Do State Laws Protecting Older Workers from Discrimination Reduce Age Discrimination in Hiring? Evidence from a Field Experiment

David Neumark *University of California, Irvine*
Ian Burn *Swedish Institute for Social Research*
Patrick Button *Tulane University*
Nanneh Chehras *Amazon*

Abstract

We conduct a résumé field experiment in all US states to study how state laws protecting older workers from age discrimination affect age discrimination in hiring for retail sales jobs. We relate the difference in callback rates between old and young applicants to states' variation in age and disability discrimination laws. These laws could boost hiring of older applicants, although they could have the unintended consequence of deterring hiring if they increase termination costs. In our preferred estimates that are weighted to be representative of the workforce, we find evidence that there is less discrimination against older men and women in states where age discrimination law allows larger damages and more limited evidence that there is less discrimination against older women in states where disability discrimination law allows larger damages. Our clearest result is that the laws do not have the unintended consequence of lowering callbacks for older workers.

1. Introduction

In the face of population aging in the United States and other countries, policy makers have focused on efforts to boost the labor supply of older workers, mainly through public pension reforms. However, many studies suggest that older workers face age discrimination in hiring (see Bendick, Jackson, and Romero 1996;

We received generous support from the Alfred P. Sloan Foundation, from the Social Security Administration through a grant to the Michigan Retirement Research Center, and from the National Institutes of Health via a postdoctoral training grant for Button (5T32AG000244-23). The views expressed are our own and not those of the foundation, the Social Security Administration, the National Institutes of Health, or Amazon. This study was approved by the Institutional Review Board

[*Journal of Law and Economics,* vol. 62 (May 2019)]
 0022-2186/2019/6202-0012$10.00

Bendick, Brown, and Wall 1999; Riach and Rich 2010; Lahey 2008a; Farber, Silverman, and von Wachter 2017; Baert et al. 2016; Neumark, Burn, and Button 2019). Although discrimination in hiring may not seem closely related to encouraging older people to work longer, it may be essential to significantly lengthen work lives, because many seniors transition to part-time or shorter-term partial retirement or bridge jobs at the end of their careers (Cahill, Giandrea, and Quinn 2006; Johnson, Kawachi, and Lewis 2014) or return to work after a period of retirement (Maestas 2010).

In this paper, we study the effects of stronger or broader laws protecting older workers from discrimination in hiring. Stronger laws include harsher penalties for being found guilty of discriminating, and broader laws extend coverage to more employers or workers. We conduct a résumé correspondence study in all 50 states and use the results, combined with coding of the states' laws, to estimate how stronger or broader antidiscrimination laws affect the callback rates of older and younger job applicants, from which we draw inferences about discrimination in hiring.

We consider disability discrimination laws and age discrimination laws. Although the latter are of course directly based on age, numerous studies argue that disability discrimination laws may be important in protecting older workers, in particular, from discrimination (see Stock and Beegle 2004; Neumark, Song, and Button 2017; Button and Khan 2018). Disabilities that can limit work, and hence also likely limit major life activities and trigger protection by disability discrimination laws, rise steeply with age, especially past age 50 or so (see, for example, Rowe and Kahn 1997); correspondingly, employers' expectations that a worker will develop a disability in the near future—posing future accommodation costs—should also rise steeply with a worker's age. Indeed, disability discrimination laws may do more to protect many older workers than do age discrimination laws. Many ailments associated with aging are classified as disabilities (Sterns and Miklos 1995). These ailments can give some older workers the option of pursuing discrimination claims under either the Age Discrimination in Employment Act (ADEA) or the Americans with Disabilities Act (ADA) or under the corresponding state laws. Potential coverage under both age and disability discrimination laws may increase protections. For example, the ADA does more to limit defenses against discrimination claims.[1] A disability discrimination claim does require

(HS 2013-9942) at the University of California, Irvine. We are grateful to Melody Dehghan, Charlotte Densham, Stephanie Harrington, Justin Sang Hwa Lee, Darren Lieu, Candela Mirat, Chengshi Wang, Helen Yu, and Grace Jia Zhou for research assistance. We thank Tom DeLeire, Sam Peltzman, and an anonymous reviewer for helpful suggestions.

[1] Unlike the Age Discrimination in Employment Act (ADEA), the Americans with Disabilities Act (ADA) does not include an exception for bona fide occupational qualifications. Exceptions arise when age is strongly associated with other factors that pose legitimate business or safety concerns (see, for example, Stock and Beegle 2004; Posner 1995; Starkman 1992). Furthermore, age-related disabilities might be judged as amenable to reasonable accommodation by employers under disability discrimination laws, which usually require reasonable accommodation of the worker, and that would make it much harder to justify an apparently discriminatory practice on the basis of business necessity (Gardner and Campanella 1991).

proving a disability, but, as we explain below, doing so can be substantially easier under some state disability discrimination laws than under the ADA.

There is existing nonexperimental research on the effects of antidiscrimination laws. Earlier research finds that the adoption of age discrimination laws increases the employment rates of older workers, possibly in part through reducing terminations (Adams 2004; Neumark and Stock 1999). More recent research finds that larger damages under state age discrimination laws boost the employment of workers incentivized to work past age 65 by Social Security reforms in the 2000s (Neumark and Song 2013). In research on disability discrimination laws, some studies find negative effects on employment (DeLeire 2000; Acemoglu and Angrist 2001; Jolls and Prescott 2004; Bell and Heitmueller 2009), and some find no effects (Houtenville and Burkhauser 2004; Hotchkiss 2004; Beegle and Stock 2003), but some, including more recent studies, point to positive effects (Kruse and Schur 2003; Button 2018; Ameri et al. 2018; Armour, Button, and Hollands 2018). However, with the exception of Ameri et al. (2018), none of these studies of either age discrimination laws or disability discrimination laws use direct measures of discrimination to ask whether the laws reduce discrimination.

In contrast, to garner evidence on whether stronger age and disability discrimination laws are likely to increase the hiring of older workers, we conduct a large-scale résumé correspondence study covering all 50 states. The correspondence study provides direct measures of discrimination in hiring by state as reflected in differences in callback rates for job interviews. The use of all of the states allows us to fully capture the variation in state age and disability discrimination laws.[2] We code features of state age discrimination laws that extend beyond the federal ADEA and of state disability discrimination laws that extend beyond the ADA to study the relationships between the states' laws and the measures of age discrimination in hiring from the field experiment.[3] Our focus is on discrimination against job applicants ages 64 to 66, who are at or near the age of retirement—an age range that is particularly germane to whether age discrimination hinders policy reforms to encourage potential retirees to work longer and whether antidiscrimination laws can help.

We find some evidence that state age discrimination laws that allow larger damages than the federal ADEA reduce hiring discrimination against older workers—as measured by differences in callback rates. We also find some evidence that state disability discrimination laws that allow larger damages than the federal ADA reduce discrimination against older workers, although for women only. Other fea-

[2] This is a substantial expansion from the 12 cities in 11 states studied in Neumark, Burn, and Button (2019), although we limit the analysis to retail jobs, whereas that study considers résumés for three other types of jobs. The limited state-level variation in Neumark, Burn, and Button (2019) is not enough to identify the effects of features of age and disability discrimination laws separately, which motivates our expansion to all states in the present study.

[3] We are aware of only two other papers that look at variation in experimental evidence on discrimination across jurisdictions with different antidiscrimination laws: Tilcsik (2011), a study of discrimination against gay men, and Ameri et al. (2018), a study of discrimination against individuals with disabilities.

tures of the laws are not associated with differences in relative callback rates of older versus younger job applicants.

It may seem natural to expect that the effects of stronger antidiscrimination protections for older workers or workers with disabilities will—if anything—increase the hiring of older workers. However, the laws may be ineffective at reducing age discrimination in hiring. Enforcement relies in large part on potential rewards to plaintiffs' attorneys. In hiring cases, it is difficult to identify a class of affected workers, which inhibits class-action suits and substantially limits awards. In addition, economic damages can be small in hiring cases because one employer's action may extend a worker's spell of unemployment only modestly.

In contrast, terminations can entail substantial lost earnings and pension accruals, and it is much easier to identify a class of workers affected by an employer's terminations. If the principal effect of laws protecting older workers from discrimination is to make it costlier to terminate older workers, then the laws could have the unintended consequence of deterring the hiring of older workers (Bloch 1994; Posner 1995). This hypothesis is explored most fully by Oyer and Schaefer (2002) with regard to the effects of the Civil Rights Act of 1991 on the employment of blacks and females. The act expanded the rights of plaintiffs bringing discrimination claims—increasing damages when intentional discrimination could be established, increasing damages for discrimination in terminations of employment, and establishing a right to a jury trial, which is viewed as more likely to lead to outcomes favorable to plaintiffs.[4]

Oyer and Schaefer (2002) report evidence consistent with the Civil Rights Act of 1991 leading to less hiring from the groups for which protections from discrimination ostensibly increased, evidence of this kind of unintended consequence of antidiscrimination laws.[5] However, on the basis of our experimental measures of discrimination in hiring—reflected in relative callback rates—we find no evidence indicating that laws protecting older workers from discrimination have the unintended consequence of making older job applicants less attractive to employers. Rather, as noted above, the evidence if anything points in the direction of the laws reducing hiring discrimination against older workers.

One important caveat is that the variation in age discrimination that we measure is cross-sectional, not longitudinal, as there have been very few changes in state antidiscrimination laws in recent decades (see Neumark and Song 2013; Neumark, Song, and Button 2017; Button, Armour, and Hollands 2018a, 2018b).[6] Thus, our evidence could in principle reflect other factors correlated with em-

[4] Oyer and Schaefer (2002) focus on race and sex because the Civil Rights Act of 1991 had much less impact on cases brought under the ADEA.

[5] Lahey (2008b) reaches a similar conclusion regarding age discrimination laws, based on differences between state and federal laws (in some states). However, Neumark (2009) argues that Lahey's evidence more likely indicates that stronger age discrimination laws in fact boosted the employment of protected workers.

[6] This is documented in Online Appendix OA. For an interesting example of correspondence study evidence collected before and after a policy change (in the context of hiring differences related to race and criminal backgrounds), see Agan and Starr (2018).

ployers' decisions about callbacks for older workers versus younger workers and with antidiscrimination laws. However, the callback outcomes we measure are responses to very similar résumés in a single industry and do not reflect, for example, the decisions of older workers to apply for jobs or population differences between older and younger workers. Coupled with the fact that the antidiscrimination laws we study have been in effect for many years (typically decades), there are no obvious candidate explanations for a spurious relationship between the discrimination laws and the callback differences we measure.

In addition, our estimated effects of antidiscrimination laws are sensitive to how we weight the data, an issue that arises in our study because the data collection in the field experiment oversampled job ads from some cities (states) and undersampled them from others. We argue that the preferred estimates—summarized above—are based on weighting the data to be representative of the underlying distribution of employment. More generally, though, the sensitivity to weighting implies that field experiments on discrimination—especially when relating measured discrimination to local variation in laws or other factors—should use a large number of cities or states so that researchers can avoid nonrepresentative results. We also echo suggestions in Solon, Haider, and Wooldridge (2015) that researchers should present both weighted and unweighted estimates and explore any differences in estimates.

2. Correspondence Study Evidence on Discrimination

Experimental audit or correspondence (AC) studies of hiring are generally viewed as the most reliable means of inferring labor market discrimination (for example, Fix and Struyk 1993), because the group differences they (potentially) detect are very hard to attribute to nondiscriminatory factors. It is true that AC studies do not directly distinguish between taste discrimination and statistical discrimination. However, both are illegal under US law.[7] Indeed, the last decade has witnessed an explosion of AC studies of discrimination.[8] Most recent research—including the present paper—uses the correspondence study method, which creates fake applicants (on paper, or electronically in more recent work) and captures callbacks for job interviews. Correspondence studies are preferred because they can involve the collection of far larger samples than audit studies, which use in-person interviews. And correspondence studies avoid experimenter effects that can influence the behavior of the applicants used in audit studies (Heckman and Siegelman 1993).

A potential downside of correspondence studies is that they examine callbacks for job interviews, rather than job offers (which, in audit studies, can be observed). However, there is evidence pointing to discrimination in callbacks also reflecting discrimination in hiring. In an audit study of age discrimination, Bendick, Brown,

[7] Neumark, Burn, and Button (2019) devote considerable attention to weighing these alternative explanations of evidence consistent with age discrimination, analyses we do not delve into here.

[8] For example, see Stijn Baert, Register of Correspondence Experiments on Hiring Discrimination since 2005 (http://users.ugent.be/~sbaert/research_register.htm).

and Wall (1999) find that 75 percent of the discrimination occurred at the interview selection stage. Riach and Rich (2002) discuss studies of ethnic discrimination in audit studies by the International Labour Organization in which 90 percent of discrimination occurred at the interview selection stage. Neumark, Bank, and Van Nort (1996) find similar evidence in an audit study of sex discrimination in restaurant hiring. Nonetheless, while we think that this evidence justifies interpreting evidence on callbacks as informative about hiring, we are careful to interpret our results in terms of callback rates, given that we do not have direct evidence on hiring (job offers).

3. The Experimental Design

This paper significantly extends the correspondence study of age discrimination by Neumark, Burn, and Button (2019), which focuses on improving the basic evidence one could obtain on age discrimination in hiring (based on callbacks). One concern was the sensitivity of results to the common practice of giving older and younger applicants similarly low levels of labor market experience, consistent with the usual AC study paradigm of making applicants identical except with respect to the characteristic in question. Although the absence of relevant experience commensurate with an older applicant's age could be a negative signal, estimates of age discrimination were generally not sensitive to giving older workers more realistic résumés, including for the retail sales jobs on which we focus in this paper. In addition, the résumés were designed to implement a method developed in Neumark (2012) to address Heckman's critique of AC study evidence (Heckman and Siegelman 1993; Heckman 1998), namely, that experimental estimates of discrimination can be biased if the groups studied have different variances in unobservable characteristics. The same résumé design is used in the present study.

The key difference in the field experiment used in the present study is to extend the data collection to obtain estimates of age discrimination in all 50 states, which is critical to our goal of estimating the effects of state antidiscrimination laws. At the same time, the extensive resources required to extend the experiment to all 50 states necessitated focusing on retail sales and omitting some of the occupations included in the previous study. This limitation implies that the evidence should be regarded as a case study that may not generalize to other low-skill jobs.[9] In addition, because the prior evidence indicates that in retail sales there is no difference in measured age discrimination whether high-experience or low-experience résumés are used for older applicants, in this paper we use low-experience résumés that match those of younger applicants. This simplified the résumé creation because a long work history did not have to be developed for the

[9] Audit or correspondence studies typically use a very limited number of jobs. For example, Farber, Silverman, and von Wachter (2017) focus on age discrimination against women in administrative assistant jobs.

older applicants, which would be challenging given that the work histories had to be tailored to the local market.

3.1. Methods

The core analysis uses models for callbacks (C) as a function of dummy variables for age (S for older or senior) and observable characteristics from the résumés ($\boldsymbol{X}$). The latent variable model (for C, denoted C^*) is

$$C_i^* = \alpha + \gamma S_i + \boldsymbol{X}_i \delta + \varepsilon_i. \tag{1}$$

In this basic model, the null hypothesis of no discrimination against older workers implies that γ equals zero. We always estimate the model separately for the male and female job applicants we created, in accordance with the evidence from Neumark, Burn, and Button (2019) that older women experience stronger age discrimination.

The key contribution of this paper is to estimate the effects of state antidiscrimination laws affecting older workers on relative callbacks of older versus younger applicants. We do this by modifying equation (1) to include interactions between the dummy variable for older applicants and a vector of dummy variables for state laws (discussed in detail below). To ensure that we estimate the independent effects of each of the variations in state antidiscrimination protections, we simultaneously estimate the effects of the different antidiscrimination laws that we study, because the presence or absence of different features of state laws are correlated across states.

Adding to equation (1) a subscript to denote states and defining $\boldsymbol{A}_s$ as the vector of dummy variables capturing state antidiscrimination laws, we augment the model to be

$$C_{is}^* = \alpha + \gamma S_{is} + S_{is} \times \boldsymbol{A}_s \boldsymbol{\gamma}' + \boldsymbol{X}_{is} \delta + \varepsilon_{is}, \tag{2}$$

where $\boldsymbol{X}$ includes the state dummy variables. Because we include state dummy variables, we do not include the main effects of the state antidiscrimination laws. Excluding the state dummy variables and including the main effects of the laws would generate a less-saturated model, whereas the models we estimate allow more flexibly for differences in baseline callback rates for younger workers across states. Of course, we have to assume that state-by-age interactions can be excluded from the model to estimate the interactive effects of interest.

The key empirical question is whether stronger or broader state antidiscrimination laws are associated with differences in the relative callback rate of older workers, captured in the vector $\boldsymbol{\gamma}'$. Given that C_{is}^* is a latent variable for the propensity of a callback, for which we see the corresponding dichotomous outcomes, we estimate probit models. The vector $\boldsymbol{\gamma}'$ measures the effects of the interactions in $S_{is} \times \boldsymbol{A}_s$, which are difference-in-differences estimates of the effects of each feature of antidiscrimination laws. However, the calculation of marginal effects of interaction variables in probit (or logit) models is more complicated than in linear models (Ai and Norton 2003). In Online Appendix OB, we detail how we use a probit model to estimate the difference-in-differences effects.[10]

3.2. Résumés

Our overarching strategy in designing the résumés for our study is to use as much empirical evidence as possible to guide decisions about how to create the résumés so as to minimize decisions that might limit the external or comparison validity of the results. Much of this evidence comes from a large sample of publicly available résumés that we downloaded from a popular national job-hunting website and then scraped to convert the résumé information to data. We also rely on public-use data in choosing the résumés' features.[11]

We use an age range for young workers similar to that in other studies (29–31 years old) but compare results to older workers near the retirement age (64–66 years old). This latter age range is interesting in light of policy efforts to prolong the work lives of potential retirees. We convey age on the résumés via the year of high school graduation. Using the ages of our artificial job applicants, we chose common names (by sex) for the corresponding birth cohorts using data from the Social Security Administration, choosing first and last names that are most likely to signal that the applicant is Caucasian. In response to each job ad, we sent out a quadruplet of résumés consisting of a young and an old male applicant and a young and an old female applicant.

The résumé database verified that there are older applicants for retail sales positions, and they apply for jobs online. In addition, data from the Current Population Survey's Tenure Supplement show a sizable representation of low-tenure older workers in the occupations comprising retail sales (retail salespersons and cashiers in the census's occupational classification). The data further show that retail sales capture appreciable shares of new hiring of older workers (especially for the types of low-skill retail jobs that we use in the experiment) and that the share of older workers among all hires is particularly high in retail as compared with other industries.

As noted above, we use cities in all 50 states to maximize external validity and to include all variation in state antidiscrimination laws. This contrasts sharply with most previous experimental studies of discrimination, which typically use one or perhaps two cities (Pager 2007; Neumark 2018).

Because low-skill workers have low geographic mobility (Molloy, Smith, and Wozniak 2011), we target the résumés to retail jobs in specific cities, with the job and education history on each résumé matching the originating city of the ad to

[10] Results are very similar using linear probability models. These estimates are described in more detail below. However, we need the probit specification to address the Heckman critique.

[11] The discussion here is brief. Additional details are provided in Neumark, Burn, and Button (2019), although there are some differences because that paper presents a more complex study with additional occupations, additional résumé types, and so on. With regard to the résumé creation, we do not do anything in the current paper that extends beyond what was done in Neumark, Burn, and Button (2019), but in some cases what we do is more limited.

which we apply. This need to customize résumés to the city in which a job application is submitted is the reason we limited the analysis to retail sales jobs.

We constructed realistic job histories for the résumés using the jobs held by retail job applicants in the résumé database we scraped and the résumé-characteristic randomizer program created by Lahey and Beasley (2009). We chose job turnover rates on the basis of secondary data for retail trade from the Bureau of Labor Statistics Job Openings and Labor Turnover Survey. We used the résumé randomizer to produce a large number of job histories and then selected a smaller set that looked the most realistic when compared with the résumés found on the job-hunting website. From this sample of acceptable histories, we created four job histories for each city and added employers' names and addresses randomly to each job in our final job histories using information about employers present in each city at the relevant dates, relying mainly on national chains that have stores in many cities.

Half the résumé quadruplets we sent out included higher skill levels, and half did not.[12] For résumés reflecting higher skills, we include seven possible attributes, five of which are chosen randomly (so that they are not perfectly collinear in the résumés indicating higher skill levels). Five of the seven attributes are general: a bachelor of arts degree, fluency in Spanish as a second language, an employee-of-the-month award at the most recent job, one of three volunteer activities (food bank, homeless shelter, or animal shelter), and an absence of typographical errors. Two are specific to retail sales: familiarity with Microsoft Office and programs used to monitor inventory (VendPOS, AmberPOS, and Lightspeed).

Each of the résumés in the quadruplet was randomly assigned a different résumé template, which ensured that the résumés looked different. Most other characteristics were randomly and uniquely assigned to a résumé in each quadruplet to further ensure that the applicants were distinguished from each other.[13]

3.3. Applying for Jobs

We identify jobs to apply for using a common job-posting website. Research assistants read the posts regularly during the data collection period to select jobs for the study, using a well-specified set of criteria. We used Python code to automate and randomize the application process for the job ads selected.[14] Our sample size results from an explicit ex ante data collection plan that covered 2 academic quarters during which we collected as much data as the available job ads would

[12] This variation across résumés is needed to address the Heckman critique (Neumark 2012).

[13] Other details of the résumés, including the assignment of residential addresses and schools and examples of résumé types, are provided in Neumark, Burn, and Button (2019).

[14] The code used city and date to match the applicant to the job ad data entered into a spreadsheet by the research assistants. Each day was randomly assigned a different quadruplet of résumés in terms of skill level and current employment status. Within each quadruplet, the order of résumés was randomized. The code ran every other day and added 7- to 8-hour delays between applications to the same job.

Table 1
Matching Callbacks

	Matched Positive Responses	Matched Negative Responses	No Response	Total
Voice mail	1,614	3	N.A.	1,617
E-mail	1,218	52	N.A.	1,270
Both	438	0	N.A.	438
All	3,270	55	11,103	14,428

Note. Each response was matched from an employer to a unique job identifier. Three voice-mail messages could not be matched to a job identifier or a résumé sent. N.A. = not applicable.

allow. No data were analyzed until the data collection was complete. During that period, we sent 14,428 applications for 3,607 jobs.

3.4. Collecting Responses

Responses to job applications could be received by e-mail or phone. We read each e-mail message and listened to each voice-mail message to record the response, using information generated by the job site and auxiliary information from the responses to match responses to job ads. Table 1 reports the distribution of responses. Each response was coded as an unambiguous positive response (for example, "Please call to set up an interview"), an ambiguous response (for example, "Please return our call. We have a few additional questions"), or an unambiguous negative response (for example, "Thank you for your interest, but the job has been filled"). To avoid having to classify subjectively the ambiguous responses, we treated them as callbacks.[15]

3.5. Disproportionate Numbers of Ads and Weighting

We applied to retail jobs in one city in each state. Under the assumption that we applied to retail jobs in proportion to the number of retail jobs in each city, the unweighted data would provide estimates representative of the universe of retail jobs in those cities—or at least those that advertise on the job-posting website we used. As it turns out, however, we obtained quite different numbers of observations by city relative to what would be expected on the basis of the number of retail jobs in the city. This occurred for several reasons. First and foremost, for some cities the website we used defines the market as the whole city, whereas for larger cities the city is divided into multiple markets, and we used a single market because of the resource constraints imposed by collecting data for cities in all 50 states.[16] Second, for a couple of very large cities in which the number of ads was extensive, the research assistants did not apply to every job, whereas for the other

[15] The ambiguous responses are 7.8 percent of all cases coded as positive callbacks.

[16] The issue of how to sample job ads from geographic areas in correspondence studies is not unique to the job-hunting website we used.

cities they applied to all of them. And third, the frequency with which employers post ads on the job-hunting website, relative to other methods of posting jobs, can vary.

Figure 1 displays information on differences in representativeness across cities. For example, we applied to a very large number of job ads in Seattle (black bar) relative to the share of retail jobs computed from the Quarterly Workforce Indicators (QWI) data for retail (gray bar).[17] In contrast, New York City has a large number of retail jobs, as does Los Angeles, but both have fewer observations in the experimental data.

If there is heterogeneity in the effects of discrimination laws across the markets we study, then the estimates can be sensitive to the weighting, and disproportionate sampling of job ads by market relative to the distribution of jobs could generate biased estimates of the average effects of these laws for the representative worker—or average partial effects (Solon, Haider, and Wooldridge 2015). Thus, in our core analyses, we reweight the data by the ratio of the percentage of employment in the QWI data (by sex) to the percentage of observations in the city's sample.[18] This makes the estimates more representative of the distribution of retail jobs by city in the QWI data.[19,20] As an alternative gauge of sensitivity of the results, we reweight the data by 1 over the share of observations in the city, giving each city equal weight and making the data representative of cities (shown by the dashed line in Figure 1).[21]

If one views the goal of our analysis as estimating descriptive statistics of the

[17] These figures are based on age ranges covering the age groups we study. We use North American Industry Classification System codes 44-45 for data at the metropolitan statistical area (MSA) level, for men and women ages 25–34, 55–64, and 65–99. The percentages indicated by the gray bars are the sums across age ranges and sexes for the MSA, divided by the totals for the MSAs used in the states. We use only the portion of the MSA in the state in question. The Quarterly Workforce Indicators (QWI) data are for quarters 1–3 of 2015. Given the lag in the release of QWI data, this is the closest we could get to the period in which the experimental data were collected (February–July 2016); note that we try to overlap quarters to capture the same seasonal pattern. For two states (Michigan and Wyoming), we use data from quarters 1–3 of 2014 since the 2015 data were not available.

[18] This is the "pweight" option in Stata, which assigns as weights the inverse of the probability that the observation is included because of the sampling and hence preserves the correct degrees of freedom.

[19] For example, the Seattle data for both men and women are weighted by .27, which reflects the overrepresentation of observations from Seattle by a factor of about 4 in the experimental data (Figure 1). The New York City data are weighted by 4.74 for men and 4.42 for women, consistent with the underrepresentation of observations from New York City in the experimental data (Figure 1).

[20] However, Solon, Haider, and Wooldridge (2015) warn that weighting in this fashion does not guarantee that the estimates match the population's average partial effect. In their example, both weighted and unweighted estimates can inconsistently estimate the population's average partial effect when the variances of groups differ. They also note how, even absent this issue, not all groups are treated, so the average effect that is estimated will always be on a particular subset of groups that happen to be treated.

[21] Figure 1 also shows that the reweighting based on equal weighting is extreme for Arkansas, South Dakota, and Wyoming owing to very small numbers of observations for these states. (The same would be true for West Virginia, but it is dropped because there are no callbacks for West Virginia and so there are perfect predictions for the probit model.) We therefore drop observations for these states, all of which have fewer than 10 observations (by sex).

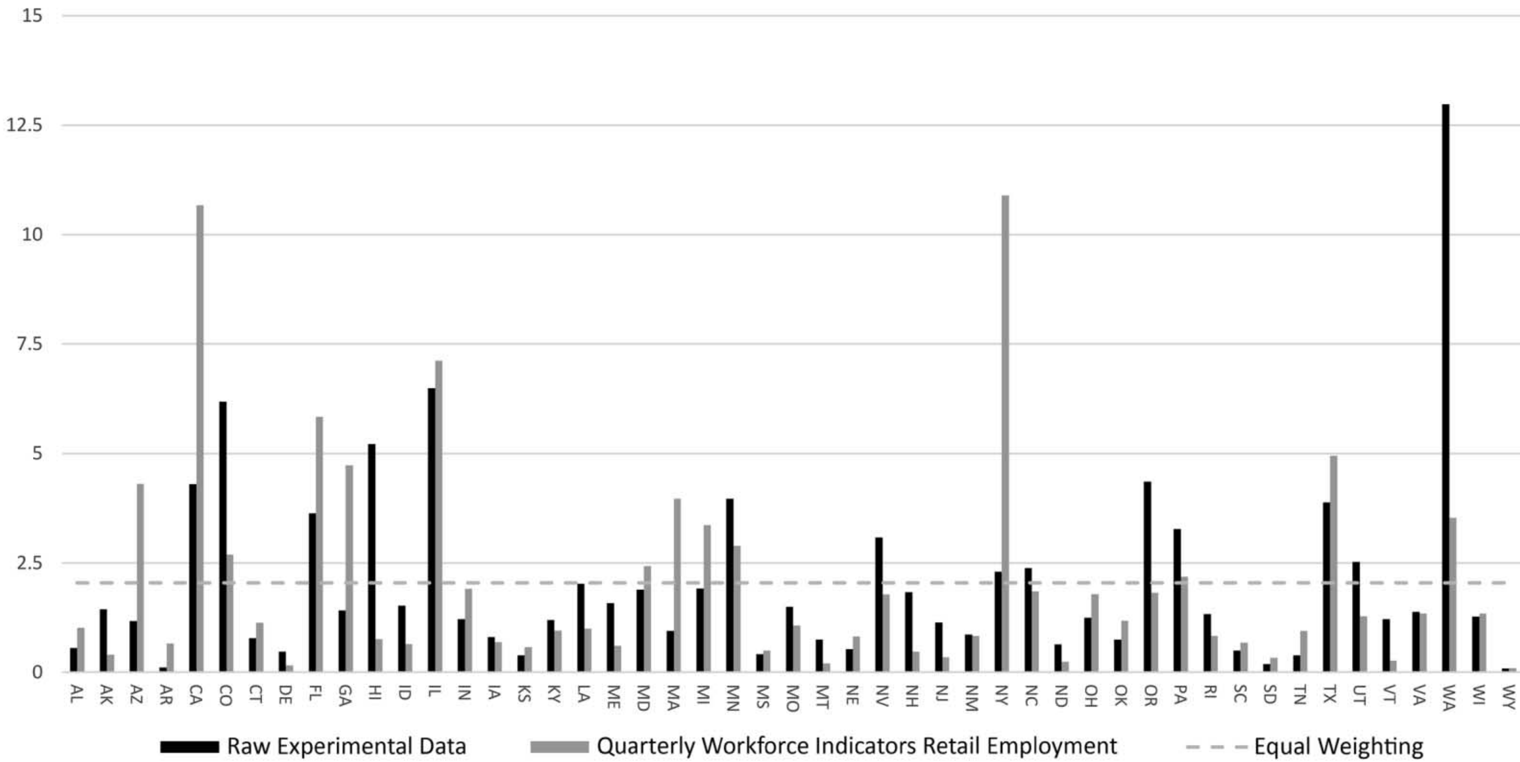

Figure 1. Percentages of observations with reweighting

population—namely, the difference in callback rates between older and younger applicants and the differences in callback rates by age under different antidiscrimination laws—then these two weighting choices are consistent with the simple advice in Solon, Haider, and Wooldridge (2015) to use weights to obtain representative estimates when the sample is unrepresentative. Between the two choices, we view reweighting the data to be representative of retail jobs to be more meaningful, as it provides estimates of the differences faced, on average, by workers applying to those jobs. Solon, Haider, and Wooldridge (2015) also argue that researchers should present both weighted and unweighted estimates to gauge whether there are heterogeneous effects (or other misspecification of the model). Our two differently weighted sets of estimates might be expected to reveal this heterogeneity if it is important, as the weighting of different states varies quite dramatically. In contrast, if the callback rate differences are similar across states, then we should obtain similar estimates whether we weight the data to be representative of jobs or of states. In practice, as detailed below, weighting does affect the statistical conclusions we draw.[22]

4. Coding of Antidiscrimination Laws

Our coding of age discrimination laws and disability discrimination laws entailed extensive background research on state statutes and their histories culled from legal databases including Lexis-Nexis, Westlaw, and Hein Online and many other sources (for example, case law, secondary sources, law journal articles, state offices, unpassed bills, and jury instructions).[23] The laws as of 2016 are reported in Table 2.[24] Further details that underlie this coding are reported in Online Appendix OA.

[22] Most previous correspondence studies focus on estimating differences in callback rates (rather than differences associated with laws), and most use a small number of cities. In a review of the experimental literature, Pager (2007, p. 120) describes a small number of studies that use more than one location and states that none in her survey used more than two. In a more recent study that provides a comprehensive survey of experiments related to discrimination against older workers (Baert et al. 2016), none of the studies covered focus on differences across jurisdictions, except, at most, to compare mean differences in callback rates. And none discuss issues of representativeness of the cities in their sample or consider the issue of weighting. The same is true of Neumark, Burn, and Button (2019), which focuses on estimating the age gap in callback rates (using data from 12 cities in 11 states). We note, though, that our conclusion regarding the age difference in callback rates is not sensitive to the weighting. As documented below, we always find significantly lower callback rates for older applicants; the result that is a bit more sensitive is the magnitude for older versus younger men. Of the two papers that focus on differences across jurisdictions, neither reweights the data to be representative. Ameri et al. (2018) report results from applying to jobs across the United States, but they do not report differences across states or discuss representativeness or weighting. Tilcsik (2011) applied to jobs in seven states but does not reference representativeness across states or weighting. Neumark, Burn, and Button (2019) do some analysis by state (11 states) to test the effects of the laws tested in this paper, but their study was not designed to test these laws. They did not employ weighting either.

[23] Earlier coding of these laws was performed for the analysis in Neumark and Song (2013) and Neumark, Song, and Button (2017); those papers also report some analyses of the effects of the laws using nonexperimental data. We update the coding to reflect laws that were in effect at the time of our data collection (2016).

[24] Table 2 reveals that the distribution of stronger protections across states does not reflect the usual pattern related to generosity of social programs, minimum wages, and so on. For example, some southern states have among the strongest antidiscrimination protections.

We focus on the two aspects of age discrimination laws that past research suggests are important. The first is the minimum firm size cutoff for the law to apply. For example, in Florida, a worker at a firm that employs fewer than 15 employees is not covered under state law. On the contrary, all employees in Colorado are covered by state law because it is applicable to all firms with at least one employee. We use a firm size cutoff of fewer than 10 workers to capture state laws that extend to substantially smaller firms (the minimum for the ADEA to apply is 20). The smaller firm size cutoff may be important because older workers are more likely to be employed at smaller firms (Neumark and Song 2013). The second is whether compensatory or punitive damages are allowed; such damages are not allowed under federal law.[25] We characterize a state law that specifies a lower firm size cutoff as a broader law and one that allows larger damages as a stronger law.

State disability discrimination laws are sometimes stronger or broader than the federal ADA in three principal ways that are captured in Table 2. As with age discrimination laws, there is a minimum firm size to which disability discrimination laws apply. The minimum size for the ADA to apply is 15 workers; in our analysis, we distinguish states with a minimum size lower than 10 workers, the same as for age discrimination laws. In fact, Table 2 shows that there is virtually no independent variation in whether the minimum firm size is lower for age discrimination or disability discrimination laws. This differs in only a handful of states, and some are states with low populations and without much data (for example, Arkansas and South Dakota). Consequently, we code a single dummy variable for whether the firm size cutoff is lower than 10 workers for the age discrimination law, the disability law, or both.

There is also variation in damages through higher or uncapped compensatory and punitive damages relative to the capped damages available under the ADA. We distinguish states with larger damages than the ADA; we base this classification on punitive rather than compensatory damages since punitive damages are likely to drive large judgments.

Finally, state laws vary in terms of the definition of disability—a different dimension of the breadth of antidiscrimination laws. Most states adopt the ADA definition, either explicitly or via case law. Under the ADA and similar state laws, plaintiffs need to prove that they have a condition that "substantially limits one or more major life activities" (42 U.S.C. sec. 12102 [1]). This has proved difficult,

[25] US Equal Employment Opportunity Commission, Remedies for Employment Discrimination (https://www.eeoc.gov/employers/remedies.cfm). Some states require proof of intent to discriminate in order for compensatory or punitive damages to be awarded, whereas others require willful violation. Because the federal law allows additional liquidated, nonpunitive damages (double back pay and benefits) when there is willful violation, the question of whether the state requires intent or willful violation may seem to be potentially relevant in deciding whether a state law offers greater protection. However, willful violation is a much stricter standard than intent (Moberly 1994). Moreover, compensatory or punitive damages are almost certainly greater than liquidated damages, and they can be much greater. As a consequence, a state law that provides compensatory or punitive damages, regardless of whether this requires proof of intent or willful violation, clearly entails stronger remedies than the federal law.

which has caused plaintiffs to lose many cases (Colker 1999).[26] However, some states use a laxer definition, changing a key part of the definition of disability from "substantially limits one or more major life activities" to either "materially limits" (Minnesota) or just "limits" (California) (see Button 2018). Other states vary the definition of disability by requiring that the disability be "medically diagnosed" without regard to whether the impairment limits major life activities (Long 2004); the definition in those states is the broadest. Table 2 includes information on both dimensions of the definition of disability, and we use both in our analysis.

5. Results

5.1. Basic Callback Rates

Figure 2 displays information on callback rates by age and by sex. This radar chart shows, for each state, the difference between the callback rate for older and younger applicants, by sex. The dashed circle represents equal callback rates for older and younger applicants, so the concentration of data points inside that circle indicates that, for most states and for both sexes, callback rates are lower for older applicants. The evidence across states is remarkably consistent, as the callback rate is higher for young applicants for most state-sex pairs, and usually notably so. Moreover, this consistency in the callback rate difference is particularly evident for women, for whom there is only one state (Maine) for which the callback rate for older applicants is higher. There are eight such states for men, but many are states with a very small number of observations; the exceptions are Florida and North Carolina, although their estimated differences in callback rates are very small.[27] Nonetheless, there is some variation across states, which suggests that studies that include only a few cities or states—which is the norm for existing correspondence studies (Pager 2007; Neumark 2018)—could generate results that are unrepresentative.

Table 3 reports aggregate descriptive information on raw differences in callback rates by age and statistical tests of whether callback rates are independent of age;[28] here, as in our preferred specifications, the data are weighted to be representative of retail jobs in the cities in our experiment. Because West Virginia has no positive responses and is dropped from the probit analysis that follows, it is also excluded from Table 3 so that the samples are the same. The exclusion has virtually no impact of the estimates in Table 3. For males, we find strong overall evidence of age discrimination, with callback rates statistically significantly lower by 6.11 percentage points (26.47 percent) for older workers compared with younger workers. The evidence for females similarly points to age discrimina-

[26] Even with the broadening of the definition of disability with the ADA Amendments Act of 2008, proving coverage is not easy for many conditions.

[27] Across all the job ads, 3.6 percent of the observations are from Florida, and 2.4 percent from North Carolina. For the other six states, these percentages range from .08 percent to 1.2 percent.

[28] This test treats the observations as independent. In the regression (probit) analyses that follow, the standard errors are clustered appropriately.

Table 2
State Disability and Age Discrimination Laws, 2016

State (City)	Age Discrimination Laws: Minimum Firm Size	Age Discrimination Laws: Larger Damages than ADEA	Disability Discrimination Laws: Minimum Firm Size	Disability Discrimination Laws: Larger Damages than ADA	Disability Discrimination Laws: Broader Definition of Disability
Alabama (Birmingham)	20	No	No law	No law	No law
Alaska (Anchorage)	1	Yes	1	Yes	No
Arizona (Phoenix)	15	No	15	No (no punitive damages)	No
Arkansas (Little Rock)	No law	No law	9	No (same as ADA)	No
California (Los Angeles)	5	Yes	5	Yes (uncapped)	Yes (limits only)
Colorado (Denver)	1	No	1	No (same as ADA)	No
Connecticut (Hartford)	3	No	3	No	Yes
Delaware (Wilmington)	4	Yes	4	No (same as ADA)	No
Florida (Miami)	15	Yes	15	No (punitive capped at $100,000)	No
Georgia (Atlanta)	1	No	15	No (no punitive)	No
Hawaii (Honolulu)	1	Yes	1	Yes (uncapped)	No
Idaho (Boise)	5	Yes	5	No (punitive capped at $10,000)	No
Illinois (Chicago)	15	Yes	1	No (no punitive)	Yes
Indiana (Indianapolis)	1	No	15	No (no punitive)	No
Iowa (Des Moines)	4	Yes	4	No (no punitive)	No
Kansas (Wichita)	4	Yes	4	No (no punitive; damages capped at $2,000)	No
Kentucky (Louisville)	8	Yes	15	No (no punitive)	No
Louisiana (New Orleans)	20	Yes	15	No (no punitive)	No
Maine (Portland)	1	Yes	1	Yes	No
Maryland (Baltimore)	1	Yes	1	No (same as ADA; no punitive in Baltimore County for employers < 15)	No
Massachusetts (Boston)	6	Yes	6	Yes (uncapped)	No
Michigan (Detroit)	1	Yes	1	No (no punitive)	No
Minnesota (Minneapolis)	1	Yes	1	No (punitive capped at $25,000)	Yes (materially limits only)

Mississippi (Jackson)	No law	No law	No law	No law	No law
Missouri (Kansas City)	6	Yes	6	Yes (uncapped)	No
Montana (Billings)	1	Yes	1	No (no punitive)	No
Nebraska (Lincoln)	20	No	15	No (no punitive)	No
Nevada (Las Vegas)	15	No	15	No (no punitive)	No
New Hampshire (Manchester)	6	Yes	6	No (no punitive)	No
New Jersey (Trenton)	1	Yes	1	Yes (uncapped)	Yes
New Mexico (Albuquerque)	4	Yes	4	No (no punitive)	No
New York (New York)	4	Yes	4	No (no punitive)	Yes
North Carolina (Charlotte)	15	No	15	No (no punitive)	No
North Dakota (Bismarck)	1	No	1	No (no damages)	No
Ohio (Columbus)	4	Yes	4	Yes (uncapped)	No
Oklahoma (Oklahoma City)	1	No	1	No (no punitive)	No
Oregon (Portland)	1	Yes	6	Yes (uncapped)	No
Pennsylvania (Pittsburgh)	4	No	4	No (no punitive)	No
Rhode Island (Providence)	4	Yes	4	Yes (uncapped)	No
South Carolina (Columbia)	15	No	15	No (same as ADA)	No
South Dakota (Sioux Falls)	No law	No law	1	No (no punitive)	No
Tennessee (Memphis)	8	Yes	8	No (no punitive)	No
Texas (Houston)	15	Yes	15	No (same as ADA)	No
Utah (Salt Lake City)	15	No	15	No (no punitive)	No
Vermont (Burlington)	1	Yes	1	Yes (uncapped)	No
Virginia (Virginia Beach)	6	No	1	No (no punitive)	No
Washington (Seattle)	8	Yes	8	No (no punitive)	Yes
West Virginia (Charleston)	12	No	12	Yes (uncapped)	No
Wisconsin (Milwaukee)	1	No	1	No (no punitive)	No
Wyoming (Cheyenne)	2	No	2	No (no punitive)	No

Sources. Data on age discrimination laws are from Neumark and Song (2013), and data on disability discrimination laws are from Neumark, Burn, and Button (2017), but both are updated to 2016.

Note. For Maryland, minimum firm size is given as 1 because this is the case for Baltimore County, our data source; the minimum is 15 for the rest of the state. Alaska has uncapped compensatory damages and punitive damages capped above the Americans with Disabilities Act (ADA) levels; Maine exceeds the ADA cap for firms of 201 or more employees. For Connecticut the evidence favors punitive damages not being available, and compensatory damages were definitely not available. ADEA = Age Discrimination in Employment Act.

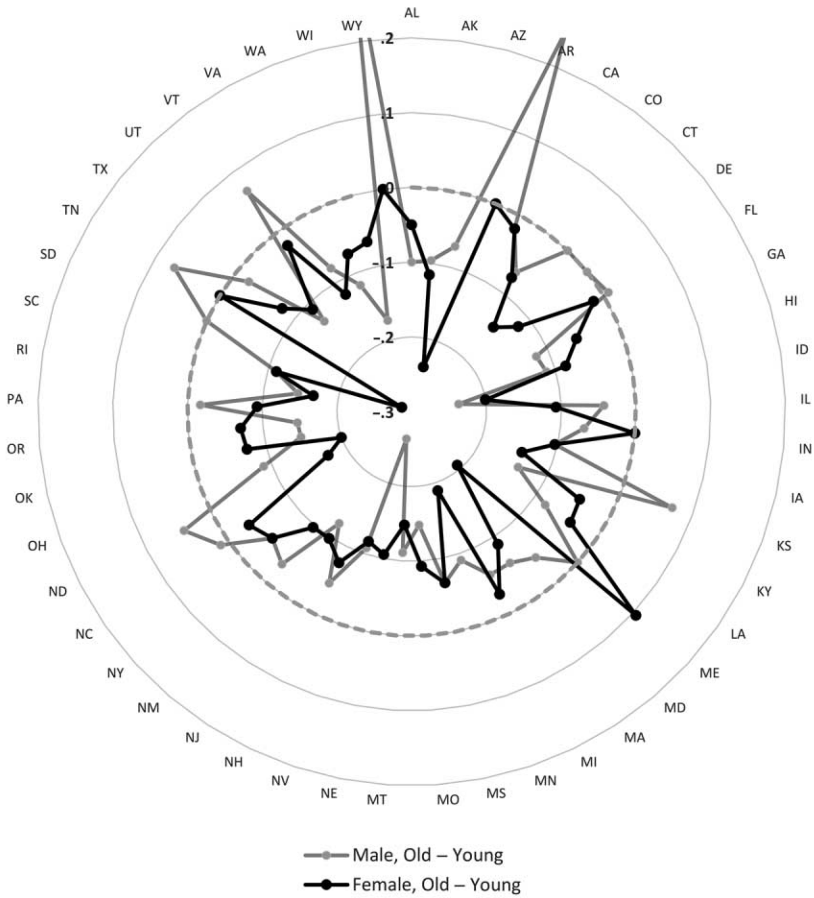

Figure 2. Relative callback rates

tion. The absolute and relative differences are larger (8.06 percentage points and 31.60 percent, respectively). These results are similar to those in Neumark, Burn, and Button (2019), although there the callback differential was larger for women (about 10 percentage points versus 6 percentage points for men).[29]

5.2. Multivariate Estimates

Table 4 reports the results of probit estimates for callbacks (equation [1]). The random assignment of age to résumés implies that the controls should not affect the estimated differences associated with age, and that is reflected here, as the es-

[29] Note that the callback rates at all ages are higher for women than for men. Similarly, Neumark, Burn, and Button (2019), Bertrand and Mullainathan (2004), and Button and Walker (2019) do not find discrimination against women in retail.

Table 3
Percentages of Callbacks by Age

	Young (29–31)	Old (64–66)	Age Independent of Callbacks (p-Value)	Difference	
				Absolute	%
Males:					
No	76.92	83.03		−6.11	−26.47
Yes	23.08	16.97			
Test of independence			.00		
Females:					
No	74.49	82.55		−8.06	−31.60
Yes	25.51	17.45			
Test of independence			.00		

Note. For each city, observations are weighted by the ratio of retail employment in the Quarterly Workforce Indicators data by sex to the number of observations in the sample. The P-values are for F-statistics for weighted tests of independence. There were no positive responses for West Virginia, so it is not included. The very small number of observations for Arkansas, South Dakota, and Wyoming are also excluded. $N = 7{,}184$ males and 7,184 females.

timates in Table 4 are very similar to those in Table 3, with an estimated shortfall in callbacks of 6.4–6.7 percentage points for older men and 8.1–8.3 percentage points for older women.[30] Given that the additional controls for résumé features make essentially no difference to the estimates, nor should they, going forward we use the more parsimonious specifications in columns 1 and 3.

5.3. Adding State Antidiscrimination Laws

We next turn to the main contribution of this paper: the estimation of the effects of state antidiscrimination laws protecting older workers on callback rates for older relative to younger workers. This analysis is based on equation (2) and the additional calculations described with reference to equations (OB1)–(OB4) in Online Appendix OB. Our first estimates are reported in Table 5 and use our preferred weighting by retail jobs (as in Tables 3 and 4).

The main effects of the variable Old refer to states where the federal law binds, and the interactions with the features of the antidiscrimination laws capture the differential in the relative callback rate where there is a stronger or broader state law along the dimension considered. We find no statistically significant evidence that cutoffs for smaller firms reduce discrimination against older job applicants (which would be reflected in positive coefficient estimates). The estimates for men are negative, and the estimates for women are positive, but they are quite small relative to other estimates discussed next; none are statistically insignificant.

The estimates for larger damages under age discrimination laws are consis-

[30] In Table 4 and subsequent tables, the standard errors are clustered at the age-state level, because the policy variation we study when we estimate the effects of state antidiscrimination laws on callbacks varies by state and by age. Absent this consideration, one might want to cluster at the level of the résumé or the job ad or use multiway clustering. Neumark, Burn, and Button (2019) verify that these alternatives have virtually no effect on the standard errors.

Table 4
Probit Estimates for Callbacks by Age

	Males		Females	
	(1)	(2)	(3)	(4)
Old (64–66)	−.064**	−.067**	−.081**	−.083**
	(.006)	(.006)	(.007)	(.007)
Controls for résumé features	No	Yes	No	Yes

Note. Marginal effects are reported, computed as the discrete change in the probability associated with the dummy variable, evaluating other variables at their means. The callback rate for young workers (29–31) is 23.08 percent for men and 25.51 percent for women. For each city, the observations are weighted by the ratio of retail employment in the Quarterly Workforce Indicators data by sex to the number of observations in the sample. Standard errors are clustered at the age-state level. All regressions include controls for the state, the order in which applications were submitted, current employment or unemployment, and attributes. Résumé features include template; e-mail script; e-mail format; script subject, opening, body, and signature; and file name format. $N = 7{,}184$ males and 7,184 females.

** Significantly different from 0 at the 1 percent level.

tently positive and larger, which indicates a reduction in the callback differential by 3.4–3.9 percentage points for men and 2.7 percentage points for women. The estimates are statistically significant at the 5 percent level for men and the 10 percent level for women. In contrast, the estimates for larger damages under disability discrimination laws are smaller (especially for men) and statistically insignificant.

Table 5 also reports estimates of the effects of the two alternative broader definitions of disability—the broader medical only definition and the definition that adds the two states (California and Minnesota) with relatively intermediate definitions of limits when compared with the ADA. The effect of a broader definition can, of course, cut two ways. On the one hand, it can extend protections and increase hiring (as reflected in callbacks). On the other hand, it could make employers warier of hiring an older worker who might suffer a decline in health and more easily become subject to state disability discrimination protections because of the broader definition. The estimates are always positive but in two of four cases are quite small (less than .01), and only the specification for the more expansive definition for men is significant. Thus, we do not view these estimates as providing a clear case of an effect of broader definitions of disability.

The estimates reported thus far are based on data weighted to make the estimates representative of retail jobs in the cities in our experiment. We next report estimates of the same models as in Table 5 using two alternative weighting schemes. We first (Table 6) reweight by the inverse of the proportion of observations in each city, which, by weighting cities equally, makes the estimates representative of the cities in our study. We then (Table 7) report unweighted estimates, which reflect the over- and undersampling of jobs by city discussed above. The key difference in Tables 6 and 7 is that the estimated effects of larger damages under age discrimination laws on the relative callback rate for older applicants diminish and become statistically insignificant (although the estimates remain

Table 5

Probit Estimates for the Effects of State Age and Disability Antidiscrimination Laws on Callbacks

	Males		Females	
	(1)	(2)	(3)	(4)
Old (64–66)	−.057**	−.056**	−.091**	−.092**
	(.011)	(.011)	(.018)	(.018)
Old (64–66) × Firm Size Cutoff < 10	−.016	−.021	.014	.014
	(.013)	(.013)	(.016)	(.016)
Old (64–66) × Age Larger Damages	.039*	.034*	.027+	.027+
	(.013)	(.013)	(.016)	(.016)
Old (64–66) × Disability Larger Damages	.002	.011	.017	.021
	(.013)	(.014)	(.013)	(.014)
Old (64–66) × Broader Disability Definition (Medical Only)	.008		.009	
	(.012)		(.013)	
Old (64–66) × Broader Disability Definition (Medical or Limits)		.021+		.017
		(.011)		(.013)

Note. The callback rate for young workers (29–31) is 23.08 percent for men and 25.51 percent for women. For each city, the observations are weighted by the ratio of retail employment in the Quarterly Workforce Indicators data by sex to the number of observations in the sample. All regressions include controls for the state, the order in which applications were submitted, current employment or unemployment, and attributes. $N = 7{,}184$ males and 7,184 females.

+ Significantly different from 0 at the 10 percent level.

* Significantly different from 0 at the 5 percent level.

** Significantly different from 0 at the 1 percent level.

positive). The estimated positive effects of a broader definition of disability also diminish, although this positive effect was not as strong anyway.[31]

The estimated effects of larger damages under disability discrimination laws for women vary relatively little; across Tables 5–7 they are in a tight range between

[31] Examination of Figures 1 and 2 and Table 2 provides some suggestive information on why and how the results might be sensitive to the weighting. For example, Figure 1 shows that Washington (Seattle) is severely down weighted based on the QWI data, while California (Los Angeles) and New York (New York) are severely up weighted. Figure 2 shows that the age-related callback difference is relatively large for men in Washington, whereas the age-related relative callback rates for California and New York are similar for men and women and a bit smaller than the average. Table 2 shows that California, New York, and Washington all have larger damages for age discrimination claims. Thus, we would expect down weighting Washington and up weighting California and New York to strengthen the evidence that larger damages for age discrimination claims reduce measured age discrimination, and conversely we would expect omitting the reweighting to weaken this evidence; this is consistent with what we find in Tables 6 and 7 compared with Table 5. To explore this further, we reestimated the weighted models excluding California, New York, and Washington. For women, the principal effect should be the exclusion of California and New York weakening the evidence that larger damages for age discrimination claims reduce measured age discrimination. The results confirm this expectation, as the estimated coefficient (standard error) on the interaction between Old (64–66) and Age Larger Damages declines to .015 (.016). For men, the prediction is not as clear, because excluding the data for Washington should push the estimates in the other direction (strengthening the evidence). On net, though, the estimated effect of larger damages for age discrimination claims in reducing measured discrimination again becomes a bit weaker than in Table 5 (.031, with a standard error of .014). Thus, for both men and women we get the same qualitative result from dropping these three states from the reweighted estimates as from dropping the reweighting.

Table 6
Probit Estimates for Callbacks, with Cities Weighted Equally

	Males		Females	
	(1)	(2)	(3)	(4)
Old (64–66)	−.091**	−.091**	−.091**	−.092**
	(.016)	(.016)	(.012)	(.012)
Old (64–66) × Firm Size Cutoff < 10	.021	.020	−.006	−.006
	(.020)	(.020)	(.014)	(.014)
Old (64–66) × Age Larger Damages	.020	.020	.004	.005
	(.017)	(.018)	(.014)	(.014)
Old (64–66) × Disability Larger Damages	−.005	−.005	.022	.021
	(.016)	(.017)	(.016)	(.016)
Old (64–66) × Broader Disability Definition (Medical Only)	.015		−.003	
	(.018)		(.011)	
Old (64–66) × Broader Disability Definition (Medical or Limits)		.016		−.008
		(.015)		(.011)

Note. Observations are weighted by 1 divided by the share of observations in the city. The callback rate for young workers (29–31) is 27.12 percent for men and 30.48 percent for women. The callback rate for young workers is 23.08 percent for men and 25.51 percent for women. All regressions include controls for the state, the order in which applications were submitted, current employment or unemployment, and attributes. $N = 7{,}184$ males and 7,184 females.
** Significantly different from 0 at the 1 percent level.

Table 7
Probit Estimates for Callbacks, with Unweighted Cities

	Males		Females	
	(1)	(2)	(3)	(4)
Old (64–66)	−.062**	−.062**	−.087**	−.088**
	(.016)	(.016)	(.012)	(.011)
Old (64–66) × Firm Size Cutoff < 10	−.013	−.015	−.009	−.007
	(.016)	(.017)	(.012)	(.011)
Old (64–66) × Age Larger Damages	.011	.007	.006	.009
	(.014)	(.015)	(.012)	(.012)
Old (64–66) × Disability Larger Damages	−.013	−.007	.020	.018
	(.014)	(.016)	(.013)	(.012)
Old (64–66) × Broader Disability Definition (Medical Only)	−.014		.006	
	(.014)		(.015)	
Old (64–66) × Broader Disability Definition (Medical or Limits)		−.0003		−.002
		(.016)		(.014)

Note. The callback rate for young workers (29–31) is 25.00 percent for men and 28.37 percent for women. All regressions include controls for the state, the order in which applications were submitted, current employment or unemployment, and attributes. $N = 7{,}184$ males and 7,184 females.
** Significantly different from 0 at the 1 percent level.

.017 and .022 and are marginally significant. Finally, across Tables 5–7 there is no statistically significant evidence of adverse effects of antidiscrimination laws related to either age or disability on the hiring of older workers.[32]

The implication of these different conclusions is that the effects of larger damages under age discrimination laws in reducing discrimination against older workers is stronger in the states for which weights are higher in the weighted data relative to the unweighted or equally weighted data. Figure 1 shows which states are significantly up weighted in the weighted data, including California, Florida, Georgia, Massachusetts, and New York (most dramatically). The states that are significantly down weighted include Colorado, Hawaii, Oregon, and Washington. We maintain that the weighted data are most relevant to asking how laws protecting older workers from discrimination affect the representative worker.

Nonetheless, an important implication of the variation in some of the results depending on the weighting is that antidiscrimination laws may have different impacts in different states (or cities). Thus, researchers have to be cautious in assuming that results generalize to all states, and studies of small numbers of states (or markets defined in other ways) may be unreliable. This same caution is likely also warranted in studies of the effects of laws intended to protect against discrimination along dimensions other than age.

5.4. Heterogeneous Effects

One other suggestion Solon, Haider, and Woolridge (2015) make when estimates are sensitive to weighting is to explore sources of heterogeneous effects to see if the sensitivity to weighting is due to those effects. We explore this using linear probability models for simplicity, given the large set of interactions involved. In particular, we consider heterogeneous effects with respect to the age structure of the population: the share 65 and older. This could be relevant in light of the Becker (1957) model of consumer discrimination, which seems especially likely to be relevant, if at all, for retail hiring.

The variable Percentage Old is the share of the population 65 and older by MSA, defined from 0 to 100, from the Census Bureau's 2016 American Community Survey 5-year estimates. We added interactions of the share of the population 65 and older with the dummy variable for older workers, and with the interactions of the dummy variable for older workers and the two damages vari-

[32] As noted above, the results are very similar using a linear probability model. For example, for the reweighted estimates corresponding to Table 5, we still find statistically significant evidence that larger damages under age discrimination claims increase the relative callback rate for older males, with similar estimated effects (.028–.034 versus .034–.039 in Table 5). For women, the estimates are a bit smaller and become insignificant (.017–.018 versus .027 in Table 5). For the results corresponding to Table 6, the estimates remain positive but are insignificant (for men, .017 versus .020 in Table 6; for women, .005 versus .004–.005 in Table 6). For the results corresponding to Table 7, the estimates remain positive but are insignificant (for men, .005–.009 versus .007–.011 in Table 7; for women, .004–.007 versus .006–.009 in Table 6).

Table 8
Linear Probability Estimates for Callbacks, with Alternative Weighting

	Males			Females		
	Weighted by Retail	Equally Weighted	Unweighted	Weighted by Retail	Equally Weighted	Unweighted
Old (64–66)	−.066**	−.088**	−.059**	−.104**	−.092**	−.086**
	(.011)	(.016)	(.015)	(.022)	(.014)	(.013)
Percentage Old × Old (64–66)	.010**	.008+	.010**	−.008	−.003	−.002
	(.003)	(.005)	(.003)	(.005)	(.004)	(.003)
Old (64–66) × Firm Size Cutoff < 10	−.031*	.009	−.026	.005	−.009	−.012
	(.012)	(.019)	(.017)	(.019)	(.015)	(.013)
Old (64–66) × Age Larger Damages	.025+	.014	.003	.015	.003	.004
	(.013)	(.017)	(.015)	(.018)	(.014)	(.012)
Percentage Old × Old (64–66) × Age Larger Damages	−.003	−.003	.000	.020**	.009	.011+
	(.005)	(.009)	(.007)	(.007)	(.007)	(.005)
Old (64–66) × Disability Larger Damages	.007	−.006	−.004	.016	.019	.017
	(.011)	(.017)	(.015)	(.014)	(.015)	(.013)
Percentage Old × Old (64–66) × Disability Larger Damages	−.016*	−.006*	−.013+	−.011	.005	−.001
	(.007)	(.009)	(.007)	(.012)	(.011)	(.009)
Old (64–66) × Broader Disability Definition (Medical or Limits)	.015	.006	.008	.012	−.001	.008
	(.012)	(.017)	(.018)	(.018)	(.016)	(.015)

Note. The main effects of the damages variables in the interactions are absorbed in the state fixed effects, but the interactions are not. Percentage Old is demeaned before forming interactions, so the main effects are evaluated at the sample means. All regressions include controls for the state, the order in which applications were submitted, current employment or unemployment, and attributes.

+ Significantly different from 0 at the 10 percent level.
* Significantly different from 0 at the 5 percent level.
** Significantly different from 0 at the 1 percent level.

ables (for which our key findings emerge).[33] The question is whether adding these heterogeneous treatment effects leads to estimates that are more robust across different weighting schemes.[34] The results are reported in Table 8.

In brief, there is some evidence of heterogeneous effects. First, for men, when the share of older workers is higher, the age difference in the callback rate is significantly lower; this is consistent with less age discrimination against men when the customer base is older. Second, for men, when the share of older workers is higher, there is some evidence of a negative effect of larger damages for disability discrimination on the relative callback rate for older applicants. Conversely, for women, when the share of older workers is higher, there is more evidence that larger damages for age discrimination do more to increase the relative callback rate for older applicants. It is hard to know how to interpret these opposite-signed effects.

Most important, though, is the question of whether allowing for heterogeneous effects reduces the sensitivity of the average effects to weighting. For one of our key results—the effect of larger damages for age discrimination on the relative callback rate for older males—it is positive and significant only in Table 5, which uses the weights representative of workers. On the other hand, with the heterogeneous effects, there is now somewhat more consistent evidence that, for women, larger damages for age discrimination increase the relative callback rate for older applicants in states with older populations.

5.5. Addressing the Heckman Critique

Finally, we turn to estimates that are intended to eliminate the bias identified by the Heckman critique. We discuss our methodology in Online Appendix OB, and this critique as applied to age discrimination is discussed in depth in Neumark, Burn, and Button (2019). As shown in the Online Appendix, correcting for this bias (while using our preferred weighting as in Table 5) eliminates the evidence of positive effects of larger damages under age discrimination laws. However, in this case, we find some statistically significant evidence of positive effects of larger damages under disability discrimination laws on the relative callback rate for older women, evidence that, in Tables 5–7, was in the same direction but only marginally significant.

6. Discussion and Conclusions

We provide evidence from a correspondence study field experiment on age discrimination in hiring for retail sales jobs. The experiment provides direct esti-

[33] Note that we also need to interact the share of older workers with the main damages variables, because unlike the main effects, they are not absorbed in the state fixed effects.

[34] As a simple example, if the age difference in relative callback rates is larger in more populous states, then in a model with a homogeneous effect of age, we will find a larger difference in callback rates if we weight by population than if we weight states equally. But if we interact the age dummy variable with population size, we will see evidence of heterogeneous effects that are not sensitive to weighting.

mates of age discrimination in hiring captured as differences in callback rates. We conduct the experiment in labor markets in all 50 states (which turns out to be important given the role of weighting). Our key focus is the empirical relationship between the measures of age discrimination by state—the difference in callback rates between old and young applicants—and variation across states in laws protecting older workers from discrimination. The identifying variation comes from state laws that are stronger than the federal laws.

We study both age discrimination and disability discrimination laws. While age discrimination laws explicitly target discrimination against older workers, we argue that it is also natural to expect disability discrimination laws to do more to protect older workers than younger workers from labor market discrimination, and hence they act as a second type of law that affects the treatment of older relative to younger workers.

As in past studies, we find evidence of hiring discrimination against older men and stronger evidence of hiring discrimination against older women. The key new evidence, however, concerns the relationship between hiring discrimination against older workers and states' variation in age and disability discrimination laws. We find some evidence that stronger laws protecting older workers from discrimination, providing for larger damages, boost callback rates for older relative to younger job applicants, which is consistent with reducing age discrimination in the labor market. In particular, we find some evidence of less discrimination against both older men and older women in states where the law allows larger damages in age discrimination claims, and we find some evidence of less discrimination against older women in states where the law allows larger damages in disability discrimination claims.

This evidence is not robust to all of the estimations we report. However, we find this evidence for age discrimination laws for both men and women when we weight the data to make the estimates representative of the universe of retail jobs in the cities included in our experiment. We also find this evidence for disability discrimination laws for women when we also adjust for potential biases in correspondence study estimates and when we allow for heterogeneous effects of laws protecting older workers from discrimination.

We do not find evidence that features that make states' laws broader to cover more older workers affect age discrimination in hiring. In particular, we do not find effects of a lower firm size cutoff for state age or disability discrimination laws or of broader definitions of disability under disability discrimination laws.

Finally, we find no consistent evidence indicating that stronger or broader laws protecting older workers from discrimination reduce callbacks to older workers (consistent with deterring the hiring of older workers). This evidence contrasts with the argument that these kinds of antidiscrimination protections principally increase termination costs and hence lead to the unintended consequence of deterring the hiring of older workers.

A potential limitation of our evidence is that it is based on cross-sectional relationships between measured discrimination and antidiscrimination laws. This is unavoidable in our context: the laws we study are long-standing, and our key innovation is to use a correspondence study to obtain direct measures of discrimination against older job applicants, which can generate only contemporaneous evidence. Nonetheless, the absence of this unintended adverse effect of laws protecting older workers from discrimination, based on our experimental evidence, is largely consistent with nonexperimental evidence on age discrimination laws and a good deal of recent nonexperimental evidence on disability discrimination laws. It bolsters the empirical case against the idea that discrimination protections have the unintended consequence of deterring the hiring of protected groups, at least with respect to older workers.

The results of our analysis also have implications for the design of experimental correspondence studies of discrimination, especially studies for which regional variation in estimated discrimination is important, like, in our case, for inferring the effects of state antidiscrimination laws. In particular, there is some evidence of heterogeneous effects of these laws across cities or states, which argues for using a large number of cities or states, so that results do not reflect idiosyncrasies of narrow regions,[35] and ensuring that the estimates are representative of the population sampled. Given that correspondence studies of labor market discrimination sample job ads in a manner similar to what we did, instead of using more standard sampling methods reflected in traditional secondary data sources, one can either weight (as we do) or try to build this representativeness into the experiment and data collection. Finally, we note that, in our study, this sensitivity pertains only to the estimation of the effects of antidiscrimination laws; the evidence of lower callback rates for older workers is highly robust and not sensitive to weighting.

References

Acemoglu, Daron, and Joshua D. Angrist. 2001. Consequences of Employment Protection? The Case of the Americans with Disabilities Act. *Journal of Political Economy* 109:915–57.

Adams, Scott J. 2004. Age Discrimination Legislation and the Employment of Older Workers. *Labour Economics* 11:219–41.

Agan, Amanda, and Sonja Starr. 2018. Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment. *Quarterly Journal of Economics* 131:191–235.

Ai, Chunrong, and Edward C. Norton. 2003. Interaction Terms in Logit and Probit Models. *Economics Letters* 80:123–29.

Ameri, Mason, Lisa Schur, Meera Adya, F. Scott Bentley, Patrick McKay, and Douglas Kruse 2018. The Disability Employment Puzzle: A Field Experiment on Employer Hiring Behavior. *Industrial and Labor Relations Review* 71:329–64.

Armour, Philip, Patrick Button, and Simon Hollands. 2018. Disability Saliency and Discrimination in Hiring. *American Economic Association: Papers and Proceedings* 108:262–66.

[35] As counterexamples, Tilcsik (2011) examines seven jurisdictions, and Agan and Starr (2018) examine two.

Baert, Stijn, Jennifer Norga, Yannick Thuy, and Marieke Van Hecke. 2016. Getting Grey Hairs in the Labour Market. An Alternative Experiment on Age Discrimination. *Journal of Economic Psychology* 57:86–101.

Becker, Gary S. 1957. *The Economics of Discrimination.* Chicago: University of Chicago Press.

Beegle, Kathleen, and Wendy A. Stock. 2003. The Labor Market Effects of Disability Discrimination Laws. *Journal of Human Resources* 38:806–59.

Bell, David, and Axel Heitmueller. 2009. The Disability Discrimination Act in the UK: Helping or Hindering Employment among the Disabled? *Journal of Health Economics* 28:465–80.

Bendick, Marc, Jr., Lauren E. Brown, and Kennington Wall. 1999. No Foot in the Door: An Experimental Study of Employment Discrimination against Older Workers. *Journal of Aging and Social Policy* 10(4):5–23.

Bendick, Marc, Jr., Charles W. Jackson, and J. Horacio Romero. 1996. Employment Discrimination against Older Workers: An Experimental Study of Hiring Practices. *Journal of Aging and Social Policy* 8(4):25–46.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review* 94:991–1013.

Bloch, Farrell. 1994. *Antidiscrimination Law and Minority Employment.* Chicago: University of Chicago Press.

Button, Patrick. 2018. Expanding Employment Discrimination Protections for Individuals with Disabilities: Evidence from California. *Industrial and Labor Relations Review* 71:365–93.

Button, Patrick, Philip Armour, and Simon Hollands. 2018a. A Comprehensive Analysis of the Effects of U.S. Disability Discrimination Laws on the Employment, Earnings, and Social Security Use of the Disabled Population. Unpublished paper. RAND Corporation, RAND Center for the Study of Aging, Santa Monica, CA.

———. 2018b. Do State Disability Discrimination Laws Increase Employment for Individuals with Disabilities? Unpublished paper. RAND Corporation, RAND Center for the Study of Aging, Santa Monica, CA.

Button, Patrick, and Mashfiqur R. Khan. 2018. Do Stronger Employment Discrimination Protections Decrease Reliance on Social Security Disability Insurance? Evidence from the Social Security Reforms. Unpublished manuscript. Tulane University, Department of Economics, New Orleans.

Button, Patrick, and Brigham Walker. 2019. Employment Discrimination against Indigenous Peoples in the United States: Evidence from a Field Experiment. Working Paper No. 25849. National Bureau of Economic Research, Cambridge, MA.

Cahill, Kevin E., Michael D. Giandrea, and Joseph F. Quinn. 2006. Retirement Patterns from Career Employment. *Gerontologist* 46:514–23.

Colker, Ruth. 1999. The Americans with Disabilities Act: A Windfall for Defendants. *Harvard Civil Rights–Civil Liberties Law Review* 34:99–162.

DeLeire, Thomas. 2000. The Wage and Employment Effects of the Americans with Disabilities Act. *Journal of Human Resources* 35:693–715.

Farber, Henry S., Dan Silverman, and Till M. von Wachter. 2017. Factors Determining Callbacks to Job Applications by the Unemployed: An Audit Study. *RSF: The Russell Sage Foundation Journal of the Social Sciences* 3:168–201.

Fix, Michael E., and Raymond J. Struyk. 1993. *Clear and Convincing Evidence: Measure-*

ment of Discrimination in America. Washington, DC: Urban Institute Press.

Gardner, Russell H., and Carolyn J. Campanella. 1991. The Undue Hardship Defense to the Reasonable Accommodation Requirement of the Americans with Disabilities Act of 1990. *Labor Lawyer* 7:37–51.

Heckman, James J. 1998. Detecting Discrimination. *Journal of Economic Perspectives* 12(2):101–16.

Heckman, James J., and Peter Siegelman. 1993. The Urban Institute Audit Studies: Their Methods and Findings. Pp. 187–258 in *Clear and Convincing Evidence: Measurement of Discrimination in America,* edited by Michael E. Fix and Raymond J. Struyk. Washington, DC: Urban Institute Press.

Hotchkiss, Julie L. 2004. A Closer Look at the Employment Impact of the Americans with Disabilities Act. *Journal of Human Resources* 39:887–911.

Houtenville, Andrew J., and Richard V. Burkhauser. 2004. Did the Employment of People with Disabilities Decline in the 1990s, and Was the ADA Responsible? Research brief. Cornell University, Rehabilitation Research and Training Center for Economic Research on Employment Policy for Persons with Disabilities, Ithaca, NY.

Johnson, Richard W., Janette Kawachi, and Eric Lewis. 2014. Older Workers on the Move: Recareering in Later Life. AARP Public Policy Institute, Washington, DC.

Jolls, Christine, and J.J. Prescott. 2004. Disaggregating Employment Protection: The Case of Disability Discrimination. Working Paper No. 10740. National Bureau of Economic Research, Cambridge, MA.

Kruse, Douglas, and Lisa Schur. 2003. Employment of People with Disabilities following the ADA. *Industrial Relations* 42:31–66.

Lahey, Joanna N. 2008a. Age, Women, and Hiring: An Experimental Study. *Journal of Human Resources* 43:30–56.

———. 2008b. State Age Protection Laws and the Age Discrimination in Employment Act. *Journal of Law and Economics* 51:433–60.

Lahey, Joanna N., and Ryan A. Beasley. 2009. Computerizing Audit Studies. *Journal of Economic Behavior and Organization* 70:508–14.

Long, Alex B. 2004. State Anti-discrimination Law as a Model for Amending the Americans with Disabilities Act. *University of Pittsburgh Law Review* 65:597–653.

Maestas, Nicole. 2010. Back to Work: Expectations and Realizations of Work after Retirement. *Journal of Human Resources* 45:718–48.

Moberly, Michael D. 1994. Reconsidering the Discriminatory Motive Requirement in ADEA Disparate Treatment Cases. *New Mexico Law Review* 24:89–124.

Molloy, Raven, Christopher L. Smith, and Abigail Wozniak. 2011. Internal Migration in the United States. *Journal of Economic Perspectives* 25(3):173–96.

Neumark, David. 2009. The Age Discrimination in Employment Act and the Challenge of Population Aging. *Research on Aging* 31:41–68.

———. 2012. Detecting Discrimination in Audit and Correspondence Studies. *Journal of Human Resources* 47:1128–57.

———. 2018. Experimental Evidence on Labor Market Discrimination. *Journal of Economic Literature* 56:799–866.

Neumark, David, Roy J. Bank, and Kyle D. Van Nort. 1996. Sex Discrimination in Restaurant Hiring: An Audit Study. *Quarterly Journal of Economics* 111:915–41.

Neumark, David, Ian Burn, and Patrick Button. 2019. Is It Harder for Older Workers to Find Jobs? New and Improved Evidence from a Field Experiment. *Journal of Political Economy* 127:922–70.

Neumark, David, and Wendy A. Stock. 1999. Age Discrimination Laws and Labor Market Efficiency. *Journal of Political Economy* 107:1081–1110.

Neumark, David, and Joanne Song. 2013. Do Stronger Age Discrimination Laws Make Social Security Reforms More Effective? *Journal of Public Economics* 108:1–16.

Neumark, David, Joanne Song, and Patrick Button. 2017. Does Protecting Older Workers from Discrimination Make It Harder to Get Hired? Evidence from Disability Discrimination Laws. *Research on Aging* 39:29–63.

Oyer, Paul, and Scott Schaefer. 2002. Sorting, Quotas, and the Civil Rights Act of 1991: Who Hires When It's Hard to Fire? *Journal of Law and Economics* 45:41–68.

Pager, Devah. 2007. The Use of Field Experiments for Studies of Employment Discrimination: Contributions, Critiques, and Directions for the Future. *Annals of the American Academy of Political and Social Science* 609:104–33.

Posner, Richard A. 1995. *Aging and Old Age*. Chicago: University of Chicago Press.

Riach, Peter A., and Judith Rich. Field Experiments of Discrimination in the Market Place. *Economic Journal* 112:F480–F518.

———. 2010. An Experimental Investigation of Age Discrimination in the English Labor Market. *Annals of Economics and Statistics*, nos. 99–100, pp. 169–85.

Rowe, John W., and Robert L. Kahn. 1997. Successful Aging. *Gerontologist* 37:433–40.

Solon, Gary, Steven J. Haider, and Jeffrey M. Wooldridge. 2015. What Are We Weighting For? *Journal of Human Resources* 50:301–16.

Starkman, Paul E. 1992. Alleging a "Pattern or Practice" under ADEA: An Analysis of the Impact and Problems of Proof. *Labor Lawyer* 8:91–123.

Sterns, Harvey L., and Suzanne M. Miklos. 1995. The Aging Worker in a Changing Environment: Organization and Individual Issues. *Journal of Vocational Behavior* 47:248–68.

Stock, Wendy A., and Kathleen Beegle. 2004. Employment Protections for Older Workers: Do Disability Discrimination Laws Matter? *Contemporary Economic Policy* 22:111–26.

Tilcsik, András. 2011. Pride and Prejudice: Employment Discrimination against Openly Gay Men in the United States. *American Journal of Sociology* 117:586–626.