Shepherd. 2008. "The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets." *Annual Review of Sociology* 34: 181. doi:10.1146/annurev.soc.33.040406.131740.
- Phelps, Edmund S. 1972. "The Statistical Theory of Racism and Sexism." *The American Economic Review* 62 (4): 659–661. doi:10.2307/1806107.
- Quillian, Lincoln. 2014. "Does Segregation Create Winners and Losers? Residential Segregation and Inequality in Educational Attainment." *Social Problems* 61 (3): 402–426. doi:10.1525/sp.2014.12193.
- Riach, P. A., and J. Rich. 2002. "Field Experiments of Discrimination in the Market Place." *Economic Journal* 112 (483): F480–F518. doi:10.1111/1468-0297.00080.
- Rich, Judith. 2014. *What Do Field Experiments of Discrimination in Markets Tell Us? A Meta Analysis of Studies Conducted since 2000*, IZA Discussion Paper No. 8584. Bonn: Institute for the Study of Labor (IZA).
- Rogers, William. 1994. "Regression Standard Errors in Clustered Samples." *Stata Technical Bulletin* 3 (13): 19–23.
- Rooth, Dan-Olof. 2010. "Automatic Associations and Discrimination in Hiring: Real World Evidence." *Labour Economics* 17 (3): 523–534. doi:10.1016/j.labeco.2009.04.005.
- Ross, Stephen, and George Galster. 2005. *Fair Housing Enforcement and Changes in Discrimination between 1989 and 2000: An Exploratory Study*. Working Paper of the University of Connecticut 2005-16. University of Connecticut, Department of Economics.
- Ross, Stephen, and Margery Austin Turner. 2005. "Housing Discrimination in Metropolitan America: Explaining Changes between 1989 and 2000." *Social Problems* 52 (2): 152–180. doi:10.1525/sp.2005.52.2.152.
- Stanley, Tom D. 2008. "Meta-regression Methods for Detecting and Estimating Empirical Effects in the Presence of Publication Selection." *Oxford Bulletin of Economics and Statistics* 70 (1): 103–127. doi:10.1111/j.1468-0084.2007.00487.x.
- Stanley, Tom D., and Hristos Doucouliagos. 2012. *Meta-regression Analysis in Economics and Business*. London: Routledge.
- Stanley, Tom D., and Hristos Doucouliagos. 2015. "Neither Fixed Nor Random: Weighted Least Squares Meta-analysis." *Statistics in Medicine* 34 (13): 2116–2127. doi:10.1002/sim.6481.
- Stanley, Tom D., and Stephen B. Jarrell. 1989. "Meta-regression Analysis: A Quantitative Method of Literature Surveys." *Journal of Economic Surveys* 3 (2): 161–170. doi:10.1111/j.1467-6419.1989.tb00064.x.
- Tang, J. L., and J. L. Y. Liu. 2000. "Misleading Funnel Plot for Detection of Bias in Meta-Analysis." *Journal of Clinical Epidemiology* 53 (5): 477–484. doi:10.1016/s0895-4356(99)00204-8.
- Thernstrom, Stephan, and Abigail Thernstrom. 1997. *America in Black and White: One Nation Indivisible*. New York: Simon and Schuster.
- Turner, Margery Austin, Fred Freiberg, Erin Godfrey, Carla Herbig, Diane K. Levy, and Robin R. Smith. 2002. *All Other Things Being Equal: A Paired Testing Study of Mortgage Lending Institutions*. Report prepared for the U.S. Department of Housing and Urban Development. Washington, DC: The Urban Institute.
- Turner, Margery A., Rob Santos, Diane K. Levy, Doug Wissoker, Claudia Aranda, and Rob Pitingolo. 2013. *Housing Discrimination Against Racial and Ethnic Minorities* 2012. Report prepared for the U.S. Department of Housing and Urban Development. Washington, DC: The Urban Institute.
- Valfort, Marie-Anne. 2018. *Anti-Muslim Discrimination in France: Evidence from a Field Experiment*. IZA Discussion Paper No. 11417. Bonn: Institute for the Study of Labor.

- Wasmer, Etienne. 2005. "Housing Market Discrimination, Housing Regulations and Intermediaries." Conference Proceedings "Expanding the Frontiers of Economics: Session Interactions Between Urban Housing and Labor Markets", 33. Philadelphia: American Economic Association.
- Weichselbaumer, Doris. 2015. "Testing for Discrimination against Lesbians of Different Marital Status: A Field Experiment." *Industrial Relations* 54 (1): 131–161. doi:10.1111/irel.12079.
- Weichselbaumer, Doris, and Rudolf Winter-Ebmer. 2005. "A Meta-analysis of the International Gender Wage Gap." *Journal of Economic Surveys* 19 (3): 479–511. doi:10.1111/j.0950-0804.2005.00256.x.
- Yinger, John. 1998. "Housing Discrimination is Still Worth Worrying About." *Housing Policy Debate* 9 (4). doi:10.1080/10511482.1998.9521322.
- Zhao, Bo, Jan Ondrich, and John Yinger. 2006. "Why Do Real Estate Brokers Continue to Discriminate? Evidence from the 2000 Housing Discrimination Study." *Journal of Urban Economics* 59 (3): 394–419. doi:10.1016/j.jue.2005.12.001.
- 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): 1115–1134. doi:10.1080/1369183X.2015.1133279.