



Employment Project, American Civil Liberties Union of Massachusetts, Union of Minority Neighborhoods, Charles Hamilton Houston Institute for Race and Justice at Harvard Law School, Civil Rights Clinic at Howard University School of Law, Fair Housing Center of Greater Boston, Massachusetts Law Reform Institute, Justice at Work, Inc., The National Workrights Institute, Blacks in Law Enforcement of America, and NAACP Boston, amici curiae in support of appellants.

Richard Pianka, ATA Litigation Center, and Prasad Sharma on brief for American Trucking Associations, Inc., amicus curiae in support of appellees.

Mark A. de Bernardo, Joseph E. Schuler, and Jackson Lewis LLP on brief for The Council for Employment Law Equity, amicus curiae in support of appellees.

Mark A. de Bernardo, Matthew F. Nieman, and Jackson Lewis LLP on brief for The Institute for a Drug-Free Workplace, amicus curiae in support of appellees.

Peter A. Biagetti and Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C., on brief for Psychemedics Corporation, amicus curiae in support of appellees.

---

May 7, 2014

---





their hair.<sup>2</sup> Under a provision of a collective bargaining agreement between the department and the police officers' union known as Rule 111, the department selected a private company, Psychmedics Corporation, to analyze employees' hair for the presence of chemicals indicating exposure to five substances: cocaine, marijuana, opiates, PCP, and amphetamines.

When Psychmedics reported that an individual's test results indicated exposure to cocaine, a licensed physician selected by the department checked to see whether the individual had been administered "cocaine hydrochloride . . . during a medical procedure." As an additional exculpatory safeguard, the individual could elect to have a "safety-net" test of a different hair sample. During much of the period in which the plaintiffs tested positive, the safety-net tests were significantly more sensitive than the initial tests in detecting the presence of cocaine and its chemical by-products.

If an employee tested positive, and was not exonerated by either the medical review or the safety-net test, the department terminated the employee unless he or she agreed to seek rehabilitation for drug abuse and to accept an unpaid suspension of 45 work days while undergoing treatment. Before a termination became final, however, Massachusetts law required the department to

---

<sup>2</sup> The drug testing regime underwent changes in 2007, and the claims at issue here relate only to the period between 1999 and 2006, inclusive.





two different groups (e.g., black employees and white employees) presuming that members of the two groups have the same likelihood of receiving a given outcome (e.g., a promotion). See Paul Meier, Jerome Sacks, and Sandy L. Zabell, What Happened in Hazelwood: Statistics, Employment Discrimination, and the 80% Rule, 1984 Am. Bar Found. Res. J. 139, 147 (1984). Statisticians are well aware that this assumption of equal opportunity, even if true, does not mean that the two groups will experience exactly equal outcomes: random variation will often create differences. To assess the likelihood that an observed difference in outcomes resulted from mere chance, statisticians calculate the probability of observing a difference equal to or greater than that which actually occurred, assuming equal opportunity.<sup>5</sup> They call this probability the "p-value." FJC Reference Manual at 250. Statisticians usually apply the label "statistically significant" to the observed differential outcomes if the p-value is less than five percent, see Fudge v. City of Providence Fire Dep't, 766 F.2d 650, 658 n.8 (1st Cir. 1985), although they sometimes use a different cut-off, such as one percent, FJC Reference Manual at 251-52.

Essentially, a finding of statistical significance means that the data casts serious doubt on the assumption that the

---

<sup>5</sup> Because the parties have not raised the issue, we do not discuss here the distinction between "one-tailed" and "two-tailed" tests for statistical significance. See Palmer v. Shultz, 815 F.2d 84, 92-96 (D.C. Cir. 1987) (discussing the issue at length).

disparity was caused by chance. When statisticians find a disparity between racial groups to be statistically significant, they are willing to reject the hypothesis that members of the groups truly had an equal chance of receiving the outcome at issue. Id.

Statistical significance and p-value are often connected with a third concept, "standard deviation."<sup>6</sup> In disparate impact cases, standard deviation serves as another way of measuring the amount by which the observed disparity in outcomes differs from the average expected result given equal opportunity, e.g., equal rates of promotion for black and white employees. A difference of 1.96 standard deviations generally corresponds to a p-value of five percent, while a difference of three standard deviations generally corresponds to a p-value of approximately 0.5%. FJC Reference Manual at 251 n.101 As the Supreme Court observed in a case involving allegations of discriminatory jury selection, "[a]s a general rule . . . , if the difference between the expected value and the observed number is greater than two or three standard deviations, then the hypothesis that the jury drawing was random

---

<sup>6</sup> We, like other courts, use the term "standard deviation" to describe a measure that statisticians often label "standard error." FJC Reference Manual at 251 n.101.



Year	# Tested/# Positive		Standard Deviation
	Black	White	
1999	521/15	1491/10	3.43
2000	537/4	1467/3	1.35
2001	530/3	1404/3	0.81
2002	532/15	1375/4	4.41
2003	529/6	1260/3	2.01
2004	522/4	1260/4	1.92
2005	529/3	1289/1	1.43
2006	522/5	1289/2	1.95
1999 to 2006	4222/55	10,835/30	7.14 <sup>8</sup>

This evidence does not establish that the differences in outcomes were large. It shows, instead, the extent to which we can be confident that the differences in outcomes, whether large or small, were not random. To the extent the facts make it appropriate to consider the eight-year aggregate data as a single sample, we can be almost certain that the difference in outcomes associated with race over that period cannot be attributed to chance alone. Nor can randomness be viewed as other than a very

---

<sup>8</sup> The standard deviation of 7.14 for all years combined is far greater than the average of the standard deviations in the individual years for the same reason that the odds of a coin landing on tails in thirty out of forty flips is far less than the odds of getting three tails in four flips. We discuss below the department's perfunctory objection that these drug test outcomes, unlike flips of a fair coin, are not independent events.

unlikely explanation for results in at least three of the years viewed in isolation.

### **C. Alleged False Positives and Racial Bias of Hair Testing**

In addition to presenting a statistical demonstration that the racial differential in outcomes very likely did not result from chance, the plaintiffs also sought to prove that differences in the chemical and physical characteristics of hair, also associated with race, may have accounted for the observed differential in outcomes. The parties presented sharply conflicting evidence on this claim. The plaintiffs' experts argue that hair tests are relatively unreliable and note that the federal government has refused to authorize hair testing in drug screening of federal employees and employees of private industries for which the government regulates drug testing. The plaintiffs' experts also opine that black individuals tend to have higher levels of melanin in their hair, and that melanin causes cocaine and associated chemicals called cocaine metabolites to bind to hair at a higher rate. They assert that cocaine in the form of an "aerosolized powder," which forms after someone has snorted or smoked it, will "deposit on any nearby surface including the hair of non-users." These deposits, the plaintiffs say, can become incorporated into the hair in such a way that current hair testing methods cannot distinguish from the effects of actual drug use. Such incorporation might be particularly likely where a person has

undergone cosmetic hair treatments more common in the black community. The plaintiffs do not claim on appeal, though, that each of them was exposed to cocaine prior to their tests.

The department's experts counter that hair testing has been validated by numerous scientific studies. They also dispute that there is any scientific evidence of racial bias in hair testing. In particular, they point to studies showing that the relative rates of positive drug tests for black and white individuals remain materially constant across different methods of testing, specifically hair testing, urine testing, and blood testing, and note that the plaintiffs' experts do not claim that results from urine and blood testing are racially skewed. The department also questions any correlation between positive test results and melanin levels. Asserting that many Caucasians and most Asian-Americans have melanin levels as high or higher than those of the plaintiffs, the department points out that no Asian-American in the department has ever tested positive, nor has any officer in the department's Drug Control Unit or Evidence Management Unit, where officers would be most likely to be exposed to cocaine on the job.

#### **D. Procedural History**

The plaintiffs initiated this suit in state court in July 2005, and the department removed it to federal court soon after. Several years of discovery followed, culminating in summary



Trust, 487 U.S. 977, 994 (1988). A plaintiff must then show that the identified practice "causes a disparate impact on the basis of race." 42 U.S.C. § 2000e-2(k)(1)(A)(i). The Supreme Court has most recently described a prima facie showing of disparate impact as "essentially a threshold showing of a significant statistical disparity . . . and nothing more." Ricci v. DeStefano, 557 U.S. 557, 587 (2009). See also Fudge v. City of Providence Fire Dep't, 766 F.2d 650, 658 n.8 (1st Cir. 1985) (holding that a prima facie case of disparate impact can be established where "statistical tests sufficiently diminish chance as a likely explanation").

- 1. There is no genuine dispute that there is a statistically significant correlation between outcomes of the department's drug testing program and race.**

In the district court, and in their opening brief on appeal, the plaintiffs made clear that the employment practice they challenge is "the Hair Test," defined by the common elements of the drug tests used by the department between 1999 and 2006, inclusive. The department does not dispute that this practice constitutes a "particular employment practice" as required by the statute. 42 U.S.C. § 2000e-2(k)(1)(A)(i).

Having identified the challenged employment practice, the plaintiffs presented evidence that the results of this practice had a statistically significant correlation with race. As their threshold for statistical significance, the plaintiffs chose a p-value of five percent, or 1.96 standard deviations, the threshold

most commonly used by social scientists. Most federal courts have also settled on this threshold in analyzing statistical showings of disparate impact.<sup>9</sup> Because the department does not challenge this convention, we accept it here without ruling on its general applicability.

Using the five percent threshold, the plaintiffs showed that, in at least three of the eight years during the relevant period, the differential between positive results for black and white employees was statistically significant. Moreover, when the data from the eight years is aggregated, the distribution in test results for black employees deviated by more than seven standard deviations from the expected norm. The department does not meaningfully challenge the raw math behind these calculations of

---

<sup>9</sup> See Tabor v. Hilti, Inc., 703 F.3d 1206, 1223 (10th Cir. 2013) (describing a statistical significance threshold of "two or three standard deviations"); Chin v. Port Auth. of N.Y. & N.J., 685 F.3d 135, 145, 153-54 (2d Cir. 2012) ("[S]tatistical significance at the five-percent level is generally sufficient . . . ."); Stagi v. Nat'l R.R. Passenger Corp., 391 F. App'x 133, 140, 144-45 (3d Cir. 2010) (holding that the threshold is generally "a probability level at or below 0.05, or at 2 to 3 standard deviations or greater"); Paige v. California, 233 F. App'x 646, 648 (9th Cir. 2007) (accepting 1.96 standard deviations as the threshold for statistical significance); Adams v. Ameritech Servs., Inc., 231 F.3d 414, 424 (7th Cir. 2000) ("Two standard deviations is normally enough to show that it is extremely unlikely (that is, there is less than a 5% probability) that the disparity is due to chance, giving rise to a reasonable inference that the hiring was not race-neutral . . . ."); Anderson v. Zubieta, 180 F.3d 329, 340 (D.C. Cir. 1999) (indicating that "disparities . . . exceed[ing] 1.96 standard deviations under a two-tailed test of statistical significance" are sufficient to establish a prima facie of disparate impact).

statistical significance. Instead, the department raises three methodological objections.

First, the department's lawyers claim that the employees who opted to avoid termination in the wake of a positive test result by undergoing drug rehabilitation or resigning were "correctly identified as using illicit drugs" and therefore must be excluded from the plaintiffs' statistical analysis. The department's own experts provide no support for this argument penned by counsel, nor do counsel venture to explain how altering the raw numbers in this way would produce any material difference in the plaintiffs' ultimate statistical results. This argument also lacks any logical foundation that we can identify without the benefit of expert testimony. The plaintiffs identify as the challenged employment practice, and therefore subject to statistical analysis, the test used to identify which officers have used drugs, i.e., the test used to identify which officers will have to choose between termination and a suspension/rehabilitation regimen. The plaintiffs must show, then, that this selection process produces identifications that are not randomly distributed by race. The accuracy of that identification process, as determined ex post, is a different matter, perhaps relevant to the business necessity defense as discussed below, but not relevant to

the statistical showing of a disparate impact in the identifications themselves.<sup>10</sup>

Second, the department argues that plaintiff Clararise Bristow cannot rely on the statistical analysis of the outcomes of hair testing for department employees because she was not an employee. Rather, she received a conditional offer of employment, but that offer was retracted when she failed a drug test. While this difference is material to our analysis of Bristow's due process claim, her status as a tested applicant rather than a tested officer, as described by the department, is immaterial to her disparate impact claim. She took the same hair test as that given to the incumbent officers. The only difference cited by the department--that she was not eligible for rehab when she failed the test--is not a difference that would plausibly affect application of the hair test to her. Given that the department claims no other respect in which Bristow is not similarly situated to those other people who were tested, there is no basis for precluding her from relying on the plaintiffs' statistical analysis regarding the impact of the test she took.

Third, the department objects to the plaintiffs' aggregation of data from the first eight years of the drug testing

---

<sup>10</sup> For the same reason, we reject the department's argument that its evidence of the reliability of hair testing, if accepted by the factfinder, would preclude the plaintiffs from even making a prima facie showing of a statistically significant disparity.

program. Some form of aggregation (albeit certainly not covering all eight years) may be necessary to sustain the claims of those plaintiffs who tested positive in a year in which the disparity was not statistically significant when looking at the data from that year alone. The department says that any aggregation was improper because many of the same individuals were tested in different years, cutting against any assumption that the test results for the different years were independent events. After not raising this argument in its memorandum supporting its motion for summary judgment in the district court, the department now relies solely on a vague, one-sentence footnote from an expert report, which offers no analysis of the actual magnitude and effect of the claimed lack of independence in year-to-year results. Whatever the merit of this argument,<sup>11</sup> the department did not sufficiently develop it in the district court to rely on it now.

Apart from floating on appeal the three foregoing arguments that we find inadequately supported and preserved, the department raises no other basis for questioning the existence of a statistically significant correlation between race and drug test results for department employees. While an amicus brief questions

---

<sup>11</sup> A demonstration that the statistical analysis was skewed by a lack of independence in the year-to-year samples would be complex, implicating such factors as layoff and hiring practices and the probability that a person who tested negative in one year will test negative in a later year, as compared to the probability of a negative result for someone first tested in the later year.

whether the statistical analysis considered and addressed possible confounding variables (such as, for example, age) the department's experts make no attempt to explain away the differentials on such grounds, nor does the department argue that such an explanation could negate a prima facie showing of disparate impact. Rather, the department's rebuttal to the plaintiffs' prima facie showing rests on the argument adopted by the district court: even a showing of a statistically significant disparity is insufficient if the size of the impact is not sufficiently large, or "practically significant," as measured by the so-called four-fifths rule. We discuss that argument in the next section of this opinion.

**2. Title VII does not require plaintiffs to prove that the observed differential is "practically significant" in order to establish a prima facie case of disparate impact.**

We turn now to the department's argument, adopted by the district court, that even a statistically significant racial skew in outcomes does not constitute a disparate impact unless the racial differential is also sufficiently large, or "practically significant." The department correctly points out that, with a large enough set of data, even very small differences can be statistically significant. See FJC Reference Manual at 252. For example, if you were to flip a coin a million times, and the coin were to land on tails exactly 50.1% of the time, the deviation from the expected result of 50% tails and 50% heads would be statistically significant, even though it amounts to just one flip

per thousand. Recognizing this possibility, statisticians acknowledge that not all statistically significant results are practically significant, meaning "practically meaningful or important." E.g., Xitao Fan, *Statistical Significance and Effect Size in Education Research: Two Sides of a Coin*, 94 J. Educ. Res. 275, 277 (2001). According to the Federal Judicial Center's reference manual on scientific evidence, "[w]hen practical significance is lacking--when the size of a disparity is negligible--there is no reason to worry about statistical significance." FJC Reference Manual at 252.

The department therefore argues that courts in disparate impact cases should ask not simply whether a disparity is nonrandom, but also whether it is sufficiently large. Under this view, liability may not be justified, for example, where a program grants promotions to 9.1% of black employees and 9.9% of white employees, even if the imbalance is statistically significant. Cf. *Waisome v. Port Auth. of N.Y. & N.J.*, 948 F.2d 1370, 1376 (2d Cir. 1991) (finding no disparate impact where, "though the disparity was found to be statistically significant, it was of limited magnitude").

As a gauge for measuring practical significance, the department proposes the "four-fifths rule," a rule of thumb developed by the Equal Employment Opportunity Commission (EEOC). The four-fifths rule provides that where an employment practice



practical significance is lacking, relying on the four-fifths rule as a measure of practical significance.

In advocating for the adoption of a practical significance requirement, the department does not press any argument based on the text of Title VII. Indeed, the statutory language provides little aid in deciding whether a nonrandom difference is enough to make a prima facie showing of disparate impact, or whether the difference must be large. For example, Merriam Webster's dictionary defines "disparate" by using terms such as "fundamentally different" and "markedly distinct," yet it also lists as a synonym "different." Webster's New Collegiate Dictionary 329 (8th ed. 1977); Merriam-Webster's Collegiate Dictionary 360 (11th ed. 2003) (offering the same definitions and synonym). The sparse and divided case law from other circuits also fails to offer any clear answers.<sup>13</sup>

Several factors nevertheless do favor the district court's conclusion that the size of a race-based differential in outcomes matters, in some manner, in assessing disparate impact claims. Of understandable importance to the district court, the EEOC's guidelines are reasonably read as interpreting Title VII to

---

<sup>13</sup> The various approaches of the other circuits is reflected in the cases cited in footnote 9 above, as well as Waisome, 948 F.2d at 1376, and Apsley v. Boeing Co., 691 F.3d 1184, 1200-01 (10th Cir. 2012).



difficulty of concluding with confidence that an alternative practice will truly lessen the already small effect.

Acknowledging the foregoing arguments favoring a requirement that a difference in results associated with race be practically significant and not only statistically significant, we also confront powerful pragmatic arguments against adopting such a requirement. To begin, the concept of practical significance is impossible to define in even a remotely precise manner. We are aware of no test generally accepted by statisticians that we might employ to gauge practical significance (as we employ, for example, the notion that a p-value less than five percent provides good reason to presume that a difference in outcomes is not the result of chance). With no objective measure of practical significance, the label may mean that simply the person applying it views a disparity as substantial enough that a plaintiff ought to be able to sue over it. Courts would find it difficult to apply such an elusive, know-it-when-you-see-it standard, let alone instruct a jury on how to do so, and parties may find it impossible to predict results.

This case illustrates these difficulties. In trying to find a measure of practical significance, the district court turned to the four-fifths rule. Although the four-fifths rule may serve as a helpful benchmark in certain circumstances, both the Supreme Court and the EEOC have emphasized that courts should not treat the



Joint Bd. of Elec. Indus., 164 F.3d 89, 97 (2d Cir. 1998) (selection rate is the rate at which applicants pass a hiring requirement), the four-fifths rule detects no disparate impact: 82% of black employees survived, which is more than four-fifths of 100%, the rate at which white employees survived. Yet the policy in this hypothetical illustration undoubtedly has a very significant and disproportionate effect on black officers.

This illustration highlights several flaws in the four-fifths rule. First, to apply the rule in cases involving the selection of current employees for employment consequences such as termination, courts must resolve the rule's ambiguity regarding whether the "selection rate" is the rate at which employees were selected for termination or the rate at which employees survived termination.<sup>15</sup> This choice can be decisive. In the above example, if a court took the "selection rate" as the rate at which employees were fired, the four-fifths rule would indicate a disparate impact, because the 0% firing rate for white employees is less than four-fifths of the 18% firing rate for black employees. Construing the four-fifths rule in this manner, however, would lead to a different

---

<sup>15</sup> The same dilemma can occur in cases involving hiring. Consider an employer who considers only job applicants who have never been convicted of a crime: the four-fifths rule does not specify whether the "selection rate" should be the rate at which applicants are excluded from consideration, or the rate at they are included. See Green v. Missouri Pac. R.R. Co., 523 F.2d 1290, 1295 (8th Cir. 1975) (choosing the former approach, before the four-fifths rule was promulgated).

problem: the rule would detect a disparate impact even if just one employee were fired (a 0/1200 firing rate for white employees would be less than four-fifths of a 1/500 firing rate for black employees), a result that seems clearly incorrect.

Second, and relatedly, the consequences of the four-fifths rule vary in a seemingly arbitrary way depending on the magnitude of the selection rates at issue. In the example above, the policy leads to the firing of 90 black officers, or 18% of the population of black employees, but this disparity is not actionable under the four-fifths rule. Yet, if the police department provided a raise to just 1% of its white employees (12 of 1200 employees) and 0.6% of its black employees (3 of 500 black employees), this would qualify as actionable disparate impact under the four-fifths rule, even though vastly fewer black employees were affected (and less severely) than in the original scenario.

Conversely, the four-fifths rule makes no distinction between an employment practice that leads to the firing of one of nine black employees and a practice that leads to the firing of 100 of 900 black employees. In either case, the same percentage of black employees is affected. Yet, the larger sample permits a much stronger inference that a disparity is non-random and can be expected to persist through future uses of the practice.

Notwithstanding these limitations, the four-fifths rule may serve important needs in guiding the exercise of agency

discretion, or in serving as a helpful rule of thumb for employers not wanting to perform more expansive statistical examinations.<sup>16</sup> The rule itself has some practical utility. There is simply nothing in that utility, however, to justify affording decisive weight to the rule to negate or establish proof of disparate impact in a Title VII case. Having previously rejected a plaintiff's reliance on the four-fifths rule where a small sample size precluded a showing of statistical significance, Fudge, 766 F.2d at 658 n.10, we reject here the defendant's reliance on the four-fifths rule to parry a proper statistical proof of a nonrandom distribution in a case with a large sample size.

Our rejection of the four-fifths rule as suitable to trump a showing of statistical significance leaves us with no statute, regulation, or case law proposing any other mathematical measure of practical significance. Nor as a matter of theory would we expect to find any single measure of the size of the impact to determine its practical significance. To fully assess practical significance, one must consider the qualitative nature and weight of the impact. See, e.g., Steve Goodman, The Dirty Dozen: Twelve

---

<sup>16</sup> The fact that the four-fifths rule is only a rule of thumb that does not always work does not mean that it can never provide evidence of a nonrandom disparity. Thus, in Ricci v. DeStefano, 557 U.S. 557 (2009), the Supreme Court could cite pass rates of 37.5% for black and Hispanic candidates and 64% for white candidates as supporting what all parties conceded was a disparate impact. Id. at 586-87. Nothing in that citation supported use of the test (again described as a "rule of thumb") to trump a more scientific calculation of the actual statistical deviation.

P-Value Misconceptions, 45 *Seminars in Hematology* 135, 136-37 (2008). For example, just a small percentage increase in mild side effects might not render a drug unsafe as compared to an alternative drug, but the same percentage increase in fatalities could well justify the elimination of the drug. So, too, an employment practice that provides one part of a multi-part promotion test has less practical impact on each employee than does, for example, a practice that calls for the labeling and firing of police officers as illegal drug users.

Ultimately, we find any theoretical benefits of inquiring as to practical significance outweighed by the difficulty of doing so in practice in any principled and predictable manner. We therefore conclude that a plaintiff's failure to demonstrate practical significance cannot preclude that plaintiff from relying on competent evidence of statistical significance to establish a prima facie case of disparate impact.

Our confidence in rejecting a practical significance requirement is bolstered by the fact that two other requirements to be met by a plaintiff in a Title VII disparate impact case indirectly secure most of the advantages that might be gained were it possible to fashion a principled and predictable direct test of practical significance. First, the very need to show statistical significance will eliminate small impacts as fodder for litigation in many instances because proving that a small impact is

statistically significant generally requires large sample sizes, which are often unavailable. See, e.g., Fudge, 766 F.2d at 657-59. Second, even in cases like this one, in which the data is available, the subsequent steps required to successfully recover on a disparate impact theory offer an additional safeguard. An employer may rebut a prima facie case of disparate impact by showing that its use of the challenged practice is "job related for the position in question and consistent with business necessity," 42 U.S.C. § 2000e-2(k)(1)(A)(i), and a plaintiff can prevail in the face of demonstrated business necessity only by proving a failure to adopt an alternative practice that would satisfy the department's legitimate business needs "without a similarly undesirable racial effect." Albemarle Paper Co. v. Moody, 422 U.S. 405, 425 (1979). See 42 U.S.C. § 2000e-2(k)(1)(c) (adopting case law prior to June 4, 1989, "with respect to the concept of 'alternative employment practice'").

Proving that an alternative practice will not have the impact identified by a plaintiff when that impact is small leaves little margin for error and will often require extensive data. A plaintiff who subjects a defendant's job-related practice to the sensitivity of a large sample analysis can fairly be required to show through statistical evidence, and with equal confidence, that the proffered alternative practice will have a smaller impact,



3. We decline to decide in the first instance whether the drug testing program is "job-related . . . and consistent with business necessity" and whether the plaintiffs have offered an adequate alternative.

Once a plaintiff has made a prima facie showing of a disparate impact, the burden shifts to the employer to show that "the challenged practice is job related for the position in question and consistent with business necessity." 42 U.S.C. § 2000e-2(k)(1)(A)(i). If the employer makes such a showing, a plaintiff has one final path to success, by proving the existence of an "alternative employment practice," 42 U.S.C. § 2000e-2(k)(1)(A)(ii), defined in case law as a different "test or selection device[], without a similarly undesirable racial effect," which "also serve[s] the employer's legitimate interest . . . .," Albemarle Paper Co. v. Moody, 422 U.S. 405, 425 (1975). If a plaintiff makes such a showing, and the employer "refuses to adopt such alternative employment practice," then the plaintiff prevails. 42 U.S.C. § 2000e-2(k)(1)(A)(ii).

The department invites us to consider on this appeal whether it has established that its hair testing program satisfies the business necessity defense under the disparate impact provisions of Title VII. Supreme Court decisions illustrate that

the defense has two main components.<sup>17</sup> First, the department must show that its program aims to measure a characteristic that constitutes an "important element[] of work behavior." Albemarle Paper Co., 422 U.S. at 431; see also Dothard v. Rawlinson, 433 U.S. 321, 331 (1977) (holding that an employer satisfied this requirement by showing that the challenged practice measured a characteristic "essential to effective job performance"). Second, the department must show that the outcomes of the drug testing program are "predictive of or significantly correlated with" the characteristic described above. Albemarle Paper Co., 422 U.S. at 431.<sup>18</sup>

Here, the plaintiffs have not disputed that abstention from illegal drug use is an important element of police officer behavior. They have admitted that the department has a "legitimate purpose of ensuring a drug-free workplace." What remains to be determined, then, is whether the results of the department's drug testing regime are "predictive of or significantly correlated with"

---

<sup>17</sup> According to the "Purposes" section of the 1991 statute that added the disparate impact provision to Title VII, Congress aimed to "codify the concepts of 'business necessity' and 'job related' enunciated by the Supreme Court in Griggs . . . and in other Supreme Court decisions prior to Wards Cove Packing Co. [decided in 1989]." Pub L. No. 102-166, § 3(2), 105 Stat. 1071, 1071 (1991).

<sup>18</sup> Our articulation of this two-part test is substantially the same as the three-part inquiry adopted by the Ninth Circuit. See Association of Mexican-Am. Educators v. California, 231 F.3d 572, 585 (9th Cir. 2000) (en banc).

drug use. The plaintiffs have asserted that hair testing is not "sufficiently reliable to be job-related and justified by business necessity." But they have presented little if any evidence that could allow a jury to conclude that the drug test is so unreliable that its results have no significant correlation with drug use. Indeed, even their own evidence, if believed, would offer cause to question the accuracy of only some of the reported results, without indicating that there is a relatively large number of false positives compared to the size of the police force. On the other hand, the department, not the plaintiffs, carries the burden to prove that the program's results are significantly correlated with actual drug use.

In their reply brief, the plaintiffs argue that, for the purposes of proving the business necessity defense, the department must separately defend more than a dozen different versions of the test used over eight years. Such an argument flies in the face of the plaintiffs' position that multiple years of test results are properly aggregated as arising from a single practice in establishing a prima facie disparate impact claim. See Fudge v. City of Providence Fire Dep't, 766 F.2d 650, 657 (1st Cir. 1985). We reject the plaintiffs' about-face as both unpreserved and unfairly inconsistent with their assumption that they were all subjected to a single challenged practice.

Given that this case has already spanned many years (as did the post-termination administrative process), we are tempted to accept the department's invitation to assess whether genuine issues of material fact remain concerning its business necessity defense. In view of the size of the record, though, and the fact that the district court judge who has presided over this case has not yet parsed that record to assess business necessity or its rejoinder, we decline to do so in the first instance. Federal appellate courts have discretion in deciding whether to take up questions not considered below, but they generally should not do so. Singleton v. Wulff, 428 U.S. 106, 120 (1976). We see no reason to depart from that general practice here.<sup>19</sup>

With the business necessity question left open for further consideration, we have no occasion to consider whether the plaintiffs' evidence will prove sufficient to show that "the employer refuses to adopt an available alternative employment practice that has less disparate impact and serves the employer's legitimate needs." Ricci v. DeStefano, 557 U.S. 557, 578 (2009) (citing 42 U.S.C. §§ 2000e-2(k)(1)(A)(ii) and (C)). We reiterate, however, our statement above concerning the manner in which the

---

<sup>19</sup> In declining to decide the issues in the first instance, we do not suggest that the district court must reopen the record to allow further discovery or expert reports. The district court retains its customary discretion to manage the case, and we expect that it will give due weight to the fact that each party has already had ample time to put its best foot forward.

plaintiffs must prove that any such alternative practice would produce a smaller racial disparity in outcomes than does the department's current system.

#### **B. Due Process**

The plaintiffs contend that the drug testing program not only caused a disparate impact but also violated their rights under the Fourteenth Amendment of the United States Constitution. The Due Process Clause of the Fourteenth Amendment prohibits states from "depriv[ing] any person of life, liberty, or property, without due process of law." The plaintiffs contend that they were not afforded sufficient process in connection with their terminations (or other adverse action against them), and seek damages under 42 U.S.C. § 1983.

##### **1. Nine of the plaintiffs had a constitutionally-protected interest in their employment.**

When a public employee can be fired only for cause under state law, that employee has a property interest in continued employment. Gilbert v. Homar, 520 U.S. 924, 928-29 (1997). Here, the department concedes that the seven fired police officers had a property interest in their employment. The department also does not dispute that the fired cadet had a property interest in her employment. And the department does not dispute that Rachelle Couch, who signed the settlement agreement and continues to work for the department, was deprived of a property interest, presumably the approximately two months of pay she lost during her suspension.



Bristow received an unconditional offer of employment, and indeed begun work as an officer, her job would have been terminable with or without cause for six months. We have previously observed that probationary employees in Massachusetts do not have a property interest in their continued employment. See Brennan v. Hendrigan, 888 F.2d 189, 195 (1st Cir. 1989); see also Dasey v Anderson, 304 F.3d 148, 156-61 (1st Cir. 2002). Consequently, even had Bristow begun to work, she would have had no cognizable property interest in continued employment during the entirety of her probationary period. A fortiori, having not begun work, Bristow also had no cognizable property interest based on the job offer alone. We make no comment on whether the conditional offer of a job not subject to a probation period might, on other facts, be sufficient to grant a due process right to its recipient.

**2. The department provided sufficient process.**

The Supreme Court has held that "[t]he fundamental requirement of due process is the opportunity to be heard at a meaningful time and in a meaningful manner." Mathews v. Eldridge, 424 U.S. 319, 333 (1976) (internal quotation marks omitted). Although the precise contours of this guarantee vary depending on the circumstances, public employees are ordinarily entitled to notice of the reasons for a proposed termination, an explanation of the evidence supporting those reasons, and an opportunity to give their side of the story at a pre-termination hearing. See

Cleveland Bd. of Educ. v. Loudermill, 470 U.S. 532, 546 (1985); Calderón-Garnier v. Rodríguez, 578 F.3d 33, 38 (1st Cir. 2009). When a state employee in the ordinary course terminates another employee who has a property interest in his or her job, the state normally cannot satisfy due process solely through post-termination process. See Loudermill, 470 U.S. at 542; Cotnoir v. Univ. of Maine Sys., 35 F.3d 6, 12 (1st Cir. 1994) ("Where an employee is fired in violation of his due process rights, the availability of post-termination grievance procedures will not ordinarily cure the violation."). As the Supreme Court has explained, a pre-termination hearing should provide "a meaningful opportunity to invoke the discretion of the decisionmaker," both as to the facts supporting the termination and as to its broader appropriateness. Loudermill, 470 U.S. at 543.

Here, there is no dispute that employees subject to termination for a positive drug test were offered an opportunity to contest their impending firing at a pre-termination disciplinary hearing. Although the plaintiffs claim that some of those who were fired never had such a hearing, they concede that all were offered one. And, indeed, Massachusetts state law mandates that public employees, before being fired, receive notice of the reasons for their termination and the opportunity to protest their firing at an evidentiary hearing. Mass. Gen. Laws ch. 31, § 41. This statutory process allowed employees to present their side of the story,

through both testimony and evidence, and even to cross-examine adverse witnesses. The plaintiffs lodge only one complaint about the process, claiming that some of the fired employees were barred from presenting evidence of drug tests taken by the employees outside of the department's program. In one case, the department's chief hearing officer refused to accept results from an outside hair test in part because there was not sufficient information regarding the reliability of the test. She also refused in at least two cases to accept results from a urine test, calling urine tests and hair tests "apples and oranges," presumably because urine tests have a much shorter window of detection.

While we do not necessarily agree that the evidence was unworthy of consideration, we find no constitutional violation in its exclusion. Even in criminal trials, the state has some leeway in crafting and applying reasonable evidentiary rules. See United States v. Scheffer, 523 U.S. 303, 308 (1998); Montana v. Egelhoff, 518 U.S. 37, 43 (1996). The Supreme Court has ruled, for example, that a state did not violate the Due Process Clause where it barred a defendant accused of drug use from presenting a polygraph test indicating that he truthfully denied the charge. See Scheffer, 523 U.S. at 312. The department here was entitled to make a similar determination as to evidence that was possibly exculpatory but arguably unreliable or irrelevant.

Even if exclusion of the evidence rose to the level of a

constitutional violation, that error would have been adequately addressed through the extensive civil service appeals process, in which terminated employees were permitted to present evidence of outside drug tests. This process involved full hearings before the impartial Massachusetts Civil Service Commission. Although we have explained that some pre-termination process is required before the state fires an employee, an extensive post-termination appeal system lessens the need for an elaborate pre-termination process. See Mard v. Town of Amherst, 350 F.3d 184, 192 (1st Cir. 2003) (finding a "limited" pre-termination hearing constitutionally sufficient given "the availability of more rigorous post-deprivation procedures"). Together, the pre-termination disciplinary hearings and the post-termination appeals process easily satisfied the state's obligation to provide due process.<sup>21</sup>

Finally, the district court correctly granted summary judgment on the due process claim of plaintiff Rachelle Couch, the employee who chose an unpaid suspension rather than termination.

---

<sup>21</sup> The plaintiffs complain that the civil service appeals process took a long time, and indeed it did, as much as ten years for some. Yet, the Civil Service Commission explained in an opinion applying to many of the plaintiffs that "the unusual delay in bringing these appeals to hearing is due, in significant part, to the decisions of the [employees]," for example in asking for continuances. We cannot find a constitutional flaw in delay where the plaintiffs contributed significantly to that delay and do not attempt to demonstrate that any delay would have occurred absent their own decisions. The plaintiffs also say that two of the fired employees "were not a part of the Civil Service Commission proceedings," but because the plaintiffs do not provide any further explanation we must assume that this lack of participation was voluntary.



§ 12114(b)(3), (a). Importantly, though, the statute explicitly excludes from protection any individual who is currently using drugs, whether addicted or not, when the employer acts on the basis of such use. 42 U.S.C. § 12114(a). All of this means, in a nutshell, that in order to survive summary judgment dismissing their claim under the ADA, the plaintiffs must provide a factual basis upon which the jury could find that the department fired them either because they were addicts or because it perceived they were addicts, rather than because, as a result of the drug tests, it believed them to be currently using illegal drugs. See Raytheon Co. v. Hernandez, 540 U.S. 44, 52-53 (2003) (holding that, in a disparate treatment claim under the ADA such as this one, plaintiffs must show that their perceived disability "actually motivated the employer's decision").<sup>22</sup>

The plaintiffs made no such showing. To the contrary, the evidence is that the department trained its efforts at directly identifying users, whether addicts or not. And if the test results may have caused the department to form an erroneous view of any plaintiff, that view--to the extent it motivated termination--was that the plaintiff was a drug user, not that the plaintiff was an addict. Nor did the department accept any defense that an officer

---

<sup>22</sup> Although the distinction may be subtle, the plaintiffs do not allege a disparate impact claim under the ADA, even if one might have been available. See Raytheon, 540 U.S. at 52 (making clear that disparate impact claims can be pursued under the ADA).

used illegal drugs only once, or was otherwise not addicted. Indeed, the very existence of the rehabilitation program shows that the department was willing to retain employees who were addicted as long as they are not users. On this record, no jury could reasonably conclude that the department was motivated by a perception that plaintiffs were addicted to drugs. See Lopez v. Pac. Mar. Ass'n, 657 F.3d 762, 764 (9th Cir. 2011) (affirming grant of summary judgment to an employer on an ADA claim because "the triggering event for purposes of [rejection of a job application] is a failed drug test, not an applicant's drug addiction"); Salley v. Circuit City Stores, Inc., 160 F.3d 977, 981 (3d Cir. 1998) (holding that no jury could find that the plaintiff was fired for his drug addiction rather than misconduct, whether or not the misconduct was caused by addiction).

#### **D. Failure to Train and Supervise**

The plaintiffs also claim that the department violated their constitutional rights due to its failure to sufficiently train and supervise the staff carrying out the program. The failure to train theory offers a way for plaintiffs to hold a municipality liable under 42 U.S.C. § 1983 for the acts of its employees. See Hayden v. Grayson, 134 F.3d 449, 456 (1st Cir. 1998). Typically, a plaintiff first identifies a constitutional violation (often related to police misconduct) and then attempts to show that the violation was the direct result of poor training or

supervision of municipal employees, stemming from "deliberate indifference to the rights of persons with whom the [employees] come into contact." Id. To prevail, the plaintiff must show that the constitutional violation had a "direct causal link" to the deficiency in training. Canton v. Harris, 489 U.S. 378, 385 (1989).

Here, the plaintiffs focus extensively on evidence of poor training but fail to produce any evidence that would allow a jury to reasonably conclude that this poor training caused a constitutional violation. Indeed, the plaintiffs fail to clearly identify in their brief any constitutional provision allegedly violated as a result of poor training. At best, the plaintiffs simply postulate without analysis violations of the Due Process Clause and Equal Protection Clause.

As to the first, we have explained above that the plaintiffs were not deprived of liberty or property without due process of law. Consequently, there can be no liability for such a violation, whether on a failure to train theory or otherwise.

As to the second, and even assuming that plaintiffs' evidence of disparate impact under Title VII would have sufficed to survive summary judgment on an equal protection claim,<sup>23</sup> the plaintiffs fail to point to any evidence indicating that training

---

<sup>23</sup> The plaintiffs raised such a claim below but declined to pursue it on appeal.

so poor as to constitute actionable indifference caused them to test positive. We therefore affirm the grant of summary judgment to the department on the failure to train claim.

#### **IV. Conclusion**

The plaintiffs have proven beyond reasonable dispute a prima facie case of disparate impact under Title VII, while the department has proffered an uncontested legitimate need to identify those few of its members who use illegal drugs. What remains to be assessed by the district court is whether the department's drug testing program advances that goal and, if so, whether the plaintiffs can carry their burden of proving a failure to adopt an available alternative that meets the department's legitimate needs while reducing the disparate impact on black employees of the department.

For these reasons, we vacate the district court's grant of summary judgment on the Title VII claims; we reverse the district court's denial of partial summary judgment for all plaintiffs on the question of whether they have proved a prima facie case of disparate impact under Title VII; we otherwise affirm the district court opinion; and we remand for further proceedings consistent with this opinion. The district court will decide at the time of final judgment whether costs of this appeal are to be

shifted in favor of a finally prevailing party under any applicable statute.

So ordered.