Age Discrimination in Hiring:
Evidence from Age-Blind vs. Non-Age-Blind Hiring Procedures

Appendix B: Data Appendix (Online)

Data on hires

One data source is the employee roster covering employees by location, at new restaurants that opened during the period covered by the data (2010-2016). The employee roster identifies hires at these locations around the time of the location opening. I define as hired workers those who were hired at, rehired at, or transferred to the location of a new restaurant, starting from the time the first hire, rehire, or transfer (to) was made at a location, and ending a year after the location opening date, hence capturing the hiring that occurred when a location opened. I focus only on hourly employees, excluding managers.[1]

Of course, I am ultimately interested in whether job applicants are offered jobs, since the job offer is the decision the employer makes. If there are many individuals included in the employee roster who I could not match to an application, this would imply that there is missing applicant data on that person. If this happened a lot, we would have to be concerned that the application data are not complete. In addition, I need the applicant data to capture applicant characteristics. Thus, I use the employee roster to assess the completeness of the application data for each restaurant location. I discarded a small number of locations that have somewhat incomplete application data. After doing this, there is a small number of individuals in the employee roster but not the application data, whom I discard.

I use the employee roster for two other purposes. First, as discussed below, I sometimes rely on the roster for data on the ages of applicants. Second, while for most applicants I have

[1] There are sometimes multiple observations at a location – for example, when an employee was a new hire and then was rehired at the same location. I use the initial hire.

information on whether a job offer was made, this information is sometimes missing, and if a person shows up on the employee roster but I do not have information on a job offer, I code them as having been offered a job.[2]

Data on applicants

For each restaurant location, the application data cover the earliest applications prior to the restaurant opening through one year after the restaurant opened. (This pertains to both the electronic and paper applications.)

Coding age

I obtain information on ages of applicants from different sources depending on the available data. When data are available from more than one source, I am able to verify that these sources are very consistent with each other. First, dates of birth are available from employees in the employee roster. I always use this information as definitive when I have it. The applicant data do not contain date of birth. Date of birth for most applicants was acquired by submitting information on applicants to a company called Accurint, which has a proprietary algorithm to identify peoples' dates of birth based on other information about those people, including name, address, home phone number and/or cell phone number, and more.[3] If Accurint was able to find the applicant in the data sources to which they match, Accurint provided us with the date of birth of the best match. The dates of birth were either a full date of birth (containing a month, day, and year) or a partial date of birth (missing day or month). Accurint was able to provide at least a

---

[2] Most of the job offer coding is based on job offers. Moreover, the data do not permit a reliable determination of whether offers were made, accepted, rejected, etc. (and, for what it is worth, the company never argued that the age differences were explained by older applicants receiving as many offers but rejecting more of them).

[3] For the paper applications, to the extent that there are Social Security numbers, these were provided to Accurint. Email addresses were available for electronic applications. Accurint relies on multiple public records data sources, such as Department of Motor Vehicle databases, as well as some private information sources.

partial date of birth for 86.96% of electronic applicants and 81.09% of paper applicants (Appendix Table B1).

To check the date of birth results provided by Accurint, I compared the dates of birth Accurint provided for employees to the company data (the employee roster). The year matches exactly for 94.17% of cases, and within plus or minus one year for 97.23% of cases. (When I looked at exact birth date, rather than only year, the exact match rate was 86.99%.) There is a small number of larger discrepancies, but these are very infrequent. Based on these results, I am highly confident in using the date of birth information identified by Accurint.[4]

For the electronic applications, there was also a method to recover an estimate of date of birth for some applicants for whom Accurint could not provide a date of birth or who did not appear on the employee roster. In particular, in the on-line application process, applicants were asked and sometimes report their year of high school graduation. Given that it is very common to graduate high school at or very near age 18, year of high school graduation can convey fairly accurate information on an applicant's age. To check this, I compared age based on year of high school graduation to age from the employee roster or from Accurint (when only the latter was available).[5] The match rate on age in years (calculating age as the year of application minus the year of graduation plus 18) was over 99%. I therefore used year of high school graduation when it was reported but age information was missing from the other sources. Appendix Table B2 shows the source of date of birth information for the electronic and paper applications, as well as

[4] Moreover, the high level of agreement between the employee data and the data returned by Accurint (with over 94% matching on year of birth) mitigates any concerns about bias attributable to lower measurement error for those who were actually hired (from using the employee roster). Instead, this evidence bolsters the reliability of the age information for those not in the employee roster.

[5] There is no equivalent check on the birth dates in the data for paper applications, because year of high school graduation is not reported often in the paper applications.

the percentages for which age could not be assigned.

One other question I can address using the employee roster is whether there is any pattern to the ages of people that Accurint had difficulty coding. I do this by comparing the age distribution for employees for whom Accurint returns a date of birth to the age distribution for employees for whom Accurint does not return a date of birth. As shown in Appendix Figure B1, these distributions look quite similar (albeit with lower frequency for the latter group). That is, there is no apparent tendency for Accurint to fail to return dates of birth for young vs. old employees.

I determined the youngest an applicant could be on the day they applied. For the applicants who were matched to a full date of birth, I know their exact age when they applied for a position with the company. For the applicants who were not matched to a full date of birth, I used the information that was available to provide a lower bound on the age of the applicant.[6] If the date of birth for an applicant was missing the month, December was used for the month of birth. If the date of birth for an applicant was missing the day of the month they were born, I use the last day of the month.

<u>Application outcomes and applicant characteristics</u>

For the electronic applications, the applicant database from the company includes information on interviews and job offers, capturing the dates that each applicant was interviewed and extended a job offer (if they were). There are some people who appear on the employee roster for whom job offer information is not recorded, cases that I recode as having received an offer at the same location (since clearly they were hired). This required matching the names and

[6] When I do not know the exact date of birth, there is a small possibility that an applicant is incorrectly assigned to the under 40 age group (as opposed to the 40 or over group). However, this misclassification error creates, if anything, a bias *against* finding an age difference in job offer rates.

restaurant locations between the applicant data and the employee roster – matches that are not always perfect because of things like different spelling of names or shortening of names (e.g., Tim vs. Timothy).[7]

In this hiring process, managers were instructed to move only one application through the system if an applicant submitted an application for more than one position. Thus, for applicants who submitted more than one application, I keep the application that went farthest in the hiring process, so that I will not miss a job offer if it occurred.[8]

For paper applications, I was provided with the paper applications filled out in person, with identifiers for restaurant location. The applicants first filled out a paper application for a pre-screening interview, with a limited number of questions. I was also provided with some booklets capturing the interviews and assessments that could follow the pre-screening interview, which include, most importantly, information on whether a job was offered. The interview assessment results ask open-ended questions about work experiences, covering topics such as handling stress, teamwork, and customer service. In principle the candidate is given a score on each of these, but recording of these scores is very incomplete.

The applications and the interview booklets were coded to make them machine-readable, working with an outside vendor (Bluestar) to ensure very high accuracy, including extensive

---

[7] I use the "reclink" fuzzy matching algorithm in Stata. The algorithm determines the best match for each record in one of the files. It computes a score between 0 and 1 (higher match scores reflect a higher probability that the two records are a match, and 1 indicates a perfect match) and proposes a match if two records have a match score between 0.6 and 1. Matches are based on the name on the application and the name of the employee on the roster. Matches were conditional on the employee and the electronic application originating at the same location (because it seems far less likely that an applicant and employee record with similar names at different locations are the same person). If the match between an employee and an application was scored 0.95 or better, it was coded as a match.

[8] If two applications stopped at the same stage (most common when they were both rejected without being interviewed), I randomly chose one and the other was discarded (using a random number generator), so as to not count a person's application as rejected more than once.

review of the coding in process. Any data that were difficult to read were indicated as missing or potentially problematic, and in my analysis I rely only on data that could be accurately coded without guesswork. Moreover, even if there is some coding error (and, like with all datasets, this cannot be ruled out) there is no reason to believe the small number of errors that might have entered the data via this coding are correlated with age in a way that would bias estimated differences in outcomes with respect to age.

I use data from the initial paper applications as control variables in my analysis of the paper applications. Those variables I capture and use in my analysis (which include all but a short work history, which cannot be coded consistently), include, among others: meeting minimum age requirements; legal work status; whether the person is a current employee of the company; availability of reliable transportation; record of felony convictions; shift availability; high school diploma; and previous company experience. (See Appendix Table A1.)

I did not use the information in the interview booklets to construct controls for the regression analyses for three reasons. First, there are very few booklets. I was able to match booklets to only about 13% of paper applications. Second, there is a lot of missing information in the booklets. For example, interview scores, both the total score and the individual components, are missing for 66% of booklets. Likewise, assessment results are missing for 26.2% of booklets. Third, it appears that the booklets for applicants who were given a job offer were provided by the company at a much higher rate than the booklets for applicants who were not given a job offer. The majority of booklets, 51%, are coded as receiving a job offer, with 49% not coded as receiving a job offer.[9] However, I do capture job offers from the booklets, when the information is available.

[9] In contrast, the percentage of paper applicants hired, in the paper application data I use, ranges from 4.60% to 17.07%.

Given that the interview booklets had job offer information, I had to match the interview booklets to the paper applications, again using names and location. The fuzzy matching algorithm described above was used. I coded an applicant as receiving a job offer in any of the following cases: if there was a date of offer recorded; if the rate of pay was recorded; if the position was accepted; if there was a start date given; or if there was an orientation date given. Job offers were also coded as having been given if the name on the application appeared in the employee roster at the same location, using the same matching procedure as for the electronic applications.

If the same applicant (same first name, middle name, and last name) submitted more than one paper application to the same location, only one paper application was kept. Unlike for the electronic applications, there is no way to select the application that went farthest, so I simply randomly select one of the applications.

For the electronic applications, there is a much more-detailed set of control variables available. These are listed in Appendix Table A1, and include things such as: willingness to be trained in different jobs, to stay late, to work multiple shifts; prior job experience; availability of transportation; past felony conviction; and assessment results based on the on-line applications (indicated in Appendix Table A1 under both “three-category ranking” and “numerical score”).

Data used in study

Appendix Table B3 lists, for each restaurant location, information on applications, job offers, and other features of the data, as well as an indication of how I analyze the data from each location (i.e., which type of analysis). Although this requires some explanation, I use the following criteria for which observations I include from which restaurants for which analyses. First, the analysis of electronic applications data includes all of the restaurants with only electronic applications and all of the restaurants with both types of applications. Second, the analysis of paper applications data includes data from restaurants using paper applications as long as the data do not indicate a very high share of applicants hired (implying incomplete application data), or there are simply very few paper applications (which happens for some restaurants for which there are electronic applications).

To begin, columns (2) and (3) list the number of electronic and paper applications. Most locations have exclusively one kind of application process, although in a few cases there is a mix, because restaurants switched from the paper system to the electronic system during the hiring period for that location's opening.[10]

Columns (4)-(9) provide additional information on hires and applicants. In particular, they focus on two issues. First, do I have application data for individuals that are actually hired? I answer this by asking whether there are applications for all or almost all people who are listed on the employee roster as a hire during the relevant time period and at a relevant location. As column (5) shows, the percentage (share) of hires not represented among the applications is typically in the 6-30% range, although it is much higher in a few cases, such as Locations 7 and 15, which have 59.35% and 66.47% hires missing from applicant data, respectively. Overall, there is application data for 82.51% of the individuals hired. This is not necessarily a problem, as I can still garner information on job offers for the applicants, and as long as there is not systematic difference in the ages of those hired for whom there is or is not application data, no bias is introduced.

Second, is the set of applications I have potentially biased, or does it provide a reliable sample of the applicant pool? Columns (7) and (8) point to more problematic issues for a few

[10] Restaurants did not use both methods simultaneously.

restaurants. Here, I report the share of applicants that are hired, and the share of paper applicants that are hired. In general, the share of applicants that are hired ranges between around 2% and 17%. However, for three locations – Location 2, Location 3, and Location 5 – this share is much higher, between 41% and 68% (column (8)). Thus, it appears that for these stores I am missing data on a very large share of applicants who were *not* hired. Moreover, for Location 12, although the share of applicants hired overall is not high (8.28%), the share of paper applicants hired is very high (84.92%). This reflects that I have almost no paper applications that did not result in a hire; based on the share of applicants that are hired in other locations (around 2 to 17%), for this location I am missing paper applications from many people who were not hired. This type of missing data is more problematic, because it can arise from the company not retaining applications from a large share of applicants who were not offered jobs. Given that the company could have discarded applications from the group it tended not hire, this type of missing data would more plausibly create bias in estimated age differences if they existed in the hypothetical complete data.

Appendix Table B3 then shows, in columns (9) and (10), the data I use for each location, if I do, and why. First, all of the restaurants with only electronic applications are included in the analysis of electronic applications. In addition, for restaurants with both electronic applications and paper applications, I include the electronic applications from those restaurants in the analysis of electronic applications. Restaurants for which electronic applications are included in the analysis of electronic applications are indicated in column (10) with either “EA” or “EA and PA” (in the few cases where I have and use both types of applications).

Second, among the restaurants with only paper applications, those that do not have an inordinately high share of applicants hired (column (9)) are included in the analysis of paper applications. In addition, for restaurants with a sizable number of paper applications, even if there are also electronic applications, I include the paper applications from those restaurants in the analysis of paper applications (with one exception noted below). Restaurants for which all or some applications are included in the analysis of paper applications are indicated in column (10) with either "PA" or "EA and PA."

There are some restaurant locations I will explain in more detail. One location is Location 12, where although the share of applicants hired overall is not high (8.28%), the share of *paper* applicants hired is very high (84.92%). I cannot rely on the paper applications for reasons explained above, but I have no reason to believe the electronic applications are unreliable; hence, as indicated in column (9), only the electronic applications are used. For Location 15, although the share of hires missing from the applicant data is very high (66.47%), I can study the electronic applicant data that does not have a high share of hires; hence, column (9) indicates "EA." Finally, for Location 7, although the share of hires missing from applicant data is very high (59.35%), the share of the applicant pool that are hires (column (7)) and the share of paper applicant pool that are hires (column (8)) are both very low; hence, I use both types of applications.

Of the 12,578 total paper applications, only one application per applicant per location is kept. After removing duplicate applications, the number of paper application is 11,377. I also excluded applications from locations not used in the analyses of paper applications. I further restrict the sample to only applications where I have information on date of birth. Of the total 79,805 electronic applications, only one application per applicant is kept per location. This reduces the number of applications to 51,800. From the 51,800 applications, only applicants where I have information on date of birth are used.

Appendix Table B1: Success Rate of Accurint Finding Ages of Applicants

	Electronic applications	Paper applications
Total applications	51,800 [100%]	10,388 [100%]
Number of applications Accurint found date of birth	45,043 [86.96%]	8,424 [81.09%]
Number of applications Accurint did not find date of birth	6,757 [13.04%]	1,963 [18.90%]

Note: Accurint was credited with finding an age if they reported the year of birth for an applicant. The first number in each cell is the number of applicants. The share of total applications is reported below.

Appendix Table B2: Source of Age Data by Application Type

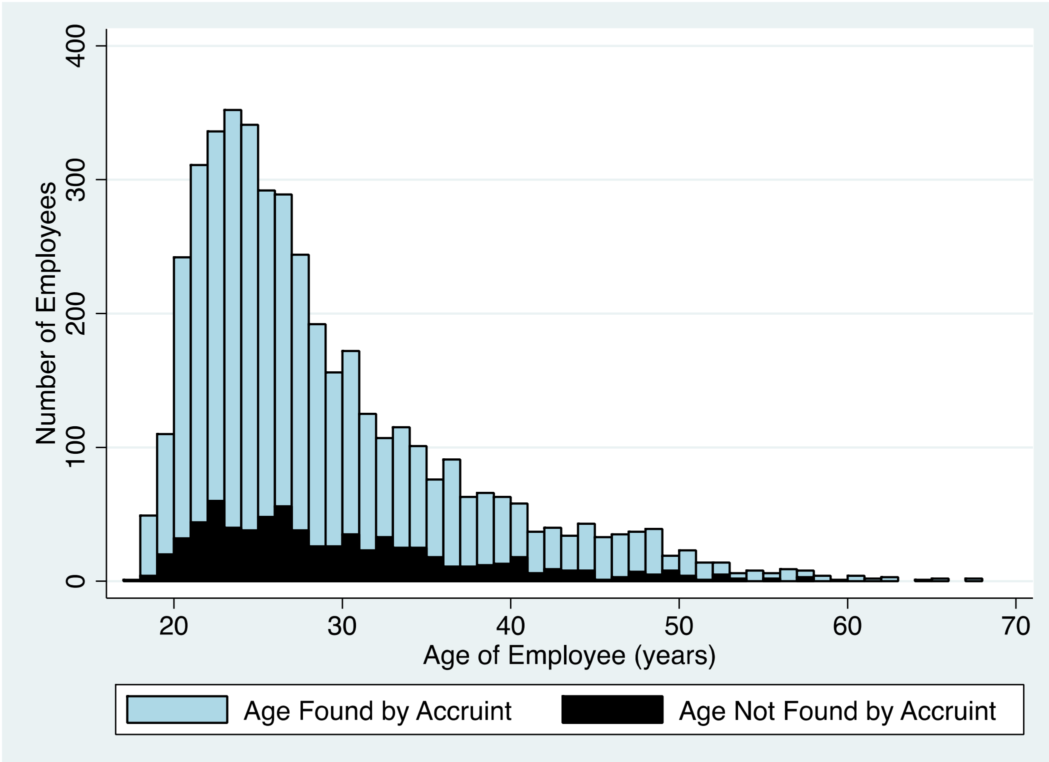
A. Electronic applications	Number of applications	Share of applications
Total number of electronic applications	51,800	100.00%
Number of applications where age was found	47,667	92.02%
Age found using employee roster (primary)	4,111	
Age found using Accurint (secondary)	41,508	
Age found using year of HS graduation (tertiary)	2,048	
Number of applications where age was not found	4,133	7.98%
B. Paper applications	Number of applications	Share of applications
Total number of paper applications for restaurants included in the analysis	10,388	100.00%
Number of applications where age was found	8,485	90.56%
Age found using employee roster (primary)	981	
Age found using Accurint (secondary)	7,504	
Number of applications where age was not found	1,903	9.44%

Note: If an application could be matched to the employee, the age was determined using the reported date of birth in the employee roster. If the applicant did not appear in the employee roster, the date of birth reported by Accurint was used to determine the age. For electronic applicants who reported a high school year of graduation and whose age was not identified using the employee roster or Accurint, the year of high school graduation was used to estimate the age. All ages are determined on the date the applicant applied. Applications with no recorded date of application are excluded.

Appendix Table B3: Treatment of Restaurants Locations and Paper and Electronic Applications from Restaurants in Analysis

Location	Electronic applicants	Paper applicants	Total hires	Share of hires missing from applicant data	Share paper applications	Share of applicant pool that are hires	Share of paper applicant pool that are hires	Analysis: hiring out of applicants	Comments
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	392	1179	184	9.78%	75.05%	10.57%	13.31%	EA and PA	
2	0	329	189	29.10%	100.00%	40.73%	40.73%		High % of paper applicants hired, suggests many paper applications missing
3	0	212	177	18.08%	100.00%	68.40%	68.40%		High % of paper applicants hired, suggests many paper applications missing
4	0	2531	151	5.96%	100.00%	5.61%	5.61%	PA	
5	0	314	177	22.60%	100.00%	43.63%	43.63%		High % of paper applicants hired, suggests many paper applications missing
6	386	1071	190	22.63%	73.51%	10.09%	11.48%	EA and PA	
7	849	891	214	59.35%	51.21%	5.00%	4.60%	EA and PA	
8	157	854	156	19.87%	84.47%	12.36%	13.93%	EA and PA	
9	0	2763	174	13.22%	100.00%	5.47%	5.47%	PA	
10	0	943	194	17.01%	100.00%	17.07%	17.07%	PA	
11	1290	1337	191	28.80%	50.89%	5.18%	5.83%	EA and PA	
12	1564	126	167	16.17%	7.46%	8.28%	84.92%	EA	High % of paper applicants hired, suggests many paper applications missing
13	3376	0	166	6.02%	0.00%	4.62%	N/A	EA	
14	3409	0	154	13.07%	0.00%	3.93%	N/A	EA	
15	1158	8	170	66.47%	0.69%	4.89%	100.00%	EA	High % of paper applicants hired, suggests many paper applications missing EA only, excluding the 8 paper applications
16	4385	6	181	7.73%	0.14%	3.80%	16.67%	EA	EA only, excluding the 6 paper applications
17	2641	0	237	7.59%	0.00%	8.29%	N/A	EA	
18	2345	1	231	12.99%	0.04%	8.57%	100.00%	EA	EA only, excluding the 1 paper applications
19	3327	0	189	12.17%	0.00%	4.99%	N/A	EA	
20	3597	3	204	13.73%	0.08%	4.89%	0.00%	EA	EA only, excluding the 3 paper applications
21	4839	0	265	8.30%	0.00%	5.02%	N/A	EA	
22	1642	0	250	20.00%	0.00%	12.18%	N/A	EA	
23	3240	0	207	13.53%	0.00%	5.52%	N/A	EA	
24	1346	0	183	13.66%	0.00%	11.74%	N/A	EA	
25	3352	0	201	6.47%	0.00%	5.61%	N/A	EA	
26	10387	0	182	7.73%	0.00%	1.62%	N/A	EA	
27	4740	0	216	7.41%	0.00%	4.22%	N/A	EA	
28	5258	0	163	16.56%	0.00%	2.59%	N/A	EA	
29	1064	0	213	28.64%	0.00%	14.29%	N/A	EA	
30	1985	0	181	23.76%	0.00%	6.95%	N/A	EA	
31	5896	0	202	10.89%	0.00%	3.05%	N/A	EA	
32	1382	0	194	17.53%	0.00%	11.58%	N/A	EA	
33	2985	0	198	13.13%	0.00%	5.76%	N/A	EA	
34	1620	10	198	14.65%	0.61%	10.37%	50.00%	EA	EA only, excluding the 10 paper applications
35	1193	0	177	9.60%	0.00%	13.41%	N/A	EA	
All locations	79,805	12,578	6,726	17.49%	13.62%	6.01%	12.00%		

Appendix Figure B1: Distribution of Employee Age on Employee Roster by Whether Accurint Returned a Date of Birth



Note: There are 6,726 employees on the employee roster. For the 5,100 employees for whom I am able to match to an application in the sample of applications used in the analyses, I calculate the age of employees rounded to the nearest year.