NET WORK LISER

Discussion Paper Series

IZA DP No. 18788

July 2026

The State of Rental Discrimination: A Meta-Analytical Comparison of Five Discrimination Grounds in the Housing Market

Pieter-Paul Verhaeghe
Vrije Universiteit Brussel

Louise Devos
Ghent University

Abel Ghekiere
Vrije Universiteit Brussel

Stijn Baert
Ghent University,
University of Antwerp,
Université Catholique de Louvain
and IZA@LISER

Louis Lippens
Ghent University and IZA@LISER

IZA DP No. 18788 July 2026

The State of Rental Discrimination: A Meta-Analytical Comparison of Five Discrimination Grounds in the Housing Market*

Abstract

This meta-analysis compares patterns of rental discrimination across five grounds: ethno-racial origin, gender, sexual orientation, disability, and social origin. We analyze 475 field experimental estimates from 31 countries, published between 2002 and 2024. In the random-effects pooled estimates, discrimination is substantial across all grounds, with the highest levels for ethno-racial origin, social origin, and disability, and smaller penalties for men and same-sex couples. For several grounds, the pooled effect attenuates toward no discrimination after correction for small-study publication bias. Marked heterogeneity exists, including a clear ethno-racial hierarchy with the strongest penalties for Arab, North African or Turkish applicants. Contrary to prior work, discrimination is higher in Europe than in North America for ethno-racial origin and sexual orientation. There is no consistent decline over time, except for social origin. Real estate agents discriminate less than private landlords, but only for ethno-racial origin. Mode-of-contact effects are mixed. Overall, discrimination appears persistent and context-dependent, calling for more targeted and effectively enforced policy interventions.

JEL classification

C93, J14, J15, J16, J71, R21

Keywords

meta-analysis, discrimination, housing, ethnic origin, gender, disability, social origin, sexual orientation

Corresponding author

Pieter-Paul Verhaeghe
Pieter-Paul.Verhaeghe@vub.be

* This work was funded by the European Union (ERC, project number 101170164 MARIS). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. The work was also financially supported by the Belgian Federal Science Office (BELSPO–LAMIE project, grant number B1/233/P3/LAMIE) and by FWO (Research Foundation – Flanders, grant number 12AM824N).

1. Introduction

More than seventy-five years ago, the Universal Declaration of Human Rights enshrined the right to adequate housing and the principle of non-discrimination as fundamental human rights. Discrimination refers to the unequal and adverse treatment of people based on a protected ground, such as ethno-racial origin, disability, or gender, among others. It can be considered a behavioral component of broader systems of exclusion, such as racism, ableism or sexism (Verhaeghe 2025). Research has shown that housing discrimination can have both individual and societal consequences, including higher housing costs and search-related stress, increased economic inequality, and spatial segregation (Yinger 1997; Squires 2007). Notwithstanding these international legal provisions and the far-reaching consequences of discrimination, much empirical evidence indicates persistent rental discrimination, especially against racialized groups (e.g. Auspurg et al. 2019; Quillian et al. 2020), disabled people (e.g. Aranda 2015; Verhaeghe et al. 2016), gay males and transgender individuals (e.g. Flage 2021; Abbate et al. 2024), and housing voucher holders (e.g. Phillips 2017).

Gradually, the main research instrument for measuring discriminatory behavior in the housing market has become field experiments, such as audit studies and correspondence tests (Gaddis 2018; Verhaeghe 2022). In these experiments, two types of applicants apply to real housing advertisements. They are identical in all relevant aspects, except for specific personal characteristics that constitute the grounds of discrimination under scrutiny. One applicant belongs to the test group (e.g. a racialized minority or a same-sex couple), and the other to the control group (e.g. the ethnic or racial majority, or a heterosexual couple). As the assignment to these groups is random, any systematic disadvantage experienced by the test group is taken as evidence of discrimination. Given the experimental design and the method's application in real-world advertisements, such field experiments are considered the 'golden standard' for measuring discrimination (Gaddis 2018).

In this study, we conduct a meta-analysis of field experimental studies on rental discrimination with the overall goal of comparing patterns of discrimination across five grounds: ethno-racial origin, gender, disability, sexual orientation, and social origin. This later ground refers to adverse treatment because of someone's socio-economic status (e.g. having a working-class job, being unemployed or getting a housing voucher). This meta-analysis contributes to literature in at least three ways.

First, we complement previous meta-analyses of rental discrimination (e.g. Flage 2018; Auspurg et al. 2019; Quillian et al. 2020; Flage 2021; Gaddis et al. 2021) by covering the five discrimination grounds for which most field experimental studies have been conducted. More precisely, we collected data from 114 unique field experimental studies across 31 countries, encompassing 475 estimates of rental discrimination from 2002 to 2024. Former meta-analyses have typically concentrated on one or two discrimination grounds, most often ethno-racial origin or sexual orientation, occasionally intersecting with gender. This study adopts a broader comparative perspective by analyzing rental discrimination across five grounds. We assess the extent to which discrimination patterns observed for one ground generalize to

others. Moreover, to the best of our knowledge, this is the first meta-analysis that examines rental discrimination based on disability and social origin.

Second, we specifically compare discrimination ratios across four contextual factors: cross-national variation, temporal evolution, differences between types of housing providers (private owners versus real estate agencies), and application mode (telephone versus e-mail or online applications). At a theoretical level, this provides insight into the extent to which contextual theories of discrimination (Quillian & Midtbøen 2021) are generalizable across grounds. For policymakers, this helps to identify whether policy interventions aimed at reducing discrimination can be designed generically or need to be tailored to specific groups or contexts.

Third, correspondence research on rental discrimination has grown rapidly (Gaddis 2018; Verhaeghe 2022). Extending the temporal scope provides additional data points from a larger number of studies, allowing us to re-examine earlier findings. Prior meta-analyses, for example, found no significant differences in ethnic, gender, or sexual orientation discrimination between the United States and Europe (Flage 2018; Auspurg et al. 2019; Flage 2021) and reported a downward trend in ethnic discrimination (Auspurg et al. 2019; Quillian et al. 2020). This paper reassesses these results using more recent evidence and more refined categorizations of potentially discriminated groups. Moreover, we discuss time trends in the five grounds of discrimination mentioned above. Finally, we provide this updated data in open access, in accordance with the FAIR principles, enabling scholars to use and potentially further enrich it with additional information.

2. Previous research on rental discrimination

Five meta-studies have already applied formal meta-analytic methods to examine rental discrimination. Covering 25 studies on ethnic and gender discrimination in OECD countries between 2006 and 2017, Flage's (2018) meta-analysis provides an important overview. Similarly, Auspurg et al.'s (2019) meta-analysis draws on 46 audit and correspondence studies on ethnic discrimination in the housing market, conducted between 1973 and 2015 in the United States, Canada, and Europe. Racial discrimination in the United States is reviewed by Quillian et al. (2020), based on 16 studies spanning the period from 1976 to 2016. In the same vein, Gaddis et al. (2021) focus on 15 studies addressing rental discrimination against Black and Hispanic Americans in the United States for the period 2006-2021. Finally, 11 studies on discrimination based on sexual orientation, carried out between 2006 and 2017 in eight OECD countries, are encompassed in Flage's (2021) latest review.

These meta-analyses indicate persistent rental discrimination against racialized groups in both North America and Europe, even when publication bias is taken into account. There are, however, ethno-racial differences in the level of discrimination. From these meta-studies, it appears that Black, Arab, Muslim, and Turkish applicants systematically face more discrimination than Hispanic, Asian, and Eastern European applicants (Flage 2018; Auspurg et al. 2019; Quillian et al. 2020). The evidence is inconclusive regarding the comparison between

Black and Arab or Muslim applicants. Quillian and colleagues (2020) reported the highest discrimination ratios against Arab applicants in the United States, but this finding was based on only two studies, and they did not perform a significance test for the difference with Black applicants. Auspurg et al.'s (2019) meta-analysis indicates that Arab and Muslim applicants are more targeted by discrimination in OECD countries than Blacks. Still, it was unclear to what extent this finding holds for both Europe and North America.

From a gender perspective, Flage (2018) found that women are generally more favored in the rental housing market than men. This finding depends, however, significantly on the ethno-racial background of candidates: rental discrimination against males is larger for racialized applicants than for their counterparts from the majority group. Moreover, the gender gap in discrimination appears to be the highest among the Arab/Muslim group.

With respect to sexual orientation discrimination, Flage's meta-analysis from 2021 showed that there was moderate but significant discrimination against homosexual candidates on the rental housing market, but not against lesbian candidates. However, some more recent studies found no evidence for discrimination against (white) homosexual applicants anymore (e.g. Schwegman 2019; Abbate et al. 2024).

No meta-analyses have yet been conducted on studies on rental discrimination based on social origin and disability, even though several studies have found profound discrimination against candidates who are unemployed or hold working-class jobs (e.g. Andersson et al. 2012; Carlsson & Eriksson 2014).[1] The evidence for the role of the housing voucher program in the United States is more mixed, with some studies finding discrimination (e.g. Moore 2017; Phillips 2017; Faber & Mercier 2025), while others do not (e.g. Chan & Fan 2023). Regarding disability, previous research indicates rental discrimination against applicants with visual impairments (e.g. Verhaeghe et al. 2016), deafness (e.g. Aranda 2015), wheelchair use (e.g. Aranda 2015), or mental disability (e.g. Hammel 2017). In fact, all published studies on disability discrimination that we have identified have consistently found discrimination in the rental housing market.

In sum, although previous meta-analyses have examined field experimental evidence on ethno-racial, gender, and sexual orientation discrimination, this study is the first to systematically compare these grounds. Moreover, it extends the comparison by incorporating two less-studied grounds and broadens the scope of analysis by including more recent data.

3. Contextual variation in rental discrimination

Aside from quantitatively charting how much rental discrimination occurs against whom, previous research has also investigated contextual variation in the level of discrimination. Here, we focus on four of these contextual factors: 1) cross-national variation, 2) evolution

[1] From a legal point of view, realtors and property owners may select rental applicants based on the level of their income, but not on its source. In other words, they may assess whether applicants can afford the rent given their income, but they may not differentiate between income derived from labor, social benefits, or from higher- versus lower-status occupations.

over time, 3) differences between housing providers, and 4) the mode of application. Quillian and Midtbøen (2021) distinguish three types of contextual theories to explain patterns of discrimination. First, historical theories explain current-day levels of discrimination by countries' or regions' historical legacies, such as colonialism, slavery, migration history and racist or sexist policies, among others. Second, situational theories emphasize contemporary social or political factors, such as the rise of the radical right, terrorist attacks or economic cycles, to explain variation in discrimination. Institutional theories, finally, focus on organizational, administrative, and legal practices that may affect discrimination levels.

Against the background of these contextual theories, one might expect cross-national variation in rental discrimination. Countries may differ considerably in their levels of discrimination, reflecting differences in policies, migration and colonial histories, or housing market characteristics. Only three studies appear to have empirically examined cross-national patterns in rental discrimination: Flage's (2018), Auspurg et al.'s (2019) and Flage's (2021) meta-analyses. In contrast to the contextual theories, none of these studies found significant differences in ethno-racial, gender, or sexual orientation discrimination between the United States and Europe. Auspurg and colleagues (2019) showed, however, that there was significantly less rental discrimination against ethno-racial minorities in Canada compared with the United States and Europe. No study has yet investigated the cross-national differences in rental discrimination based on disability and social origin.

In addition, one might also expect changes over time in the level of rental discrimination. Especially, the situational and institutional theories predict temporal variation due to policy changes, political transitions, and focal events. In addition, the so-called modernization theory (Bell 1972) claims that socio-economic positions would be increasingly allocated based on merit (achievement) rather than ascribed characteristics, such as ethno-racial origin, gender, sexual orientation or disability. This proposition is reflected in contemporary popular narratives that portray society as 'meritocratic' or 'color-blind'. When this reasoning is applied to the rental market, long-term rental discrimination should decline. However, at the same time, the radical right appears to be on the rise in both Europe and the United States, which might result in more exclusionary attitudes (e.g. Ruisch & Ferguson 2022), and ultimately more rental discrimination (or at least a slowdown in the long-term decline).

Several studies have already examined time trends in ethno-racial discrimination in the rental housing market. Large-scale audit testing conducted by the Department of Housing and Urban Development (HUD) in the United States shows a steady decline in discrimination against Black Americans between 1977 and 2012 (Turner et al. 2013). Auspurg et al. (2019) examined the time trends since the 1970s in Europe and the United States. Their results suggest declining levels of rental discrimination, especially during the last two decades. However, their study does not clarify whether housing discrimination has also declined for Arab or Muslim people (in Europe) or only for Black Americans (in the United States). In a recent meta-analysis, Quillian et al. (2020) examined time trends in discrimination in the United States, finding that the more exclusionary forms of discrimination (e.g. receipt of no response or the advertised unit supposedly becoming 'unavailable') significantly declined

between 1976 and 2016. However, more subtle forms of differential treatment (e.g. the number of units recommended or inspected) have decreased but remain high. Although time trends have been extensively studied concerning ethnic discrimination, those for other discrimination grounds remain unresearched.

As a third contextual factor, we examine differences in rental discrimination across types of housing providers, more specifically, private owners and real estate agencies. Drawing on institutional theory, discrimination can be understood as embedded in a broader professional environment (Hirsh & Kornrich 2008; Verstraete & Verhaeghe 2020). In contrast to private owners, who typically rent out property only occasionally, rental practices among real estate agencies tend to be more standardized and professionalized. This higher degree of standardization may reduce individual discretion and, consequently, bias and discrimination. Moreover, in most countries, real estate agencies are subject to stronger institutional regulations and enforcement mechanisms against discrimination than private owners. Finally, one can argue that realtors are less risk-averse than private owners because they are not letting out their own property. Among existing meta-analyses of rental discrimination, only Flage (2018) explicitly examined the role of provider type and found that private owners indeed discriminate significantly more against ethnic minorities than real estate agencies. However, it remains unclear to what extent this pattern generalizes to other discrimination grounds.

Finally, the mode of application may constitute an additional relevant contextual factor. We distinguish between rental applications submitted by telephone and those submitted in writing via e-mail or online forms. Hogan and Berry (2011) argue that phone interactions are synchronous, allowing rental candidates and housing providers to engage directly and reciprocally. As a result, housing providers cannot simply 'ghost' an application by phone or leave it unanswered. They must instead explicitly reject a candidate, which may cause discomfort and violate politeness norms. In a similar vein, Heylen and Van den Broeck (2016) argue that e-mail and online applications afford housing providers greater anonymity, thereby reducing social accountability and making discriminatory behavior easier. For these reasons, we expect lower levels of rental discrimination in phone-based applications compared to applications submitted via e-mail or online forms. Flage's (2021) meta-analysis is the only study to have systematically examined this contextual variable and indeed found that field experiments conducted by phone reported less discrimination against same-sex couples than those conducted via e-mail. For other discrimination grounds, no systematic analyses have yet been conducted.

4. Data and methods

4.1 Scope and study selection

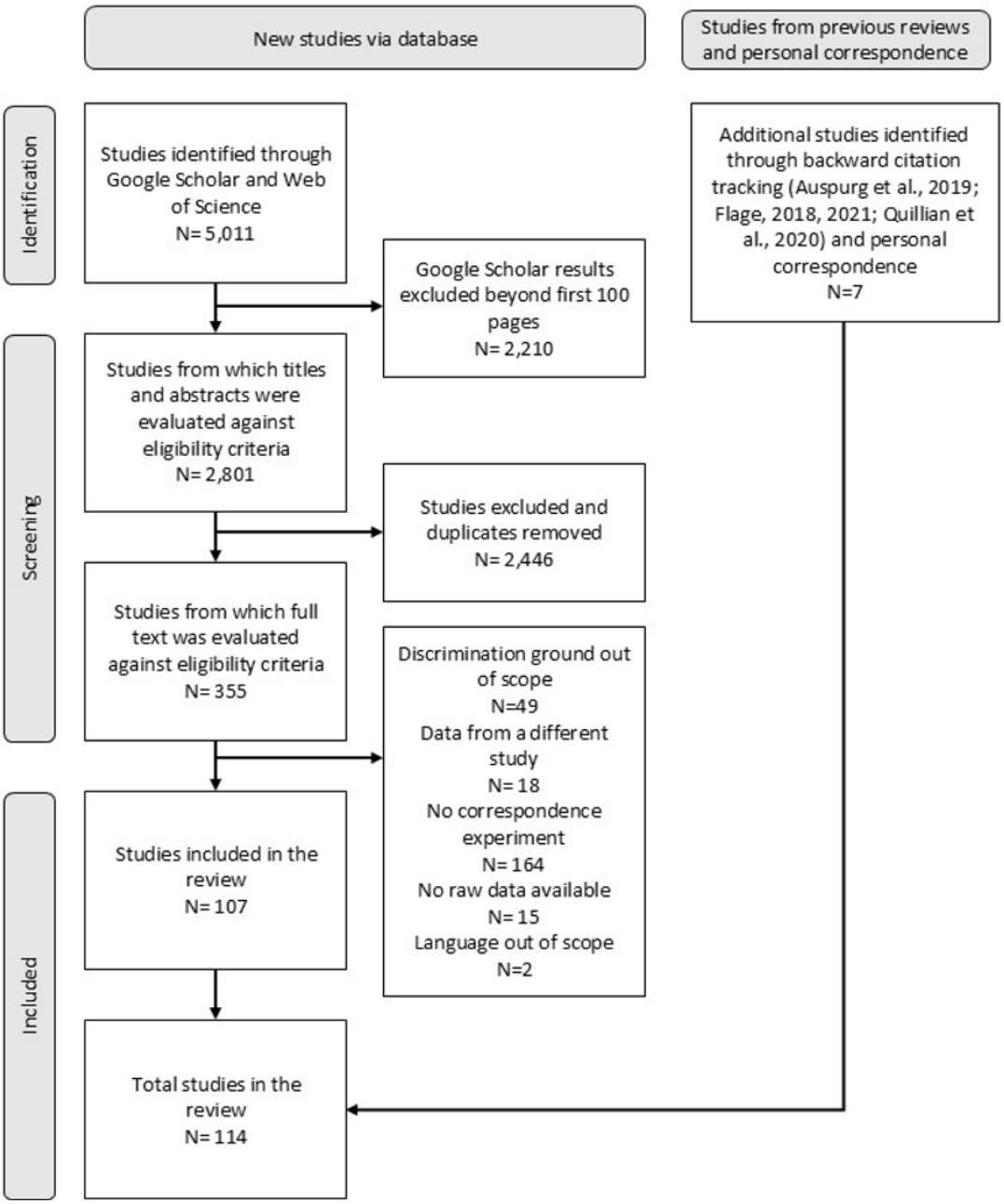
For this meta-analysis, we compiled a database of field experiment studies on rental discrimination. The inclusion criteria are summarized in Appendix A1, using a variation of

Richardson et al.'s (1995) Population, Intervention, Comparison, Outcome eligibility framework. The meta-analysis is restricted to studies which were published or posted between January 2002 and December 2024. The starting point of data collection is the publication of the first national-level Housing Discrimination Study in the United States (Turner et al. 2002), which marked the broader adoption of more rigorous, standardized methodologies in housing discrimination research. We only covered studies written in English, French, German or Dutch, consistent with the authors' language proficiencies. Eligible field experiments are restricted to correspondence tests and telephone audits. In-person audits were excluded because of their lower degree of standardization, stemming from potential verbal and physical differences between testers. Within these confines, we kept all field experiments on rental discrimination, including short-term rentals and roommate searching. We focused on five grounds of discrimination in the housing literature: ethno-racial origin, gender, disability, sexual orientation, and social origin. We only retained studies that reported raw counts or response rates that allowed us to reconstruct the outcome measure, capturing discriminatory differences in an invitation to view a dwelling or any other positive initial response from the housing provider.[2]

We systematically searched Google Scholar and Web of Science to identify relevant studies. A detailed overview of the search strings is provided in Appendix A2. As an additional identification strategy, we screened the reference lists of prior meta-studies on housing and incorporated relevant studies identified through personal correspondence. The study selection process is summarized in the PRISMA flow diagram in Figure 1. The final sample consists of 114 studies, from which 475 individual effect sizes on rental discrimination were extracted. When a study compared more than one test group (e.g. both wheelchair users and blind candidates were compared with the control group), we recorded separate effect sizes for each test group.

[2] Continuous indicators, such as the number of recommended units, were rarely reported outside the United States and were therefore excluded to ensure cross-national consistency.

Figure 1. PRISMA flow diagram for selecting field experimental studies



Note: PRISMA flow diagram showing study selection (adapted from Page et al. 2021). Numbers at each stage reflect the study count.

4.2. Variables

From each field experimental study, we captured several variables (Table A3 in the Appendix). The main outcome variable is the log-transformed discrimination ratio. For each field experimental comparison *k*, we registered the number of rental applications performed by both test and control groups (n_k^{Test} and $n_k^{Control}$) and their number of positive responses (y_k^{Test} and $y_k^{Control}$). Two types of positive responses are used in studies: general positive responses (e.g. requests to call back or provide additional information) and invitations to view a dwelling.

Across discrimination grounds, the mean positive response rates for the test and control groups equate to 44% and 53%, respectively. Subsequently, we calculate discrimination ratios by dividing the test group's positive response rate by that of the control group. The discrimination ratio is a risk ratio of binary outcomes (Karlson & Quillian 2023). In this case, a ratio of 1 indicates equal treatment, with lower values indicating greater rental discrimination against applicants from the test group *k*. As the discrimination ratio reflects a relative difference, a value of 0.80 indicates that test group applicants receive 20% fewer positive responses than control group applicants, aggregated at the field experiment level. In line with standard meta-analytic procedures, discrimination ratios were log-transformed before synthesis. The resulting approximate normal distribution of effects improves the validity of inverse-variance weighting and, therefore, enables more reliable estimation of pooled effects and standard errors (Borenstein et al. 2009). Equation 1 represents the calculation of the log-transformed discrimination ratio (DR_k) for effect size *k* with its standard error defined in Equation 2.

$$\ln(DR_k) = \ln\left(\frac{y_k^{Test}/n_k^{Test}}{y_k^{Control}/n_k^{Control}}\right) \tag{1}$$

$$SE_{\ln(DR_k)} = \sqrt{\frac{1}{y_k^{Test}} - \frac{1}{n_k^{Test}} + \frac{1}{y_k^{Control}} - \frac{1}{n_k^{Control}}} \tag{2}$$

Effect sizes with zero positive responses in either the test or the control group are excluded without continuity correction, since the log discrimination ratio and its delta-method standard error are undefined in those cases; this affects a small number of rows and does not materially shift the per-ground pooled estimates.

The analysis includes five main independent variables capturing the discrimination grounds analyzed in this study. First, the ethno-racial origin of the rental applicants is assessed through a variable with six categories: 'Arab, North African or Turkish', 'Black, Central African and African American', 'Asian' (mainly Indian and Chinese), 'Eastern European and other White', 'Hispanic or Southern American', and 'Other'. This categorization is based on processes of racialization along phenotypic, cultural, and religious boundaries. The category 'Arab, North African and Turkish' groups applicants with origins in Muslim-majority countries. The category 'Eastern European and other White' mainly refers to Roma, Polish, Romanian and Slovakian candidates. The 'Other' category comprises a heterogeneous set of ethno-racial groups with insufficient effect sizes to warrant separate analyses (e.g., mixed origin or Australian minority groups). The specific classification of countries across these categories can be found in the open-access dataset published with this study. Second, gender discrimination is captured with two applicant categories: 'Male' and 'Female'. Third, the variable 'sexual orientation' distinguishes between applicants with 'Different-sex' and 'Same-sex' relationships. Fourth, the discrimination ground of social origin is operationalized as a binary variable distinguishing applicants with 'lower' and 'higher' social status. Both categories are heterogeneous, but the

first concerns people in more precarious socio-economic positions, such as jobless, working-class, or low-skilled applicants, or housing voucher holders. Contrarily, the latter refers to their middle- or upper-class counterparts. Fifth, for disability, we differentiated between 'No disability', 'Mental disability', and 'Physical disability'. The latter category comprises blind candidates (with or without assistance dogs), wheelchair users or deaf applicants. Table 1 shows that the bulk of the 475 effect sizes in our meta-analysis relates to ethno-racial discrimination (69.68%), followed by gender (9.05%), sexual orientation (8.42%) and social origin (7.16%) discrimination. The minority of effect sizes (5.68%) concerns disability discrimination.

In addition, we included four variables that may capture contextual variation in rental discrimination. First, for each study, we have the specific country where the field experiment took place. For parsimony, this was subsequently reduced to a geographic variable with three categories: 'European countries' (65.89% of the effect sizes), 'North American countries' (30.53%), and 'Other countries' (mainly Asian and South American countries; 3.58%). Effect size descriptives at the country level are reported in Appendix A4.

The second contextual variable concerns the temporal variation in discrimination. For each field experimental study, we assembled the start and end years of the data collection. In the analyses, we averaged both, resulting in a continuous time variable ranging from 1999 to 2023. This slight discrepancy vis-à-vis the temporal scope of the eligibility criteria (i.e. studies published between 2002 and 2024) is a logical outcome of the lag between data collection and publication. There has been a steady increase in the number of field experiments on rental discrimination, which was even surging for ethno-racial discrimination (Figure A1 in the Appendix).

Third, we distinguished between types of housing providers: 'Private owners', 'Real estate agencies', and 'Mixed or unknown'. The latter category consists of studies with insufficient information about the specific housing provider or in which the authors did not distinguish between providers (i.e. 52.00% of the effect sizes). The fourth and final variable relates to the application mode, in which 'E-mail or online applications' are distinguished from 'Telephone or mixed applications', comprising, respectively, 91.16% and 8.84% of the effect sizes in our meta-analysis.

Remaining recorded variables are related to the design of the field experiment, more specifically whether researchers used a matched design (typically sending multiple applicants to the same advertisement) or unmatched assignment of applications; the level of detail given in the rental applications (e.g. additional information, such family size, smoking behavior, possession of pets); whether it concerned a short-term (Airbnb) rental advertisement or not; whether the study was peer-reviewed or not; the type of positive response recorded (invitation to view the property vs. any positive reaction, including phone or e-mail contact); and the language of the study (English, French, German or Dutch; French and German studies are grouped into 'Other' due to small sample sizes). When the experiment was not about gender discrimination, we also captured the gender of the applicants separately: 'Male', 'Female', 'Both' (when discrimination was recorded for both genders but could not be split by

gender) or 'Unknown'. In the same vein, we recorded the employment status and educational level of the applicant, when the experiment was not about social origin discrimination. Finally, when the experiment tested ethno-racial discrimination, we registered whether the minority applicant was a first-generation migrant, a second-generation migrant, or of unknown generational status.

4.3. Meta-analytic strategy

We estimated pooled mean log discrimination ratios for each discrimination ground using two complementary methods: random-effects (RE) and unrestricted weighted least squares (UWLS) meta-analysis. The RE estimator is our main estimator; UWLS estimates, which correct for small-study publication bias (*infra*), are reported alongside as a robustness check. The RE approach assumes the data generation shown in Equation 3, where μ is the mean true log discrimination ratio; $u_{s(k)}$ is the random intercept capturing between-study (*s*) and within-study between-effect (*k*) heterogeneity;[3] τ_s^2 quantifies the total between-study variance; and ε_{k} is the sampling error with known variance v_{k}.

$$y_k = \mu + u_{s(k)} + \varepsilon_{\mathrm{k}}, \quad u_{s(k)} \sim \mathcal{N}(0, \tau_s^2), \quad \varepsilon_{\mathrm{k}} \sim \mathcal{N}(0, v_{\mathrm{k}}) \tag{3}$$

The RE pooled estimate in Equation 4 is the intercept from the multilevel model in Equation 3, estimated by restricted maximum likelihood (REML). REML is typically more reliable at estimating heterogeneity when it is high (often when there are few studies or effects). Conservatively, we use CR2 small-sample bias-corrected study-clustered standard errors with Satterthwaite degrees-of-freedom adjustment (Pustejovsky & Tipton 2022).

$$\hat{\mu} = \frac{\sum w_k \ln DR_k}{\sum w_k}, \quad w_k = 1/(v_{\mathrm{k}} + \tau_s^2) \tag{4}$$

UWLS-based pooled estimates are obtained as the intercept from the bivariate regression in Equation 5 with inverse-variance weights $1/v_k$ and study-clustered standard errors (Stanley and Doucouliagos, 2014). Via UWLS, we apply a Precision Effect Test (PET) and Precision Effect Estimate with Standard Errors (PEESE) switching rule by inserting either the standard error of $\ln(DR_k)$ SE^1 (PET) or the sampling variance SE^2 (PEESE) into the regression equation as an additional regressor (Stanley and Doucouliagos, 2014). Following the canonical Stanley & Doucouliagos (2014) rule, extended to the multivariate setting per Reed (2023), we report the PEESE estimate when the bias-corrected effect from the PET fit — the predicted log discrimination ratio at SE = 0 with covariates at their sample means — is significantly non-null at a two-sided $p < 0.10$. Otherwise, we report the PET variant. The expected direction is negative (discrimination) for all grounds except gender, where preferential treatment of women is common in housing audits and a positive direction is used.

[3] Within-study clusters are defined as treatment effects depending on the same control (reference) group.

$$y_k = \mu + \beta_k^{SE} SE^p + \varepsilon_k, \quad \varepsilon_k \sim \mathcal{N}(0, v_k) \tag{5}$$

To identify effect heterogeneity in rental discrimination, we rely on the canonical meta-regression specification in Equation 6, where X_k collects the moderators listed in Section 4.2 (i.e. country, year, type of housing provider, and application mode), and $u_{s(k)} \sim \mathcal{N}(0, \tau_s^2)$ captures residual between-study and between-effect heterogeneity not explained by the covariates or the sampling error ε_k.

$$\ln DR_k = \beta_0 + X_k' B + u_{s(k)} + \varepsilon_k, \quad u_{s(k)} \sim \mathcal{N}(0, \tau_s^2), \quad \varepsilon_k \sim \mathcal{N}(0, v_k) \tag{6}$$

The target coefficient vector B in the heterogeneity analysis is estimated using three complementary approaches: random-effects meta-regression (RE-MRA), unrestricted weighted least squares meta-regression (UWLS-MRA), and Bayesian hierarchical meta-regression (BH-MRA). The three estimators described below differ in how they decompose the residual $u_{s(k)}$ and account for within-study dependence. For all three estimators, no model is fitted when the sum of the estimated parameters, i.e. the intercept, fixed-effect and random-effect components, and any interaction terms, exceeds the number of unique studies (i.e. parameter-based estimation constraint). This restriction generally applies to the discrimination ground 'Disability', which contributes just 14 studies. Hence, we do not perform meta-regressions focusing solely on disability discrimination. RE-MRA is treated as our main specification and discussed alongside the UWLS-MRA specification, which adjusts for small-study publication bias. For the test-group fixed effect, the reference category follows the default factor ordering applied during data wrangling; for ethno-racial origin specifically, the reference category is 'Arab, North African, Turkish' so that coefficients on the other test groups represent contrasts against the most-discriminated-against subgroup. Results from the Bayesian alternative estimator are reported in the Appendix as a robustness check.

First, RE-MRA is estimated by restricted maximum likelihood. Residual heterogeneity $u_{s(k)}$ is decomposed into a nested random-effects structure, i.e. a two-level structure when multiple distinct within-study clusters are present, and a single study-level structure otherwise, that accounts for correlation among multiple effect sizes from the same study (Viechtbauer, 2010). As in the pooled meta-analytic estimate approach described above, standard errors use the conservative CR2 small-sample bias-corrected sandwich estimator with Satterthwaite degrees-of-freedom approximation for inference (Pustejovsky & Tipton, 2022).

Second, UWLS-MRA is the 'unrestricted' alternative because the used weights, i.e. the inverses of the observations' sampling variance $1/v_k$, do not depend on estimated between-study heterogeneity, which may be contaminated by small-sample publication bias (Stanley & Doucouliagos, 2012). Between-study variance is implicitly accounted for through the clustered standard errors. Publication bias is modeled following the PET and PEESE switching mechanism as described above.

Third, BH-MRA is the Bayesian alternative that explicitly accounts for the metadata's multilevel structure. In the primary specification, it decomposes $u_{s(k)}$ into crossed random intercepts for each study and test group and is estimated using a Student-t likelihood, which reduces sensitivity to outlying effect sizes. In addition, we introduce a random intercept for the test group, assuming that group-specific discrimination effects exist around a common mean per treated group. The estimated residual standard deviation accounts for variance not captured by the design-based variance, much as the sampling error in the RE-MRA specification. Priors are weakly informative following $\mathcal{N}(0, 1)$ on all fixed effect coefficients and the intercept, and half-Cauchy, $HC(0, 0.5)$, on variance components. Each model uses four chains of 10,000 draws, with the first 2,000 as warmup iterations. Convergence is assessed using the potential scale reduction factor ($\hat{R}$ < 1.01) and bulk effective sample size (ESS ≥ 400 per chain; Vehtari et al. 2021).

Finally, we performed several sensitivity and exploratory analyses in the Appendix. Two sensitivity specifications substitute (a) country fixed effects for the regional geography indicator (with the United States as the reference) and (b) replace the continuous time variable with year-midpoint fixed effects. Next, two exploratory interaction analyses test whether the North American effect varies by candidate group type and by year, respectively.

4.4. Publication bias

In our meta-analysis, the sample of studies is not a random draw from all housing correspondence studies ever conducted, but a potentially biased sample of published studies. This publication bias can operate through selective publication, in which studies with significant or large (but possibly imprecise) effects are more likely to be published, even as preprints. The insights of this exercise inform which estimator to trust more. For example, if we detect statistical anomalies consistent with small-study publication bias, UWLS estimation should be more accurate and informative than RE estimation. We assessed the occurrence of this publication bias through four complementary diagnostics: 1) funnel plot asymmetry tests, 2) p-curve, 3) p-uniform*, and 4) step-function selection models. All diagnostics, except the funnel plot asymmetry tests, are applied to the full sample, to each ground, and to each test group; asymmetry tests are less accurate under high effect heterogeneity, as expected at the ground level. Following Harrer et al. (2021), grounds with fewer than ten effect sizes are excluded from ground-level analyses. Publication bias details and statistics are reported in the Appendix and discussed throughout the Results section where appropriate.

5. Results

5.1. Publication bias

We assessed publication bias through four complementary diagnostics applied to the full meta-dataset and to each discrimination ground separately (see Section 4 for details and Appendix A5 to A8). Grounds with fewer than 10 observations were excluded from subset-

level analyses. Taken together, the four diagnostics yield a mostly consistent picture. We detect small-study effects and funnel asymmetry, most prominently in the ethno-racial origin literature, and the step-function selection model diagnostics confirm statistically significant publication selection across the full meta-dataset and by ethno-racial origin and gender. Yet, for gender, the selection model signal is not corroborated by the regression-based funnel tests, indicating that any selection bias on that ground may operate at the significance threshold rather than as a small-study effect. In contrast, *p*-uniform* rejects the hypothesis of no true effect across all five discrimination grounds, indicating that the observed pattern of significant results is inconsistent with a null underlying effect, even under the assumption of selective reporting. The strong right-skewness of the *p*-curve is further inconsistent with *p*-hacking as the primary driver of the observed statistical significance. In other words, the housing discrimination effects documented in our meta-dataset generally reflect genuine underlying phenomena but are, to some extent, affected by imprecise effects. Therefore, we opted to provide (small-study) imprecision-corrected estimates separately via the PET and PEESE frameworks applied within the UWLS meta-analyses and -regressions in the main text. The Bayesian alternative estimator is reported in the Appendix.

5.2. Discrimination by ground and test group

Table 1 shows the estimated pooled discrimination ratios (DR) by ground, which allows us to assess the magnitude of discrimination for each ground. For the specific test groups, we report pooled discrimination ratios only when there are at least 10 effect sizes (as recommended by Harrer et al. 2021). Pooled discrimination ratios with 95% confidence intervals are visualized in Figure 2.

For the RE specification, the highest discrimination is observed for social origin, with a pooled discrimination ratio of 0.766. Those with a less beneficial origin receive, across worldwide conducted tests, about 23% fewer positive responses. This level is closely followed by ethno-racial discrimination with a ratio of 0.789. In other words, racialized applicants receive 21% fewer positive responses than their majority counterparts. Next, candidates with a physical or mental disability have a 19% lower chance of receiving a positive response than their counterparts without a disability. For the other grounds, sexual orientation and gender, we found significantly lower levels of discrimination. Same-sex couples have a 4% lower chance of receiving a positive response (DR = 0.958) than different-sex candidates. Regarding gender discrimination, we find a positive ratio of 1.040, indicating discrimination against male candidates, although this ratio is only marginally statistically significant.

The picture slightly changes when we consider the (small-study) bias-corrected pooled estimates from the UWLS specification. Generally, DR estimates shrink toward 1, indicating no discrimination. Rental discrimination based on social origin, for example, diminishes from 23% to about 14% fewer positive responses, and the estimated effect becomes imprecise, where the pooled effect is no longer statistically distinguishable from a null effect. The same applies to sexual orientation (DR = 0.984). However, UWLS has known limitations: when only a few

studies are included in the meta-analysis, when effects are uniformly based on small samples, or when heterogeneity in the estimates is high, the method tends to overcorrect (Stanley, 2017). Nevertheless, even though UWLS estimates might be too conservative, they remain meaningful given the publication bias uncovered in Section 5.1, particularly regarding ethno-racial origin. Specifically, the DR of ethno-racial origin shrinks from 0.789 to 0.957 (95% CI = [0.886, 1.033]). After correction, the pooled bias-corrected effect is no longer statistically distinguishable from a null effect. A similar pattern is observed for disability, where the UWLS-corrected DR shrinks from 0.813 to 0.984 (95% CI = [0.963, 1.005]) and likewise straddles the 'no discrimination' threshold.

Table 1. Pooled discrimination ratios by discrimination ground and test group

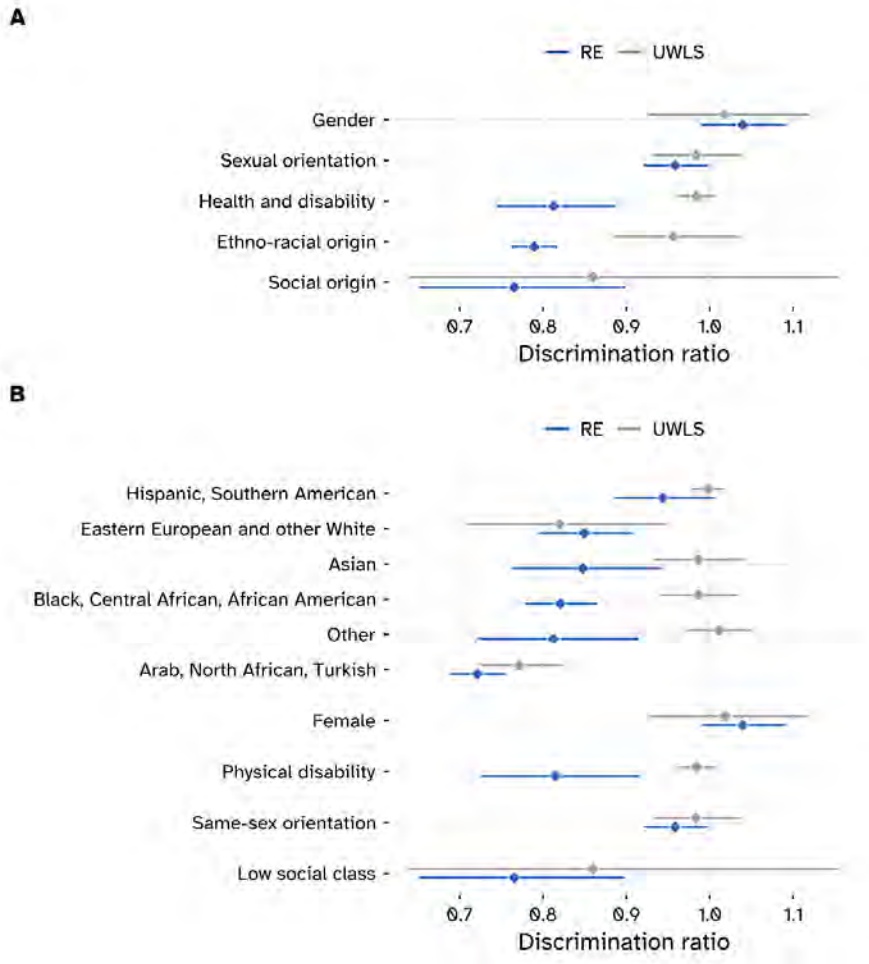
Ground and test groups	*k*	*N*	RE Pooled DR [95% CI]	τ^2	I^2	UWLS Pooled DR [95% CI]	UWLS variant
Ethno-racial origin	331	94	0.789 [0.763, 0.816]	0.03	0.93	0.957 [0.886, 1.033]	**PET**
Arab, North African or Turkish	113	53	0.721 [0.689, 0.754]	0.03	0.86	0.771 [0.722, 0.823]	**PEESE**
Asian	48	11	0.848 [0.763, 0.942]	0.02	0.82	0.986 [0.933, 1.042]	**PET**
Black, Central African, or African American	62	32	0.821 [0.779, 0.864]	0.02	0.89	0.987 [0.943, 1.033]	**PET**
Eastern European and other White	56	23	0.850 [0.796, 0.906]	0.02	0.89	0.819 [0.709, 0.947]	**PEESE**
Hispanic or Southern American	32	13	0.944 [0.887, 1.005]	0.01	0.85	0.998 [0.979, 1.016]	**PET**
Other	20	13	0.813 [0.723, 0.914]	0.04	0.94	1.011 [0.975, 1.049]	**PET**
Gender	43	28	1.040 [0.992, 1.090]	0.01	0.68	1.019 [0.928, 1.119]	**PET**
Disability	27	14	0.813 [0.746, 0.886]	0.03	0.96	0.984 [0.963, 1.005]	**PET**
Physical disability	21	12	0.815 [0.727, 0.914]	0.04	0.97	0.984 [0.963, 1.007]	**PET**
Sexual orientation	40	18	0.958 [0.922, 0.997]	<0.01	0.33	0.984 [0.933, 1.037]	**PET**
Social origin	34	23	0.766 [0.653, 0.898]	0.14	0.97	0.859 [0.640, 1.153]	**PET**

Notes: k (number of effects), N (number of studies), RE (random-effects), DR (discrimination ratios), UWLS (unrestricted weighted least squares), PEESE (precision-effect estimate with standard errors), PET (precision-effect test). The RE column reports the random-effects pooled estimate with a nested random-effects structure (clustering within studies). The UWLS column reports the small-sample bias-corrected pooled estimate from the preferred PET or PEESE specification; the selected variant is indicated in the rightmost column. PEESE is selected when the PET fit's bias-corrected effect (i.e., predicted value at SE = 0 with covariates at their sample means) is significantly non-null at a two-sided $p < 0.10$; otherwise, PET is reported (Stanley & Doucouliagos, 2014; Reed, 2023). For gender, the anticipated direction is positive (DR > 1); for all other grounds, it is negative (DR < 1). Discrimination ratios and their confidence intervals were back-transformed from the original log specification. The heterogeneity indicators τ^2 and I^2 are the between-study variance and the proportion of total variance attributable to between-study heterogeneity.

We found that all racialized groups have worse chances of receiving a positive response compared to their counterparts (RE specification). Applicants of Arab, North African or Turkish origin appear to be the most discriminated against, with a pooled discrimination ratio of 0.721 or about 28% fewer positive responses. This applicant group is significantly more discriminated against than all other ethno-racial minority groups on the rental housing market. Under canonical UWLS bias correction, the discrimination level for this group remains substantial (DR = 0.771, 95% CI = [0.722, 0.823] – i.e. about 23% fewer positive responses). Furthermore, Black and Asian candidates are also disadvantaged in their rental applications, with 18% and 15%

fewer chances of getting a positive response, respectively. Under UWLS bias correction, these effects shrink to about 1% each (DR_{Asian}= 0.986, 95% CI = [0.933, 1.042]; DR_{Black} = 0.987, 95% CI = [0.943, 1.033]) and are no longer statistically distinguishable from no discrimination. Effect heterogeneity remains high (I^2 = 0.82–0.89, cf. infra), and the correction may be over-conservative in subgroups with relatively few studies. In addition, candidates of Eastern European or other white origin face a comparable level of discrimination after UWLS correction (DR = 0.819, 95% CI = [0.709, 0.947]), with the bias-corrected estimate close to the RE estimate and statistically distinguishable from no discrimination. Finally, Hispanic or Southern American applicants appear to be the least discriminated against, with a marginally statistically significant discrimination ratio of 0.944 at the 10% significance level. After UWLS bias correction, the bias-corrected estimate moves further toward the 'no discrimination' threshold (DR = 0.998, 95% CI = [0.979, 1.016]). They are also significantly less discriminated against in comparison with the other ethno-racial test groups (except for the 'Eastern European and other White origin' group).

Figure 2. Pooled discrimination ratios by ground and test group



Notes: Panel A: discrimination ratios by ground, Panel B: discrimination ratios by test group. Blue-filled diamonds show random-effects (RE) pooled estimates; gray-filled diamonds show publication bias-corrected UWLS estimates (PET or PEESE; see Table 1 for the decision rule). Horizontal lines denote 95% confidence intervals. In Panel A, grounds are ordered by the RE point estimates, while in Panel B, test groups are ordered by the RE point estimates per ground. k = 475.

Results for disability discrimination can be further detailed for physical disability. Candidates who indicate a physical disability have approximately 18% less chance of receiving a positive response under the RE specification (95% CI = [0.727, 0.914]). After UWLS bias correction, the estimate attenuates substantially (DR = 0.984, 95% CI = [0.963, 1.007]) and is no longer statistically distinguishable from no discrimination, consistent with the strong funnel asymmetry signal for the disability literature reported in Appendix Table A5. Due to a low number of effect sizes, we cannot reliably estimate a pooled discrimination ratio for mental disability (k = 6).

Table 1 also presents two measures of statistical heterogeneity: τ^2 and I^2. Higher τ^2 values indicate that true effects vary more across studies; higher I^2 values indicate that a greater proportion of total variance reflects between-study heterogeneity rather than sampling error. Although τ^2 values are modest to low in absolute terms, I^2 values extend up to 0.97, reflecting that the high precision of the underlying field experiments leaves little of the total variance attributable to sampling error. This finding suggests that the true discrimination ratios are disparate within discrimination grounds – especially for social origin, ethno-racial origin, and disability – and treatment groups. We further explore this heterogeneity by testing study-level contextual moderators using meta-regression techniques in the next section.

5.3. Contextual heterogeneity in discrimination

Table 2 reports the coefficients of the meta-regression analyses on the pooled discrimination ratios for the full sample, combining all discrimination grounds. Table 3 presents the same analyses, but separately by ground. Robustness and sensitivity analyses are in Tables A9–A16. Given the specification of the discrimination ratio in Equation 1, positive coefficients generally indicate greater equality in treatment and thus lower levels of discrimination. The exception is gender, where the ratio is defined with female applicants as the minority group, so positive coefficients indicate a relative disadvantage for male applicants. We focus primarily on findings that are consistent in sign and significance across all three estimators (i.e. the random-effects specification, RE-MRA; the unrestricted weighted least squares specification, UWLS-MRA; and the Bayesian specification BH-MRA in Appendix A9-10), and note where results diverge.

There are several contextual variations in discrimination levels. First, studies conducted in North America exhibit, on average, lower discrimination ratios than European studies (β = 0.102, p = 0.029, RE-MRA estimate in Table 2). Discrimination levels are approximately 10% higher in Europe than in North America. This finding is consistent across statistical specifications (RE-MRA, UWLS-MRA and BH-MRA). Moreover, subsequent analyses show no significant interaction between region and time (Table A11), indicating a stable regional effect of more discrimination in Europe than in North America. Separate analyses by discrimination ground, shown in Table 3, indicate that these regional differences in discrimination are observed only for ethno-racial origin (β = 0.160, p = 0.003, RE-MRA estimate) and tentatively for sexual orientation (β = 0.158, p = 0.103, RE-MRA estimate; β = 0.163, p = 0.023, UWLS-MRA estimate). These coefficients indicate that discrimination levels against racialized candidates

and same-sex couples are about 16% higher in Europe than in North America. Again, there are no relevant significant interaction effects with time (Table A12). For the other discrimination grounds, we do not have evidence of regional variation between Europe and North America. However, these comparisons are imprecise due to limited between-region variation in the smaller subsamples. Supplementary analyses further calculate country-specific DRs (Appendix A13). It appears that the highest levels of ethno-racial discrimination can be found in Czechia, Greece, Austria and Sweden, while the lowest levels in Canada, Iceland, the United States and Switzerland. These estimates should, however, be considered with much caution, given the often very small number of effect sizes on which they are based.

Second, there is no evidence of a continuous-time trend in discrimination in the field experiments published between 2002 and 2024. This pattern holds across all statistical specifications for the full sample (Table 2) and across individual discrimination grounds (Table 3). The only exception is discrimination based on social origin, for which we observe a linear decrease over time in the publication bias-adjusted UWLS-MRA specification (β = 0.061, p = 0.030). In additional analyses, we examined potential non-linear time trends by distinguishing 5-year periods but again found no evidence of temporal variation (Appendix A14-A15).

Third, studies conducted with real estate agencies yield lower discrimination ratios than those conducted with private owners (β = 0.178, p < 0.001, RE-MRA estimate in Table 2). From Table 3, it appears that this pattern only holds for ethno-racial discrimination (β = 0.234, p < 0.001, RE-MRA estimate) and not for the other grounds. Depending on the statistical specification, ethno-racial discrimination levels are circa 19% (UWLS-MRA) to 23% (RE-MRA) lower among realtors.

Fourth, telephone and mixed-mode audits are associated with higher discrimination ratios than email-based studies, though this effect reaches conventional significance only in the publication bias-adjusted UWLS-MRA specification (β = −0.098, p = 0.003 in Table 2). This result appears to be mainly attributable to higher discrimination against ethno-racial groups (β = −0.111, p = 0.022, UWLS-MRA estimate in Table 3) and female candidates (β = −0.174, p = 0.022, UWLS-MRA estimate) in these modes. However, these findings should be interpreted with caution, as the evidence is not consistent across all model specifications.

Finally, the additional context variables exhibit several interesting patterns in Table 3. The rental advantage of female vis-à-vis male candidates diminishes among unemployed candidates, resulting in more discrimination against unemployed women compared to unemployed men (β = −0.331, p = 0.005, RE-MRA estimate in Table 3). This finding is consistent across estimators. Furthermore, for ethno-racial and sexual orientation discrimination, positive response ratios are higher for female candidates, indicating less discrimination against racialized women and lesbian couples compared with their male and homosexual counterparts. Yet, this pattern is apparent only in the RE-MRA and BH-MRA specifications (Table A10), but not in the publication bias-adjusted UWLS-MRA specification. Moreover, ethno-racial discrimination is smaller for second-generation migrants than for first-generation migrants (β = 0.229, p = 0.012, RE-MRA estimate), with the effect consistent in sign and significance across estimators.

Table 2. Meta-regression of rental discrimination on contextual factors, full meta-sample

	RE-MRA	UWLS-MRA
Intercept	−1.041 (8.031)	—
$SE_{\ln DR}$	—	−0.976*** (0.264)
Region *(ref.: Europe)*		
North America	0.102* (0.045)	0.100** (0.035)
Other	−0.045 (0.125)	−0.100 (0.059)
Year	0.000 (0.004)	0.002 (0.003)
Type of housing provider *(ref.: Private owner)*		
Real estate agency	0.178*** (0.046)	0.137** (0.046)
Mixed or unknown	0.085 (0.049)	0.044 (0.044)
Mode of application *(ref.: E-mail or online)*		
Telephone or mixed	−0.077 (0.050)	−0.098** (0.032)
Candidate employment status *(ref.: Employed)*		
Unemployed	−0.368 (0.137)	−0.267*** (0.044)
Mixed or unknown	−0.131 (0.067)	−0.021 (0.031)
Candidate gender *(ref.: Male)*		
Female	0.065* (0.021)	−0.004 (0.027)
Both	0.074 (0.045)	0.059* (0.027)
Unknown or NA	0.076 (0.133)	0.105** (0.039)
Extra controls and fixed effects		
Study language	Yes	Yes
Peer-reviewed status	Yes	Yes
Airbnb indicator	Yes	Yes
Callback type	Yes	Yes
Matched design	Yes	Yes
Application detail level	Yes	Yes
Candidate education	Yes	Yes
Test group	Yes	Yes
UWLS variant	—	PET
k	475	475
N	114	114
Pseudo-R^2	21.7%	—
Adj. R^2	—	53.6%

Notes: The dependent variable is the natural log of the positive response ratio, where more negative values indicate stronger discrimination against the test group. RE-MRA is a multilevel random-effects meta-regression (restricted maximum likelihood, random intercepts at the study and cluster levels) with cluster-robust CR2 standard errors. UWLS-MRA is an unrestricted weighted least squares meta-regression with heteroskedasticity-robust standard errors. $SE_{\ln DR}$ denotes the standard error of the outcome; it serves as publication bias correction term under the PET estimator. The UWLS variant row reports the selected bias-correction specification (i.e., the PET fit's bias-corrected effect at SE = 0 with covariates at their sample means is significantly non-null at two-sided p < 0.10; Stanley & Doucouliagos, 2014; Reed, 2023). Dashes (—) indicate that a variable is ineligible for a given discrimination ground or is excluded because the degree-of-freedom constraint binds ($N < 1 + p + q$), with *p* number of fixed-effect parameters and *q* the number of random-effects components. Test group enters RE-MRA as a covariate and UWLS-MRA as a fixed effect. *k* = number of effect sizes; N = number of independent studies. Pseudo-R^2 for RE-MRA measures the variance explained relative to an intercept-only model with the same random-effects structure. Adj. R^2 for UWLS-MRA is the within-estimator adjusted R^2. RE-MRA significance levels are based on CR2 Satterthwaite degrees of freedom, which may be very small for sparse subgroups; cluster-robust standard error in parentheses. * p < 0.05, ** p < 0.01, *** p < 0.001.

Table 3. Meta-regression of rental discrimination on contextual factors by ground

	RE-MRA			UWLS-MRA			
	Ethno-racial origin	**Gender**	**Sexual orientation**	**Ethno-racial origin**	**Gender**	**Sexual orientation**	**Social origin**
Intercept	−6.827 (9.974)	5.222 (8.240)	28.097 (20.247)	—	9.912 (9.113)	25.351 (27.932)	−120.832* (52.315)
$SE_{\ln DR}$	—	—	—	−1.168*** (0.306)	0.806 (0.633)	−0.348 (0.812)	−3.400 (2.302)
Region *(ref.: Europe)*							
North America	0.160** (0.051)	−0.038 (0.070)	0.158 (0.039)	0.157** (0.047)	−0.040 (0.064)	0.163* (0.065)	−0.104 (0.241)
Other	−0.092 (0.169)	—	0.418 (0.132)	−0.052 (0.068)	—	0.419*** (0.104)	−0.066 (0.264)
Year	0.003 (0.005)	−0.003 (0.004)	−0.014 (0.010)	0.004 (0.004)	−0.005 (0.005)	−0.013 (0.014)	0.061* (0.026)
Type of housing provider *(ref.: Private owner)*							
Real estate agency	0.234*** (0.038)	−0.064 (0.109)	0.016 (0.077)	0.186*** (0.049)	−0.169 (0.099)	−0.021 (0.101)	−0.337 (0.295)
Mixed or Unknown	0.128** (0.043)	−0.064 (0.048)	0.116 (0.053)	0.059 (0.047)	−0.064 (0.047)	0.052 (0.145)	−0.380 (0.216)
Mode of application *(ref.: E-mail or online)*							
Telephone or Mixed	−0.013 (0.053)	−0.182 (0.067)	−0.041 (0.202)	−0.111* (0.048)	−0.174* (0.072)	−0.030 (0.093)	0.386 (0.212)
Candidate employment status *(ref.: Employed)*							
Unemployed	−0.104 (0.149)	−0.331** (0.072)	—	−0.231*** (0.066)	−0.351*** (0.082)	—	—
Mixed or Unknown	−0.006 (0.081)	−0.010 (0.049)	0.341 (0.085)	−0.004 (0.036)	0.018 (0.054)	0.327** (0.084)	−0.419** (0.126)
Candidate gender *(ref.: Male)*							
Female	0.051* (0.018)	—	0.125* (0.050)	0.010 (0.027)	—	0.060 (0.051)	0.162 (0.145)
Both	0.043 (0.036)	—	—	0.023 (0.033)	—	—	0.542** (0.178)
Unknown or NA	0.111* (0.050)	—	—	−0.003 (0.039)	—	—	0.386 (0.247)
Candidate migrant generation *(ref.: First generation)*							
Second generation	0.229* (0.007)	—	—	0.227*** (0.015)	—	—	—
Unknown or NA	0.154 (0.160)	—	—	0.155** (0.057)	—	—	—
Extra controls or fixed effects							
Study language	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Peer-reviewed status	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Airbnb indicator	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Callback type	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Matched design	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Application detail level	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Candidate education	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Test group	Yes	No	No	Yes	No	No	No
UWLS variant	—	—	—	PET	PET	PET	PET
k	331	43	40	331	43	40	34
N	94	28	18	94	28	18	23
Pseudo-R^2	19.6%	18.4%	0.0%	—	—	—	—
Adj. R^2	—	—	—	61.6%	43.4%	<0.1%	55.8%

Notes: See Table 2

6. Conclusion and discussion

This meta-analysis examines patterns of rental discrimination across the five most studied grounds of discrimination: ethno-racial origin, gender, sexual orientation, disability, and social origin. While prior meta-analyses covered only one or two of these grounds (e.g. Flage 2018; Auspurg et al. 2019; Quillian et al. 2020; Flage 2021; Gaddis et al. 2021), our meta-study is the first to compare patterns of rental discrimination across five grounds systematically, and to assess discrimination on the grounds of disability and social origin. Furthermore, building on contextual theories of discrimination, we investigate the extent to which rental discrimination follows common patterns across these five grounds. Specifically, we analyze variation in discrimination across countries, time periods, types of housing providers, and modes of application. For these purposes, we collected data from 114 unique field experimental studies across 31 countries, encompassing 475 estimates of rental discrimination studies from 2002 to 2024. The findings reveal several striking similarities across discrimination grounds, alongside a limited number of meaningful differences.

First, this meta-analysis shows that the level of rental discrimination varies significantly across grounds, although some estimates are substantively attenuated after controlling for publication bias. In our primary random-effects specification, we observe the highest discrimination ratios for ethno-racial origin, social origin, and disability. On average, applicants of racialized origin receive 21% fewer positive responses than majority applicants, candidates signalling a socially deprived background receive 23% fewer replies, and applicants with a mental or physical disability face a 19% reduction. While ethno-racial discrimination has been widely documented in the literature (e.g. Flage 2018; Auspurg et al. 2019; Quillian et al. 2020; Gaddis et al. 2021), discrimination against disabled and socially disadvantaged applicants has received comparatively limited attention, despite effect sizes of similar magnitude. In addition, male applicants and same-sex couples also experience discrimination in the rental housing market, albeit to a very minor extent. To our knowledge, this study provides the first meta-analytic synthesis of discrimination based on disability and social origin.

Second, for each of these five discrimination grounds, there is substantial underlying heterogeneity across field experimental studies, partly because the test groups they cover differ. This observation is very clear with respect to ethno-racial discrimination. Based on our primary random-effects specification, all racialized groups face rental discrimination. Still, the magnitude varies substantially: applicants of Arab, North African, or Turkish origin experience the highest level of discrimination worldwide (28% fewer positive responses), followed by Black (18%), Eastern European (15%), and Asian (15%) candidates. Hispanic or Southern American applicants face comparatively lower levels of discrimination, with a mere 6% less positive responses, which is only marginally statistically significant. This ethno-racial hierarchy in the housing market is consistent with previous meta-analyses on rental discrimination (Flage 2018; Auspurg et al. 2019; Quillian et al. 2020) and resembles patterns observed in the labor market (Lippens et al. 2023). However, only Auspurg et al. (2019) also found significantly higher

levels of discrimination against Arab or Muslim applicants than against Black applicants, a finding that we corroborate in our meta-study. In addition, we found that ethno-racial discrimination is smaller for second-generation migrants than for first-generation migrants.

Third, we found that regional contexts matter for rental discrimination. Ethno-racial discrimination is considerably higher in Europe than in North America; for sexual orientation, the same pattern emerges, though more tentatively. Discrimination against racialized applicants is approximately 16% higher in the 'Old World', with a similar magnitude, though greater uncertainty, for same-sex couples. This finding contrasts with previous meta-analyses on ethnic discrimination (Flage 2018; Auspurg et al. 2019), which did not identify such regional differences. These regional and national differences in discrimination levels might be explained by contextual theories (Quillian and Midtbøen 2021), which focus on the role of countries' historical legacies and situational and institutional factors. However, the challenge for future research is to empirically test these theories and explain why countries and regions differ in their discrimination patterns.

Fourth, except for social origin, we find no time trends in rental discrimination in field experimental studies published between 2002 and 2024, either in North America or Europe; discrimination based on social origin appears to be decreasing over time. This finding is noteworthy for several reasons. Previous meta-analyses reported a decline in ethno-racial discrimination over time (Auspurg et al. 2019; Quillian et al. 2020), but they covered different, longer periods (1973-2015 and 1976-2016, respectively). Their findings are consistent with the so-called HUD studies in the United States, which documented a decrease in the most egregious forms of racial discrimination during the 1990s (Ross & Turner 2005; Turner et al. 2013). However, both meta-analyses also reported a marked drop in ethno-racial discrimination in the final decade they covered, a pattern that is not observed in this meta-analysis. A potential explanation for these divergent findings is that this study accounted for more covariates (e.g. type of housing provider, gender/education/employment of applicants, migrant generation). The absence of any (declining) time trend is also striking, given the gradual strengthening of anti-discrimination policies in European countries since the 2000 EU Racial Equality Directive (Yavçan & Gorgerino 2025). However, meta-analyses on the labor market also found no robust decrease in discrimination in hiring over time (Lippens et al. 2023; Quillian & Lee 2023) and even indications of an increase in discrimination against foreign-born applicants (Quillian & Midtbøen 2026).

Fifth, consistent with the findings of Flage (2018), real estate agencies appear to discriminate much less against racialized groups than private landlords. One possible explanation is that this segment of the rental market tends to be more regulated and professionalized (Hirsh & Kornrich 2008; Verstraete & Verhaeghe 2020). Moreover, realtors may be less risk-averse than private landlords because they do not rent out their own property. However, this meta-analysis is the first to show that this difference between housing providers appears to matter only for ethno-racial discrimination and not for discrimination based on gender, social origin, disability, or sexual orientation.

Finally, we hypothesized that rental applications submitted via e-mail or online forms would result in greater discrimination, as it is easier to ignore or dismiss 'unwanted' candidates than over the phone. However, in some model specifications, we find the opposite: telephone and mixed-mode audits reveal higher levels of discrimination against ethno-racial groups and female candidates, albeit only when we control for small-sample publication bias.

These findings have important implications for policymakers. Once again, we provide clear and convincing evidence for rental discrimination. Moreover, despite the anti-discrimination policies of the last few decades, discrimination has not declined. This discrepancy calls into question the efficacy of current policy efforts to address rental discrimination. In addition, our study reveals considerable variation across grounds and contexts of discrimination, suggesting that a more tailored approach, rather than a general one, is needed to address them. While all forms of discrimination must be condemned, limited resources should be prioritized toward tackling discrimination based on ethno-racial origin – particularly by private owners – as well as discrimination based on disability and social origin. Generalization beyond the sampled contexts is constrained by the geographic distribution of the underlying field experiments: studies from the United States, Belgium and the United Kingdom together account for approximately 59% of effect sizes (Table A4), and South American, Asian and Oceanian evidence remains thin.

References

Abbate, N., Berniell, I., Coleff, J., Laguinge, L., *et al.* (2024) 'Discrimination against gay and transgender people in Latin America: a correspondence study in the rental housing market', *Labour Economics*, 87: 102486.

Andersson, L., Jakobsson, N., Kotsadam, A. (2012) 'A field experiment of discrimination in the Norwegian housing market: gender, class, and ethnicity', *Land Economics*, 88/2: 233-240.

Aranda, C. (2015) 'Targeting disability discrimination: findings and reflections from the national study on housing discrimination against people who are deaf and people who use wheelchairs', *Cityscape*, 17/3: 103-122.

Auspurg, K., Schneck, A., Hinz, T. (2019) 'Closed doors everywhere? A meta-analysis of field experiments on ethnic discrimination in rental housing markets', *Journal of Ethnic and Migration Studies*, 45/1: 95-114.

Bell, D. (1972) 'Meritocracy and equality', *The Public Interest*, 29: 29-68.

Carlsson, M., Eriksson, S. (2014) 'Discrimination in the rental market for apartments', *Journal of Housing Economics*, 23: 41-54.

Chan, E., Fan, Y. (2023) 'Housing discrimination in the low-income context: evidence from a correspondence experiment', *Journal of Housing Economics*, 59: 101889.

Faber, J., Mercier, M. (2025) 'Multidimensional discrimination in the online rental housing market: implications for families with young children', *Housing Policy Debate*, 35/2: 161-184.

Flage, A. (2018) 'Ethnic and gender discrimination in the rental housing market: evidence from a meta-analysis of correspondence tests, 2006–2017', *Journal of Housing Economics*, 41: 251-273.

Flage, A. (2021) 'Discrimination against same-sex couples in the rental housing market: a meta-analysis', *Economics Bulletin*, 42/2: 643-653.

Gaddis, S. (2018) 'An introduction to audit studies in the social sciences'. In: Gaddis, S. (ed.) *Audit studies: behind the scenes with theory, method, and nuance*, pp. 3-44. Cham: Springer.

Gaddis, S., Larsen, E., Crabtree, C., Holbein, J. *et al.* (2021) 'Discrimination against Black and Hispanic Americans is highest in hiring and housing contexts: a meta-analysis', working paper.

Hammel, J. (2017) *Rental housing discrimination on the basis of mental disabilities: results of pilot testing*. Washington, DC: US Government Printing Office.

Harrer, M., Cuijpers, P., Furukawa, T., Ebert, D. (2021) *Doing meta-analysis with R*. Boca Raton: CRC Press.

Heylen, K., Van den Broeck, K. (2016) 'Discrimination and selection in the Belgian private rental market', *Housing Studies*, 31/2: 223-236.

Hirsh, C., Kornrich, S. (2008) 'The context of discrimination: workplace conditions, institutional environments, and sex and race discrimination charges', *American Journal of Sociology*, 113/5: 1394–1432.

Hogan, B., Berry, B. (2011) 'Racial and ethnic biases in rental housing: an audit study of online apartment listings', *City & Community*, 10/4: 351–372.

Karlson, K., Quillian, L. (2023) 'How (not) to use risk ratios in sociological research', *Socius*, 9: 23780231231192394.

Lancee, B. (2021) 'Ethnic discrimination in hiring: comparing groups across contexts', *Journal of Ethnic and Migration Studies*, 47/6: 1181–1200.

Lippens, L., Baert, S., Ghekiere, A., Verhaeghe, P.P., Derous, E. *et al.* (2022) 'Is labour market discrimination against ethnic minorities better explained by taste or statistics?', *Journal of Ethnic and Migration Studies*, 48/17: 4243-4276.

Lippens, L., Vermeiren, S., Baert, S. (2023) 'The state of hiring discrimination: a meta-analysis of (almost) all recent correspondence experiments', *European Economic Review*, 151: 104315.

Moore, M. (2017) *Rental market discrimination in the housing choice voucher program*. PhD dissertation.

Page, M., McKenzie, J., Bossuyt, P., Boutron, I., *et al.* (2021) 'The PRISMA 2020 statement: an updated guideline for reporting systematic reviews', *BMJ*, n71.

Phillips, D. (2017) 'Landlords avoid tenants who pay with vouchers', *Economics Letters*, 151: 48-52.

Pustejovsky, J., Tipton, E. (2022) 'Meta-analysis with robust variance estimation: expanding the range of working models', *Prevention Science*, 23: 425-438.

Quillian, L., Lee, J. (2023) 'Trends in racial and ethnic discrimination in hiring in six Western countries', *PNAS*, 120/6: e2212875120.

Quillian, L., Lee, J., Honoré, B. (2020) 'Racial discrimination in the US housing and mortgage lending markets: a quantitative review of trends, 1976–2016', *Race and Social Problems*, 12/1: 13-28.

Quillian, L., Midtbøen, A. (2021) 'Comparative perspectives on racial discrimination in hiring: the rise of field experiments', *Annual Review of Sociology*, 47/1: 391-415.

Quillian, L., Midtbøen, A. (2026) 'The racialized penalties of immigrant origin: meta-analytic evidence on hiring discrimination in twelve countries', *Social Forces*, soag028.

Reed, W. R. (2023). *Univariate Regression Tests for Publication Bias: Why Do We Do Them*. MAER-Net blog post.

Richardson, W., Wilson, M., Nishikawa, J., Hayward, R. (1995) 'The well-built clinical question: a key to evidence-based decisions', *ACP Journal Club*, 123/3: A12-A13.

Ross, S., Turner, M. (2005) 'Housing discrimination in metropolitan America: explaining changes between 1989 and 2000', *Social Problems*, 52: 152-180.

Ruisch, B., Ferguson, M. (2022) 'Changes in Americans' prejudices during the presidency of Donald Trump', *Nature Human Behaviour*, 6/5: 656-665.

Schwegman, D. (2019) 'Rental market discrimination against same-sex couples: evidence from a pairwise-matched e-mail correspondence test', *Housing Policy Debate*, 29/2: 250-272.

Stanley, T. (2017) 'Limitations of PET-PEESE and other meta-analysis methods', *Social Psychological and Personality Science*, 8/5: 581-591.

Stanley, T., Doucouliagos, H. (2012) *Meta-regression analysis in economics and business*. London: Routledge.

Stanley, T., Doucouliagos, H. (2014) 'Meta-regression approximations to reduce publication selection bias', *Research Synthesis Methods*, 5/1: 60-78.

Turner, M., Levy, D., Wissoker, D. *et al.* (2013) *Housing discrimination against racial and ethnic minorities 2012*. Washington, DC: Urban Institute.

Turner, M., Ross, S., Galster, G., Yinger, J. (2002) *Discrimination in metropolitan housing markets: national results from Phase I HDS*. Washington, DC: Urban Institute.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. et al. (2021) 'Rank-normalization, folding, and localization: an improved R̂ for assessing convergence of MCMC', *Bayesian Analysis*, 16/2.

Verhaeghe, P.P. (2022) 'Correspondence studies'. In: Zimmermann, K. F. (ed.) *Handbook of labor, human resources and population economics*, pp. 1-19. Cham: Springer.

Verhaeghe, P.P. (2025) 'Five traditions of quantitative research on racism: their research objects, methods, assumptions and relationships', *Sociology Compass*, 19/12: e70141.

Verhaeghe, P.P., Van der Bracht, K., Van de Putte, B. (2016) 'Discrimination of tenants with a visual impairment on the housing market: empirical evidence from correspondence tests', *Disability and Health Journal*, 9/2: 234-238.

Verstraete, J., Verhaeghe, P.P. (2020) 'Ethnic discrimination upon request? Real estate agents' strategies for discriminatory questions of clients', *Journal of Housing and the Built Environment*, 35/3: 703-721.

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

Yavçan, B., Gorgerino, M. (2025) *MIPEX 2025: a roadmap for inclusive policy in the EU*. Brussels: Migration Policy Group.

Yinger, J. (1997) ‘Cash in your face: the cost of racial and ethnic discrimination in housing’, *Journal of Urban Economics*, 42/3: 339-365.

Appendix

Table A1. Eligibility criteria for study inclusion

Criterium	Details
Timing and language	Studies published or posted between January 2002 and December 2024 in English, French, German or Dutch.
Study type	Field experiments conducted in the rental housing market, specifically correspondence (e-mail/online) tests and telephone audits.
Population	Applicants from five test groups and their counterparts from control groups.
Discrimination grounds	Applicants assigned to a test group that signals ethno-racial origin, gender, disability, social origin, or sexual orientation, and their counterparts from the control group.
Outcome	An invitation to view a rental dwelling or another broad positive initial response from the housing provider, with raw counts of positive responses enabling computation of the discrimination ratio (or derivable equivalent).
Comparison	Positive responses to the applicants from the test group compared with their counterparts in the control group.

Note: Eligibility criteria are based on the PICO (Population, Intervention, Comparison, Outcome) framework by Richardson et al. (1995).

Table A2. Search strategy on Google Scholar and Web of Science to identify field experimental studies on rental discrimination

Concept	Search terms
Field experimental studies	"audit stud*" OR "audit design" OR "audit experiment*" OR "audit test*" OR "audit trial*" OR "field stud*" OR "field design" OR "field experiment*" OR "field test*" OR "field trial*" OR "correspondence stud*" OR "correspondence design" OR "correspondence experiment*" OR "correspondence test*" OR "correspondence trial*"
Housing context	"rental housing" OR "rental market" OR tenancy OR landlord OR lettings OR "housing application" OR housing
Discrimination concept	discriminat* OR “uneq* treat*” OR “ineq* treat*” OR “differ* treat*” OR “disparate treat*” OR “bias* treat*” OR “disadvant* treat*” OR “unfav* treat*” OR “prejudic* treat*” OR “unfair treat*”
Five discrimination grounds	(ethnic* OR race OR racial* OR nationality OR origin OR migrant* OR arab* OR maghreb* OR "middle eastern*" OR asian* OR african* OR american* OR indian* OR european* OR indigenous OR hispanic OR latin*) OR (*gender* OR *sex* OR wom* OR female*) OR (“sexual orient*“OR “sexual preference*” OR “sexual nature” OR queer OR gay OR lesbian OR lgbt*) OR (disab* OR impair* OR disorder OR blind* OR deaf* OR wheelchair OR autism) OR (wealth* OR affluen* OR rich OR fortun* OR prosperity OR poverty OR elite OR “low class” OR “high class” OR caste OR "housing voucher")

Notes: The concept blocks were combined using the Boolean operator AND. On Google Scholar, equivalent search strings were also applied in French, Dutch and German.

Table A3. Descriptive statistics of variables extracted from primary studies

Variable	k	Percent	Mean	SD	Min	Median	Max
Panel A. Continuous Variables							
Applications: Minority Group	475	—	518.38	962.87	9	244	8476
Applications: Majority Group	475	—	561.70	1008.81	2	252	8476
Callbacks: Minority Group	475	—	223.11	448.02	1	85	4848
Callbacks: Majority Group	475	—	288.22	532.96	1	119	5086
Positive Response Rate: Minority Group	475	—	0.44	0.18	0.04	0.42	0.96
Positive Response Rate: Majority Group	475	—	0.53	0.18	0.07	0.50	1.00
Log Positive Response Ratio	475	—	−0.22	0.27	−1.92	−0.18	0.41
Start Year of Data Collection	475	—	2015.63	4.76	1999	2016	2023
End Year of Data Collection	474	—	2016.07	4.81	1999	2017	2024
Panel B. Categorical Variables							
Ground							
Ethno-racial origin	331	69.68	—	—	—	—	—
Arab, North African, Turkish	113	34.14	—	—	—	—	—
Black, Central African, African American	62	18.73	—	—	—	—	—
Eastern European and other White	56	16.92	—	—	—	—	—
Asian	48	14.50	—	—	—	—	—
Hispanic, Southern American	32	9.67	—	—	—	—	—
Other	20	6.04	—	—	—	—	—
Gender	43	9.05	—	—	—	—	—
Female	43	100.00	—	—	—	—	—
Sexual orientation	40	8.42	—	—	—	—	—
Same-sex orientation	40	100.00	—	—	—	—	—
Social origin	34	7.16	—	—	—	—	—
Low social class	34	100.00	—	—	—	—	—
Disability	27	5.68	—	—	—	—	—
Physical disability	21	77.78	—	—	—	—	—
Mental disability	6	22.22	—	—	—	—	—
Region (aggregated)							
Europe	313	65.89	—	—	—	—	—
North America	145	30.53	—	—	—	—	—
Other	17	3.58	—	—	—	—	—
Application Detail Level							
Standard	255	53.68	—	—	—	—	—
Extensive	220	46.32	—	—	—	—	—
Type of Housing Provider							
Mixed or Unknown	247	52.00	—	—	—	—	—
Private owner	126	26.53	—	—	—	—	—
Real estate agent	102	21.47	—	—	—	—	—
Matched Design							
Yes	273	57.47	—	—	—	—	—
No	202	42.53	—	—	—	—	—
Mode of Application							
Email	433	91.16	—	—	—	—	—

Telephone or Mixed	42	8.84	—	—	—	—	—
Callback Type							
Positive reaction	348	73.26	—	—	—	—	—
Viewing invitation	127	26.74	—	—	—	—	—
Airbnb Indicator							
No	468	98.53	—	—	—	—	—
Yes	7	1.47	—	—	—	—	—
Candidate Gender							
Male	213	44.84	—	—	—	—	—
Female	139	29.26	—	—	—	—	—
Both	74	15.58	—	—	—	—	—
Unknown or NA	49	10.32	—	—	—	—	—
Candidate Education							
Various, Unknown or NA	420	88.42	—	—	—	—	—
Higher education	55	11.58	—	—	—	—	—
Candidate Employment							
Mixed, Unknown or NA	321	67.58	—	—	—	—	—
Employed	150	31.58	—	—	—	—	—
Unemployed	4	0.84	—	—	—	—	—
Candidate Migrant Generation							
Unknown or NA	434	91.37	—	—	—	—	—
First generation	22	4.63	—	—	—	—	—
Second generation	19	4.00	—	—	—	—	—
Peer-reviewed Study							
Yes	288	60.63	—	—	—	—	—
No	187	39.37	—	—	—	—	—
Study Language							
English	371	78.11	—	—	—	—	—
Dutch	94	19.79	—	—	—	—	—
Other	10	2.11	—	—	—	—	—

Notes: Abbreviations and acronyms used: *k* (number of effects), *SD* (standard deviation).

Table A4. Effect sizes by country

Country	k	Percent
Panel A. Region: Europe		
Austria	12	2.53
Belgium	96	20.21
Czechia	6	1.26
Finland	7	1.47
France	29	6.11
Germany	21	4.42
Greece	3	0.63
Iceland	6	1.26
Ireland	7	1.47
Italy	11	2.32
Netherlands	14	2.95
Norway	8	1.68
Poland	5	1.05
Portugal	2	0.42
Serbia	2	0.42
Spain	8	1.68
Sweden	21	4.42
Switzerland	5	1.05
United Kingdom	50	10.53
Panel B. Region: North America		
Canada	10	2.11
United States	135	28.42
Panel C. Region: Other		
China	3	0.63
India	2	0.42
Japan	1	0.21
Turkey	1	0.21
Australia	1	0.21
Argentina	2	0.42
Bahamas	1	0.21
Colombia	2	0.42
Ecuador	2	0.42
Peru	2	0.42

Notes: Abbreviations and acronyms used: *k* (number of effects).

Table A5. Publication bias: Egger's and Peters' tests for funnel plot asymmetry

Egger's and Peters' tests evaluate funnel plot asymmetry by regressing a standardized effect measure on study precision and test whether the intercept differs from zero (Egger et al. 1997; Peters et al. 2006). Specifically, Egger's test regresses the effect estimate on its standard error. A non-zero intercept indicates that smaller, less precise studies report systematically larger effects, consistent with selective reporting (Egger et al. 1997). Peters' test applies an approach adapted for binary outcomes, regressing the effect estimate on the inverse of the total sample size with inverse-variance weighting and likewise testing whether the intercept is zero (Peters et al. 2006). The PET and PEESE switching mechanism described in Section 4.3 of the main manuscript builds a publication-bias adjustment directly into the UWLS-MRA estimates. Unlike Egger's test, which evaluates funnel-plot asymmetry through the slope of the precision regression, the canonical rule tests the bias-corrected effect itself and selects PEESE only when a non-null effect remains after small-study bias correction.

Egger's regression test indicates significant funnel asymmetry for the ethno-racial origin ground (t = −9.933, p < 0.001), with less precise studies showing systematically larger discrimination effects in absolute magnitude. Significant asymmetry is also detected for disability (t = −4.316, p < 0.001). By contrast, Egger's test yields non-significant results for gender (t = 0.388, p = 0.700), sexual orientation (t = −0.986, p = 0.331), and social origin (t = −1.073, p = 0.291). Peters' test does not detect significant asymmetry within any individual ground (all per-ground p > 0.10), though at the full meta-sample level it is significant (t = 2.281, p = 0.023). This pattern, with Peters' test significant only at the pooled level, suggests that any selection on sample size operates across the discrimination literature as a whole rather than uniformly within each ground.

Ground and Test Group	***t***	***p***
Panel A. Egger's test		
Ethno-racial origin	−9.933	<0.001
Arab, North African, Turkish	−3.029	0.003
Black, Central African, African American	−7.427	<0.001
Eastern European and other White	0.232	0.818
Asian	−3.338	0.002
Hispanic, Southern American	−2.740	0.010
Other	−4.872	<0.001
Gender	0.388	0.700
Sexual orientation	−0.986	0.331
Social origin	−1.073	0.291
Disability	−4.316	<0.001
Physical disability	−3.752	0.001
Mental disability	—	—
Panel B. Peters' test		
Ethno-racial origin	1.639	0.101
Arab, North African, Turkish	1.311	0.190
Black, Central African, African American	−0.762	0.446
Eastern European and other White	0.333	0.739
Asian	1.842	0.065
Hispanic, Southern American	−0.218	0.828
Other	−0.420	0.675

Gender	0.272	0.785
Sexual orientation	−1.040	0.298
Social origin	1.125	0.260
Disability	−0.416	0.678
Physical disability	−0.596	0.552
Mental disability	—	—

Notes: Presented *t*-statistics relate to slope from a regression of the effect estimate on its standard error (Egger's test) or on the inverse total sample size (Peters' test). A significant test statistic indicates funnel plot asymmetry. M-dashes (—) indicate that tests were not computed due to an insufficient number of effects (required: $k \geq 10$). Significance threshold: $p < 0.05$.

References:

Egger, M., Davey Smith, G., Schneider, M. and Minder, C. (1997) 'Bias in meta-analysis detected by a simple, graphical test', *BMJ*, 315/7109: 629-634.

Peters, J., Sutton, A., Jones, D., Abrams, K. et al. (2006) 'Comparison of two methods to detect publication bias in meta-analysis', *JAMA*, 295/6: 676-680.

Table A6. Publication bias: P-curve results

Furthermore, the p-curve examines whether statistically significant p-values cluster just below conventional thresholds or are instead right-skewed, as expected when a true effect is present (Simonsohn et al. 2014). A flat or left-skewed distribution of significant p-values is consistent with p-hacking or selective reporting.

The *p*-curve analysis revealed strong right-skewness of statistically significant *p*-values across the full meta-sample (z = −50.185, p < 0.001, k = 218 significant effects) and within each ground, including ethno-racial origin (z = −43.582, p < 0.001, k = 177), social origin (z = −25.845, p < 0.001, k = 18), disability (z = −9.688, p < 0.001, k = 16), sexual orientation (z = −3.591, p < 0.001, k = 4), and gender (z = −3.658, p < 0.001, k = 4; significant effects expected in favor of women). A right-skewed *p*-curve indicates that statistically significant results concentrate sufficiently below the .05 threshold rather than clustering just beneath it, inconsistent with pervasive *p*-hacking as the primary explanation for the observed significance. For gender and sexual orientation, the analysis rested on only four significant effects in the expected direction, and the results should be interpreted with caution. Specifically for gender, the expected direction was positive (PRR > 1), consistent with the sign convention applied throughout: in this ground, female applicants received equal or preferential treatment relative to male applicants in the majority of audits.

Ground and Test Group	***z***	***p***	***k***	k_{sig}	**Side**
Overall	−50.185	<0.001	475	218	Left
Ethno-racial origin	−43.582	<0.001	331	177	Left
Arab, North African, Turkish	−32.359	<0.001	113	82	Left
Black, Central African, African American	−15.361	<0.001	62	40	Left
Eastern European and other White	−24.151	<0.001	56	23	Left
Asian	−10.630	<0.001	48	11	Left
Hispanic, Southern American	−3.245	0.001	32	8	Left
Other	−8.151	<0.001	20	13	Left
Gender	−3.658	<0.001	43	4	Right
Sexual orientation	−3.591	<0.001	40	4	Left
Social origin	−25.845	<0.001	34	18	Left
Disability	−9.688	<0.001	27	16	Left
Physical disability	−9.422	<0.001	21	12	Left
Mental disability	—	—	6	4	—

Notes: Abbreviations used: k (number of effects), k_{sig} (number of statistically significant effects). Presented z-statistics are the continuous *p*-curve statistic, which tests the skewness of the distribution of significant *p*-values at the 5% significance threshold (Simonsohn et al., 2014). Significant negative values indicate right-skewness consistent with evidential value under a publication process that does not report selectively. The last column indicates the expected direction of the pooled effect. M-dashes (—) indicate that tests were not computed due to an insufficient number of significant effects (required: $k_{sig} \geq 3$, $k \geq 10$). Significance threshold: p < 0.05.

Reference: Simonsohn, U., Nelson, L. and Simmons, J. (2014) 'P-curve: a key to the file drawer', *Journal of Experimental Psychology: General,* 143/2: 534-547.

Table A7. Publication bias: P-uniform* results

Relatedly, the p-uniform* estimator tests for selective reporting by modeling the distribution of test statistics conditional on statistical significance and on the expected direction of the effect (van Aert & van Assen 2026). It is estimated using maximum likelihood.

The likelihood-ratio statistic of the *p*-uniform* test, testing the presence of a true underlying effect, was significant across all five discrimination grounds: ethno-racial origin ($\chi^2(1)$ = 166.416, p < 0.001), social origin ($\chi^2(1)$ = 11.866, p < 0.001), sexual orientation ($\chi^2(1)$ = 9.706, p = 0.002), gender ($\chi^2(1)$ = 9.780, p = 0.002), and disability ($\chi^2(1)$ = 3.948, p = 0.047). Specifically for gender, the test was conducted on the right tail, consistent with female applicants receiving equal or preferential treatment in this ground. The test on the full meta-dataset likewise rejected the null ($\chi^2(1)$ = 176.921, p < 0.001). These results indicate that the concentration of significant results in the meta-dataset cannot plausibly be attributed entirely to the selective reporting of null findings.

Ground and Test Group	$\chi^2(1)$	***p***	***k***	**Side**
Overall	176.921	<0.001	475	Left
Ethno-racial origin	166.416	<0.001	331	Left
Arab, North African, Turkish	46.343	<0.001	113	Left
Black, Central African, African American	28.225	<0.001	62	Left
Eastern European and other White	40.019	<0.001	56	Left
Asian	31.907	<0.001	48	Left
Hispanic, Southern American	5.398	0.020	32	Left
Other	4.116	0.042	20	Left
Gender	9.780	0.002	43	Right
Sexual orientation	9.706	0.002	40	Left
Social origin	11.866	0.001	34	Left
Disability	3.948	0.047	27	Left
Physical disability	2.846	0.092	21	Left
Mental disability	—	—	6	—

Notes: Abbreviations used: *k* (number of effects). Presented χ^2-statistics with 1 degree of freedom are the likelihood ratio test statistic testing deviations from the theoretically expected direction of rental discrimination using the conditional distribution of statistically significant *p*-values at the 5% significance threshold (van Aert & van Assen, 2026). Significant values indicate that the data are inconsistent with a true zero effect after accounting for the selective reporting of statistically significant results. The last column indicates the expected direction of the pooled effect. M-dashes (—) indicate that tests were not computed due to an insufficient number of effects (required: $k \geq 10$). Significance threshold: p < 0.05.

Reference: van Aert, R. and van Assen, M. (2026) 'Correcting for publication bias in a meta-analysis with the p-uniform* method', *Psychonomic Bulletin & Review*, 33/3.

Table A8. Publication bias: Step-function selection model results

Finally, a step-function selection model assigns a higher probability of publication to results that cross a one-sided significance threshold (α = 0.025) to assess whether the pooled estimate would shift under a publication process that selects on statistical significance (Vevea & Hedges 1995).

The likelihood-ratio test (LRT) of the step-function selection model, testing the null hypothesis of no publication selection (two-sided $p < .05$; one-sided $p < .025$), was highly significant for the full meta-dataset ($\chi^2(1) = 22.922$, $p < 0.001$) and for ethno-racial origin ($\chi^2(1) = 10.698$, $p = 0.001$), confirming statistically detectable publication selection in the most prominent strand of the rental discrimination literature. The LRT also reached significance for gender ($\chi^2(1) = 5.679$, $p = 0.017$). This result is notable given that neither Egger's nor Peters' test detected funnel asymmetry in that ground and suggests that any selection in the gender literature may operate at the significance threshold rather than through a small-study effect. For disability ($\chi^2(1) = 2.463$, $p = 0.117$), social origin ($\chi^2(1) = 2.177$, $p = 0.140$), and sexual orientation ($\chi^2(1) = 0.272$, $p = 0.602$), the LRT did not indicate significant publication selection. However, statistical power to detect selection is limited in these grounds, given the smaller numbers of primary studies, so a non-significant result should not be interpreted as strong evidence against the presence of selection bias.

Ground and Test Group	$\chi^2(1)$	***p***	***k***
Overall	22.922	<0.001	475
Ethno-racial origin	10.698	0.001	331
Arab, North African, Turkish	1.040	0.308	113
Black, Central African, African American	1.396	0.237	62
Eastern European and other White	4.113	0.043	56
Asian	2.999	0.083	48
Hispanic, Southern American	0.142	0.706	32
Other	2.286	0.131	20
Gender	5.679	0.017	43
Sexual orientation	0.272	0.602	40
Social origin	2.177	0.140	34
Disability	2.463	0.117	27
Physical disability	2.827	0.093	21
Mental disability	—	—	6

Notes: Abbreviations used: *k* (number of effects). Presented χ^2-statistics with 1 degree of freedom are the likelihood ratio test statistic from a step-function selection model testing the relative probability that a non-significant result vs a significant result (at one-sided $p = 0.025$) is published for a given group of studies (Vevea & Hedges, 1995). Significant values are consistent with the selective reporting of statistically significant results. M-dashes (—) indicate that tests were not computed due to an insufficient number of effects (required: $k \geq 10$). Significance threshold: $p < 0.05$.

Reference: Vevea, J. and Hedges, L. (1995) 'A general linear model for estimating effect size in the presence of publication bias', *Psychometrika*, 60/3: 419-435.

Table A9. Bayesian hierarchical meta-regression (BH-MRA): Full meta-sample

	Full sample
Intercept	−1.110 [−14.368, 11.899]
Region *(ref.: Europe)*	
North America	0.092 [0.024, 0.161]
Other	−0.065 [−0.200, 0.066]
Year	0.000 [−0.006, 0.007]
Type of housing provider *(ref.: Private owner)*	
Real estate agent	0.165 [0.090, 0.240]
Mixed or Unknown	0.069 [0.006, 0.134]
Mode of application *(ref.: Email or online)*	
Telephone or Mixed	−0.032 [−0.109, 0.044]
Candidate employment status *(ref.: Employed)*	
Unemployed	−0.299 [−0.546, −0.038]
Mixed or Unknown	−0.042 [−0.117, 0.033]
Candidate gender *(ref.: Male)*	
Female	0.042 [0.002, 0.081]
Both	0.078 [0.017, 0.138]
Unknown or NA	0.135 [−0.005, 0.266]
Extra controls	
Study language	Yes
Peer-reviewed status	Yes
Airbnb indicator	Yes
Callback type	Yes
Matched design	Yes
Application detail level	Yes
Candidate education	Yes
k	474
N	114
j_{study}	114
j_{group}	11
τ_{study}	0.094 [0.068, 0.121]
τ_{group}	0.087 [0.049, 0.152]
σ	0.072 [0.048, 0.096]
Bayes R^2	0.308 [0.258, 0.357]

Notes: Abbreviations used: *k* (number of effects), *N* (number of studies), *j* (number of levels). BH-MRA is a Bayesian hierarchical meta-regression. The dependent variable is the natural log of the positive response ratio, where more negative values indicate stronger discrimination against the minority group. Point estimates are posterior means; values in brackets are 95% credible intervals. BH-MRA includes study-level (j_{study}) and test group-level (j_{group}) grouping terms. τ_{study} and τ_{group} are the posterior standard deviations of the respective random intercepts; σ is the posterior residual standard deviation. Bayes R^2 is the ratio of explained to total posterior variance (Gelman et al., 2019). BH-MRA used a Student-*t* likelihood to handle outlying effect sizes robustly and was estimated via Markov chain Monte Carlo (MCMC). Four chains were run for 10,000 iterations each (2,000 of which were warmup iterations), yielding 32,000 post-warmup draws. Priors are standard normal-distributed on the intercept and all fixed-effect coefficients, and half-Cauchy distributed with a scale of 0.5 on random-effect standard deviations and σ. All parameters reached convergence with $\hat{R} < 1.01$ and $ESS \geq 400$ (following Vehtari et al., 2021).

Table A10. Bayesian hierarchical meta-regression (BH-MRA): Estimates by discrimination ground

	Ethno-racial origin	Gender	Sexual orientation	Social origin
Intercept	−6.814 [−22.747, 8.830]	5.739 [−17.292, 29.340]	29.125 [−32.288, 87.938]	−43.283 [−168.566, 89.445]
Region *(ref.: Europe)*				
North America	0.154 [0.070, 0.239]	−0.037 [−0.188, 0.119]	0.144 [−0.091, 0.372]	0.063 [−0.573, 0.689]
Other	−0.066 [−0.251, 0.117]	—	0.184 [−0.859, 1.231]	0.170 [−0.568, 0.893]
Year	0.003 [−0.005, 0.011]	−0.003 [−0.014, 0.009]	−0.015 [−0.044, 0.016]	0.022 [−0.044, 0.084]
Type of housing provider *(ref.: Private owner)*				
Real estate agent	0.223 [0.150, 0.295]	−0.058 [−0.272, 0.177]	0.037 [−0.421, 0.493]	0.370 [−0.357, 1.037]
Mixed or Unknown	0.108 [0.038, 0.180]	−0.066 [−0.216, 0.085]	−0.053 [−1.103, 1.002]	0.045 [−0.518, 0.676]
Mode of application *(ref.: Email or online)*				
Telephone or Mixed	−0.010 [−0.103, 0.084]	−0.184 [−0.445, 0.067]	−0.073 [−0.640, 0.484]	−0.014 [−0.837, 0.731]
Candidate employment status *(ref.: Employed)*				
Unemployed	−0.137 [−0.489, 0.208]	−0.314 [−0.688, 0.064]	—	—
Mixed or Unknown	−0.031 [−0.121, 0.057]	−0.007 [−0.185, 0.166]	0.296 [−0.077, 0.657]	−0.404 [−0.862, 0.005]
Candidate gender *(ref.: Male)*				
Female	0.049 [0.011, 0.086]	—	0.063 [−0.042, 0.167]	−0.033 [−0.359, 0.278]
Both	0.051 [−0.025, 0.127]	—	0.174 [−0.847, 1.184]	0.252 [−0.396, 0.823]
Unknown or NA	0.076 [−0.199, 0.354]	—	—	0.302 [−0.656, 1.187]
Candidate migrant generation *(ref.: First generation)*				
Second generation	0.231 [0.138, 0.324]	—	—	—
Unknown or NA	0.134 [0.013, 0.252]	—	—	—
Extra controls				
Study language	Yes	Yes	Yes	Yes
Peer-reviewed status	Yes	Yes	Yes	Yes
Airbnb indicator	Yes	Yes	Yes	Yes
Callback type	Yes	Yes	Yes	Yes
Matched design	Yes	Yes	Yes	Yes
Application detail level	Yes	Yes	Yes	Yes
Candidate education	Yes	Yes	Yes	Yes
k	331	42	40	34
N	94	27	18	23
j_{study}	94	27	18	23
j_{group}	6	—	—	—
τ_{study}	0.116 [0.089, 0.147]	0.077 [0.008, 0.159]	0.060 [0.002, 0.193]	0.243 [0.025, 0.500]
τ_{group}	0.094 [0.043, 0.211]	—	—	—
σ	0.047 [0.023, 0.070]	0.047 [0.003, 0.112]	0.078 [0.009, 0.158]	0.160 [0.019, 0.335]

Bayes R^2	0.382 [0.327, 0.437]	0.443 [0.262, 0.629]	0.390 [0.231, 0.512]	0.581 [0.310, 0.822]

Notes: Abbreviations used: *k* (number of effects), *N* (number of studies), *j* (number of levels). BH-MRA is a Bayesian hierarchical meta-regression. The dependent variable is the natural log of the positive response ratio, where more negative values indicate stronger discrimination against the minority group. Point estimates are posterior means; values in brackets are 95% credible intervals. BH-MRA includes study-level (j_{study}) and test group-level (j_{group}) grouping terms. τ_{study} and τ_{group} are the posterior standard deviations of the respective random intercepts; σ is the posterior residual standard deviation. Test group random intercepts are only included when multiple test groups exists within a ground. Bayes R^2 is the ratio of explained to total posterior variance. BH-MRA used a Student-*t* likelihood to handle outlying effect sizes robustly and was estimated via Markov chain Monte Carlo (MCMC). Four chains were run for 10,000 iterations each (2,000 of which were warmup iterations), yielding 32,000 post-warmup draws. Priors are standard normal-distributed on the intercept and all fixed-effect coefficients, and half-Cauchy distributed with a scale of 0.5 on random-effect standard deviations and σ. All parameters reached convergence with $\hat{R} < 1.01$ and $ESS \geq 400$ (following Vehtari et al., 2021).

Table A11. RE-MRA with region–year interaction: Full meta-sample

	Full sample
Intercept	6.423 (10.462)
Region *(ref.: Europe)*	
North America	−16.592 (16.683)
Other	−126.796 (85.982)
Year	−0.003 (0.005)
Type of housing provider *(ref.: Private owner)*	
Real estate agent	0.184*** (0.044)
Mixed or Unknown	0.087 (0.049)
Mode of application *(ref.: Email or online)*	
Telephone or Mixed	−0.043 (0.051)
Candidate education *(ref.: Higher educated)*	
Various or Unknown	0.053 (0.049)
Candidate employment status *(ref.: Employed)*	
Unemployed	−0.362 (0.147)
Mixed or Unknown	−0.129 (0.069)
Candidate gender *(ref.: Male)*	
Female	0.063* (0.021)
Both	0.055 (0.046)
Unknown or NA	0.062 (0.131)
Interaction between Region and Year	
North America * Year	0.008 (0.008)
Other * Year	0.063 (0.043)
k	475
N	114
τ^2	0.031
Pseudo R^2	0.228
Other controls	Test group; Matched design; Application detail level; Callback type; Airbnb indicator, Peer-Reviewed status, Study language

Notes: Abbreviations used: *k* (number of effects), *N* (number of studies). RE-MRA is a random-effects meta-regression estimated via restricted maximum likelihood (REML). The dependent variable is the natural log of the positive response ratio, where more negative values indicate stronger discrimination against the minority group. Point estimates are mean regression coefficients; values in parentheses are cluster-robust standard errors with small-sample correction (CR2; Tipton & Pustejovsky, 2015). τ^2 is the estimated between-study heterogeneity variance. Pseudo R^2 reflects the proportion of variance explained by the model's predictors. The interaction terms test whether the temporal trend in discrimination differs across regions. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A12. RE-MRA with region–year interaction: Estimates by discrimination ground

	Ethno-racial origin	Gender	Social origin
Intercept	−2.254 (13.192)	9.931 (18.390)	−96.980 (96.615)
Region *(ref.: Europe)*			
North America	−4.488 (18.626)	−9.045 (20.164)	395.657 (284.845)
Other	−511.065*** (39.789)		−118.487 (141.821)
Year	0.001 (0.007)	−0.005 (0.009)	0.047 (0.048)
Type of housing provider *(ref.: Private owner)*			
Real estate agent	0.241*** (0.036)	−0.061 (0.113)	0.674 (0.511)
Mixed or Unknown	0.124** (0.043)	−0.066 (0.052)	0.167 (0.493)
Mode of application *(ref.: Email or online)*			
Telephone or Mixed	−0.018 (0.052)	−0.169 (0.089)	0.288 (0.360)
Candidate education *(ref.: Higher educated)*			
Various or Unknown	−0.019 (0.081)	0.022 (0.031)	
Candidate employment status *(ref.: Employed)*			
Unemployed	−0.118 (0.141)	−0.347** (0.076)	
Mixed or Unknown	−0.018 (0.079)	−0.010 (0.053)	−0.315** (0.005)
Candidate gender *(ref.: Male)*			
Female	0.050* (0.018)		−0.031 (0.075)
Both	0.055 (0.035)		0.359 (0.566)
Unknown or NA	0.113* (0.049)		0.438 (0.664)
Candidate Migrant Generation *(ref.: First generation)*			
Second generation	0.229* (0.007)		
Unknown or NA	0.151 (0.155)		
Interaction between Region and Year			
North America * Year	0.002 (0.009)	0.004 (0.010)	−0.196 (0.141)
Other * Year	0.253*** (0.020)		0.059 (0.070)
k	331	43	34
N	94	28	23
τ^2	0.022	0.009	0.144
Pseudo R^2	0.244	0.097	0.000
Other controls	Test group; Matched design; Application detail level; Callback type; Airbnb indicator, Peer-Reviewed status, Study language	Matched design; Application detail level; Callback type; Airbnb indicator, Peer-Reviewed status, Study language	Matched design; Application detail level; Callback type; Airbnb indicator, Peer-Reviewed status, Study language

Notes: Abbreviations used: *k* (number of effects), *N* (number of studies). RE-MRA is a random-effects meta-regression estimated via restricted maximum likelihood (REML). The dependent variable is the natural log of the positive response ratio, where more negative values indicate stronger discrimination against the minority group. Point estimates are mean regression coefficients; values in parentheses are cluster-robust standard errors with small-sample correction (CR2; Tipton & Pustejovsky, 2015). τ^2 is the estimated between-study heterogeneity variance. Pseudo R^2 reflects the proportion of variance explained by the model's predictors. The interaction terms test whether the temporal trend in discrimination differs across regions. M-dashes (—) indicate that a coefficient was not estimable for that ground due to insufficient variation in the predictor. Grounds with insufficient degrees of freedom for reliable estimation (sexual orientation, health and disability) are excluded from the table. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A13. Discrimination ratios by discrimination ground and country

Ground	Country	k	Studies	UWLS (bias-corrected) [95% CI]	RE-MRA [95% CI]
Ethno-racial origin	Austria	12	1	0.69 [0.64, 0.73]	0.64 [0.52, 0.79]
	Belgium	62	21	0.79 [0.74, 0.86]	0.77 [0.72, 0.81]
	Canada	7	2	0.92 [0.88, 0.97]	0.94 [0.75, 1.18]
	China	2	1	0.81 [0.79, 0.83]	0.87 [0.67, 1.13]
	Czechia	6	1	0.58 [0.55, 0.60]	0.64 [0.48, 0.86]
	Finland	6	2	0.79 [0.73, 0.86]	0.84 [0.69, 1.02]
	France	24	7	0.81 [0.75, 0.87]	0.76 [0.68, 0.83]
	Germany	17	9	0.86 [0.79, 0.93]	0.78 [0.71, 0.86]
	Greece	3	1	0.57 [0.53, 0.61]	0.63 [0.50, 0.79]
	Iceland	4	1	0.94 [0.81, 1.08]	0.84 [0.58, 1.21]
	Ireland	4	1	0.70 [0.66, 0.75]	0.75 [0.56, 1.02]
	Italy	8	1	0.75 [0.70, 0.80]	0.76 [0.56, 1.03]
	Netherlands	7	5	0.80 [0.70, 0.90]	0.78 [0.65, 0.93]
	Norway	4	1	0.88 [0.82, 0.94]	0.79 [0.62, 1.01]
	Poland	4	1	0.85 [0.77, 0.93]	0.71 [0.54, 0.95]
	Spain	5	2	0.81 [0.76, 0.87]	0.77 [0.64, 0.91]
	Sweden	10	5	0.72 [0.64, 0.83]	0.65 [0.57, 0.74]
	Switzerland	4	3	0.86 [0.79, 0.93]	0.96 [0.81, 1.14]
	United Kingdom	42	5	0.81 [0.74, 0.90]	0.86 [0.77, 0.96]
	United States	97	22	0.91 [0.88, 0.93]	0.87 [0.82, 0.92]
Gender	Belgium	7	3	0.97 [0.88, 1.08]	0.96 [0.83, 1.11]
	Iceland	2	1	0.99 [0.88, 1.11]	0.98 [0.66, 1.47]
	Italy	2	1	1.04 [1.00, 1.08]	1.04 [0.80, 1.35]
	Norway	2	1	1.12 [1.12, 1.12]	1.12 [0.85, 1.48]
	Sweden	7	3	1.15 [0.99, 1.32]	1.18 [1.01, 1.39]
	United Kingdom	4	2	1.00 [0.81, 1.23]	0.99 [0.80, 1.23]
	United States	9	8	1.04 [0.90, 1.20]	1.02 [0.92, 1.13]
Disability	Belgium	16	7	0.81 [0.72, 0.92]	0.79 [0.69, 0.89]
	Netherlands	2	1	1.04 [0.69, 1.56]	0.75 [0.48, 1.20]
	United States	7	4	0.88 [0.80, 0.96]	0.86 [0.72, 1.02]
Sexual orientation	Argentina	2	1	0.94 [0.94, 0.95]	0.93 [0.78, 1.12]
	Belgium	4	2	1.00 [0.98, 1.02]	1.00 [0.87, 1.15]
	Canada	2	1	0.91 [0.86, 0.96]	0.92 [0.77, 1.11]
	Colombia	2	1	0.96 [0.80, 1.15]	0.90 [0.63, 1.28]
	Ecuador	2	1	1.22 [0.99, 1.50]	1.13 [0.78, 1.65]
	Germany	2	1	0.94 [0.92, 0.96]	0.95 [0.80, 1.12]
	Ireland	2	1	0.82 [0.80, 0.84]	0.80 [0.65, 0.98]
	Netherlands	4	3	0.90 [0.87, 0.93]	0.88 [0.74, 1.05]
	Peru	2	1	0.94 [0.80, 1.12]	0.88 [0.63, 1.25]
	Portugal	2	1	0.86 [0.81, 0.91]	0.84 [0.66, 1.06]
	Serbia	2	1	0.87 [0.80, 0.94]	0.89 [0.75, 1.06]
	Sweden	2	2	0.98 [0.96, 1.00]	0.99 [0.84, 1.17]
	United States	12	5	0.97 [0.88, 1.07]	1.01 [0.94, 1.08]
Social origin	Belgium	7	5	0.82 [0.58, 1.17]	0.80 [0.56, 1.15]
	France	3	2	0.75 [0.62, 0.92]	0.60 [0.34, 1.03]

India	2	1	0.40 [0.30, 0.52]	0.45 [0.22, 0.90]
Norway	2	1	0.92 [0.92, 0.92]	0.92 [0.45, 1.88]
Spain	2	2	0.77 [0.62, 0.95]	0.87 [0.48, 1.57]
Sweden	2	1	0.57 [0.44, 0.76]	0.73 [0.33, 1.64]
United Kingdom	4	2	0.76 [0.46, 1.25]	1.05 [0.61, 1.81]
United States	10	7	0.67 [0.46, 0.97]	0.68 [0.51, 0.92]

Notes: Abbreviations and acronyms used: *k* (number of effects), Studies (number of distinct audit studies), RE-MRA (random-effects meta-regression), UWLS (unrestricted weighted least squares). Country-level estimates are derived from models fitted simultaneously across all countries; the United States serves as the reference category. The UWLS column reports the publication-bias-corrected country-level estimate from the preferred PET or PEESE specification; PEESE is selected when the PET fit's bias-corrected effect (i.e., predicted value at SE = 0 with covariates at their sample means) is significantly non-null at a two-sided $\alpha = 0.10$; otherwise PET. The expected direction is negative for all grounds except gender, where it is positive. The RE-MRA column reports the country-level estimate from a random-effects meta-regression with nested random effects (clustering within studies). Positive response ratios and their confidence intervals were back-transformed from the original log specification; values below 1 indicate that the minority or stigmatised group receives fewer positive responses than the majority or non-stigmatised group. M-dashes (—) indicate that country-level estimation was not possible due to model convergence failure or a collinear country dummy. Only countries contributing at least two effect sizes ($k \geq 2$) are included.

Table A14. RE-MRA with five-year period categories: Full meta-sample

	Full sample
Intercept	-0.423 (0.131)
Region *(ref.: Europe)*	
North America	0.112* (0.045)
Other	-0.036 (0.126)
Five-year periods *(ref.: 2002-2005)*	
2006-2010	0.158 (0.064)
2011-2015	0.139 (0.070)
2016-2020	0.092 (0.068)
2021-2024	0.131 (0.088)
Type of housing provider *(ref.: Private owner)*	
Real estate agent	0.184** (0.049)
Mixed or Unknown	0.090 (0.054)
Telephone or Mixed	-0.085 (0.052)
Candidate education *(ref.: Higher educated)*	
Various or Unknown	0.064 (0.051)
Candidate employment status *(ref.: Employed)*	
Unemployed	-0.342 (0.161)
Mixed or Unknown	-0.126 (0.069)
Candidate gender *(ref.: Male)*	
Female	0.065* (0.021)
Both	0.072 (0.045)
Unknown or NA	0.054 (0.145)
k	475
N	114
τ^2	0.032
Pseudo R^2	0.220
Other controls	Test group; Matched design; Application detail level; Callback type; Airbnb indicator, Peer-Reviewed status, Study language

Notes: Abbreviations used: *k* (number of effects), *N* (number of studies). RE-MRA is a random-effects meta-regression estimated via restricted maximum likelihood (REML). The dependent variable is the natural log of the positive response ratio, where more negative values indicate stronger discrimination against the minority group. This specification replaces the continuous year variable in the main RE-MRA model (Table 2) with five-year period indicators, using 2002–2005 as the reference category, to probe non-linearity in the time trend. Point estimates are mean regression coefficients; values in parentheses are cluster-robust standard errors with small-sample correction (CR2; Tipton & Pustejovsky, 2015). τ^2 is the estimated between-study heterogeneity variance. Pseudo R^2 reflects the proportion of variance explained by the model's predictors. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A15. RE-MRA with five-year period categories: Estimates by discrimination ground

	Ethno-racial origin	Gender	Social origin
Intercept	-0.710* (0.198)	0.152 (0.094)	0.175 (1.004)
Region *(ref.: Europe)*			
North America	0.161** (0.050)	0.015 (0.073)	0.535 (0.476)
Other	-0.065 (0.177)	0.202 (0.183)	
Five-year periods *(ref.: 2002-2005)*			
2006-2010	0.143 (0.064)	-0.063 (0.074)	-0.230 (0.698)
2011-2015	0.181 (0.068)	-0.100 (0.079)	-0.536 (0.641)
2016-2020	0.135 (0.070)	-0.103 (0.073)	-0.612 (0.275)
2021-2024	0.136 (0.088)	0.028 (0.091)	
Type of housing provider *(ref.: Private owner)*			
Real estate agent	0.238*** (0.038)	-0.074 (0.113)	0.645 (0.544)
Mixed or Unknown	0.142** (0.045)	-0.073 (0.054)	-0.118 (0.427)
Telephone or Mixed	-0.023 (0.054)	-0.265 (0.096)	0.286 (0.285)
Candidate education *(ref.: Higher educated)*			
Various or Unknown	-0.008 (0.084)	0.011 (0.034)	
Candidate employment status *(ref.: Employed)*			
Unemployed	-0.082 (0.149)	-0.240 (0.102)	
Mixed or Unknown	-0.005 (0.083)	-0.005 (0.049)	-0.315* (0.006)
Candidate gender *(ref.: Male)*			
Female	0.051* (0.018)	-0.062 (0.058)	
Both	0.053 (0.041)	0.317 (0.359)	
Unknown or NA	0.078 (0.067)	0.922 (0.396)	
Candidate Migrant Generation *(ref.: First generation)*			
Second generation	0.229* (0.008)		
Unknown or NA	0.155 (0.158)		
k	331	43	34
N	94	28	23
τ^2	0.023	0.008	0.115
Pseudo R^2	0.198	0.124	0.149
Other controls	Test group; Matched design; Application detail level; Callback type; Airbnb indicator, Peer-Reviewed status, Study language	Matched design; Application detail level; Callback type; Airbnb indicator, Peer-Reviewed status, Study language	Matched design; Application detail level; Callback type; Airbnb indicator, Peer-Reviewed status, Study language

Notes: Abbreviations used: *k* (number of effects), *N* (number of studies). RE-MRA is a random-effects meta-regression estimated via restricted maximum likelihood (REML). The dependent variable is the natural log of the positive response ratio, where more negative values indicate stronger discrimination against the minority group. This specification replaces the continuous year variable in the ground-specific RE-MRA models (Table 3) with five-year period indicators, using 2002–2005 as the reference category. Point estimates are mean regression coefficients; values in parentheses are cluster-robust standard errors with small-sample correction (CR2; Tipton & Pustejovsky, 2015). τ^2 is the estimated between-study heterogeneity variance. Pseudo R^2 reflects the proportion of variance explained by the model's predictors. M-dashes (—) indicate that a coefficient was not estimable for that ground due to insufficient variation in the predictor or degrees-of-freedom constraints. Grounds with insufficient degrees of freedom for reliable estimation (sexual orientation, health and disability) are excluded from the table. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A16. RE−MRA with country fixed effects: Full meta-sample and ethno-racial origin ground

	Full sample	Ethno-racial origin
Intercept	−4.051 (9.987)	−2.247 (11.146)
Year	0.002 (0.005)	0.001 (0.006)
Type of housing provider *(ref.: Private owner)*		
Real estate agent	0.212*** (0.048)	0.262*** (0.042)
Mixed or Unknown	0.112* (0.043)	0.183*** (0.042)
Mode of application *(ref.: Email or online)*		
Telephone or Mixed	−0.011 (0.046)	0.025 (0.044)
Candidate employment status *(ref.: Employed)*		
Unemployed	−0.442 (0.076)	−0.332* (0.149)
Mixed or Unknown	−0.156* (0.061)	−0.040 (0.078)
Candidate gender *(ref.: Male)*		
Female	0.066** (0.020)	0.051* (0.017)
Both	0.063 (0.055)	0.021 (0.033)
Unknown or NA	0.067 (0.168)	0.079 (0.101)
Candidate migrant generation *(ref.: First generation)*		
Second generation		0.228* (0.008)
Unknown or NA		0.140 (0.142)
Extra controls		
Country	Yes	Yes
Study language	Yes	Yes
Peer-reviewed status	Yes	Yes
Airbnb indicator	Yes	Yes
Callback type	Yes	Yes
Matched design	Yes	Yes
Application detail level	Yes	Yes
Candidate education	Yes	Yes
Test group	Yes	No
k	475	331
N	114	94
τ^2	0.027	0.018
Pseudo-R^2	0.329	0.375

Notes: Abbreviations used: *k* (number of effects), *N* (number of studies). RE-MRA is a random-effects meta-regression estimated via restricted maximum likelihood (REML). The dependent variable is the natural log of the positive response ratio, where more negative values indicate stronger discrimination against the minority group. This specification replaces the aggregated region indicator in the main RE-MRA model (Table 2) with country fixed effects (reference category: United States). Point estimates are mean regression coefficients; values in parentheses are cluster-robust standard errors with small-sample correction (CR2; Tipton & Pustejovsky, 2015). τ^2 is the estimated between-study heterogeneity variance. Pseudo R^2 reflects the proportion of variance explained by the model's predictors. Country fixed effect coefficients are omitted from the table for conciseness. Regressions for other discrimination grounds are not shown due to degrees-of-freedom constraints or convergence issues under the full country fixed effects specification. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Figure A1. Evolution over time in the number of effect sizes k per discrimination ground