

UC Irvine

UC Irvine Previously Published Works

Title

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

Permalink

<https://escholarship.org/uc/item/0jj4v77k>

Journal

The Journal of Law and Economics, 62(2)

ISSN

0022-2186

Authors

Neumark, David

Burn, Ian

Button, Patrick

et al.

Publication Date

2019-05-01

DOI

10.1086/704008

Peer reviewed



HHS Public Access

Author manuscript

J Law Econ. Author manuscript; available in PMC 2020 May 01.

Published in final edited form as:

J Law Econ. 2019 May ; 62(2): 373–402. doi:10.1086/704008.

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

David Neumark [Professor of Economics],

University of California, Irvine, NBER, IZA, and CESifo.

Ian Burn [Assistant Professor],

Swedish Institute for Social Research, Stockholm University

Patrick Button [Assistant Professor of Economics],

Tulane University, NIH/NIA Postdoctoral Scholar, RAND Corporation, NBER, and IZA.

Nanneh Chehras

Economist, Amazon

Abstract

We conduct a resume field experiment in all U.S. 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 state 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 lower discrimination against older women in states where disability discrimination law allows larger damages. Our clearest result is that these laws do not have the unintended consequence of lowering callbacks for older workers.

Introduction

In the face of population aging in the United States and other countries, policymakers have focused on efforts to boost the labor supply of older workers, mainly through public pension reforms. However, several studies suggest that older workers face age discrimination in hiring.¹ Although discrimination in hiring may not seem closely related to encouraging older

³Unlike the ADEA, the ADA does not include an exception for bona fide occupational qualifications (BFOQs). BFOQ exceptions arise when age is strongly associated with other factors that pose legitimate business or safety concerns (e.g., 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, making it much harder to justify an apparently discriminatory practice on the basis of business necessity (Gardner and Campanella, 1991).

¹See Bendick, Jackson, and Romero (1997); Bendick, Brown, and Wall (1999); Riach and Rich (2010); Lahey (2008a); Farber, Silverman, and von Wachter (2017); Baert et al. (2016), and Neumark, Burn, and Button (2019). For the most recent evidence, see Neumark et al. (2019), as well as the studies reviewed there, in Neumark (2018), and in Baert et al. (2016).

people to work longer, it may actually 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 et al., 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 resume correspondence study in all 50 states, and use the results, combined with coding of state laws, to estimate how stronger or broader anti-discrimination 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 as well as 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.² 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 (e.g., Rowe and Kahn, 1997); correspondingly, employer expectations that a worker will develop a disability in the near future – posing future accommodation costs – should also rise steeply with age. Indeed, disability discrimination laws may do more to protect many older workers than do age discrimination laws. Many ailments associated with aging have become classified as disabilities (Sterns and Miklos, 1995). These ailments can give some older workers an option of pursuing discrimination claims under either the Age Discrimination in Employment Act (ADEA) or the Americans with Disabilities Act (ADA) – or 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. A disability discrimination claim does require 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 non-experimental on the effects of anti-discrimination laws. Earlier research found that the adoption of age discrimination laws increased 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 boosted 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; Stock and Beegle, 2004), but 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 these laws reduce discrimination.

²See Stock and Beegle (2004), Neumark, Song, and Button (2017), and Button and Khan (2018).

In contrast, to garner evidence on whether stronger age and disability discrimination laws are likely to increase hiring of older workers, we conduct a large-scale resume 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. And the use of all states allows us to fully capture the variation in state age and disability discrimination laws.⁴ 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 these state laws and the measures of age discrimination in hiring from the field experiment.⁵ Our focus is on discrimination against job applicants aged 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 anti-discrimination 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 against older workers, although for older women only. Other features of these laws are not associated with differences in relative callback rates or older versus younger job applicants.

It may seem natural to expect that the effects of stronger anti-discrimination protections for older workers or workers with disabilities will – if anything – increase the hiring of older workers. However, these 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 employers' terminations. If the principal effect of laws protecting older workers from discrimination is to make it costlier to terminate older workers, then these laws could have the unintended consequence of deterring hiring of older workers (Bloch, 1994; Posner, 1995). This hypothesis was explored most fully by Oyer and Schaefer (2002), with regard to the effects of the Civil Rights Act of 1991 (CRA91) on the employment of blacks and females. CRA91 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.⁶

⁴This is a substantial expansion from the 12 cities in 11 states studied in Neumark et al. (2019), although we limit the analysis to retail jobs – whereas the previous study sent out resumes for three other types of jobs. The limited state-level variation in Neumark et al. (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.

⁵We are aware of only two other papers that look at variation in experimental evidence on discrimination across jurisdictions with different anti-discrimination laws – Tilcsik's (2011) study of discrimination against gay men, and Ameri et al.'s (2018) study of discrimination against individuals with disabilities.

Oyer and Schaefer (2002) report evidence consistent with CRA91 leading to less hiring from the groups for which protections from discrimination ostensibly increased – evidence of this kind of unintended consequence of anti-discrimination laws.⁷ However, based on 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 consequences of making older job applicants less attractive to employers. Rather, as noted above, the evidence if anything points in the direction of these 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 anti-discrimination laws in recent decades.⁸ Thus, our evidence could in principle reflect other factors correlated with employer decisions about callbacks for older workers versus younger workers and with anti-discrimination laws. However, the callback outcomes we measure are responses to very similar resumes in a single industry and do not reflect, for example, decisions of older workers to apply for jobs, or population differences between older and younger workers. Coupled with the fact that the anti-discrimination 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 anti-discrimination 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 – which we 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 non-representative 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.

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 (e.g., Fix and Struyk, 1993), because the group differences they (potentially) detect are very hard to attribute to non-discriminatory factors. It is true that AC studies do not directly distinguish between taste discrimination and statistical discrimination. However, both are illegal under U.S. law.⁹ Indeed, the last decade has witnessed an explosion of AC studies of discrimination.¹⁰ Most

⁶Oyer and Schaefer (2002) focus on race and sex because CRA91 had much less impact on cases brought under the ADEA.

⁷Lahey (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 her evidence more likely indicates that stronger age discrimination laws in fact boosted the employment of protected workers.

⁸See Neumark and Song (2013), Neumark et al. (2017), and Button, Armour, and Hollands, 2018a, 2018b). This is documented in our online legal appendix (Online Appendix OB). 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).

⁹Neumark et al. (2019) devote considerable attention to weighing these two alternative explanations of evidence consistent with age discrimination – analyses we do not delve into here.

recent research – including the present paper – use 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 collect far larger samples than audit studies, which use in-person interviews. And correspondence studies avoid “experimenter effects” that can influence the behavior of the actual applicants used in audit studies (Heckman and Siegelman, 1993).

A potential downside of correspondence studies is that we observe callbacks for job interviews, rather than actual job offers (which, in audit studies, can be observed). However, there is evidence pointing to discrimination in callbacks also reflecting discrimination in hiring. In a previous audit study of age discrimination, Bendick et al. (1999) found that 75 percent of the discrimination occurred at the interview selection stage. And Riach and Rich (2002) discuss studies of ethnic discrimination in audit studies by the International Labor Organization (ILO), in which 90 percent of discrimination occurred at the interview selection stage. Neumark (1996) finds similar evidence in an audit study of sex discrimination in restaurant hiring. Nonetheless, while we think 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 actual hiring (job offers).

The Experimental Design

This paper significantly extends the correspondence study of age discrimination by Neumark et al. (2019). That study focused 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 similar lower levels of labor market experience, consistent with the usual AC study paradigm of making applicants identical except with respect to the group 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 resumes, including for the retail sales jobs on which we focus in this paper. In addition, the resumes were designed to implement a method developed in Neumark (2012) to address the Heckman 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 of unobservables. The same resume 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 anti-discrimination laws. At the same time, the extensive resources required to extend the experiment to all 50 states necessitated focusing only on retail sales, omitting some of the occupations included in the previous study. This limitation implies that the evidence should be regarded as a case study, which may not generalize to other low-skill jobs.¹¹ In addition, because the prior evidence indicated that in

¹⁰For example, see the registry maintained by Stijn Baert (http://users.ugent.be/~sbaert/research_register.htm, viewed November 7, 2018).

retail sales there was no difference in measured age discrimination whether we used high-experience or low-experience resumes for older applicants, in this paper we use low-experience resumes that match those of younger applicants. This simplified the resume creation because a long work history did not have to be developed for the older applicants, which would be challenging given that the work histories have to be tailored to the local market.

Methods—The core analysis uses models for callbacks (C) as a function of dummy variables for age (S for older/senior) and observables from the resumes (X). The latent variable model (for C , denoted C^*) is

$$C_i^* = \alpha + \gamma S_i + X_i \delta + \varepsilon_i. \quad (1)$$

In this basic model, the null hypothesis of no discrimination against older workers implies that $\gamma = 0$. We always estimate the model for separately for the male and female job applicants we created, based on the evidence from Neumark et al. (2019) that older women experienced stronger age discrimination.

The key contribution of this paper is to estimate the effects of state anti-discrimination 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 these state laws (discussed in detail below). To ensure that we estimate the independent effects of each of the variations in state anti-discrimination protections, we simultaneously estimate the effects of the different anti-discrimination laws that we study, because the presence or absence of different features of state laws are correlated across states.

Adding to equation (1) an 's' subscript to denote states, and defining A_s as the vector of dummy variables capturing state anti-discrimination laws, we augment the model to be

$$C_{is}^* = \alpha + \gamma S_{is} + S_{is} \cdot A_s \gamma' + X_{is} \delta + \varepsilon_{is}, \quad (2)$$

where X includes the state dummy variables. Because we include state dummy variables, we do not include the main effects of the state anti-discrimination 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 anti-discrimination laws are associated with differences in the relative callback rate of older workers, captured in the vector γ' . Given that C_{is}^* is a latent variable for the propensity of a callback, for which we

¹¹AC studies typically use a very limited number of jobs. For example, Farber et al. (2017) focus only on age discrimination against women in administrative assistant jobs.

see the corresponding dichotomous outcomes, we estimate probit models. The vector γ' measures the effects of the interactions in $S_{is} \cdot A_s$, which are difference-in-differences estimates of the effects of each feature of anti-discrimination 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 OA, we detail how we use the probit model estimates to estimate the difference in differences.¹²

Resumes—Our overarching strategy in designing the resumes for our study was to use as much empirical evidence as possible to guide decisions about how to create the resumes, 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 resumes that we downloaded from a popular national job-hunting website, and then scraped to convert the resume information to data. We also rely on public-use data in choosing resume features.¹³

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

The resume database verified that there are older applicants in retail sales, and they apply for jobs online. In addition, data from the Current Population Survey (CPS) Tenure Supplement show a sizable representation of low-tenure older workers in the occupations that make up retail sales (retail salespersons and cashiers in the Census 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 to other industries.

As noted above, we use cities in all 50 states to maximize external validity and to include all variation in state anti-discrimination 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 resumes to retail jobs in specific cities, with the job and education history on each resume matching the city from which the job ad to which we apply

¹²Results were very similar using linear probability models. These estimates are described in more detail later. However, we need the probit specification to address the Heckman critique.

¹³The discussion here is brief. Many additional details are provided Neumark et al. (2019), including the online appendix to that paper, although there are some differences because that paper presents a more complex study with additional occupations, additional resume types, etc. With regard to resume creation, we do not do anything in the current paper that extends beyond what was done in Neumark et al. (2019), but in some cases what we do is more limited.

originates. This need to customize resumes to the city in which a job application is submitted is the reason we limited the analysis in this paper to retail sales jobs.

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

Half the resume quadruplets we sent out included higher skills, and half did not.¹⁴ For higher-skilled resumes, there are seven possible skills, five of which are chosen randomly (so that they are not perfectly collinear within the higher-skilled resumes). Five of the seven skills are general: a Bachelor of Arts degree; fluency in Spanish as a second language; an “employee of the month” award on the most recent job; one of three volunteer activities (food bank, homeless shelter, or animal shelter); and an absence of typographical errors. Two skills are specific to retail sales, including Microsoft Office and programs used to monitor inventory (VendPOS, AmberPOS, and Lightspeed).

Each of the four resumes in the quadruplet was randomly assigned a different resume template, which ensured that all four resumes looked different. Most other characteristics were randomly and uniquely assigned to each resume in each quadruplet to further ensure that the applicants were distinguished from each other.¹⁵

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.¹⁶ Our sample size resulted from an explicit ex ante data collection plan that covered two academic quarters, in which we collected as much data as the available job ads would allow. No data were analyzed until the data collection was complete. During that time period, we sent 14,428 applications to 3,607 jobs.

Collecting Responses—Responses to job applications could be received by email or phone. We read each email and listened to each voicemail to record the response, using information generated by the job site and auxiliary information from the responses to match

¹⁴This skill variation across resumes is needed to address the Heckman critique (Neumark, 2012).

¹⁵Other details of the resumes, including the assignment of residential addresses and schools, as well as examples of resume types, are provided in Neumark et al. (2019, and online appendix).

¹⁶The code matched the job ad data entered into a spreadsheet by the research assistants to the applicant based on city and date. Each day was randomly assigned a different quadruplet of resumes in terms of skill levels, and currently employed or unemployed. Within each quadruplet the order of resumes was randomized. The code ran every other day and added 7- to 8-hour delays between applications to the same job.

responses to job ads. Table 1 reports the distribution of responses by phone or email. Each response was coded as an unambiguous positive response (e.g., “Please call to set up an interview”), an ambiguous response (e.g. “Please return our call, we have a few additional questions”), or an unambiguous negative response (e.g. “Thank you for your interest, but the job has been filled”). To avoid having to classify subjectively the ambiguous responses, they were treated as callbacks.¹⁷

Disproportionate Numbers of Ads and Weighting—We applied to retail jobs in one city in each state (these are listed in Table 2, discussed below). Under the assumption that we applied to the number of retail jobs in proportion to the actual number of retail jobs in each city, the unweighted data would provide estimates representative of the universe of retail jobs in these cities – or at least those that advertise on the job-posting website we use. As it turns out, however, we obtained quite different numbers of observations by city relative to what would be expected based on 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.¹⁸ Second, for a couple of very large cities where the number of ads was huge, the research assistants did not apply to every job, whereas for the other cities they applied to all of them. And third, the frequency with which employers post ads on the job-hunting website we use in this study, 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, WA (black bar), relative to the share of retail jobs computed from the Quarterly Workforce Indicators data for retail (gray bar).¹⁹ In contrast, New York, NY has a large number of retail jobs, as does Los Angeles, CA, but 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 actual distribution of jobs could generate biased estimates of the effects of the average effects of these laws for the representative worker – or “average partial effects” (Solon et al., 2015). Thus, in our core analyses, we reweight the data by the ratio of the percent of employment in the QWI data (by sex) to the percent of observations in the experimental sample in the city.²⁰ This will make the estimates more representative of the distribution of retail jobs by city in the QWI data.^{21,22} As an alternative to gauge sensitivity of the results, we reweight the data by one over the share of observations in the

¹⁷The ambiguous responses are 7.8% of all cases coded as positive callbacks.

¹⁸The issue of how to sample job ads from geographic areas in correspondence studies is not unique to the job board we used.

¹⁹These figures are based on age ranges covering the age groups we study. The data are NAICS codes 44–45, for the MSA-level data, for ages 25–34, 35–44, 45–54, and 55–64, and for men and women. The percentages reported 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 covered by the graph. We use only the portion of the MSA in the state in question. The QWI data are for 2015Q1–Q3. 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 we try to overlap quarters to capture the same seasonal pattern. For two states (Michigan and Wyoming) we have to use data from 2014Q1–Q3, since the 2015 data were not yet available.

²⁰This 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.

city, giving each city equal weight and making the data representative of cities (shown by the horizontal dashed line in Figure 1).²³

If one views the goal of our analysis as estimating population descriptive statistics – namely, the difference in callback rates between older and younger applicants, and the differences in callback rates by age under different anti-discrimination laws – then these two weighting choices are consistent with the simple advice, in Solon et al. (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 these jobs. Solon et al. (2015) also argue that researchers should present both weighted and unweighted estimates, to gauge whether there are heterogeneous effects (or other model misspecification). 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.²⁴

Coding of Anti-Discrimination 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, as well as many other sources (e.g., case law, secondary sources, law journal articles, state offices, unpassed bills, and jury instructions).²⁵ The laws as of 2016 are reported in Table 2.²⁶ Further details that underlie this coding are reported in Online Appendix OB.

²¹For example, the Seattle data for both men and women are weighted by 0.27, reflecting the overrepresentation of Seattle observations by a factor of about four in the experimental data (Figure 1). And the New York City data are weighted by 4.74 for men and 4.42 for women, consistent with the underrepresentation of New York City observations in the experimental data (Figure 1).

²²However, Solon et al. (2015) warn that weighting in this fashion does not guarantee that the estimates match the population average partial effect. In their example, both weighted and unweighted estimates can inconsistently estimate the population average partial effect when the variances of different 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 which happen to be treated.

²³Figure 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 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).

²⁴Most previous correspondence studies only focus on estimating differences in callback rates (rather than differences associated with laws), and most use a small number of cities. In her review of the experimental literature, Pager (2007, p. 120) describes only 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 our prior paper (Neumark et al., 2019), which focused solely on estimating the age gap in callback rates (based on 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 actual magnitude for older vs. younger men.

Of the two papers that focus on differences across jurisdictions, none have reweighted the data to be representative. Ameri et al. (2018) applied 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, and also does not reference representativeness across states or weighting. Neumark et al. (2019) did some analysis by state (there were 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.

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 who works at a firm that employs fewer than 15 employees is not covered under the Florida 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.²⁷ 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. Like with age laws, there is a minimum firm size to which disability discrimination laws apply. The minimum firm size for the ADA to apply is 15; in our analysis, we distinguish states with a firm-size minimum lower than 10, the same as for age discrimination laws. In fact, Table 2 shows that there is virtually no independent variation in whether the firm-size minimum is lower for age discrimination or disability discrimination laws. There is only a handful of states where this differs, and some are small states without much data (e.g., Arkansas and South Dakota). Consequently, we code a single dummy variable for whether the firm-size cutoff is lower than 10 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 anti-discrimination 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. Code §12102 (1)). This has proven difficult, leading plaintiffs to lose many cases (Colker, 1999).²⁸ 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

²⁵Earlier coding of these laws was done for the analysis in Neumark and Song (2013) and Neumark et al. (2017); these papers also report some analyses of the effects of these laws using non-experimental data. We updated these laws to those that were active at the time of our data collection (2016).

²⁶Table 2 reveals that the distribution of stronger protections across states does not reflect the usual pattern related to generosity of social programs, minimum wages, etc. For example, some southern states have among the strongest anti-discrimination protections.

²⁷See United States Equal Employment Opportunity Commission (2002). 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, non-punitive 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, whether or not this requires proof of intent or willful violation, clearly entails stronger remedies than the federal law.

²⁸Even with the broadening of the definition of disability with the ADA Amendments Act of 2008 (ADAAA), proving coverage is not easy for many conditions.

limits” (Minnesota) or just “limits” (California) (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 disability definition in these states is the broadest. Table 2 includes information on both dimensions of the definition of disability, and we use both in our analysis.

Results

Basic Callback Rates

Figure 2 displays information on callback rates by age and by sex. This is a radar chart, which shows, for each state, the difference between the callback rate for older and younger applicants, for men (gray) and women (black). 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-by-sex pairs, and usually notably so. Moreover, this consistency in the callback rate difference is particularly evident for women, where 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, for which the numbers of observations are higher, although for these two states the estimated differences in callback rates are very small.²⁹ 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;³⁰ here, as in our preferred specifications, the data are weighted to be representative of retail jobs in the cities in our experiment. In Panel A, for males, we find strong overall evidence of age discrimination, with callback rates statistically significantly lower by 6.11 percentage points for older workers compared to younger workers, or 26.47% lower. The evidence in Panel B, for females, similarly points to age discrimination. The absolute and relative differences are larger (8.06 percentage points and 31.60%, respectively). These results are similar to those in Neumark et al. (2019), although there the callback differential was larger for women (about 10%, versus 6% for men).³¹

Multivariate Estimates

Table 4 reports the results of probit estimates for callbacks (equation (1)), showing marginal effects. In each case, we first report results with controls for the state, the order in which applications were submitted, current employment/unemployment, and skills. We then add

²⁹Across all the job ads, 3.6 percent of the observations come from Florida, and 2.4 percent from North Carolina. For the other six states, these percentages range from 0.08 percent to 1.2 percent.

³⁰This test treats the observations as independent. In the regression (probit) analyses that follow, the standard errors are clustered appropriately.

³¹Note that the callback rates at all ages are higher for women than for men. Similarly, Neumark et al. (2019), Bertrand and Mullainathan (2004), and Button and Walker (2019) did not find discrimination against women in retail.

controls for an extensive set of resume features listed in the table notes. The random assignment of age to resumes implies that the controls should not affect the estimated differences associated with age, and that is reflected here, as the estimates 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.³² Given that the additional resume feature controls make essentially no difference to the estimates, nor should they, going forward we use the more parsimonious specifications in columns (1) and (3).

Adding State Anti-Discrimination Laws

We next turn to the main contribution of this paper – the estimation of the effects of state anti-discrimination 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 (OA1-OA4) in Online Appendix OA. Our first estimates are reported in Table 5; these estimates use our preferred weighting by retail jobs (as in Tables 3 and 4).

The main effects of “Old” refer to states where the federal law binds, and the interactions with the features of the anti-discrimination 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 lower firm-size cutoffs reduce discrimination against older job applicants (which would be reflected in positive coefficient estimates). The two estimates for men are negative, and the two estimates for women are positive, but they are quite small relative to other estimates discussed next, and none are statistically insignificant.

The estimates for larger damages under age discrimination laws are consistently positive and larger – indicating a reduction in the callback differential by 3.4 to 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.

The last two rows of the table report estimates of the effects of the two alternative broader definitions of disability – the broader medical-only definition, and the definition that adds in the two states (California and Minnesota) with relatively “intermediate” definitions based on broader definitions of “limits” than 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 health decline and become subject to state disability discrimination protections more easily because of the broader disability definition. The estimates are always positive, but in two of four cases are quite small (less than 0.01), and

³²In this and subsequent tables, the estimates are clustered at the age-by-state level, because the policy variation we study when we estimate the effects of state anti-discrimination laws on callbacks varies by state and by age. Absent this consideration, one might want to cluster at the level of the resume or the job ad or use multi-way clustering. In Neumark et al. (2019) we verified that these alternatives have virtually no effect on the standard errors.

only significant at the 10-percent level for one specification for men – based on the second and more expansive definition of disability. 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 simply unweighted estimates, which reflect the over- and under-sampling of jobs by city discussed earlier. In Tables 6 and 7, the key difference is that the estimated effects of larger damages under age discrimination laws on the relative callback rate for older applicants diminish and becomes statistically insignificant (although the estimates remain positive). The estimated positive effects of a broader definition of disability also diminish, although this positive effect was not as strong to begin with.³³

The estimated effects of larger damages under disability discrimination laws, for women, vary relatively little, and across Tables 5–7 are in a tight range between vary between 0.017 and 0.022 and are marginally significant. Finally, across Tables 5–7 there is no evidence of adverse effects of anti-discrimination laws related to either age or disability on the hiring of older workers.³⁴

The implication of the different conclusions based on the weighting 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 get significantly upweighted in the weighted data. These include California, Florida, Georgia, Massachusetts, and New York (most dramatically). The states that get significantly downweighted include Colorado, Hawaii, Oregon, and Washington. We maintain that the weighted data are most

³³Examination of Figures 1 and 2, and Table 2, provide 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 downweighted based on the QWI data, while California (Los Angeles) and New York (New York) are severely upweighted. Figure 2 shows that the age-related callback difference is relatively large for men in Washington, whereas the age-related relative callback rates for New York and California 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, for women we might expect downweighting Washington and upweighting California and New York to weaken the evidence that larger damages for age discrimination claims reduce measured age discrimination. To explore this further, we re-estimated the weighted models for women in columns (2) and (4) of Table 5, but without California, New York, and Washington. The results show exactly this, as the estimated coefficient (standard error) on the interaction between “Old (64–66)” and “Age larger damages” declines to 0.015 (0.016). For men, we do not have as clear a prediction, because the large age-related difference in callbacks in Washington pushes things in the other direction. Again, though, when we re-estimated the weighted models for men without California, New York, and Washington, the estimated effect of larger damages for age discrimination claims in reducing measured discrimination became a bit weaker than in Table 5 (0.031, with a standard error of 0.014). Thus, for both men and women we get the same qualitative result from dropping these three states as from dropping the reweighting that makes the data representative of employment in retail.

³⁴As noted earlier, the results were very similar using a linear probability model. For example, for the reweighted estimates corresponding to Table 5, we still found statistically significant evidence that larger damages under age discrimination claims increased the relative callback rate for older males, with similar estimated effects (0.028–0.034, vs. 0.034–0.039 in Table 5). For women, the estimates were a bit smaller and became insignificant (0.017–0.018, vs. 0.027 in Table 5). For the estimates corresponding to Table 6, the corresponding estimates remained positive but insignificant (for men, 0.017, vs. 0.020 in Table 6, and for women, 0.005 vs. 0.004–0.005 in Table 6). For the estimates corresponding to Table 7, the corresponding estimates remained positive but insignificant (for men, 0.005–0.009, vs. 0.007–0.011 in Table 7, and for women, 0.004–0.007 vs. 0.006–0.009 in Table 6).

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 anti-discrimination 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.

Heterogeneous Effects

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

Specifically, we added to the model interactions between the share of the population age 65 and over with the dummy variable for older workers, and with the interactions between the old dummy variable and the two damages variables (for which our key findings emerge).³⁵ The question is whether adding these heterogeneous treatment effects leads to estimates that are more robust across different weighting schemes.³⁶ The results are reported in Table 8.

Briefly, there is some evidence of heterogeneous effects. First, for men, when the share old is higher, the age difference in callbacks rate is significantly lower, consistent with less age discrimination against men when the customer base is older. Second, for men, when the share old is high, 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 old 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 – concerning the effect of larger damages for age discrimination on the relative callback rate for older males – it remains the case that the effect is positive and significant only in Table 5, using 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

³⁵Note that we also need to interact the share old with the main damages variables, because unlike the main effects, these are not absorbed in the fixed state effects.

³⁶As 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.

age discrimination increase the relative callback rate for older applicants, for states with older populations.

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 OA, and this critique as applied to age discrimination is discussed in depth in Neumark et al. (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 on the relative callback rate for older women of larger damages under disability discrimination laws – evidence that, in Tables 5–7, was in the same direction but only marginally significant.

Conclusions and Discussion

We provide evidence from a correspondence study field experiment on age discrimination in hiring for retail sales jobs. The experiment provides direct estimates of age discrimination in hiring – captured as differences in callback rates. We conduct the experiment in labor markets in all 50 states (which turned 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 to act as a second type of law that affects 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 state 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, consistent with reducing age discrimination in the labor market. In particular, we find some evidence of lower 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 lower 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; and we find this evidence for disability discrimination laws, for women, when we also adjust for potential biases in correspondence study estimates, and also when we allow for heterogeneous effects of laws protecting older workers from discrimination.

We do not find evidence that features that make state laws broader, covering more older workers, affect age discrimination in hiring. Specifically, we do not find effects of a lower firm-size cutoff for state age or disability discrimination laws, or for 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 reduced callbacks to older workers (consistent with deterring hiring of older workers). This evidence contrasts with the argument that these kinds of anti-discrimination protections principally increase termination costs and hence lead to the unintended consequence of deterring hiring of older workers.

A potential limitation of our evidence is that it is based on cross-sectional relationships between measured discrimination and anti-discrimination 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 only generate 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 non-experimental evidence on age discrimination laws and a good deal of recent non-experimental evidence on disability discrimination laws, and hence bolsters the empirical case against the idea that discrimination protections have the unintended consequence of deterring 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 anti-discrimination laws. In particular, there is some evidence of heterogeneous effects of these laws across cities or states, which argues for (a) using a large number of cities or states, so that results do not reflect idiosyncrasies of narrow regions,³⁷ and (b) 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, rather than 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 anti-discrimination laws; the evidence of lower callback rates for older workers is highly robust and not sensitive to weighting.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

³⁷As counterexamples, Tilczik (2011) studied 7 jurisdictions, and Agan and Starr (2018) studied two.

Acknowledgments

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 Patrick Button (5T32AG000244–23). The views expressed are our own, and not those of the Foundation, the Social Security Administration, the NIH, or Amazon. This study was approved by UC Irvine’s Institutional Review Board (HS#2013–9942). We are grateful to Candela Mirat, Charlotte Densham, Chengshi Wang, Darren Lieu, Grace Jia Zhou, Helen Yu, Justin Sang Hwa Lee, Melody Dehghan, and Stephanie Harrington for research assistance. We thank Tom DeLeire, Sam Peltzman, and an anonymous reviewer for helpful suggestions.

References

- Acemoglu Daron, and Angrist Joshua D.. 2001 “Consequences of Employment Protection? The Case of the Americans with Disabilities Act.” *Journal of Political Economy*, Vol. 109, pp. 915–57.
- Adams Scott. 2004 “Age Discrimination Legislation and the Employment of Older Workers.” *Labour Economics*, Vol. 11, pp. 219–41.
- Agan Amanda, and Starr Sonja. 2018 “Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment.” *Quarterly Journal of Economics*, Vol. 131, pp. 191–235.
- Ai Chunrong, and Norton Edward C.. 2003 “Interaction Terms in Logit and Probit Models.” *Economics Letters*, Vol. 80, pp. 123–9.
- Ameri Mason, et al. 2018 “The Disability Employment Puzzle: A Field Experiment on Employer Hiring Behavior.” *Industrial and Labor Relations Review*, Vol. 71, pp. 329–364.
- Armour Philip, Button Patrick, and Hollands Simon. 2018 “Disability Saliency and Discrimination in Hiring.” *American Economic Association Papers and Proceedings*, Vol. 108, pp. 262–266.
- Baert Stijn, Norga Jennifer, Thuy Yannick, and Van Hecke Marieke. 2016 “Getting Grey Hairs in the Labour Market. An Alternative Experiment on Age Discrimination.” *Journal of Economic Psychology*, Vol. 57, pp. 86–101.
- Becker Gary S. 1957 *The Economics of Discrimination*. Chicago: University of Chicago Press.
- Beegle Kathleen, and Stock Wendy A.. 2003 “The Labor Market Effects of Disability Discrimination Laws.” *Journal of Human Resources*, Vol. 38, pp. 806–59.
- Bell David, and Heitmueller Axel. 2009 “The Disability Discrimination Act in the UK: Helping or Hindering Employment among the Disabled?” *Journal of Health Economics* Vol. 28, pp. 465–480. [PubMed: 19091434]
- Bendick Marc Jr., Brown Lauren E., and Wall Kennington. 1999 “No Foot in the Door: An Experimental Study of Employment Discrimination Against Older Workers.” *Journal of Aging & Social Policy*, Vol. 10, pp. 5–23. [PubMed: 10724770]
- Bendick Marc Jr., Jackson Charles W., and Romero J. Horacio 1997 “Employment Discrimination Against Older Workers: An Experimental Study of Hiring Practices.” *Journal of Aging & Social Policy*, Vol. 8, pp. 25–46.
- Bertrand Marianne, and Mullainathan Sendhil. 2004 “Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination.” *American Economic Review*, Vol. 94, pp. 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*, Vol. 71, pp. 365–93.
- Button Patrick, Armour Philip, and Hollands Simon. 2018a “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, Armour Philip, and Hollands Simon. 2018b “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.

- Button Patrick, and Khan Mashfiqur R. 2018 “Do Stronger Employment Discrimination Protections Decrease Reliance on Social Security Disability Insurance? Evidence from the Social Security Reforms” Unpublished paper. Tulane University, Department of Economics, New Orleans, LA.
- Button Patrick, and Walker Brigham. 2019 “Employment Discrimination Against Indigenous Peoples in the United States: Evidence from a Field Experiment” IZA Discussion Paper No. 12131. IZA Institute of Labor Economics, Bonn, Germany.
- Cahill Kevin E, Giandrea Michael D., and Quinn Joseph F. 2006 “Retirement Patterns from Career Employment.” *The Gerontologist*, Vol. 46, pp. 514–23. [PubMed: 16921005]
- Colker Ruth. 1999 “The Americans with Disabilities Act: A Windfall for Defendants.” *Harvard Civil Rights Civil Liberties Law Review*, Vol. 34, pp. 99–162.
- DeLeire Thomas. 2000 “The Wage and Employment Effects of the Americans with Disabilities Act.” *Journal of Human Resources*, Vol. 35, pp. 693–715.
- Farber Henry S., Silverman Dan, and von Wachter Till. 2017 “Factors Determining Callbacks to Job Applications by the Unemployed.” *RSF: The Russell Sage Foundation Journal of the Social Sciences*, Vol. 3, pp. 168–201.
- Fix Michael, and Struyk Raymond. 1993 *Clear and Convincing Evidence: Measurement of Discrimination in America*. Washington, DC: The Urban Institute Press.
- Gardner Russell H., and Campanella Carolyn J.. 1991 “The Undue Hardship Defense to the Reasonable Accommodation Requirement of the Americans with Disabilities Act of 1990.” *Labor Lawyer*, Vol. 7, pp. 37–51.
- Heckman James J. 1998 “Detecting Discrimination.” *Journal of Economic Perspectives*, Vol. 12, pp. 101–16.
- Heckman James, and Siegelman Peter. 1993 “The Urban Institute Audit Studies: Their Methods and Findings” In Fix and Struyk, eds., *Clear and Convincing Evidence: Measurement of Discrimination in America*. Washington, D.C.: The Urban Institute Press, pp. 187–258.
- Hotchkiss Julie L. 2004 “A Closer Look at the Employment Impact of the Americans with Disabilities Act.” *Journal of Human Resources*, Vol. 39, pp. 887–911.
- Houtenville Andrew J., and Burkhauser Richard V.. 2004 “Did the Employment of People with Disabilities Decline in the 1990s, and was the ADA Responsible?” Cornell University, Rehabilitation Research and Training Center for Economic Research on Employment Policy for Persons with Disabilities.
- Johnson Richard W. 2014 “Later Life Job Changes before and after the Great Recession.” Draft final report to AARP.
- Jolls Christine, and Prescott JJ 2004 “Disaggregating Employment Protection: The Case of Disability Discrimination.” NBER Working Paper No. 10740.
- Kruse Douglas, and Schur Lisa. 2003 “Employment of People with Disabilities Following the ADA.” *Industrial Relations*, Vol. 42, pp. 31–66.
- Lahey Joanna. 2008a “Age, Women, and Hiring: An Experimental Study.” *Journal of Human Resources*, Vol. 43, pp. 30–56.
- Lahey Joanna. 2008b “State Age Protection Laws and the Age Discrimination in Employment Act.” *Journal of Law and Economics*, Vol. 51, pp. 433–60.
- Lahey Joanna N., and Beasley Ryan A.. 2009 “Computerizing Audit Studies.” *Journal of Economic Behavior & Organization*, Vol. 70, pp. 508–14. [PubMed: 24904189]
- Maestas Nicole. 2010 “Back to Work: Expectations and Realizations of Work after Retirement.” *Journal of Human Resources*, Vol. 45, pp. 718–48. [PubMed: 24791018]
- Moberly Michael D. 1994 “Reconsidering the Discriminatory Motive Requirement in ADEA Disparate Treatment Cases.” *New Mexico Law Review*, Vol. 24, pp. 89–124.
- Molloy Raven, Smith Christopher L., and Wozniak Abigail. 2011 “Internal Migration in the United States.” *Journal of Economic Perspectives*, Vol. 25, pp. 173–96.
- Neumark David. 2018 “Experimental Evidence on Labor Market Discrimination.” *Journal of Economic Literature*, Vol. 56, pp. 799–866.
- Neumark David. 2012 “Detecting Evidence of Discrimination in Audit and Correspondence Studies.” *Journal of Human Resources*, Vol. 47, pp. 1128–57.

- Neumark David. 2009 “The Age Discrimination in Employment Act and the Challenge of Population Aging.” *Research on Aging*, Vol. 31, pp. 41–68.
- Neumark David. 1996 “Sex Discrimination in Restaurant Hiring: An Audit Study” *Quarterly Journal of Economics*, Vol. 111, pp. 915–941.
- Neumark David, Burn Ian, and Button Patrick. 2019 “Is It Harder for Older Workers to Find Jobs? New and Improved Evidence from a Field Experiment.” *Journal of Political Economy*, Vol. 127, pp 922–970.
- Neumark David, and Song Joanne. 2013 “Do Stronger Age Discrimination Laws Make Social Security Reforms More Effective?” *Journal of Public Economics*, Vol. 108, pp. 1–16.
- Neumark David, Song Joanne, and Button Patrick. 2017 “Does Protecting Older Workers from Discrimination Make It Harder to Get Hired? Evidence from Disability Discrimination Laws.” *Research on Aging*, Vol. 39, pp. 29–63. [PubMed: 28181873]
- Oyer Paul, and Schaefer Scott. 2002 “Sorting, Quotas, and the Civil Rights Act of 1991: Who Hires When It’s Hard to Fire?” *Journal of Law and Economics*, Vol. 45, pp, 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*, No. 609, pp. 104–33.
- Posner Richard A. 1995 *Aging and Old Age*. Chicago: University of Chicago Press.
- Riach Peter A., and Rich Judith. 2010 “An Experimental Investigation of Age Discrimination in the English Labor Market.” *Annals of Economics and Statistics*, No. 99/100, pp. 169–85.
- Rowe John W., and Kahn Robert L.. 1997 “Successful Aging.” *The Gerontologist*, Vol. 37, pp. 433–40. [PubMed: 9279031]
- Solon Gary, Haider Steven J., and Wooldridge Jeffrey M.. 2015 “What Are We Weighting For?” *Journal of Human Resources*, Vol. 50, pp. 301–16.
- Starkman Paul E. 1992 “Alleging a ‘Pattern or Practice’ under ADEA: An Analysis of the Impact and Problems of Proof.” *Labor Lawyer*, Vol. 8, pp. 91–123.
- Sterns Harvey L., and Miklos Suzanne M.. 1995 “The Aging Worker in a Changing Environment: Organization and Individual Issues.” *Journal of Vocational Behavior*, Vol. 47, pp. 248–68.
- Stock Wendy A., and Beegle Kathleen. 2004 “Employment Protections for Older Workers: Do Disability Discrimination Laws Matter?” *Contemporary Economic Policy*, Vol. 22, pp. 111–26.
- Tilcsik András. 2011 “Pride and Prejudice: Employment Discrimination against Openly Gay Men in the United States.” *American Journal of Sociology*, Vol. 117, 586–626.

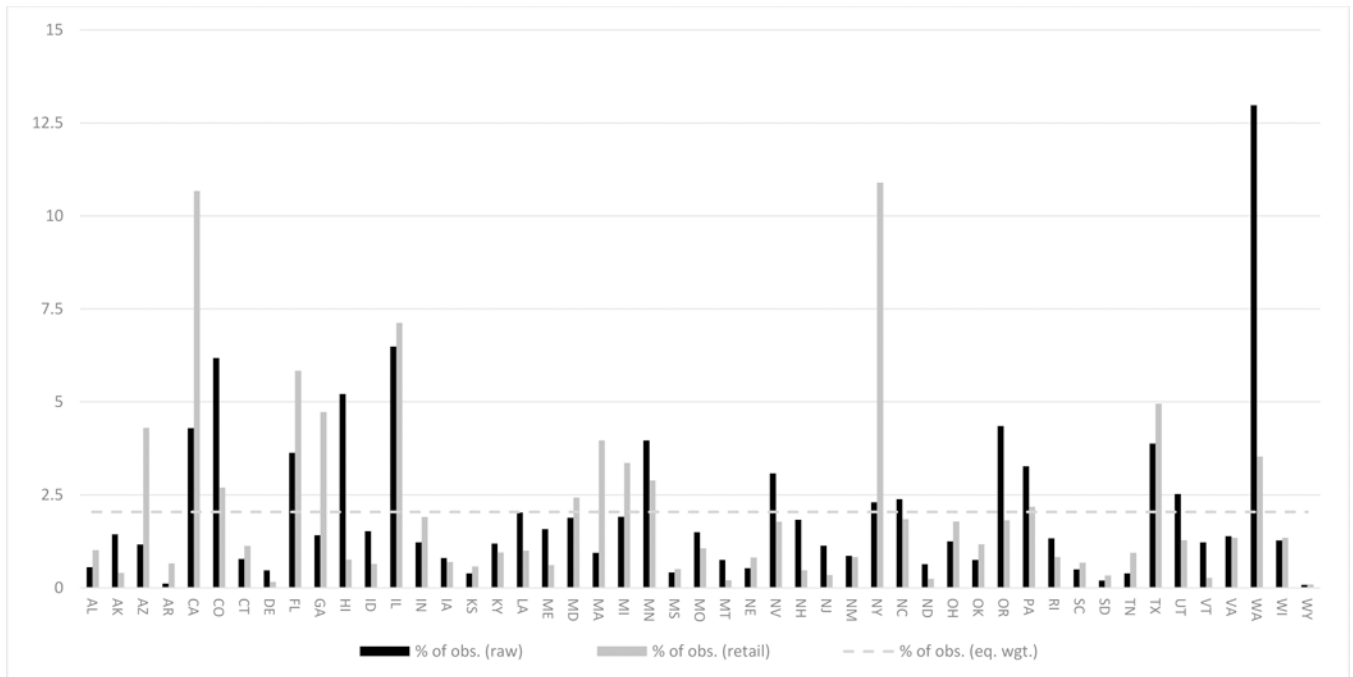


Figure 1:
 Percent of Observations by State (City), and Reweighting by QWI Data for Retail (Male and Female), and for Equal Weighting

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

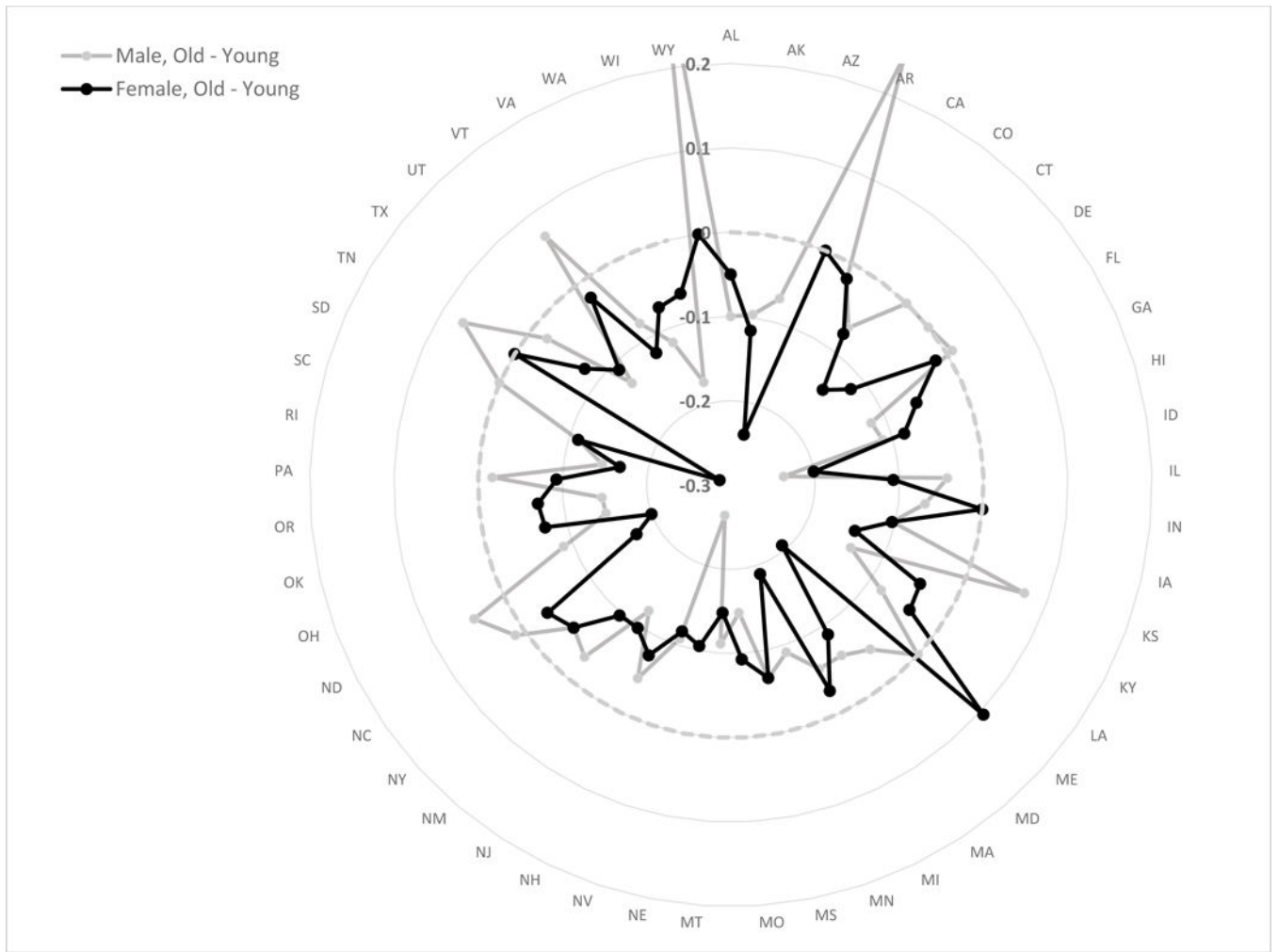


Figure 2:
Relative Callback Rates by Age (Old – Young), Male and Female Applicants, by State

Table 1:

Level of Matching of Callbacks

	Matched positive responses	No responses	Total
Voicemail	1,614	N.A.	1,614
Email	1,218	N.A.	1,218
Both	438	N.A.	438
All	3,270	11,158	14,428

Notes: There are 3,270 matched responses to 14,428 resumes that were sent out. For responses received from employers, we tried to match each response to a unique job identifier. We received three voicemails that we were unable to match to either a unique job identifier or to the resume that was sent.

Table 2:

State Disability and Age Discrimination Laws, 2016

State (City)	Age discrimination laws		Disability discrimination laws		
	Minimum firm size	Larger damages than ADEA	Minimum firm size	Larger damages than ADA	Broader (medical) 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 \$100k)	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 \$10k)	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, damages capped at \$2k)	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 damages 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 \$25k)	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

State (City)	Age discrimination laws		Disability discrimination laws		
	Minimum firm size	Larger damages than ADEA	Minimum firm size	Larger damages than ADA	Broader (medical) definition of disability
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

Notes: State laws are as of 2016. Age discrimination laws are from Neumark and Song (2013) and disability discrimination laws are from Neumark et al. (2017), but both are updated to 2016. For Maryland, under Minimum firm size, we list the value 1. This is the case for Baltimore County, from which our data come; the minimum is 15 for the rest of the state. For the states listed as “Yes” under Larger damages than ADA, but not uncapped, details are as follows: Alaska – uncapped compensatory damages, punitive damages capped above ADA levels; Maine – exceeds ADA cap for firms of 201+ employees. As discussed more in-depth in Online Appendix OB, for Connecticut the evidence favors punitive damages not being available, and compensatory damages were definitely not available.

Table 3:

Callback Rates by Age

		Young (29–31)	Old (64–66)	Absolute (percentage point) difference in callback rate for old	Percent difference in callback rate for old
<i>A. Males (N=7,184)</i>					
<i>Callback (%)</i>	No	76.92	83.03	-6.11	-26.47%
	Yes	23.08	16.97		
<i>Tests of independence (p-value), young vs. old</i>		0.00			
<i>B. Females (N=7,184)</i>					
<i>Callback (%)</i>	No	74.49	82.55	-8.06	-31.60%
	Yes	25.51	17.45		
<i>Tests of independence (p-value), young vs. old</i>		0.00			

Notes: For each city, the observations are weighted by the ratio of QWI Retail Employment, by sex, to the number of observations in the sample. The p-values reported for F-statistics for weighted tests of independence. There were no positive responses for West Virginia, so it drops out of the probit analysis in subsequent tables. We therefore also drop West Virginia from this table to have results for the same sample; this has virtually no impact on the estimates in this table. We also drop the very small number of observations for Arkansas, South Dakota, and Wyoming.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4:

Probit Estimates for Callbacks by Age, Marginal Effects

	Males		Female	
	(1)	(2)	(3)	(4)
<i>Callback estimates</i>				
Old (64–66)	-0.064 ^{***} (0.006)	-0.067 ^{***} (0.006)	-0.081 ^{***} (0.007)	-0.083 ^{***} (0.007)
<i>Controls</i>				
State, order, unemployed, skills	X	X	X	X
Resume features		X		X
<i>Callback rate for young (29–31)</i>	23.08%		25.51%	
<i>N</i>	7,184		7,184	

Notes: Marginal effects are reported, computed as the discrete change in the probability associated with the dummy variable, evaluating other variables at their means. For each city, the observations are weighted by the ratio of QWI Retail Employment, by sex, to the number of observations in the sample. Standard errors are clustered at the age-by-state level.

(^{***}) Significantly different from zero at 1-percent level,

(^{**}) 5-percent level

(^{*}) or 10-percent level.

Resume features include: template; email script; email format; script subject, opening, body, and signature; and file name format. See notes to Table 3.

Table 5:

Probit Estimates for Callbacks by Age, and Effects of State Age and Disability Anti-Discrimination Laws, Marginal Effects

	Males		Female	
	(1)	(2)	(3)	(4)
<i>Callback estimates</i>				
Old (64–66)	–0.057*** (0.011)	–0.056*** (0.011)	–0.091*** (0.018)	–0.092*** (0.018)
Old (64–66) x Age and/or disability firm-size cutoff < 10	–0.016 (0.013)	–0.021 (0.013)	0.014 (0.016)	0.014 (0.016)
Old (64–66) x Age larger damages	0.039** (0.013)	0.034** (0.013)	0.027* (0.016)	0.027* (0.016)
Old (64–66) x Disability larger damages	0.002 (0.013)	0.011 (0.014)	0.017 (0.013)	0.021 (0.014)
Old (64–66) x Broader disability definition (medical only)	0.008 (0.012)		0.009 (0.013)	
Old (64–66) x Broader disability definition (medical or limits)		0.021* (0.011)		0.017 (0.013)
<i>Controls</i>				
State, order, unemployed, skills	X	X	X	X
<i>Callback rate for young (29–31)</i>	23.08%		25.51%	
<i>N</i>	7,184		7,184	

Notes: See notes to Tables 3 and 4. For each city, the observations are weighted by the ratio of QWI Retail Employment, by sex, to the number of observations in the sample.

Table 6:

Probit Estimates for Callbacks by Age, and Effects of State Age and Disability Anti-Discrimination Laws, Marginal Effects, Cities Weighted Equally

	Males		Female	
	(1)	(2)	(3)	(4)
<i>Callback estimates</i>				
Old (64–66)	–0.091*** (0.016)	–0.091*** (0.016)	–0.091*** (0.012)	–0.092*** (0.012)
Old (64–66) x Age and/or disability firm-size cutoff < 10	0.021 (0.020)	0.020 (0.020)	–0.006 (0.014)	–0.006 (0.014)
Old (64–66) x Age larger damages	0.020 (0.017)	0.020 (0.018)	0.004 (0.014)	0.005 (0.014)
Old (64–66) x Disability larger damages	–0.005 (0.016)	–0.005 (0.017)	0.022 (0.016)	0.021 (0.016)
Old (64–66) x Broader disability definition (medical only)	0.015 (0.018)		–0.003 (0.011)	
Old (64–66) x Broader disability definition (medical or limits)		0.016 (0.015)		–0.008 (0.011)
<i>Controls</i>				
State, order, unemployed, skills	X	X	X	X
<i>Callback rate for young (29–31)</i>	27.12%		30.48%	
<i>N</i>	7,184		7,184	

Notes: See notes to Tables 3 and 4. For each city, the observations are weighted by 1 divided by the share of observations in the city.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 7:

Probit Estimates for Callbacks by Age, and Effects of State Age and Disability Anti-Discrimination Laws, Marginal Effects, Unweighted

	Males		Female	
	(1)	(2)	(3)	(4)
<i>Callback estimates</i>				
Old (64–66)	–0.062*** (0.016)	–0.062*** (0.016)	–0.087*** (0.012)	–0.088*** (0.011)
Old (64–66) x Age and/or disability firm-size cutoff < 10	–0.013 (0.016)	–0.015 (0.017)	–0.009 (0.012)	–0.007 (0.011)
Old (64–66) x Age larger damages	0.011 (0.014)	0.007 (0.015)	0.006 (0.012)	0.009 (0.012)
Old (64–66) x Disability larger damages	–0.013 (0.014)	–0.007 (0.016)	0.020 (0.013)	0.018 (0.012)
Old (64–66) x Broader disability definition (medical only)	–0.014 (0.014)		0.006 (0.015)	
Old (64–66) x Broader disability definition (medical or limits)		–0.0003 (0.016)		–0.002 (0.014)
<i>Controls</i>				
State, order, unemployed, skills	X	X	X	X
<i>Callback rate for young (29–31)</i>	25.00%		28.37%	
<i>N</i>	7,184		7,184	

Notes: See notes to Tables 3 and 4. The observations are unweighted.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Linear Probability Estimates for Callbacks by Age, and Effects of State Age and Disability Anti-Discrimination Laws, Marginal Effects, with Interactions and Alternative Weighting

Table 8:

	Males			Female		
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Weighting</i>	Table 5	Table 6	Table 7	Table 5	Table 6	Table 7
<i>Callback estimates</i>						
Old (64–66)	-0.066*** (0.011)	-0.088*** (0.016)	-0.059*** (0.015)	-0.104*** (0.022)	-0.092*** (0.014)	-0.086*** (0.013)
% Old x Old (64–66)	0.010*** (0.003)	0.008* (0.005)	0.010*** (0.003)	-0.008 (0.005)	-0.003 (0.004)	-0.002 (0.003)
Old (64–66) x Age and/or disability firm-size cutoff < 10	-0.031** (0.012)	0.009 (0.019)	-0.026 (0.017)	0.005 (0.019)	-0.009 (0.015)	-0.012 (0.013)
Old (64–66) x Age larger damages	0.025* (0.013)	0.014 (0.017)	0.003 (0.015)	0.015 (0.018)	0.003 (0.014)	0.004 (0.012)
% Old x Old (64–66) x Age larger damages	-0.003 (0.005)	-0.003 (0.009)	0.000 (0.007)	0.020*** (0.007)	0.009 (0.007)	0.011* (0.005)
Old (64–66) x Disability larger damages	0.007 (0.011)	-0.006 (0.017)	-0.004 (0.015)	0.016 (0.014)	0.019 (0.015)	0.017 (0.013)
% Old x Old (64–66) x Disability larger damages	-0.016** (0.007)	-0.006** (0.009)	-0.013* (0.007)	-0.011 (0.012)	0.005 (0.011)	-0.001 (0.009)
Old (64–66) x Broader disability definition (medical or limits)	0.015 (0.012)	0.006 (0.017)	0.008 (0.018)	0.012 (0.018)	-0.001 (0.016)	0.008 (0.015)

Notes: See notes to Tables 5–7. The other control variables are the same. The % Old is the percent of the population aged 65 and over, by MSA, from the 2016 American Community Survey 5-year estimates. The model also includes interactions between % Old and the two damages variables; the main effects of the damages variables are absorbed in the fixed state effects, but these interactions are not. The variable is demeaned before forming interactions, so the main effects measure effects evaluated at the sample means and are hence comparable to Tables 5–7. The % Old variable is defined from 0–100.