

Does Size Matter? Hiring Discrimination and Firm Size*

By Stijn Baert,ⁱ Ann-Sofie De Meyer,ⁱⁱ Yentl Moerman,ⁱⁱⁱ and Eddy Omey^{iv}

Abstract

We study the association between firm size and hiring discrimination against women, ethnic minorities and older job candidates. To this end, we merge field experimental measures on unequal treatment against these groups with firm-level data. These data enable us to assess whether discrimination varies by indicators of firm size, keeping other firm characteristics constant. In contrast with our theoretical expectations, we find no evidence for an association between firm size and hiring discrimination. On the other hand, we do find suggestive evidence for hiring discrimination being lower in respect of public or non-profit firms (compared to commercial firms).

Keywords: discrimination, hiring, firm size, gender, ethnicity, age.

JEL: J71, J23, J16, J15, J14.

Word count: 7645 words.

* **Acknowledgements.** We are grateful to Jens Agerström, Wouter De Tavernier, Gunn Elisabeth Birkelund, Ian Burn, Emmanuel Duguet, Margaret Maurer-Fazio, Laura Naegele, and Dan-Olof Rooth for their comments, which helped us to improve this study substantially.

ⁱ **Corresponding author.** Ghent University, University of Antwerp, Université catholique de Louvain, IZA, GLO, and IMISCOE. 003292643481. Stijn.Baert@UGent.be. <http://users.UGent.be/~sbaert>.

ⁱⁱ Ghent University.

ⁱⁱⁱ Ghent University.

^{iv} Ghent University.

1. Introduction

The lack of labour market integration of vulnerable groups receives much attention in both policy and academic circles (see, e.g., Lahey, 2008; OECD, 2008; OECD, 2010; Oreopoulos, 2011). During the past decade, evidence from large-scale field experiments has shown that discrimination in hiring is a key factor contributing to this poor labour market integration. In particular, substantial levels of hiring discrimination against women, ethnic minorities, and older job candidates have been measured in countries such as Australia, Belgium, China, France, Germany, Mexico, Spain, Sweden, the United Kingdom, and the United States (Albert et al., 2011; Arceo-Gomez and Campos-Vazques, 2014; Baert, 2017; Baert et al., 2015; Booth et al., 2012; Carlsson, 2010; Drydakis and Vlassis, 2010; Neumark, in press; Petit, 2007; Zhou et al., 2013). However, measuring discrimination is one thing, tackling it is another. To effectively combat hiring discrimination, one needs to understand its driving factors. In other words, to design adequate policy actions, targeted to the right employers in the right way, one has to gain insight into when individuals are discriminated in particular, i.e. into the moderators of labour market discrimination. In this study, we focus on firm size as a moderator of hiring discrimination.

From a theoretical point of view, many studies have provided arguments for a negative relationship between firm size and hiring discrimination. We distinguish between two clusters of arguments for the expectation that larger firms discriminate less. First, larger companies are expected to be more objective because of their large-scale organisation (Baert and Omey, 2015; Carlsson and Rooth, 2007; Maurer-Fazio, 2012; Wood et al., 2009). They have more often a dedicated human resource department and make more often use of standardised recruitment procedures. Moreover, they are documented as having more capacity to invite a higher number of candidates for an interview and, as a consequence, may rely less on statistical discrimination when making first hiring decisions (Arrow, 1973). Finally, as argued by a reviewer of a former version of this study, larger companies may have more accurate information about the distribution of unobserved characteristics of job applicants because they hire more people and, thereby, learn faster about the true distribution of skills than smaller firms do. Again, this may result in a lower level of statistical discrimination.

Second, larger companies are also expected to be *forced to be* more objective (Baert and Omey, 2015; Bonoli, 2017; Cornelissen and Jirjahn, 2012; Goerke and Pannenberg, 2011). Indeed, these larger firms are, on average, confronted with stronger labour unions so that discrimination is riskier. Moreover, in some countries, particular anti-discrimination policies only affect large companies. Finally, it is argued that it is more difficult to monitor a large workforce so that recruiting the most productive candidate is more important (and there is no space for decisions based simply on employer tastes or prejudices) for larger firms.¹

Also from an empirical point of view, there is suggestive evidence for this negative association between firm size and hiring discrimination. More concretely, the hiring discrimination against minorities in Germany, Sweden, and the United Kingdom measured in Carlsson and Rooth (2007), Kaas and Manger (2012), and Wood et al. (2009) is (to a diverging extent) driven by the tested vacancies from smaller firms.² This empirical evidence is, however, limited in at least two aspects. First, these studies investigated the association between hiring discrimination and firm size within the context of (one experiment focused on) one discrimination ground only, i.e. ethnicity. Thereby, it remains unsure whether this association is also relevant with respect to other forms of hiring discrimination. As a consequence, these studies' results are not very informative to policy makers aiming to tackle hiring discrimination in general in their region. Second, these studies were mainly focused on the causal measurement of the prevalence of discrimination and only in second order on its association with firm size. As a consequence, they included hardly any controls for other firm-level drivers of discrimination in their statistical analyses. Thereby, the negative association they found might be driven by an omitted variable problem.

In the present study, we are the first to study the heterogeneity in hiring discrimination by firm size with respect to multiple discrimination grounds. To this end, we merge data from three correspondence experiments, conducted to identify hiring discrimination based

¹ The latter argument is related to the well-known prediction of the seminal theory of taste-based discrimination (Becker, 1957) that firms which discriminate based on tastes or (irrational) prejudices cannot survive in a competitive market.

² On the other hand, no significant heterogeneity by firm size in ethnic discrimination was found in Maurer-Fazio (2012) for China. In addition, Baert and Omey (2015) reported that discrimination based on youth union membership does not vary with firm size in Belgium.

on gender, ethnicity, and age in Belgium, with firm-level data. Thereby, we are able to assess whether unequal treatment based on these grounds varies by different indicators of firm size, keeping other firm characteristics constant.

2. Data

2.1 Field Experimental Data

Our analyses are based on the data gathered in Baert et al. (2016a), Baert et al. (2016b), and Baert et al. (2017), which are three correspondence experiments conducted in Flanders (the Northern, Dutch speaking part of Belgium). In correspondence experiments, fictitious job applications, differing only in a randomly assigned discrimination ground, signalled through, e.g., the name or age of the applicant, are sent to genuine vacancies. By monitoring the subsequent call-back from employers, unequal treatment based on this single characteristic is identified and can be given a causal interpretation. For this reason, these experiments are often referred to as the golden standard to measure hiring discrimination (Baert, 2017; Bertrand and Mullainathan, 2004; Neumark, in press). In addition, nowadays, this kind of experiments is used to measure other success factors in getting hired such as unemployment duration, education level, and Facebook profile picture appearance (Baert, in press; Baert and Verhaest, 2014; Eriksson and Rooth, 2014; Kroft et al., 2013; Verhaest et al., in press).

In what follows, we summarise the data gathering process of the three experiments on which we build.³ First, Baert et al. (2016a) investigated the importance of employer preferences in explaining Sticky Floors, i.e. the fact that women are, compared to men, less likely to (start to) climb the job ladder. To this end, these authors performed a

³ The reader will notice that these experiments were not designed to test the association between firm size and hiring discrimination. Stated otherwise, our hypothesis of lower discrimination in larger firms was constructed *ex post*. Limitations related to *ex post* hypotheses and multiple testing are discussed in, e.g., Croson (2005) and Fink et al. (2014). In this perspective, it is relevant to highlight that the mentioned experimental data have not been used for analyses other than those mentioned in Baert et al. (2016a), Baert et al. (2016b), and Baert et al. (2017).

correspondence experiment between October 2013 and March 2014. In this experiment, pairs of job candidates with five years of work experience were sent to 248 vacancies for jobs at the same functional level as the candidates' current job and to 328 vacancies for jobs at a higher level.⁴ The male and female gender of the applicants was randomly assigned to the pair members who applied for these jobs. At the level of the full sample, no gender discrimination was found. However, female applicants received, compared to males, about 19% fewer positive reactions when they applied for a job at a higher functional level.

Second, Baert et al. (2017) conducted a correspondence experiment to measure ethnic discrimination. They sent 768 fictitious job applications in response to 384 vacancies, between December 2015 and May 2016. Five different ethnic origins were randomly assigned to these applications. One applicant always had a typically Flemish name, the other one a typically Turkish, Moroccan, Slovakian, or Ghanaian name. Job candidates with a non-Flemish name got about 32% less positive call-back.

Third, Baert et al. (2016b) conducted a correspondence experiment on age discrimination. More concretely, 1152 paired job applications, aged 38, 44, or 50, were sent to 576 vacancies, between December 2014 and May 2015. The 6 or 12 years older candidate within the pair had, on average, a 28% lower probability of receiving some sort of positive reaction.

In the present study, we build on the full datasets of the second and third experiment. With respect to the first experiment, on gender discrimination, we rely on the data for the 328 vacancies for jobs at a higher functional level only.⁵ Based on these data, we constructed two dependent variables for our analyses. First, we created an indicator variable that is 1 in case the candidate from the discriminated group (so, the female, foreign, or older candidate) got no positive reaction from the employer while the candidate from the control group (so, the male, native, or younger candidate) within the same pair got a positive reaction, and 0

⁴ All tested vacancies were matched with an occupation in the ISCO-08 (International Standard Classification of Occupations) classification system. A promotion in terms of functional level was defined as a transition to a new job at a higher ISCO-08 one-digit level (compared with the candidate's current job). We provide a discussion of the tested occupations and sectors in Section 2.3.

⁵ However, analyses including also the data for the other 248 vacancies yielded the same empirical pattern as the one described in Section 3.1.

otherwise.⁶ Second, for our alternative dependent variable, we restricted this indicator to be 1 only in case the candidate from the discriminated group was not invited to a job interview while the candidate from the control group was immediately invited (and 0 otherwise). In the remainder of this article, we will refer to these variables as ‘discrimination in positive reaction’ and ‘discrimination in invitation’.⁷

In addition, we retain for all vacancies in the three experiments an indicator of the tested occupation. More concretely, each vacancy was linked to an occupation in the ISCO-08 International Standard Classification of Occupations (ISCO-08) at the 2-digit level. This is an interesting control variable as (i) certain occupations may be more common in bigger firms and (ii) previous research has suggested that discrimination varies by factors at the occupation level.⁸ Therefore, in our analyses, we will control for occupation fixed effects. As the résumés used in the three experiments were fixed at the occupation level, these occupation fixed effects also control for candidate characteristics such as education level.⁹

Finally, based on the firm name, contact address, and contact person mentioned in the vacancy, we were able to look up the Belgian tax ID number of the firm for each vacancy. This ID number was used to merge the data for the tested vacancies with data on firm size and firm-level controls from Bel-first (for 2015).

2.2 Firm Data

The Bel-first firm data contain comprehensive information on company financials, detailed corporate structures, shareholders, subsidiaries, and legal form concerning about 2 million companies in Belgium and Luxembourg. These data are gathered by the private company Bureau Van Dijk.¹⁰ From these Bel-first data we retrieved further control variables for our

⁶ We follow the literature by defining a positive reaction as (i) an invitation for a job interview, (ii) a question to provide more information, (iii) a question to call back, or (iv) the proposition of an alternative job.

⁷ We return to alternative dependent variables in Section 3.2.

⁸ In particular, Baert et al. (2015) provide suggestive evidence for ethnic discrimination to be lower in occupations where recruitment is difficult (due to high labour market tightness).

⁹ Baert et al. (2015), Carlsson and Rooth (2007), and Wood et al. (2009) provide suggestive evidence for ethnic discrimination to be lower against high-educated candidates.

¹⁰ For more information, we refer to the website of Bureau Van Dijk: <https://www.bvdinfo.com/en-gb/our->

analysis. In total, firm-level data were included in Bel-first for 253 vacancies (72.7%) of the experiment on gender discrimination, 337 vacancies (87.8%) of the experiment on ethnic discrimination, and 427 vacancies (74.1%) of the experiment on age discrimination.¹¹

With respect to the independent variables of our analysis discussed in Section 3, we selected multiple indicators of firm size. Our benchmark variable in this respect is the number of workers in the firm in 2015 divided by 1000. However, as this variable is expected to be right-skewed, we also constructed a logarithmic transformed version of this variable. Finally, we included three indicators of firm size following the definitions in Bel-first.¹² A big firm following these definitions is defined as a firm that meets at least two of the following criteria: (i) revenue of at least 50 million euro, (ii) operating income of at least 100 million euro, (iii) at least 300 workers, and (iv) stock market listed. A medium firm meets the following conditions: (i) revenue of at least 1.5 million euro, (ii) operating income of at least 3 million euro, (iii) at least 15 workers, and (iv) not being defined as a big firm. A small firm is a firm that is not defined as a big or medium firm.

In addition, three clusters of control variables were constructed based on the Bel-first data. First, in many countries, public and non-profit firms are, more than commercial firms, expected to promote equality and to make efforts to reduce labour market discrimination (Baert and Vujić, in press; Wood et al., 2009). Therefore, based on the Bel-first firm data we constructed an indicator for public and non-profit firms. This indicator is based on the Bel-first information concerning firms' legal form. More concretely, the indicator is 1 for firms with as a legal form: (i) private non-profit association, (ii) company limited by shares with a social aim, (iii) non-profit association, (iv) public utility founding, (v) co-operative society with limited liability with a social aim, (vi) public company limited by shares, or (vii) ministries. It is 0 for all other legal forms.

Second, data are included on firms' financial health. Following Becker's (1957) aforementioned theory of taste-based discrimination, unequal treatment based on tastes

products/data/national/bel-first.

¹¹ At least with respect to our indicators of discrimination in positive reaction and interview invitation, the remaining vacancies were comparable to those that could not be linked to firms in Bel-first.

¹² Similar definitions were used (by Bureau Van Dijk) for firm data from other countries.

or (irrational) prejudices should yield a worse financial health. In addition, firms' financial health and size may be correlated. Therefore, we selected two key indicators in this respect: the firms' solvency and liquidity ratio.¹³

Third, we selected for all vacancies an indicator of the sector of the vacancy. More concretely, we selected from Bel-first the NACE-2 code for each firm. To have enough observations in each sector, we used the highest level, i.e. the A level, of this sector classification system of the European Union. This level comprises 21 clusters of economic activities going from 'Agriculture, forestry, and fishing' (code A) to 'Activities of extraterritorial organisations and bodies' (code U). In our analyses, we will include sector fixed effects as gender, ethnic, and age discrimination may differ by sector, related to, e.g., the union density in the sector (Baert and Omey, 2015; Hirsch et al., 2000).

2.3 Summary Statistics

Table 1 describes the merged dataset except for the information on the occupation and sector categorical variables. Discrimination in positive reaction is found in about one out of ten vacancies. This indicator is the highest for the experiment on gender discrimination and the lowest for the experiment on age discrimination. This is because the overall call-back rate was higher in the experiment on gender discrimination (with more high-educated profiles; see below), resulting in more variation in our discrimination measure. However, as written in Section 2.1, the level of discrimination among the vacancies *with at least one positive call-back* is higher in the experiments on ethnic and age discrimination. We return to this point in our discussion of the undertaken robustness checks in Section 3.2.

<Table 1 about here>

In the experiments on gender and ethnic discrimination, about half of the tested firms were big firms following the Bel-first definition (see Section 2.2). In the experiment on age discrimination, small, medium, and large firms are fairly evenly represented. Figure A.1, Figure A.2, and Figure A.3 (in Appendix A) provide the reader with histograms of the number of workers at the company of each vacancy in the data on gender, ethnic, and age

¹³ The solvency ratio in our data is defined as the ratio of shareholders' funds and total assets. The liquidity ratio is defined as the current assets of the firm (except for its stocks) divided by its current liabilities.

discrimination, respectively. As can be seen from these figures, our benchmark indicator of firm size, i.e. the number of workers divided by 1000, is right-skewed (with a skewness of 5.669). In contrast, its logarithmic transformation is quite symmetric (skewness of 0.393).¹⁴ In addition, the proportion of public or non-profit firms is rather low (ranging from 3.6% in the experiment on gender discrimination to 9.5% in the experiment on ethnic discrimination). Finally, on average, the included firms are in relatively good financial health, with a solvency ratio of 0.400 or lower and a liquidity ratio of around 2.

The number of vacancies tested in the three experiments by occupation and sector can be found in Table A.1 (in Appendix A). Clearly, the occupations tested in the experiment on gender discrimination were higher-skilled compared to those in the experiments on ethnic and age discrimination. This is related to (i) the focus of the first experiment (gender differences in getting jobs at a higher functional level after five years of work experience) and (ii) the fact that we excluded the vacancies from this experiment that did not imply a promotion in functional level (see Section 2.1). In addition, the vacancies in the experiment on gender discrimination were situated more often in the administrative and support services sector while the vacancies in manufacturing, construction, and wholesale were overrepresented in the other experiments. When comparing our analyses for the three subsamples by discrimination ground, these differences in setting should be kept in mind.¹⁵

3. Results

3.1 Main Analysis

To estimate the association between firm size and hiring discrimination based on the merged data, we conduct linear probability model regressions with White-corrected standard errors. These models have the advantage to be easy to interpret while Angrist and

¹⁴ Information on variable skewness is not included in Table 1.

¹⁵ More in general, this point has consequences for the generalisability of our findings, on which we elaborate in Section 4.

Pischke (2008) proved their adequate performance with binary dependent variables.¹⁶ Nevertheless, we also estimated alternative models to check the robustness of our results. We return to this point in Section 3.2.

For each subsample by discrimination ground, seven regression models are specified. In model (1), our benchmark dependent variable, i.e. discrimination in positive reaction, is regressed on our benchmark independent variable, i.e. the number of workers in the firm divided by 1000, only. In model (2), firm-level controls (the public or non-profit firm dummy, the solvency ratio, and the liquidity ratio) are added and in model (3) occupation and sector dummies are additionally included. Model (3) is our benchmark model. The sensitivity of the results of this model for alternative approaches concerning the independent variable(s) and dependent variable is tested in models (4) to (7). More concretely, we test the sensitivity of our results for using the natural logarithm of the number of workers (model (4)), indicators for medium and big firms (with small firms as the reference category; model (5)), and an indicator for big firms only (model (6)) as independent variable(s). In model (7), we use discrimination in invitation as our dependent variable.

The results of these regression analyses of the data on gender discrimination are shown in Table (2). In contrast to what was hypothesised in Section 1, the measured association is positive. So, the higher the (logarithmic transformation of the) number of workers in the firm, the higher is the probability of gender discrimination by this firm. In addition, the coefficient for the indicator of big firms (in model (5) and model (6)) is positive. However, for none of the regression models these associations between our firm size and discrimination indicators are significantly different from 0. In addition, also the economic significance of the measured associations is negligible. When the number of workers is increased with 1000 (48.8% of a standard deviation, i.e. $1/2.050$), the probability of discrimination in positive reaction increases with 0.007 (2.3% of a standard deviation; i.e. $0.007/0.304$). So, based on our data for Belgium, no evidence for an association between firm size and gender discrimination in hiring is found.

In addition, based on these data, we find suggestive evidence for a (weakly) significant association between discrimination in positive call-back and the legal form of firms: gender

¹⁶ According to Wooldridge (2010), the biggest difference between our linear probability model and a logit model is that the former model implies constant marginal effects while the latter model allows diminishing effects.

discrimination is somewhat lower in respect of public and non-profit firms. However, given the low variation in this control variable, as acknowledged in Section 2.2, the coefficients for this dummy are very imprecisely estimated. Finally, no association between the firms' financial health and their tendency to discriminate is found.

<Table 2 about here>

Table 3 tells the same story with respect to the association between firm size and ethnic discrimination. Again, for none of the specifications used a significant association is found. The results with respect to age discrimination, which are presented in Table 4, are—at least at first sight—somewhat more nuanced. When regressing discrimination in positive reaction on the number of workers divided by 1000 and firm-level controls (model (1) and model (2)), a significantly positive association between firm size and discrimination is found. However, this association loses most of its significance when, from model (3) on, occupation and sector dummies are included. Moreover, the associations becomes completely insignificant, both in statistical and in economic terms, when the other measures of firm size and discrimination are used. Therefore, we conclude that also with respect to age discrimination, our analysis does not provide (consistent) evidence for an association between hiring discrimination and firm size. In addition, again, we find evidence for a—this time highly significantly—negative association between the public or non-profit nature of a firm and its tendency to discriminate (in terms of positive call-back in broad sense).

<Table 3 about here>

<Table 4 about here>

3.2 Robustness Checks

To further test the robustness of our finding of no (significant) association between firm size and hiring discrimination—at least as based on gender, ethnicity, and age in Belgium—we conducted several additional analyses. First, we re-estimated the models discussed in Section 3.1 after restricting our research sample to (i) vacancies for which at least one of both candidates received a positive reaction¹⁷ and (ii) vacancies that were directly posted by

¹⁷ This relates to the discussion in the correspondence testing literature concerning whether situations in which

the firm (and not by an intermediary such as an employment office). Second, we run models in which the data from the three experiments were pooled and controls for the ground of discrimination were added. Third, we re-estimated our models with the individual job application and not the individual vacancy as our unit of observation (resulting in models including interactions between a dummy for ‘minority’ candidates and our independent and control variables—standard errors were corrected for clustering at the vacancy level). Fourth, we tested the robustness of our results when using logit models instead of linear probability models. Fifth, we explored alternative dependent variables with three levels instead of two levels. In particular, the level of positive discrimination (in case the ‘minority’ candidate got a positive reaction while the ‘majority’ candidate got no positive reaction) was added to our benchmark dependent variable. Also a dependent variable resulting from summing up the discrimination in positive reaction and discrimination in invitation dummies was used. Finally, we re-estimated our models using other sets of control variables. In particular, we (i) excluded the control variables related to the companies’ financial health as these controls may be affected by our dependent variable—as aforementioned, following Becker (1957), discriminatory behaviour should result in lower profits—and (ii) included differing additional controls observed in (some of) the experimental datasets, such as the gender of the contact person mentioned in the vacancy and the geographical distance between the firm and the candidate’s residence. However, none of these analyses, the results of which are available on request, led to a conclusion other than that of no (substantial) association between firm size and hiring discrimination.

4. Discussion

In this study, we were the first to investigate the association between firm size and hiring discrimination with respect to multiple discrimination grounds. To this end, we analysed field experimental measures on gender, ethnic, and age discrimination in hiring that were merged with data on the tested companies. Several regression approaches, with diverging

none of the pair members got a (positive) reaction should be seen as a non-observation or as a situation of equal treatment (Darity and Mason, 1998; Riach and Rich, 2002).

dependent, independent, and control variables, based on diverging vacancy samples, were applied. All these approaches yielded the same conclusion, i.e. that of no significant association between firm size and hiring discrimination (based on gender, ethnicity, or age). From a policy perspective, our results suggest that discrimination policies, e.g., more vigorous detection of discrimination by the government, should not particularly focus on larger or smaller firms. We return to the policy relevance of our results below.

Our results contrast to what was hypothesised based on dominant theoretical mechanisms and to the suggestive evidence on the heterogeneity of ethnic discrimination by firm size in the literature. Discussions with experts in the field yielded the following speculative arguments that may counterbalance the dominant theories. A first argument is related to principal-agent theory (Grossman, 1983; Harris and Raviv, 1979; Myerson, 1982; Ross, 1973). In larger firms, the human resource departments may act in their own best interests, following their own tastes and prejudices, instead of focussing on profit maximisation. In other words, they may be less exposed to the trade-off between taste-based discrimination and profitability (as put forward by Becker (1957)) because they are further away from where the impact of lower profitability occurs more directly (i.e. among owners or shareholders). In contrast, in smaller firms, the ownership of the firm and the recruitment decisions may fall on the same shoulders. Second, the more extensive recruiting procedures in larger firms may not only facilitate more objective hiring decisions (as argued in Section 1) but might also allow to hide unequal treatment in a more ingenious way (e.g., by formulating extensive *post hoc* arguments for recruiting a 'majority' candidate). Third, larger firms are expected to be confronted with larger groups of job candidates (as a consequence of the fact they are better known and have, often, a stronger image). As a consequence, they might feel a higher need for statistical discrimination when making first hiring decisions (Arrow, 1973).

In addition to these theoretical reasons for a zero association, one might wonder whether our analyses are not underpowered to confirm our research hypothesis. We believe this is not the case. First, a *post hoc* power analysis shows that based on the variation in the 253, 337, and 427 observations in our three subsamples, we were able to distinguish

moderate to small effects from zero effects.¹⁸ This variation also allowed us to find significant associations with respect to the legal form of the firm, i.e. a variable with substantially less variation than our firm size variables. Second, as mentioned in Section 3.1, also the economic significance of our regression coefficients of main interest was negligible. Third, the fact that we did not find a significantly negative association that would be in line with our research hypothesis has in fact nothing to do with statistical power because our coefficients pointed in the direction of an insignificantly *positive* association.

We end this article by discussing some limitations of our research and related fruitful directions for future research. First, in this article we took a step forward in the unbiased measurement of the association between firm size and hiring discrimination as, in contrast with former contributions exploring this association (as a secondary focus), we controlled in our analyses for several firm-level controls and for occupation and sector fixed effects. Still, our measures of firm size may correlate with unobserved (firm-level) determinants of hiring discrimination. As a consequence, the measured associations cannot be interpreted as causal effects. In addition, we measured the association of interest in a particular context, i.e. the vacancies tested within certain occupations and during a certain period in three field experiments in Belgium. So, our findings cannot be easily generalised to other contexts. For these reasons, we are in favour of studies exploring the association between firm size and (various forms of) hiring discrimination in other institutional settings. To take a next step forward, these studies should try to link data with exogenous variation in firm size (related to shocks or captured in instrumental variables) with field experimental data on discrimination.

Second, as argued by a reviewer of a former version of this article, the zero results reported in this study may also (partly) result from the following limitation to our data. Our data capture information at the company level. A company may consist of multiple workplaces. As a consequence, it could make a difference whether recruitment is organised

¹⁸ This *post hoc* analysis is based on the regression results of model (3) in Table 2, Table 3, and Table 4, combined with the standard deviations of our benchmark independent and dependent variables presented in Table 1. If an increase of the firm size with one standard deviation had led to a decrease in the level of gender, ethnic or age discrimination by 15.9% (i.e. $[1.96 \cdot 0.012 / 0.304] \cdot 2.050$), 14.6% (i.e. $[1.96 \cdot 0.024 / 0.294] \cdot 0.910$), or 16.6% (i.e. $[1.96 \cdot 0.019 / 0.278] \cdot 1.236$) of a standard deviation or more, respectively, a zero association could have been rejected at the 5% significance level.

centrally (*ergo*, at the firm level) or locally (at the workplace level). In the latter case, our firm size variables may contain some noise, which may lead to downward biased estimates.

Third, the insignificant association between firm size and hiring discrimination may be the result of mechanisms that cancel each other out. From a policy point of view, it would be interesting to get an insight in which of the theoretical mechanisms discussed above survive the confrontation with the empirical literature. In this respect, it would be interesting to merge data on discrimination with administrative data on firm-level human resource expenditures or survey data on the extent and structure of firm's recruitment procedures.

Fourth, while finding no evidence for an association between firm size and hiring discrimination, this study provides suggestive empirical evidence for heterogeneity in discrimination by firms' legal form. However, as in the experiments on which we built only few public and non-profit firms were tested, the estimates concerning this association were very imprecise. Therefore, we are in favour of future studies that re-analyse field experimental datasets with more variation in this respect, with a focus on the association between discrimination and legal form of the firm.

References

Albert, R., Escot, L. and Fernandez-Cornejo, J. (2011), "A field experiment to study sex, age discrimination in the Madrid labour market", *International Journal of Human Resource Management*, Vol. 22, pp. 351-375.

Angrist, J.D. and Pischke, J.-S. (2008), *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, Princeton.

Arceo-Gomez, E.O. and Campos-Vazquez, R. M. (2014), "Race and Marriage in the Labor Market: A Discrimination Correspondence Study in a Developing Country", *American Economic Review*, Vol. 104, pp. 376-380

Arrow, K. J. (1973), "The Theory of Discrimination", In Ashenfelter, O., Rees, A. (Eds.), *Discrimination in Labor Markets*, Princeton University Press, Princeton.

Baert, S. (2017), "Hiring Discrimination: An Overview of (Almost) All Correspondence Experiments Since 2005", IZA Discussion Paper Series, 10738.

Baert, S. (in press), "Facebook profile picture appearance affects recruiters' first hiring decisions", *New Media & Society*. DOI: 10.1177/1461444816687294.

Baert, S., Albanese, A., du Gardein, S., Ovaere, J. and Stappers, J. (2017), "Does Work Experience Mitigate Discrimination?", *Economics Letters*, Vol. 155, pp. 35-38.

Baert, S., Cockx, B., Gheyle, N. and Vandamme, C. (2015), "Is There Less Discrimination in Occupations Where Recruitment Is Difficult?", *ILR Review*, Vol. 68, pp. 467-500.

Baert, S., De Pauw, A.-S. and Deschacht, N. (2016a), "Do Employer Preferences Contribute to Sticky Floors?", *ILR Review*, Vol. 69, pp. 714-736.

Baert, S., Norga, J., Thuy, Y. and Van Hecke, M. (2016b), "Getting Grey Hairs in the Labour Market. A Realistic Experiment on Age Discrimination", *Journal of Economic Psychology*, Vol. 57, pp. 86-101.

Baert, S. and Omeij, E. (2015), "Hiring Discrimination against Pro-Union Applicants: The Role of Union Density and Firm Size", *Economist*, Vol. 163, pp. 263-280.

Baert, S. and Verhaest, D. (2014), "Unemployment or Overeducation: Which is a Worse Signal to Employers?", IZA Discussion Paper Series, 8312.

Baert, S. and Vujić, S. (in press), "Does it Pay to Care? Volunteering and Employment Opportunities", *Journal of Population Economics*. DOI: 10.1007/s00148-017-0682-8.

Becker, G.S. (1957), *The Economics of Discrimination*, University of Chicago Press, Chicago.

Bertrand, M. and Mullainathan, S. (2004), "Are Emily, Greg more employable than Lakisha, Jamal? A field experiment on labor market discrimination", *American Economic Review*, Vol. 94, pp. 991-1013.

Booth, A., Leigh, A. and Varganova, E. (2012), "Does ethnic discrimination vary across minority groups? Evidence from a field experiment", *Oxford Bulletin of Economics and Statistics*, Vol. 74, pp. 547-573.

Carlsson, M. (2010), "Experimental Evidence of Discrimination in the Hiring of First-and

Second-generation Immigrants”, *Labour*, Vol. 24, pp. 263-278.

Carlsson, M. and Rooth, D.-A. (2007), “Evidence of ethnic discrimination in the Swedish labor market using experimental data”, *Labour Economics*, Vol. 14, pp. 716-729.

Croson, R. (2005), “The Method of Experimental Economics”, *International Negotiation*, Vol. 10, pp. 131-148.

Darity, W.A.Jr. and Mason, P.L. (1998), “Evidence on Discrimination in Employment: Codes of Color, Codes of Gender”, *Journal of Economic Perspectives*, Vol. 12, pp. 63-90.

Drydakis, N. and Vlassis, M. (2010), “Ethnic discrimination in the Greek labour market: occupational access, insurance coverage and wage offers”, *Manchester School*, Vol. 78, pp. 201-218.

Eriksson, S. and Rooth, D.-O. (2014), “Do Employers Use Unemployment as a Sorting Criterion When Hiring? Evidence from a Field Experiment”, *American Economic Review*, Vol. 104, pp. 1014-1039.

Fink, G., McConnell, M. and Vollmer, S. (2014), “Testing for heterogeneous treatment effects in experimental data: false discovery risks and correction procedures”, *Journal of Development Effectiveness*, Vol. 6, pp. 44-57.

Grossman, S.J. and Hart, O.D. (1983), “An Analysis Of The Principal-Agent Problem”, *Econometrica*, Vol. 51, pp. 7-45.

Harris, M. and Raviv, A. (1979), “Optimal Incentive Contracts with Imperfect Information”, *Journal of Economic Theory*, Vol. 20, pp. 231-259.

Hirsh, B.T., Macpherson, D.A. and Hardy, M.A. (2000), “Occupational age structure and access for older workers”, *ILR Review*, Vol. 53, pp. 401-418.

Kaas, L. and Manger, C. (2012), “Ethnic Discrimination in Germany’s Labour Market: A Field Experiment”, *German Economic Review*, Vol. 13, pp. 1-20.

Kroft, K., Lange, F. and Notowidigdo, M.J. (2013), “Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment”, *Quarterly Journal of Economics*, Vol. 128, pp. 1123-1167.

Lahey, J. (2008), “Age, women, hiring: An experimental study”, *Journal of Human*

Resources, Vol. 43, pp. 30-56.

Maurer-Fazio, M. (2012), "Ethnic Discrimination in China's Internet Job Board Labor Market", *IZA Journal of Migration*, Vol. 1, pp. 1-24.

Myerson, R.B. (1982), "Optimal Coordination Mechanisms In Generalized Principal Agent Problems", *Journal of Mathematical Economics*, Vol. 10, pp. 67-81.

Neumark, D. (in press), "Experimental Research on Labor Market Discrimination", *Journal of Economic Literature*.

OECD (2008), *Jobs for Immigrants. Labour Market Integration in France, Belgium, the Netherlands and Portugal*, OECD Publishing, Paris.

OECD (2010), *Sickness, Disability and Work. Breaking the Barriers – A Synthesis of Findings across OECD Countries*, OECD Publishing, Paris.

Oreopoulos, P. (2011), "Why do skilled immigrants struggle in the labor market? A field experiment with thirteen thousand resumes", *American Economic Journal: Economic Policy*, Vol. 3, pp. 148-171.

Petit, P. (2007), "The effects of age, family constraints on gender hiring discrimination: A field experiment in the French financial sector", *Labour Economics*, Vol. 14, pp. 371-391.

Riach, P.A. and Rich, J. (2002), "Field experiments of discrimination in the market place", *Economic Journal*, Vol. 112, pp. 480-518.

Ross, S. (1973), "The Economic Theory of Agency: The principal's problem", *American Economic Review*, Vol. 63, pp. 134-139.

Verhaest, D., Bogaert, E., Dereymaeker, J., Mestdagh, L. and Baert, S. (in press), "Do Employers Prefer Overqualified Graduates? A Field Experiment", *Industrial Relations: A Journal of Economy and Society*.

Wood, M., Hales, J., Purdon, S., Sejersen, T. and Hayllar, O. (2009), "A test for racial discrimination in recruitment practice in British cities", DWP Research Reports, 607.

Wooldridge, J. M. (2010), *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge.

Zhou, X., Zhang, J. and Song, X. (2013), "Gender discrimination in hiring: Evidence from

19,130 resumes in China”, Mimeo.

Appendix A: Additional Figures and Tables

<Figure A.1 about here>

<Figure A.2 about here>

<Figure A.3 about here>

<Table A.1 about here>

Table 1. Data Description

	Experiment: gender discrimination (N = 253)		Experiment: ethnic discrimination (N = 337)		Experiment: age discrimination (N = 427)	
	Mean	SD	Mean	SD	Mean	SD
A. Dependent variables						
Discrimination in positive reaction	0.103	0.304	0.095	0.294	0.084	0.278
Discrimination in invitation	0.063	0.244	0.050	0.219	0.042	0.201
B. Independent variables						
#workers/1000	0.741	2.050	0.275	0.910	0.310	1.236
log(#workers)	4.027	2.451	3.910	1.759	3.386	1.993
Small firm	0.253	0.436	0.145	0.353	0.304	0.461
Medium firm	0.209	0.408	0.335	0.473	0.328	0.470
Big firm	0.538	0.500	0.519	0.500	0.368	0.483
C. Control variables						
Public or non-profit firm	0.036	0.186	0.095	0.294	0.040	0.196
Solvency ratio	0.377	0.250	0.400	0.251	0.335	0.278
Liquidity ratio	2.242	3.054	2.007	2.016	1.872	2.040

Notes. The variable ‘discrimination in positive reaction’ (‘discrimination in invitation’) is 1 in case the candidate from the discriminated group, i.e. the female, foreign, or older candidate, got no positive reaction (invitation) from the employer while the candidate from the control group, i.e. the male, native, or younger candidate within the same pair, got a positive reaction (invitation), and 0 otherwise. See Section 2.2 for a description of the independent and control variables. Besides the mentioned variables, the data also comprise occupation and sector dummies, which are described in Table A.1.

Table 2. Firm Size and Gender Discrimination: Regression Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
#workers/1000	0.008 (0.011)	0.009 (0.011)	0.007 (0.012)				0.011 (0.012)
log(#workers)				0.008 (0.009)			
Medium firm					-0.012 (0.062)		
Big firm					0.044 (0.051)	0.050 (0.039)	
Public or non-profit firm		-0.115*** (0.025)	-0.158* (0.082)	-0.159* (0.082)	-0.148* (0.082)	-0.150* (0.081)	-0.128 (0.078)
Solvency ratio		-0.013 (0.084)	0.017 (0.086)	0.020 (0.087)	0.017 (0.089)	0.019 (0.087)	-0.007 (0.069)
Liquidity ratio		-0.002 (0.004)	-0.003 (0.004)	-0.003 (0.004)	-0.002 (0.004)	-0.003 (0.004)	0.006 (0.008)
Intercept	0.097*** (0.020)	0.111*** (0.036)	0.106 (0.076)	0.075 (0.075)	0.073 (0.082)	0.068 (0.075)	-0.042 (0.043)
Occupation dummies			x	x	x	x	x
Sector dummies			x	x	x	x	x
Dependent variable: discrimination in positive reaction	x	x	x	x	x	x	
Dependent variable: discrimination in invitation							x
Observations	253	253	253	253	253	253	253

Notes. The presented statistics are linear probability model estimates with Huber-White corrected standard errors in parentheses. The variable ‘discrimination in positive reaction’ (‘discrimination in invitation’) is 1 in case the female candidate got no positive reaction (invitation) from the employer while the male candidate got a positive reaction (invitation), and 0 otherwise. See Section 2.2 for a description of the independent and control variables. The estimation results for the model’s independent variables are in bold. *** (**) (*) indicates significance at the 1% (5%) (10%) significance level.

Table 3. Firm Size and Ethnic Discrimination: Regression Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
#workers/1000	0.011 (0.027)	0.013 (0.028)	0.018 (0.024)				0.016 (0.023)
log(#workers)				-0.005 (0.011)			
Medium firm					-0.014 (0.058)		
Big firm					-0.039 (0.056)	-0.029 (0.037)	
Public or non-profit firm		-0.048 (0.042)	-0.074 (0.061)	-0.070 (0.065)	-0.069 (0.064)	-0.070 (0.064)	-0.051 (0.053)
Solvency ratio		0.035 (0.066)	0.038 (0.072)	0.045 (0.075)	0.053 (0.075)	0.051 (0.074)	0.034 (0.055)
Liquidity ratio		-0.002 (0.006)	-0.004 (0.008)	-0.004 (0.008)	-0.005 (0.008)	-0.005 (0.008)	-0.001 (0.005)
Intercept	0.092*** (0.017)	0.086*** (0.028)	0.044 (0.102)	0.118 (0.154)	0.130 (0.161)	0.121 (0.156)	0.106 (0.122)
Occupation dummies			x	x	x	x	x
Sector dummies			x	x	x	x	x
Dependent variable: discrimination in positive reaction	x	x	x	x	x	x	
Dependent variable: discrimination in invitation							x
Observations	337	337	337	337	337	337	337

Notes. The presented statistics are linear probability model estimates with Huber-White corrected standard errors in parentheses. The variable ‘discrimination in positive reaction’ (‘discrimination in invitation’) is 1 in case the foreign candidate got no positive reaction (invitation) from the employer while the native candidate got a positive reaction (invitation), and 0 otherwise. See Section 2.2 for a description of the independent and control variables. The estimation results for the model’s independent variables are in bold. *** (**) (*) indicates significance at the 1% (5%) (10%) significance level.

Table 4. Firm Size and Age Discrimination: Regression Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
#workers/1000	0.035** (0.017)	0.036** (0.017)	0.034* (0.019)				0.012 (0.013)
log(#workers)				0.007 (0.009)			
Medium firm					0.027 (0.035)		
Big firm					0.006 (0.034)	-0.009 (0.028)	
Public or non-profit firm		-0.108*** (0.027)	-0.113*** (0.042)	-0.104*** (0.039)	-0.100*** (0.038)	-0.096*** (0.037)	-0.044* (0.026)
Solvency ratio		0.027 (0.065)	0.018 (0.072)	0.023 (0.073)	0.031 (0.073)	0.034 (0.072)	0.038 (0.068)
Liquidity ratio		-0.009* (0.005)	-0.008 (0.005)	-0.008 (0.005)	-0.009* (0.005)	-0.010* (0.005)	-0.002 (0.004)
Intercept	0.073*** (0.014)	0.086*** (0.022)	0.019 (0.039)	0.000 (0.046)	0.005 (0.049)	0.019 (0.043)	-0.025 (0.027)
Occupation dummies			x	x	x	x	x
Sector dummies			x	x	x	x	x
Dependent variable: discrimination in positive reaction	x	x	x	x	x	x	
Dependent variable: discrimination in invitation							x
Observations	427	427	427	427	427	427	427

Notes. The presented statistics are linear probability model estimates with Huber-White corrected standard errors in parentheses. The variable ‘discrimination in positive reaction’ (‘discrimination in invitation’) is 1 in case the older candidate got no positive reaction (invitation) from the employer while the younger candidate got a positive reaction (invitation), and 0 otherwise. See Section 2.2 for a description of the independent and control variables. The estimation results for the model’s independent variables are in bold. *** (**) (*) indicates significance at the 1% (5%) (10%) significance level.

Table A.1. Histogram of Firm Size in Experiment on Gender Discrimination

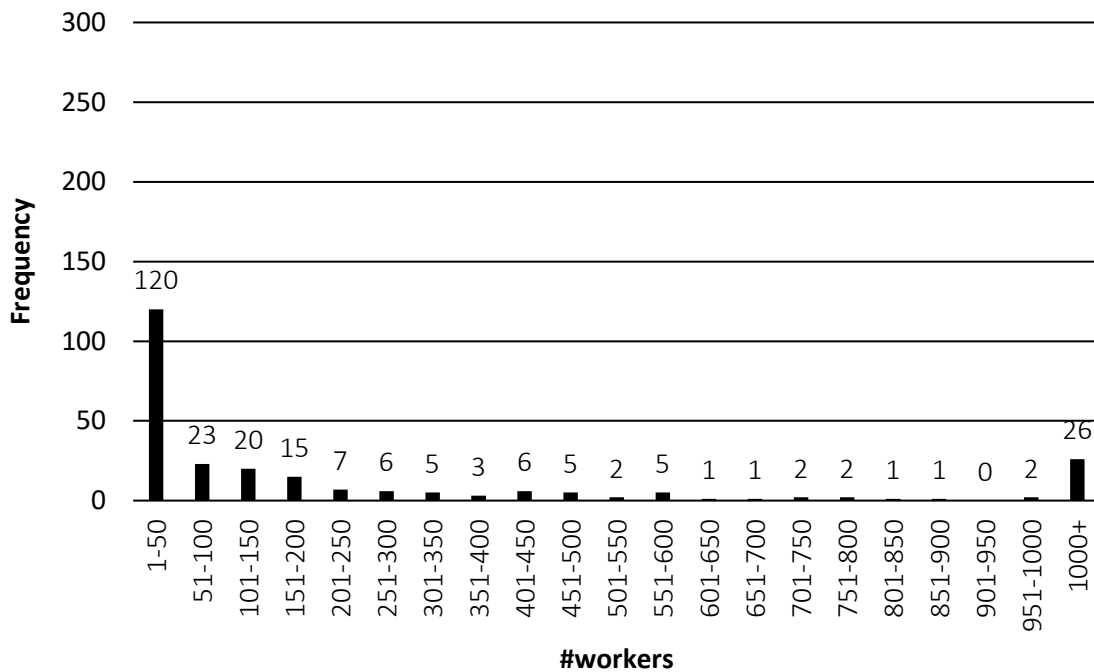


Table A.2. Histogram of Firm Size in Experiment on Ethnic Discrimination

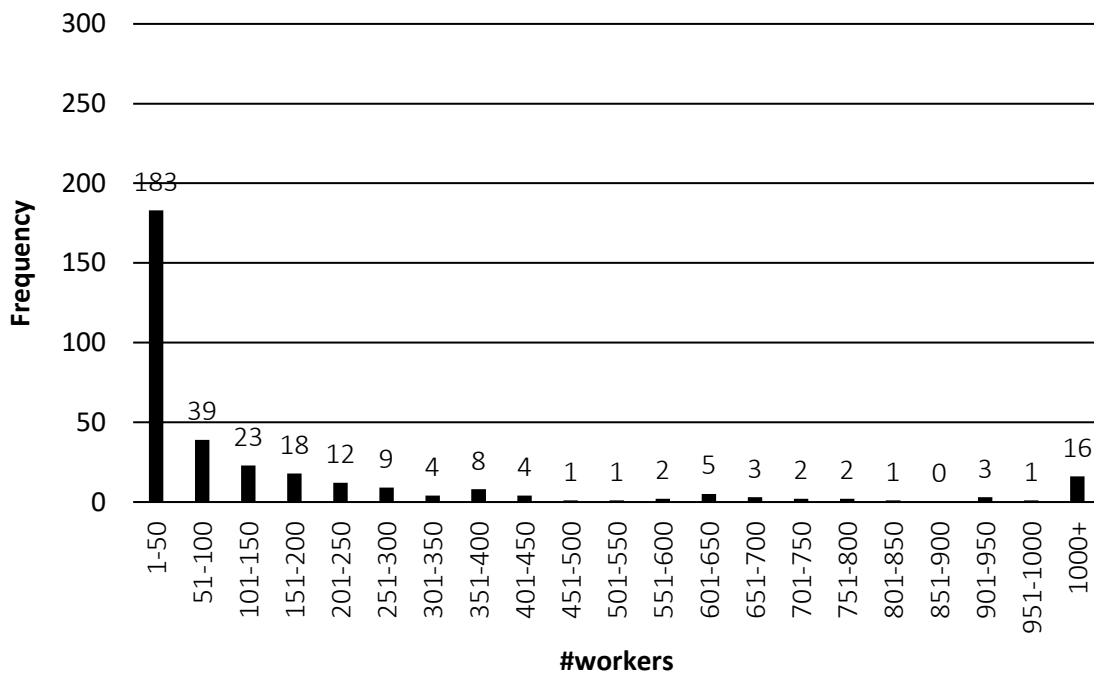


Table A.3. Histogram of Firm Size in Experiment on Age Discrimination

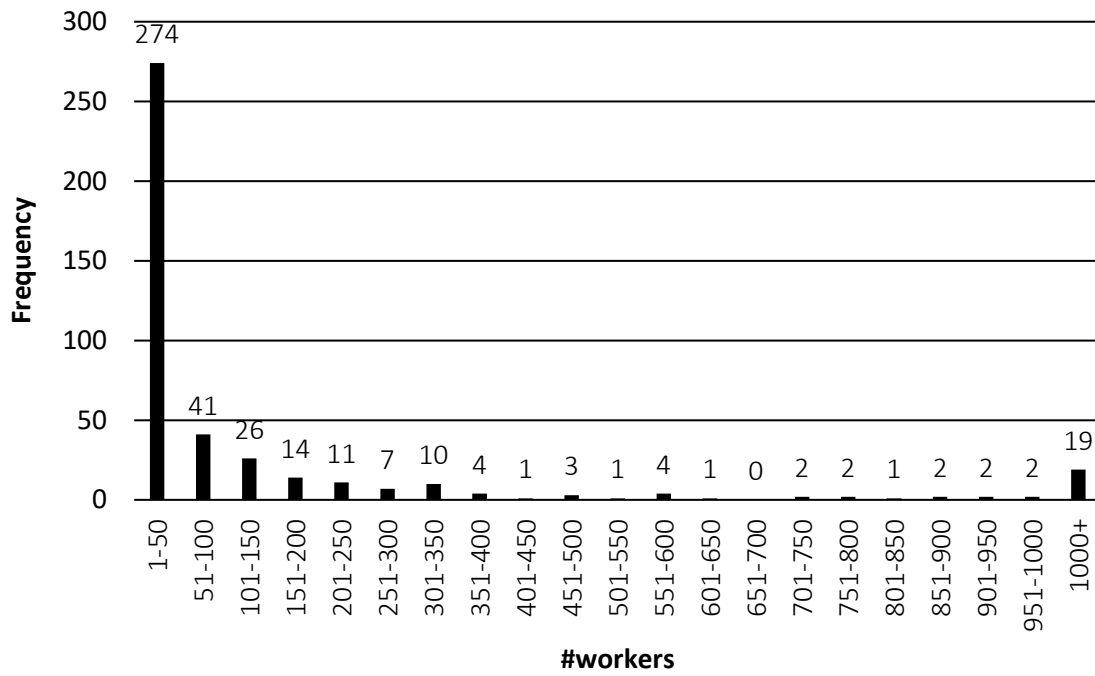


Table A.1. Tested Occupations and Sectors in Each Experiment

	Experiment: gender discrimination	Experiment: ethnic discrimination	Experiment: age discrimination
A. Number of vacancies by occupation			
Administrative and commercial managers (ISCO-08 code 12)	57	0	0
Production and specialised services managers (ISCO-08 code 13)	18	0	0
Business and administration professionals (ISCO-08 code 24)	157	0	0
Information and communications technology professionals (ISCO-08 code 25)	21	0	0
Health associate professionals (ISCO-08 code 32)	0	8	57
Business and administration associate professionals (ISCO-08 code 33)	0	116	161
Information and communications technicians (ISCO-08 code 35)	0	19	0
General and keyboard clerks (ISCO-08 code 41)	0	37	46
Personal service workers (ISCO-08 code 51)	0	12	33
Sales workers (ISCO-08 code 52)	0	18	0
Building and related trades workers, excluding electricians (ISCO-08 code 71)	0	29	0
Drivers and mobile plant operators (ISCO-08 code 83)	0	19	0
Cleaners and helpers (ISCO-08 code 91)	0	9	0
Labourers in mining, construction, manufacturing, and transport (ISCO-08 code 93)	0	70	130
B. Number of vacancies by sector			
Agriculture, forestry, and fishing (NACE code A)	0	1	3
Manufacturing (NACE code C)	20	46	94
Electricity, gas, steam, and air conditioning supply (NACE code D)	0	1	0
Water supply, sewerage, waste management, and remediation (NACE code E)	2	4	0
Construction (NACE code F)	6	46	30
Wholesale and retail trade, repair of motor vehicles, and motorcycles (NACE code G)	32	109	85
Transporting and storage (NACE code H)	10	9	15
Accommodation and food service activities (NACE code I)	0	12	24
Information and communication (NACE code J)	18	11	18
Financial and insurance activities (NACE code K)	6	4	4
Real estate activities (NACE code L)	1	1	2
Professional, scientific, and technical activities (NACE code M)	39	32	45
Administrative and support service activities (NACE code N)	106	30	77
Public administration, defence, and compulsory social security (NACE code O)	0	4	1
Education (NACE code P)	11	11	15
Human health and social work activities (NACE code Q)	1	9	11
Arts, entertainment, and recreation (NACE code R)	1	7	3