10-1-2015

The First Circuit Strikes Out in Jones v. City of Boston: A Pitch for Practical Significance in Disparate Impact Cases

Michael Stenger

THE FIRST CIRCUIT STRIKES OUT IN *JONES v. CITY OF BOSTON*:
A PITCH FOR PRACTICAL SIGNIFICANCE
IN DISPARATE IMPACT CASES

Michael Stenger*

"Do not put your faith in what statistics say until you have
carefully considered what they do not say."[1]

I. Introduction: Courts Are "Striking Out" in Determining Significance

Despite their best efforts to comply with the law, modern employers are likely to face allegations of unlawful discrimination.[2] While most employers are aware that purposeful, overt discrimination is illegal, many fail to realize the extent of liability potentially resulting from practices or policies that have the *effect* of discriminating against a protected class.[3] Employers increasingly utilize statistics in hiring practices, collecting and analyzing data about applicants and employees to determine who makes the best employee.[4] Employing "Moneyball"-type statistical analyses may be useful to employers in amassing a more productive workforce, but it

1. *Quotes About Statistics*, My Statistics: Making Sense of Data, http://www.mystatistics.info/page0/page0.html (last visited Dec. 31, 2014) (quoting William W. Watt).

2. *See* Arlene Switzer Steinfield & Andrew D. Peters, *Dealing with Employees Who Are Suing Your Company*, Corp. Counsel (Aug. 13, 2014), http://www.corpcounsel.com/id=1202666703184? (noting many companies will face discrimination or harassment suits at one time or another); *see also* U.S. EEOC, *Title VII of the Civil Rights Act of 1964 Charges*, http://www.eeoc.gov/eeoc/statistics/enforcement/titlevii.cfm (last visited Dec. 31, 2014) (providing statistics on number of employment discrimination charges filed with EEOC). In 2013, the EEOC received 67,558 complaints of illegal discrimination under Title VII, of which 12,005 led to some sort of financial liability to employers. *See id.* Plaintiffs received $255.9 million in total damages, or about $21,316 per successful plaintiff. *See id.*

3. *See* 42 U.S.C. § 2000e-2(k) (2012) (codifying disparate impact theory of discrimination).

4. *See* Jamie Konn, *Analytics and . . . Employment Law?*, Labor Dish (Sept. 15, 2014), http://www.labordish.com/2014/09/15/analytics-and-employment-law/ (describing how commercial employers have begun utilizing statistics to help make hiring decisions).

also increases employers' risk of exposure to disparate impact claims under 42 U.S.C § 1981.[5]

Since its inception in 1971, the disparate impact theory of employment discrimination has been arguably the most controversial and most inconsistently applied framework under which employers are held liable for discrimination.[6] Title VII of the Civil Rights Act of 1964 prohibits employers from discriminating on the basis of "race, color, religion, sex, or national origin . . . ."[7] Originally enacted to prevent outright acts of discrimination, Title VII was expanded to forbid employers from enacting facially neutral employment practices that disproportionately impact a protected group.[8] Much of the inconsistency in the disparate impact con-

---

5. *See id.* ("There is a greater risk of a disparate impact claim when analytics are used in the hiring stage . . . ."). *See generally* MICHAEL LEWIS, MONEYBALL: THE ART OF WINNING AN UNFAIR GAME (2004) (telling story of Oakland Athletics General Manager Billy Beane and his revolutionary use of statistics in Major League Baseball). Employers have turned to statistics as a predictor of performance, much in the same way that Billy Beane and the Oakland Athletics have utilized "sabermetrics" in evaluating baseball prospects. *See id.* "Moneyball" refers to the assessment of baseball talent from a statistical, quantitative standpoint, rather than the traditional, subjective measures used by Major League franchises for much of the sport's history. *See id.*

6. *See* Michael K. Grimaldi, *Disparate Impact After* Ricci *and* Lewis, 14 SCHOLAR 165, 166 (2011) ("The disparate-impact theory of employment discrimination has remained controversial since its inception."); Herman N. Johnson, Jr., *Disambiguating the Disparate Impact Claim*, 22 TEMP. POL. & CIV. RTS. L. REV. 433, 434 (2013) (hinting disparate impact may receive "ominous reception" in future Supreme Court cases); Joseph A. Seiner, *Plausibility and Disparate Impact*, 64 HASTINGS L.J. 287, 295 (2013) ("[D]isparate impact has been marked with controversy since its inception."); Michael Selmi, *Was the Disparate Impact Theory a Mistake?*, 53 UCLA L. REV. 701, 702 (2006) (arguing that disparate impact theory has attracted most attention and controversy within antidiscrimination law); *see also* Ricci v. DeStefano, 557 U.S. 557, 594 (2009) (Scalia, J., concurring) (suggesting that disparate impact theory may be unconstitutional under Equal Protection Clause). Justice Scalia described the difficulty employers have in enacting policies that satisfy both the disparate treatment and disparate impact provisions of Title VII:

> The difficulty is this: Whether or not Title VII's disparate-treatment provisions forbid "remedial" race-based actions when a disparate-impact violation would *not* otherwise result . . . it is clear that Title VII not only permits but affirmatively *requires* such actions when a disparate-impact violation *would* otherwise result. . . . Title VII's disparate-impact provisions place a racial thumb on the scales, often requiring employers to . . . make decisions based on (because of) [ ] racial outcomes.

Id.

7. 42 U.S.C. § 2000e-2(a)(1) ("It shall be an unlawful employment practice for an employer—to fail or refuse to hire or to discharge any individual . . . because of such individual's race, color, religion, sex, or national origin . . . .").

8. *See* Griggs v. Duke Power Co., 401 U.S. 424, 431 (1971) (interpreting Title VII to include disparate impact cause of action). The Supreme Court held that Congress's intent in enacting Title VII was "the removal of artificial, arbitrary, and unnecessary barriers to employment when the barriers operate invidiously to discriminate on the basis of racial or other impermissible classification." *Id.* As such, the Court reasoned that Congress had forbidden "giving [testing or measuring procedures] controlling force unless they are demonstrably a reasonable measure of job performance." *Id.* at 436.

text stems from the difficulty courts experience in determining whether plaintiffs have demonstrated a prima facie case.[9] In order to demonstrate a prima facie case of disparate impact discrimination, a plaintiff must show that an employment practice or policy (1) is used by an employer and (2) causes an adverse impact on a protected class.[10]

In order to prove the causation element, plaintiffs typically provide statistical analyses that compare the selection rates of majority and minority employees to demonstrate that the disparity adversely impacts the plaintiffs' class.[11] Judges must determine whether to assign weight to such statistical analyses, but they have done so in an arbitrary and unpredictable manner.[12] Further, courts have wrestled with when to properly label a given sample as statistically significant, or when to take other factors into consideration.[13]

Notably, there exists one "other" factor that some courts have incorporated into their disparate impact analyses: demonstration of practical significance.[14] Nevertheless, many courts often choose to ignore the practical implications of a statistically significant sample, despite evidence that statistical significance is not, in all circumstances, an accurate measure of adverse impact.[15] Although the Supreme Court has not yet addressed the practical significance requirement directly, several lower courts have taken

9. For discussion of a prima facie case and courts' determination of the causation element, see *infra* notes 27–88 and accompanying text.

10. *See* 42 U.S.C. § 2000e-2(k)(1)(A) (codifying requirements for plaintiff's prima facie case).

11. *See* D.H. Kaye, *Is Proof of Statistical Significance Relevant?*, 61 WASH. L. REV. 1333, 1335 (1986) (describing how courts evaluating Title VII cases have come to expect standard deviation analysis); *see also* Moultrie v. Martin, 690 F.2d 1078, 1082 (4th Cir. 1982) (holding that all cases involving racial discrimination require application of standard deviation analysis). While not all circuits have demanded a showing of statistical analysis, it is incredibly rare for plaintiffs not to offer some sort of statistical evidence to support the causation prong of their case. *See* Watson v. Fort Worth Bank & Trust, 487 U.S. 977, 994 (1988) (plurality opinion) (suggesting that plaintiffs should establish causation by offering sufficient statistical evidence).

12. For a discussion of the inconsistent weight given to statistical evidence in the lower courts, see *infra* notes 53–55 and accompanying text.

13. *See* Kaye, *supra* note 11, at 1339–40 (describing evidentiary issue of allowing uncontroverted testimony of statistical experts to determine existence of prima facie case at trial); Allan G. King, *"Gross Statistical Disparities" as Evidence of a Pattern and Practice of Discrimination: Statistical Versus Legal Significance*, 22 LAB. LAW. 271, 272 (2007) ("[L]ower courts frequently have turned to 'statistical significance' as the measuring rod."); *see also* Tracy Bateman Farrell et al., *Four-Fifths or 80% Rule*, 45A AM. JUR. 2D JOB DISCRIMINATION § 313 (describing four-fifths rule as useful benchmark in analyzing Title VII disparate impact testing cases).

14. *See, e.g.*, Waisome v. Port Auth. of N.Y. & N.J., 948 F.2d 1370, 1376–77 (2d Cir. 1991) (citing Bilingual Bicultural Coal. on Mass Media, Inc. v. FCC., 595 F.2d 621, 642 n.57 (D.C. Cir. 1978)) (holding that statistical significance is not indicative of "importance, magnitude, or *practical significance*" of presented disparity (emphasis added)).

15. *See* Kaye, *supra* note 11, at 1365 (calling into question ability of statistical significance testing to satisfy legal standards of proof).

the standard into consideration and are divided as to the necessity of a practical significance standard.[16]

Jones v. City of Boston,[17] a recent First Circuit case, illustrates the difficulties that courts experience when it comes to determining the significance of statistical analyses in the face of practical objections.[18] The First Circuit rejected the need for plaintiffs to demonstrate the practical significance of a statistically significant sample, despite strong evidence to suggest that such a showing was necessary in the case.[19]

This Note argues that the First Circuit was incorrect in rejecting the practical significance standard and proposes a workable framework for courts that incorporates both statistical and practical considerations.[20] Part II provides an overview of disparate impact and examines the analytical progression courts undergo in a disparate impact claim.[21] Part III traces the facts of *Jones* and explains the court's reasoning in rejecting practical significance as a measure of disparate impact.[22] Part IV presents arguments against the court's holding in *Jones*, while Part V proposes a uniform two-part test for incorporating significance at the prima facie stage.[23] Part VI concludes by discussing the impact of *Jones* within the First Circuit and beyond.[24]

II. Stepping Up to the Plate: The History of Disparate Impact

The disparate impact theory—though one of the most significant developments in employment discrimination jurisprudence—has evolved from a controversial history.[25] Courts have addressed the theory by em-

---

16. For a discussion of the circuit split over the practical significance standard, see *infra* notes 73–88 and accompanying text.

17. 752 F.3d 38 (1st Cir. 2014).

18. *See, e.g.*, *id.*

19. *See id.* at 53 (holding that demonstration of practical significance is not required by Title VII to prove prima facie case of disparate impact).

20. For a critical analysis of *Jones*, see *infra* notes 127–43 and accompanying text.

21. For a discussion of the history of disparate impact and practical significance, see *infra* notes 25–88 and accompanying text.

22. For a discussion of the facts of *Jones* and the court's reasoning, see *infra* notes 89–126 and accompanying text.

23. For a critical analysis of the First Circuit's holding in *Jones*, see *infra* notes 127–43 and accompanying text. For a discussion of the proposed two-part significance test, see *infra* notes 144–76 and accompanying text.

24. For a discussion of the impact of *Jones*, see *infra* notes 177–83 and accompanying text.

25. *See* Selmi, *supra* note 6, at 702–05 (describing disparate impact as controversial from its inception and having made no noticeable societal change). For a discussion of the history of disparate impact law in the Supreme Court, see *infra* notes 27–38 and accompanying text.

ploying different standards to determine the existence of disparate impact.[26]

A. *Leveling the Playing Field:* Griggs *and the History of the Disparate Impact Doctrine*

Title VII prohibits discrimination in the workplace on the basis of race, color, religion, sex, or national origin.[27] During the seven years after the enactment of Title VII, courts interpreted this prohibition to include only intentional discrimination, such as refusing to hire or firing a person *because of* their membership in a protected class.[28]

In 1971, in a landmark case, the Supreme Court expanded the meaning of Title VII to include a more subtle form of discrimination: the use of facially-neutral practices or policies that adversely impact a protected class.[29] In *Griggs v. Duke Power Co.*,[30] a group of black plaintiffs challenged Duke Power's policy that all employees must have a high school diploma and score satisfactorily on a general aptitude test, which served to preclude many black applicants from gaining employment or promotion within the company.[31] The Court held that requirements such as those implemented by Duke Power violated Title VII because they were not related to successful job performance, and they disqualified black applicants at a much higher rate than white applicants.[32] After *Griggs,*

---

26. For a discussion of the different approaches that lower courts have taken in disparate impact cases, see *infra* notes 39–88 and accompanying text.

27. *See* 42 U.S.C. § 2000e (2012).

28. *See* Gary D. Sanders, Note, *Disparate Impact Analysis May Be Applied to Prove Racial Discrimination in Subjective Employment Decisions:* Watson v. Fort Worth Bank & Trust, 20 TEX. TECH L. REV. 1381, 1383–84 (1989) (describing disparate treatment theory of discrimination). The Supreme Court noted that this type of intentional disparate treatment is "the most easily understood type of discrimination. The employer simply treats some people less favorably than others because of their race, color, religion, sex, or national origin." Int'l Bhd. of Teamsters v. United States, 431 U.S. 324, 335 n.15 (1977).

29. *See generally* Griggs v. Duke Power Co., 401 U.S. 424 (1971).

30. 401 U.S. 424 (1971).

31. *See id.* at 427–28 (providing facts of *Griggs*). Although the plaintiffs could not prove that Duke Power adopted the requirements malevolently, they were able to demonstrate that significant disparities existed between whites and blacks as to both requirements. *See id.* at 429 n.6 (describing disparities between races in passage rates on aptitude tests and in high school graduation rates). In 1960, only 12% of black males in North Carolina (Duke Power's place of business) had a high school diploma, compared to 34% of white males. *See id.* Similarly, black test-takers passed the standard aptitude tests at a much lower rate (6%) than did their white counterparts (58%). *See id.*

32. *See id.* at 429 (noting objective of Title VII to achieve equality of employment opportunities). The Court held that "practices, procedures, or tests neutral on their face, and even neutral in terms of intent, cannot be maintained if they operate to 'freeze' the status quo of prior discriminatory employment practices." *Id.* at 430. In enacting Title VII, Congress intended to remove "artificial, arbitrary, and unnecessary barriers to employment when the barriers operate invidiously to discriminate on the basis of . . . impermissible classification[s]." *Id.* at 431.

the Court continued to expand and strengthen the disparate impact theory.[33]

Nearly two decades after *Griggs*, the Court, perhaps overzealously, attempted to significantly limit its expanded interpretation in *Wards Cove Packing Co. v. Atonio*.[34] In *Wards Cove*, non-white plaintiffs alleged that their employer's hiring practices caused a racial divide between white and non-white employees.[35] In rejecting the plaintiff's disparate impact claim, the Court held that employers need only provide a business justification for apparently discriminatory hiring practices.[36] Subsequently, in passing the Civil Rights Act of 1991, Congress rejected the Court's reasoning and statutorily superseded the Court's *Wards Cove* holding.[37] Since the passage

33. *See, e.g.*, Albemarle Paper Co. v. Moody, 422 U.S. 405, 425–28 (1975) (establishing burden-shifting framework in disparate impact claims). In 1975, the Court fleshed out the prima facie case requirements and burden-shifting procedure in *Albemarle Paper*. *See id.* The Court held that after a plaintiff makes out a prima facie case, the burden of proof shifts to the employer to show that the business practice was required by a business necessity. *See id.* at 425. If the employer can demonstrate business necessity or "job relatedness," the burden shifts back to the plaintiff to prove the existence of a less-discriminatory alternative practice that the defendant could have adopted. *See id.* at 426. In *Albemarle Paper*, the Court held that "it remains open to the complaining party to show that other tests or selection devices, without a similarly undesirable racial effect, would also serve the employer's legitimate interest in 'efficient and trustworthy workmanship.'" *Id.* at 425 (quoting McDonnell Douglas Corp. v. Green, 411 U.S. 792, 801 (1973)).

34. 490 U.S. 642, 658–59 (1989) (holding that "business justification" was appropriate defense to alleged discriminatory practice and that burden of proof in rebutting employer's justification falls on plaintiffs). At the time of the decision, but before the Civil Rights Act of 1991 was passed, the *Wards Cove* holding was the subject of much academic attention and was generally regarded as a misstep by the Supreme Court in its disparate impact jurisprudence. *See, e.g.*, Michael K. Braswell, Gary A. Moore & Bill Shaw, *Disparate Impact Theory in the Aftermath of* Wards Cove Packing Co. v. Atonio*: Burdens of Proof, Statistical Evidence, and Affirmative Action*, 54 ALB. L. REV. 1, 1–2 (1989) ("The 1989 Supreme Court decision in [*Wards Cove*] is a major 'stride backwards' concerning the use of disparate impact theory . . . ." (footnotes omitted)); *see also* Grimaldi, *supra* note 6, at 172 (describing disparate impact theory as having "near-death experience" in *Wards Cove*).

35. *See Wards Cove*, 490 U.S. at 646–48 (providing factual background). The employer, an Alaskan salmon cannery, had two general types of jobs: unskilled cannery positions and non-cannery managerial positions. *See id.* at 647. Native Alaskans and Filipinos comprised a vast majority of the cannery positions, while their white coworkers held virtually all of the managerial positions. *See id.* The plaintiffs alleged that the disparity between white and non-white employees was caused by the company subjectively declining to offer non-whites the higher-paying non-cannery jobs. *See id.* at 647–48.

36. *See id.* at 658–59 (detailing business justification defense).

37. *See* Civil Rights Act of 1991, Pub. L. No. 102-166, 105 Stat. 1071 (codified in scattered sections of 42 U.S.C.) (rejecting reasoning of *Wards Cove* and adopting reasoning of *Griggs* and *Albemarle*). The amended statute read as follows:

(a) Section 703 of the Civil Rights Act of 1964 (42 U.S.C. 2000e-2) is amended by adding at the end of the following new subsection:

(k)(1)(A) An unlawful employment practice based on disparate impact is established under this title only if—

(i) a complaining party demonstrates that a respondent uses a particular employment practice that causes a disparate impact on

of the 1991 amendments, the Court has expanded the disparate impact doctrine to include subjective hiring or promotional practices and has assigned liability where one of an employer's many individual hiring practices has a disparate impact, despite the overall hiring process having no such impact.[38]

B. *Here Comes the Curveball: Lower Courts' Application of the Doctrine*

Without definitive guidance from the Supreme Court on the issue, lower courts have struggled to come up with a workable solution for the most important stage of analysis in disparate impact cases: a plaintiff's demonstration that there exists a causal connection between the selected employment practice and the alleged discriminatory impact.[39] Generally, plaintiffs and defendants turn to three arguments to establish or disprove causation: statistical significance, the four-fifths rule, and practical significance.[40]

1. *Statistical Significance*

The most common, persuasive, and widely-accepted method that plaintiffs offer to demonstrate causation is the statistical significance analy-

---

> the basis of race, color, religion, sex, or national origin and the respondent fails to demonstrate that the challenged practice is job related for the position in question and consistent with business necessity . . . .

Id.

38. *See, e.g.*, Watson v. Fort Worth Bank & Trust, 487 U.S. 977 (1988) (plurality opinion) (holding that subjective or discretionary employment policies may be challenged under disparate impact theory); *see also* Connecticut v. Teal, 457 U.S. 440 (1982). In *Teal*, a Connecticut state agency used a written examination as the first step in its promotion test, which black candidates passed at 68% the rate of white candidates. *See id.* at 443. Although the agency promoted black employees at a rate 170% higher than that of white employees, the Court still found that disparate impact existed in the agency's testing procedures because the test eliminated an impermissibly high percentage of black applicants. *See id.* at 456. The Court thus rejected the "bottom line" defense to a prima facie showing of disparate impact discrimination because Title VII guarantees equality in hiring and promotion policies and practices. *See id.* Despite the Court's expansion of the doctrine in these cases, scholars have suggested that the Court has begun to chip away at the disparate impact doctrine with its holding in *Ricci v. DeStefano*, 557 U.S. 557 (2009). *See, e.g.*, Olatunde C.A. Johnson, *The Agency Roots of Disparate Impact*, 49 HARV. C.R.-C.L. L. REV. 125, 128–33 (2014) (describing Supreme Court's view of disparate impact as skeptical in recent rulings).

39. *See* Maxwell O. Chibundu, *Delinking Disproportionality from Discrimination: Procedural Burdens as Proxy for Substantive Visions*, 23 N.M. L. REV. 87, 132 (1993) (describing how lower courts have struggled to distinguish disparate impact and disparate treatment theories of discrimination and thus have confused causation prong of both tests).

40. *See* Jennifer L. Peresie, *Toward a Coherent Test for Disparate Impact Discrimination*, 84 IND. L.J. 773, 782–88 (2009) (describing statistical significance and four-fifths tests); *see also* Waisome v. Port Auth. of N.Y. & N.J., 948 F.2d 1370 (2d Cir. 1991) (embracing defendant's argument that practical significance is appropriate measure of disparate impact).

sis, also known as a standard deviation analysis.[41] Although statistical analyses may vary in technical form, they all measure the probability that an observed disparity is the product of chance; in other words, a standard deviation analysis calculates how much a given sample differs from a different hypothetical sample which assumes equal opportunity.[42] A difference of roughly two standard deviations equates to a 5% chance that the measured disparity was a result of chance.[43]

The Supreme Court first turned to statistics as a measure of causation in the employment discrimination context in *International Brotherhood of Teamsters v. United States*.[44] In *Teamsters*, the plaintiffs alleged that their employer had discriminated against African-American and Hispanic employees and offered a statistical analysis of the disparity in hiring and promotion rates between minority and white employees.[45] The Court gave the plaintiffs' statistical evidence substantial weight, noting that statistics can be especially probative in revealing imbalances in the racial composition of a work force.[46] The Court confirmed that statistical analyses were

---

41. *See* Michael D. Lieder, Hot Issues Concerning Use of Statistical Evidence in Employment Discrimination Class Action Litigation, Sprenger Lang PLLC (2009), *available at* http://apps.americanbar.org/labor/lel-annualcle/09/materials/data/papers/103.pdf (detailing courts' evaluation of causation in terms of standard deviations).

42. *See* Peresie, *supra* note 40, at 785–87 (describing function of various significance tests); *see also* Michael J. Zimmer et al., Cases and Materials on Employment Discrimination 133 (Vicki Been et al. eds., 8th ed. 2012) (explaining probability testing generally). "[P]robability theory starts with a comparison between the 'observed' racial (or gender or age) distribution in the employer's work force and the 'expected,' that is, the racial distribution one would anticipate if race were not a factor in the selection of employees." *Id.*

43. *See* King, *supra* note 13, at 292 (citing Griffin v. Bd. of Regents of Regency Univ., 795 F.2d 1281, 1291 n.19 (7th Cir. 1986)) (explaining that 1.96 standard deviations corresponds to .05 level of statistical significance).

44. 431 U.S. 324, 339 (1977) (holding that statistical analysis was probative in demonstrating discrimination).

45. *See id.* at 337–39 (describing plaintiffs' proffered statistical analysis). The plaintiffs presented evidence that African-Americans and Hispanics made up 5% and 4% of the company's 6,472 employees respectively, and that among the sought-after position of "line driver," only 0.4% were African-American and 0.3% were Hispanic. *See id.* at 337–38.

46. *See id.* at 339 n.20 (describing utility of statistical analyses). The Court noted that statistics can often uncover covert or even unintentional discriminatory employment practices:

Statistics showing racial or ethnic imbalance are probative in a case such as this one only because such imbalance is often a telltale sign of purposeful discrimination; absent explanation, it is ordinarily to be expected that nondiscriminatory hiring practices will in time result in a work force more or less representative of the racial and ethnic composition of the population in the community from which employees are hired. . . . "Since the passage of the Civil Rights Act of 1964, the courts have frequently relied upon statistical evidence to prove a violation. . . . In many cases the only available avenue of proof is the use of racial statistics to uncover clandestine and covert discrimination by the employer or union involved."

the approved means of establishing causation in employment discrimina-
tion cases in *Hazelwood School District v. United States.*[47] Relying on a jury
selection discrimination case it had decided earlier that year, the Court
held that statistical analyses revealing a disparity of "two or three standard
deviations" reflected a "gross" statistical difference, such that illegal dis-
crimination could be inferred.[48]

The issue of statistical analyses being offered to prove causation arose
again in *Watson v. Fort Worth Bank & Trust.*[49] In a plurality opinion, the
Court noted that lower courts had not been instructed to apply the *Hazel-
wood* "two-to-three standard deviations" analysis in a mechanistic fashion.[50]
Instead, the Court advised that lower courts should analyze statistical evi-
dence on a case-by-case basis in light of the surrounding facts and circum-
stances.[51] The Court has also noted recently that statistical significance
may be enough to demonstrate a prima facie case.[52]

Although not explicitly instructed to do so, lower courts have adopted
the *Hazelwood* threshold in weighing the significance of a standard devia-
tion analysis.[53] Several circuits have even gone so far as to hold that any
sample not rising to a disparity of at least two standard deviations is inad-
missible as proof of causation.[54] Other courts prefer statistical signifi-

---

Id. (second alteration in original) (quoting United States v. Ironworkers Local 86,
443 F.2d 544, 551 (9th Cir. 1971)).

47. 433 U.S. 299 (1977).

48. *Id.* at 308 n.14; *see also id.* at 307 (citing *Teamsters,* 431 U.S. at 339) (hold-
ing that "gross" statistical disparities alone may constitute prima facie case in cer-
tain circumstances). In *Hazelwood,* the plaintiff teachers brought suit against their
school district alleging discrimination in the district's employment patterns and
practices. *See id.* at 301. The Court held that the proper groups to compare were
not the percentage of black teachers in the school district and the percentage of
black students in the district, as the district court had held, but rather the proper
comparison was the racial composition of the district's work force and the racial
composition of the "qualified public school teacher population in the relevant la-
bor market." *Id.* at 308.

49. 487 U.S. 977 (1988) (plurality opinion).

50. *See id.* at 995 n.3 (noting that Court had not suggested that set number of
standard deviations can determine prima facie case). The Court noted that it had
"emphasized the useful role that statistical methods can have in Title VII cases, but
[had] not suggested that any particular number of 'standard deviations' can deter-
mine whether a plaintiff has made out a prima facie case in the complex area of
employment discrimination." *Id.* (citing *Hazelwood,* 433 U.S. at 311 n.17).

51. *See id.* (holding that case-by-case approach is appropriate manner of ana-
lyzing standard deviations).

52. *See* Ricci v. DeStefano, 557 U.S. 557, 587 (2009) ("[A] prima facie case of
disparate-impact liability [is] essentially, a threshold showing of a significant statis-
tical disparity . . . ." (citing Connecticut v. Teal, 457 U.S. 440, 446 (1982))).

53. *See* King, *supra* note 13, at 276 ("[M]any lower courts have adopted the
Castaneda-Hazelwood criterion of 'two or three standard deviations' as a bright-line
rule.").

54. *See, e.g.,* Mems v. City of St. Paul, Dep't of Fire & Safety Servs., 224 F.3d
735, 741 (8th Cir. 2000) (rejecting small sample size due to lack of statistical signif-
icance); Bennett v. Total Minatome Corp., 138 F.3d 1053, 1062 (5th Cir. 1998)
(holding that plaintiff failed to raise inference of discrimination after failing to

cance tests as evidence of disparate impact because they quantify the certainty with which it can be inferred that the impact is not a product of chance.[55]

Despite its mathematical nature, a standard deviation analysis is not without its weaknesses.[56] The most notable weakness is its sensitivity to sample size: the larger the sample size, the more magnified differences appear in relation to standard deviation.[57] Put simply, larger sample sizes are more conducive to a finding of statistical significance, even when actual differences are in fact small.[58] In order to achieve a larger sample size—and hence a finding of statistical significance—plaintiffs typically aggregate data from numerous years in order to make their proffered sample statistically significant.[59] Courts are generally willing to accept aggregation as a tool to achieve a sample size large enough to produce statistically significant results.[60] However, some courts have also ques-

demonstrate statistical significance); Palmer v. Shultz, 815 F.2d 84, 95 (D.C. Cir. 1987) (rejecting plaintiff's statistical analysis for lack of significance).

55. *See, e.g.*, Carpenter v. Boeing Co., 456 F.3d 1183, 1202 (10th Cir. 2006) ("[A statistically significant] difference strongly indicates *some* influence on the results other than the operation of pure chance."); Moultrie v. Martin, 690 F.2d 1078, 1082 (4th Cir. 1982) (holding that disparate impact cases require standard deviation analysis).

56. For a discussion of the weaknesses of standard deviation analyses, see *infra* notes 57–61 and accompanying text.

57. *See* Peresie, *supra* note 40, at 787 ("Statistical significance is also tremendously sensitive to sample size. . . . [T]he larger the number of applicants, the smaller the magnitude of difference that will be statistically significant . . . ."); Daniel L. Rubinfeld, *Econometrics in the Courtroom*, 85 Colum. L. Rev. 1048, 1067 (1985) (noting that large samples increase likelihood of finding statistical significance); Tom Tinkham, *The Uses and Misuses of Statistical Proof in Age Discrimination Claims*, 27 Hofstra Lab. & Emp. L.J. 357, 376 (2010) (noting that large sample sizes can disguise lack of practical significance).

58. *See* Rubinfeld, *supra* note 57, at 1067 ("[*T*]-statistics are quite sensitive to the size of the sample being studied. With a sufficiently large sample, the likelihood of getting a *t*-value greater than two can get large.").

59. *See* Richard E. Biddle, *Disparate Impact Analysis with Small Samples*, Cal. Lab. & Emp. L. Q. (Fall 1995), *available at* http://www.biddle.com/documents/disparatesmall.htm (providing various aggregation methods plaintiffs use to increase sample size). The EEOC Guidelines accept several scenarios in which aggregation may be appropriate, including: aggregating data from the same test given for several jobs within the same company; aggregating data from similar tests given for the same job over many years; aggregating several years of data for one job; and aggregating data for more than one minority group. *Id.*

60. *See, e.g.*, Hazelwood Sch. Dist. v. United States, 433 U.S. 299, 313 (1977) (permitting aggregation because data for applicants from separate occupation census codes were similar enough for analytical purposes); Paige v. California, 291 F.3d 1141, 1148 (9th Cir. 2002) ("[I]t is a generally accepted principle that aggregated statistical data may be used where it is more probative than subdivided data."); Eldredge v. Carpenters 46 N. Cal. Cntys., 833 F.2d 1334, 1340 n.8 (9th Cir. 1987) (allowing analysis of aggregated nine-year sample because it was most probative in light of surrounding circumstances); *see also* 29 C.F.R. § 1607.4 (1981). The Uniform Guidelines on Employee Selection Procedures also support aggregation in certain circumstances: "Where [ ] evidence concerning the impact of a selection procedure indicates adverse impact but is based upon numbers which are too

tioned the viability of aggregation by scrutinizing its probative value in light of other considerations.[61]

2. *The Four-Fifths Rule*

As an alternative to a showing of statistical significance, plaintiffs and defendants alike frequently turn to the Equal Employment Opportunity Commission's (EEOC) "four-fifths" rule to prove or disprove the existence of disparate impact.[62] The rule dictates that a selection rate (for promotions, hires, or related employment decisions) of a protected class that is less than four-fifths (or eighty percent) of the selection rate for the group with the highest selection rate will be considered actionable by the EEOC.[63] Though the Supreme Court has labeled the four-fifths rule as nothing more than a rule of thumb for the courts, it has also noted that

___

small to be reliable, evidence concerning the impact of the procedure over a longer period of time . . . may be considered in determining adverse impact." *Id.*

61. *See, e.g.*, Apsley v. Boeing Co., 722 F. Supp. 2d 1218, 1238 (D. Kan. 2010) (rejecting plaintiffs' aggregation of data to make prima facie case of age discrimination); United States v. City of Yonkers, 609 F. Supp. 1281, 1288 (S.D.N.Y. 1984) ("A need to aggregate [data] to achieve statistical significance may lessen the probative force of the data . . . ."). In cases of employers with a large workforce, a standard deviation analysis based on aggregation of data from different departments within the company or from a number of years may not always be appropriate. *See* Peresie, *supra* note 40, at 787 ("[T]he statistical significance rule could be said to have a disparate impact on large employers because even a small disparity may achieve statistical significance.").

62. *See* 29 C.F.R. § 1607.4(D). The relevant portion of the Code of Federal Regulations notes:

> A selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact, while a greater than four-fifths rate will generally not be regarded by Federal enforcement agencies as evidence of adverse impact. Smaller differences in selection rate may nevertheless constitute adverse impact, where they are significant in both statistical and practical terms or where a user's actions have discouraged applicants disproportionately on grounds of race, sex, or ethnic group.

Id.

63. *See id.* For example, a promotion test that 50 out of 100 men (50%) pass, but that 80 out of 100 women (80%) pass, would be considered evidence of discrimination because 50% is less than four-fifths of 80% (50% < 64%). *See* Adoption of Questions and Answers to Clarify and Provide a Common Interpretation of the Uniform Guidelines on Employee Selection Procedures, 44 Fed. Reg. 11996, 11998 (Mar. 2, 1979) (providing formula for calculating four-fifths rule). To determine whether the four-fifths rule has been satisfied, one must use the following four-step process:

> (1) [C]alculate the rate of selection for each group (divide the number of persons selected from a group by the number of applicants from that group);
> (2) observe which group has the highest selection rate;
> (3) calculate the impact ratios by comparing the selection rate for each group with that of their highest group (divide the selection rate for a group by the selection rate for the highest group);

EEOC guidelines deserve a measure of respect.[64] With no clear guidance from the Supreme Court as to the weight that should be given to the rule, or when to properly apply it, circuits are split over parties using it to demonstrate or disprove impact.[65] Federal district courts have also been divided over the rule's use.[66] Despite the courts' skepticism of the four-fifths rule, several federal agencies have adopted the rule as a standard for determining when to act on a claim.[67]

---

  (4) observe whether the selection rate for any group is substantially less (i.e., usually less than 4/5ths or 80%) than the selection rate for the highest group. If it is, adverse impact is indicated in most circumstances.

Id.

 64. *See* Watson v. Fort Worth Bank & Trust, 487 U.S. 977, 995 n.3 (1988) (plurality opinion) (noting four-fifths rule "has not provided more than a rule of thumb for the courts"); *see also* Fed. Express Corp. v. Holowecki, 552 U.S. 389, 399 (2008) (describing EEOC Guidelines as reflecting "a body of experience and informed judgment to which courts and litigants may properly resort for guidance. . . . As such they are entitled to a measure of respect. . . ." (citations omitted) (internal quotation marks omitted)). Despite its history of distrust in the rule, recently the Court essentially changed the meaning of the above quote from *Watson*; in *Ricci v. DeStefano* the Court quoted only the last part of the *Watson* description, noting that the four-fifths rule "is 'a rule of thumb for the courts.'" 557 U.S. 557, 587 (2009) (quoting *Watson*, 487 U.S. at 995 n.3).

 65. *See, e.g.*, Stagi v. Nat'l R.R. Passenger Corp., 391 F. App'x 133, 139 (3d Cir. 2010) (describing four-fifths rule as "non-binding" and refusing to use it as factor to determine existence of disparate impact); Allen v. City of Chicago, 351 F.3d 306, 317 (7th Cir. 2003) (holding police department liable for disparate impact because examination pass rates of minorities fell short of four-fifths rule); Boston Police Superior Officers Fed'n v. City of Boston, 147 F.3d 13, 21 (1st Cir. 1998) (using violation of four-fifths rule as one of many factors weighing into finding of disparate impact); Waisome v. Port Auth. of N.Y. & N.J., 948 F.2d 1370, 1379–80 (2d Cir. 1991) (finding violation of four-fifths rule, in combination with statistically significant disparity, as sufficient demonstration of disparate impact); Firefighters Inst. for Racial Equal. v. City of St. Louis, Mo., 616 F.2d 350, 356–57 (8th Cir. 1980) (utilizing four-fifths rule analysis in finding existence of disparate impact); Craig v. Cnty. of L.A., 626 F.2d 659 (9th Cir. 1980) (noting that defending employment practices is more difficult where EEOC guidelines have not been followed); Moore v. Sw. Bell Tel. Co., 593 F.2d 607, 608 (5th Cir. 1979) (holding that no disparate impact existed where pass rate for black applicants was 93% of pass rate of white applicants).

 66. *See, e.g.*, Bradley v. City of Lynn, 433 F. Supp. 2d 157, 159 (D. Mass. 2006) (finding violation of Title VII where passing scores of promotion test by minorities were less than four-fifths of white test-takers). *But see, e.g.*, Johnson v. City of Memphis, Nos. 00-2608 DP, 04-2017 DP, 04-2013 DA, 2006 WL 3827481, at *21 (W.D. Tenn. Dec. 28, 2006) (finding liability despite employer's test satisfying four-fifths rule), *rev'd*, 770 F.3d 464 (6th Cir. 2014).

 67. *See* Peresie, *supra* note 40, at 781 (providing list of agencies that have adopted four-fifths rule). The Department of Labor, the Department of Justice, and the Office of Personnel Management have adopted the rule as a measure of disparate impact. *See id.* The rule has also been adopted by the Office of Federal Contract Compliance Programs. *See* Eric M. Dunleavy, DCI Consulting Grp., A Consideration of Practical Significance in Adverse Impact Analysis (July 2010), *available at* http://dciconsult.com/whitepapers/PracSig.pdf (noting that Office of Federal Contract Compliance Programs has also adopted four-fifths rule).

Using the rule as a marker for disparate impact has some noted advantages, particularly its simplicity.[68] Yet, despite being easy to understand, the rule has been extensively criticized by both courts and scholars alike as an improper measure of disparate impact.[69] The greatest disadvantage is the rule's impact on small employers: often times, the addition or subtraction of a small number of employees will have a substantial impact on the selection ratio between groups.[70] Further, large disparities in impact may still exist but remain undetectable because of the rule's strict cutoff.[71] Such unreliable results and the noted weaknesses of the rule have undoubtedly led to inconsistencies among the lower courts.[72]

3. *Practical Significance*

Another measure courts turn to when deciding disparate impact cases is practical significance, an intuitive measure of substantive impor-

---

68. *See* Peresie, *supra* note 40, at 783 (describing simplicity as primary advantage of four-fifths rule).

69. *See, e.g.*, Clady v. Cnty. of L.A., 770 F.2d 1421, 1428 (9th Cir. 1985) (describing four-fifths rule as "ill-conceived" and capable of producing anomalous results); Cormier v. P.P.G. Indus., Inc., 519 F. Supp. 211, 255 (W.D. La. 1981) (describing four-fifths rule as inappropriate test to determine disparate impact), *aff'd*, 702 F.2d 567 (5th Cir. 1983); Scott W. McKinley, *The Need for Legislative or Judicial Clarity on the Four-Fifths Rule and How Employers in the Sixth Circuit Can Survive the Ambiguity*, 37 Cap. U. L. Rev. 171, 172 (2008) (describing four-fifths rule as undermining purpose of Title VII); Elaine W. Shoben, *Differential Pass-Fail Rates in Employment Testing: Statistical Proof Under Title VII*, 91 Harv. L. Rev. 793, 805 (1978) (advocating for rejection of four-fifths rule in all contexts).

70. *See* Peresie, *supra* note 40, at 784 ("[A] small employer with a small absolute disparity . . . might face liability under the rule, while a large employer can have a much greater disparity and still comply with the four-fifths rule."); Shoben, *supra* note 69, at 806–07 (performing four-fifths analysis and finding that selection rate is more sensitive in small samples than in large ones).

71. *See* Richard E. Biddle, *Disparate Impact Reference Trilogy for Statistics*, Lab. L.J. (1995), *available at* http://www.biddle.com/documents/disparatetrilogy.htm ("Smaller differences in selection rate [i.e., differences within the 80 percent limit, such as .81 or .95, etc.] may nevertheless constitute adverse impact, where they are significant in both statistical and practical terms . . . ." (first alteration in original)). Further, the question over the appropriate "selection rate" often leads to inconsistent application of the rule. *See, e.g.*, Council 31, Am. Fed'n of State, Cnty., & Mun. Emps., AFL-CIO v. Ward, 978 F.2d 373 (7th Cir. 1992) (providing example of confusion about selection rates). In *Council 31*, the employee-plaintiffs presented evidence that white employees were fired at 35% of the rate that black employees were. *See id.* at 376. Recognizing that this was in clear violation of the four-fifths rule, the employer offered statistics regarding retention, rather than termination, of employees, emphasizing that black employees were retained at 94% of the rate that white employees were. *See id.* at 379. The court adopted the plaintiffs' presentation of the statistics, but the case is an example of how both sides in litigation attempt to frame the 80% requirement to their advantage.

72. *See* McKinley, *supra* note 69, at 176–77 (arguing that courts' inconsistent application of four-fifths rule should prevent use of rule as guideline for employers).

tance.[73] A finding of practical significance means "the magnitude of the effect being studied is not de minimis—it is sufficiently important substantively for the court to be concerned."[74] Although it is not as well-defined in the legal context as the standard deviation analysis or the four-fifths rule, some courts have examined a proffered disparity with an eye toward its real-world meaning.[75] For example, the First Circuit impliedly addressed practical significance in *Boston Chapter, NAACP, Inc. v. Beecher*.[76] In *Beecher*, the First Circuit held that statistical evidence demonstrating that African-American and Hispanic applicants performed worse on a multiple-choice test than their white counterparts, in conjunction with *other* evidence, could demonstrate a prima facie case of disparate impact.[77] The First Circuit expanded on this line of reasoning in *Fudge v. City of Providence Fire Department*,[78] holding that the proper analysis in disparate impact claims involves an "intuitive judicial judgment" as to the substantiality of a statistically significant discrepancy.[79]

---

73. *See* FED. JUDICIAL CTR., REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 292 (3d ed. 2011), *available at* http://www.fjc.gov/public/pdf.nsf/lookup/SciMan3D 01.pdf/$file/SciMan3D01.pdf (explaining practical significance).

74. *Id.* at 318; *see also* Hamer v. City of Atlanta, 872 F.2d 1521, 1524–26 (11th Cir. 1989) (describing practical significance as "the degree to which test scores relate to job performance"); Elizabeth Tippett, *Robbing a Barren Vault: The Implications of* Dukes v. Wal-Mart *for Cases Challenging Subjective Employment Practices*, 29 HOFSTRA LAB & EMP. L.J. 433, 449 n.94 (2012) ("Practical significance refers to whether the statistical differences are meaningful in the real world.").

75. *See* Frazier v. Garrison I.S.D., 980 F.2d 1514, 1526 (5th Cir. 1993) (holding 4.5% difference in selection rates was trivial where 95% of applicants were selected); Waisome v. Port Auth. of N.Y. & N.J., 948 F.2d 1370, 1379 (2d Cir. 1991) (citing Ottaviani v. State Univ. of N.Y. at New Paltz, 875 F.2d 365, 371 (2d Cir. 1989)) (noting that courts should take case-by-case approach that considers both statistical analyses and surrounding facts and circumstances); Bilingual Bicultural Coal. on Mass Media, Inc. v. FCC, 595 F.2d 621, 642 n.57 (D.C. Cir. 1978) (Robinson, J., dissenting in part) ("[S]tatistical significance is not the same as practical significance because in isolation it tells nothing about the importance or magnitude of the differences."); NAACP v. Fla. Dep't of Corr., No. 5-00-cv-100-Oc-10GRJ, 2004 U.S. Dist. LEXIS 28348, at *54 (M.D. Fla. July 7, 2004) (rejecting statistically significant sample for lack of practical significance); Jones v. Pepsi-Cola Metro. Bottling Co., 871 F. Supp. 305, 313 n.25 (E.D. Mich. 1994) (noting that employers must demonstrate both statistical and practical significance in defending discriminatory test).

76. 504 F.2d 1017 (1st Cir. 1974).

77. *See id.* at 1021 n.7 (explaining that other factors are useful in determining validity of entrance exam). The First Circuit explained that non-statistical evidence, such as the fact that Boston firefighters were often recruited by relatives and friends, can be probative in determining whether or not a standardized test had an adverse impact on minority candidates. *See id.* (explaining that relatives and friends do not "invalidate the test [but] they are merely additional reasons for inquiring into its utility").

78. 766 F.2d 650 (1st Cir. 1985).

79. *See id.* at 657–58 (holding that proffered statistical disparity was not sufficient to establish prima facie case). In *Fudge*, the plaintiffs offered evidence that a lower percentage of black applicants had passed a screening test as compared to white applicants, but they did not present any additional statistical analysis based

A practical significance consideration typically takes place after a plaintiff has introduced a statistical analysis that produced significant results.[80] Though not binding on courts, the EEOC's guidelines require evidence of both statistical and practical significance in order to prove disparate impact.[81]

One glaring difficulty in adopting a practical significance standard is its lack of a solid definition.[82] Another criticism comes in the form of subjectivity; since there is no set definition of practical significance, courts would have to perform a case-by-case analysis, which could lead to such arbitrary results as the four-fifths rule.[83] Though not directly adopted by the Supreme Court, its line of cases surrounding the causation element of the prima facie case has hinted at what amounts to a consideration of practical significance.[84]

---

on those figures. *See id.* at 657. The plaintiffs' only evidence of disparate impact was that 4% of black applicants and 13% of white applicants had passed the screening test. *See id.*

80. *See* United States v. City of N.Y., 637 F. Supp. 2d 77, 94 (E.D.N.Y. 2009) (holding that plaintiffs' showing of statistical significance was bolstered by demonstration of practical significance); EEOC v. Sears, Roebuck & Co., 628 F. Supp. 1265, 1286 (N.D. Ill. 1986) ("To determine the practical significance of statistical results, a court must look at the theories and assumptions underlying the analysis and apply common sense.").

81. *See* 29 C.F.R. § 1607.4(D) (1981) (requiring showing of both practical and statistical significance). The guidelines provide, in relevant part, that "[s]maller differences in selection rate may nevertheless constitute adverse impact, where they are significant in both statistical and practical terms or where a user's actions have discouraged applicants disproportionately on grounds of race, sex, or ethnic group." *Id.*

82. *See* DUNLEAVY, *supra* note 67, at 1–2 (noting difficulties in practical significance standard). Dunleavy, quoting an unpublished Office of Federal Contract Compliance report, identifies two potential problems with practical significance as a standard for disparate impact:

> First, any standard of practical significance is arbitrary. It is not rooted in mathematics or statistics, but in a practical judgment as to the size of the disparity from which it is reasonable to infer discrimination. Second, no single mathematical measure of practical significance will be suitable to widely different factual settings.

Id. (quoting R. SOBEL ET AL., STATISTICAL INFERENCES OF EMPLOYMENT DISCRIMINATION AND THE CALCULATION OF BACK PAY. PART I: INFERENCES OF DISCRIMINATION (1979) (unpublished OFCCP statistical standards panel report)).

Some courts have noted this as a fault so severe as not to adopt the standard. *See, e.g.*, Jones v. City of Boston, 752 F.3d 38, 50 (1st Cir. 2014) (describing concept of practical significance as incapable of being defined with precision and refusing to adopt it as standard to demonstrate disparate impact).

83. *See* FED. JUDICIAL CTR., REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 192 n.34 (2d ed. 2000) ("There is no specific percentage threshold above which a result is practically significant. Practical significance must be evaluated in the context of a particular legal issue."); *see also* King, *supra* note 13, at 291 ("[T]he standard for determining the sufficiency of statistical evidence should depend on the particular issue to which that evidence is addressed.").

84. *See* Hazelwood Sch. Dist. v. United States, 433 U.S. 299, 312 (1977) (noting that appropriate comparison figures in statistical analysis must be evaluated after further findings); Int'l Bhd. of Teamsters v. United States, 431 U.S. 324, 340

Practical significance as a measure of disparate impact has several advantages.[85] When combined with a statistical significance analysis, courts can determine both the existence of a disparity and the consequences of the disparity.[86] Further, a practical significance analysis has the effect of ruling out those cases where a statistically significant sample is mislabeled as dispositive due to its large sample size.[87] Although not capable of a precise definition, scholars have noted the utility of adopting a practical significance standard and have suggested several procedures for evaluating the practical significance of a statistically significant sample.[88]

(1977) ("We caution [ ] that statistics are not irrefutable; they come in infinite variety and, like any other kind of evidence, they may be rebutted. In short, *their usefulness depends on all of the surrounding facts and circumstances.*" (emphasis added)); *see also* Watson v. Fort Worth Bank & Trust, 487 U.S. 977, 995 n.3 (1988) (plurality opinion) (recognizing that case-by-case approach is proper in determining significance of numerical disparities).

85. For a discussion of the advantages of adopting a practical significance standard, see *infra* notes 86–88 and accompanying text.

86. *See* Peresie, *supra* note 40, at 790 (noting that statistical and practical analyses serve two separate functions).

87. *See* Tinkham, *supra* note 57, at 362 (commenting on utility of practical significance analysis). Although a disparity of two or three standard deviations is enough to raise an inference of discrimination, a statistical analysis does not determine whether that disparity is practically significant. *See id.*

88. *See, e.g.*, Biddle, *supra* note 59 (noting advantages to practical significance standard); DUNLEAVY, *supra* note 67, at 4–8 (arguing that practical significance is appropriate standard for an adverse impact analysis). One approach to quantifying practical significance requires the analyst (whether it be the court or an expert witness) to add two people from the "unfavorable" status (such as failing a drug test) to the "favorable" status (passing said test); if the change renders a once statistically significant sample as insignificant, then the results would not be practically significant. *See id.* at 4–8 (proposing approach). *But see* Kevin R. Murphy & Rick R. Jacobs, *Using Effect Size Measures to Reform the Determination of Adverse Impact in Equal Employment Litigation*, 18 PSYCHOL. PUB. POL'Y & L. 477, 483 (2012) (suggesting that this method provides only marginal protection from trivial claims). Other scholars have suggested using difference in selection rates as a measure of practical significance. *See, e.g.*, Peresie, *supra* note 40, at 801 (advocating for adoption of difference in selection rates as appropriate measure of practical significance). Peresie argues that "disparities with the same absolute magnitude should matter the same amount." *Id.* Under this theory, a selection rate of 15% men and 10% women would indicate the same disparate impact as a rate of 75% men and 70% women. *See id.*

The social sciences have also recently adopted a showing of practical significance as necessary to validate studies, and they have considered a failure to do so a deficiency, although no precise definition exists in that sphere of academia. *See, e.g.*, DUNLEAVY, *supra* note 67, at 6 (citing Sheldon Zedeck, *Instructions for Authors*, 88 J. APPLIED PSYCHOL. 35 (2003)) (requiring authors to provide, among other things, showing of practical significance).

III.   "You're Out of Here!": The First Circuit Explicitly Rejects the
Practical Significance Requirement in
Jones v. City of Boston

In *Jones*, the First Circuit refused to consider practical concerns regarding the probative value of statistical evidence.[89] The court's holding turned on its acceptance of statistical evidence as sufficient to demonstrate a prima facie case.[90] In reaching its holding, the court discounted the reasoning of the district court and refused to accept any practical considerations of the plaintiffs' proffered statistically significant sample.[91]

A.   *Covering the Bases: Facts and Procedure*

In 1999, the Boston Police Department (the "Department") began annually testing officers and cadets for illicit drug use.[92] Specifically, the Department utilized a hair test to detect the existence of five different illegal drugs.[93] Officers and cadets who failed the drug test were offered a second, more sensitive hair test and a medical review to determine if the first test was inaccurate.[94] If neither of these steps exonerated offenders, the officers were given the option to remain at the Department so long as they agreed to seek rehabilitative treatment for drug abuse and accept a forty-five day unpaid suspension; if not, the offending officer or cadet was terminated, and any conditional offer of employment to a recruit would be withdrawn.[95]

The plaintiffs were all black officers, applicants, and cadets who tested positive for the use of cocaine and failed the hair drug test.[96] They alleged that the Department's drug test violated Title VII because it caused a

89. For a thorough discussion of the *Jones* decision, see *infra* notes 92–126 and accompanying text.

90. For a discussion on the First Circuit's reasoning regarding statistical evidence, see *infra* notes 102–13 and accompanying text.

91. For a discussion on the First Circuit's rejection of the four-fifths rule and practical significance standard, see *infra* notes 114–26 and accompanying text.

92. *See* Jones v. City of Boston, 752 F.3d 38, 42 (1st Cir. 2014) (describing Department's drug testing policy). The drug test also applied to job applicants, who would have to consent to a drug test after a conditional offer of employment was extended to them. *See id.*

93. *See id.* (listing drugs that Department tested for). The drugs the Department tested for were cocaine, marijuana, opiates, PCP, and amphetamines. *See id.*

94. *See id.* (explaining "exculpatory" remedies after failed initial drug test).

95. *See id.* (describing options given to officers upon failed drug test and lack of exculpation). If officers or cadets refused to accept the Department's offer of therapy and suspension, they would be terminated. *See id.* (noting Department's rationale for firing officers or cadets). Massachusetts law required that terminated employees be provided with a written notice of reasons for their termination. *See id.* An evidentiary hearing followed, at which officers could argue that they were terminated without cause. *See id.*

96. *See id.* at 42–43 (describing plaintiffs). The plaintiffs included seven former officers fired by the Department, one former cadet who had been fired, one cadet who opted to undergo rehabilitation, and one former applicant. *See id.* at 41.

disparate impact on the basis of their race.[97] Although only a small num-
ber of black and white officers and cadets tested positive for cocaine be-
tween 1999 and 2006, black officers and cadets tested positive for cocaine
at a higher rate than did their white colleagues.[98] The district court
granted summary judgment to the Department on all claims, holding that
the plaintiffs' proffered statistics, while statistically significant, were not
practically significant and thus were insufficient to make a prima facie
showing of racial discrimination.[99]

B. *That's a Balk!: The First Circuit's Reasoning in* Jones

In reversing the district court's judgment, the First Circuit described
the disparate impact analysis in conjunction with its views on the concept
of statistical significance, and it applied those theories to the case at
hand.[100] The court then rejected the district court's reasoning and held
that Title VII did not require a showing of practical significance as to a
plaintiff's statistically significant sample.[101]

1. *"Lies, Damned Lies, and Statistics"*

The First Circuit began its analysis by briefly summarizing the dispa-
rate impact framework.[102] The court explained that, in order to make a
prima facie showing of disparate impact discrimination, a plaintiff must
identify a specific employment practice that the employer actually uses
and show that the identified practice causes a disparate impact on the ba-

---

97. *See id.* (providing background of plaintiffs' claims). The plaintiffs also
brought claims under the Due Process Clause of the Fourteenth Amendment, the
Americans with Disabilities Act, and for failure to train and supervise under 42
U.S.C. § 1983. *See id.* at 55–60. This Note will only discuss and analyze the dispa-
rate impact claim.

98. *See id.* at 42–43 (describing rates and numbers at which officers and cadets
tested positive for cocaine). Black applicants tested positive for cocaine at both a
higher percentage and a higher total number than whites. *See id.* at 43. Over the
eight years the Department utilized the test, fifty-five black officers tested positive
for cocaine as opposed to thirty white officers, even though there were over 6,000
more white officers tested during that time. *See id.* at 44. Black officers tested
positive for cocaine at a 1.3% rate, while white officers tested positive 0.3% of the
time. *See id.* at 43–44.

99. *See id.* at 41 (providing district court's holding); *see also* Jones v. City of
Boston, No. 05-11832-Gao, 2012 WL 4530594, at *3–6 (D. Mass. Sept. 28, 2012)
(detailing district court's reasoning).

100. *See Jones*, 752 F.3d at 42–45 (providing background on statistical analysis
and implications of finding statistical significance); *id.* at 46–48 (analyzing statisti-
cally significant correlation in plaintiffs' proffered sample).

101. *See id.* at 53 (providing First Circuit's holding). The court held that a
practical significance standard would be too difficult to articulate in a "principled
and predictable manner," and that failing to establish practical significance cannot
preclude a plaintiff from proving a prima facie case of disparate impact. *Id.*

102. *See id.* at 46 ("Title VII prohibits employers from utilizing 'employment
practices that cause [ ] a disparate impact on the basis of race' . . . ." (first altera-
tion in original) (quoting 42 U.S.C. § 2000e-2(k) (2012))).

sis of race.[103] Much of the court's analysis focused on the second step by noting that a disparate impact analysis essentially boils down to whether a plaintiff can offer a statistically significant disparity at the prima facie level.[104]

The First Circuit explained the significance of statistical analyses in proving the causal link between the employment practice and the alleged disparate impact.[105] Referencing Mark Twain's famous quip that, "there are three kinds of lies: lies, damned lies, and statistics," the court rejected the notion that the parties' various briefs would be of any use in interpreting the meaning of a statistical analysis, and it instead gave deference to the objective, precise calculations of statisticians.[106] Though statistical significance can be assessed by a variety of methods, the court deemed that a type of test known as a probability analysis was the most relevant for its present purposes.[107] A probability analysis, the court described, determines the likelihood that a given sample was the result of chance, assuming that in a perfect world, both groups would receive equal treatment.[108]

---

103. *See id.* ("[A] plaintiff starts by 'isolating and identifying' the employment practice being challenged." (quoting Watson v. Fort Worth Bank & Trust, 487 U.S. 977, 994 (1988) (plurality opinion))); *see also id.* (citing 42 U.S.C. § 2000e-2(k)(1)(A)(i)) (describing second step of establishing prima facie case).

104. *See id.* ("[A] prima facie showing of disparate impact [is] 'essentially a threshold showing of a significant statistical disparity . . . and nothing more.'" (second alteration in original) (quoting Ricci v. DeStefano, 557 U.S. 557, 587 (2009))); *see also id.* ("[A] prima facie case of disparate impact can be established where 'statistical tests sufficiently diminish chance as a likely explanation.'" (quoting Fudge v. City of Providence Fire Dep't, 766 F.2d 650, 658 n.8 (1st Cir. 1985))).

105. *See id.* at 43–45 (giving court's summation of statistical significance and its application to case at hand).

106. *See id.* at 43 (giving deference to precise approach of statisticians). The court posited that, without any expertise in statistical analysis, the parties would be unable to persuade the court of the actual meaning of a statistical test and only offered "unhelpful types of competing characterizations of the numbers." *Id.* The court made clear its preference for a statistician's precise approach, whereby it may be determined whether a correlation between an employment practice and race exists, and if it does, whether that correlation exists as a product of chance. *See id.*; *see also* MARK TWAIN, THE AUTOBIOGRAPHY OF MARK TWAIN 195 (Charles Neider ed., 2000) ("Figures often beguile me, particularly when I have the arranging of them myself . . . There are three kinds of lies: lies, damned lies, and statistics."). The court referenced part of the foregoing passage, which appropriately set the tone for the remainder of the court's analysis. *See Jones*, 752 F.3d at 43.

107. *See Jones*, 752 F.3d at 43 (citing Paul Meier, Jerome Sacks & Sandy L. Zabell, *What Happened in* Hazelwood: *Statistics, Employment Discrimination, and the 80% Rule*, 1984 AM. BAR FOUND. RESEARCH J. 139, 147 (1984)) (noting court's reasoning for probability analysis). The court framed the probability test to the facts of *Jones* to illustrate its usefulness in assessing statistical significance. *See id.* ("In the approach most relevant here, statisticians may compare outcomes for 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).").

108. *See id.* (observing that assumption of equal opportunity does not necessitate actual equal outcomes). The court noted that random variation may cause differences in a given outcome, and that the purpose of performing a probability

The court observed that when statisticians employ this test, they will deem results statistically significant if certain criteria are met, specifically if the tested sample deviates a certain amount from the expected outcome.[109] Statisticians, the court noted, will reject the idea that the disparity between the groups was caused by chance when the probability test produces statistically significant results.[110]

Applying this framework to the case at hand, the First Circuit found that the plaintiffs' proffered analysis was statistically significant.[111] Despite confessing its uncertainty as to whether to consider the data aggregated over eight years as a single sample, the court nevertheless held that it could "be *almost* certain that the difference in outcomes associated with race over that period cannot be attributed to chance alone."[112] The court then overruled several secondary objections raised by the Department, the most notable of which questioned the plaintiffs' aggregation of data in its proffered statistical analysis.[113]

---

analysis is to determine the likelihood that the difference from the expected outcome was the result of random chance. *See id.* (citing FED. JUDICIAL CTR., *supra* note 73, at 251). The probability of observing a difference equal to or greater than that which actually occurred is known as the "*p*-value." *See id.* at 43–44. Courts generally accept a *p*-value of 5% as evidence of statistical significance. *See id.* (citing *Fudge*, 766 F.2d at 658 n.8).

109. *See id.* (describing concept of standard deviation). Standard deviation serves as a measure of the amount by which the given sample differs from an outcome resulting from equal opportunity. *See id.* at 44 (citing *Fudge*, 766 F.2d at 658 n.8). A standard deviation of 1.96 is equivalent to a *p*-value of 0.5 and is the level at which courts will recognize statistical significance. *See id.* (citing Castaneda v. Partida, 430 U.S. 482, 496 n.17 (1977)) (noting correlation between standard deviation, *p*-value, and statistical significance). The Supreme Court noted in *Castaneda* that "[a]s a general rule for such large samples, if the difference between the expected value and the observed number is greater than two or three standard deviations, then the hypothesis . . . [of randomness] would be suspect to a social scientist." *Castaneda*, 430 U.S. at 496 n.17.

110. *See Jones*, 752 F.3d at 43 ("Essentially, a finding of statistical significance means that the data casts serious doubt on the assumption that the disparity was caused by chance.").

111. *See id.* at 44 (providing plaintiffs' statistical analysis). The court noted that the presented analysis did not indicate whether the difference in outcomes was, on its face, large, but it indicated the extent to which the outcomes were attributed to something other than chance. *See id.* at 45 ("[T]he difference in outcomes associated with race over [the eight-year] period cannot be attributed to chance alone.").

112. *Id.* (emphasis added). The court also noted that randomness could not be viewed as anything but "a very unlikely explanation for results in at least three of the years viewed in isolation." *Id.*

113. *See id.* at 47 (describing Department's objections). The Department's concern over aggregation stemmed from the fact that many individual officers had been tested in several of the years given in the sample, which called into question the assumption that the test results for different years were independent events. *See id.* The court rejected this argument, noting that a demonstration that the plaintiffs' analysis was skewed would be too complex. *See id.* at 48 n.11 (determining that Department's analysis would "implicat[e] such factors as layoff and hiring practices and the probability that a person who tested negative in one year will test

2. *Picked Off: The First Circuit Rejects the Practical Significance Standard*

Most significantly, the First Circuit rejected the Department's argument that Title VII requires a plaintiff to prove that a statistically significant sample also is practically significant in order to demonstrate a prima facie case.[114] The court began by acknowledging that several of the Department's arguments regarding sample size had merit, conceding that the larger the sample size, the more likely it is that even small differences between groups can be statistically significant.[115] Further, the court recognized that statisticians regularly assess the practical significance of large, statistically significant samples due to the problems associated with those outcomes.[116] The court also referenced the EEOC's guidelines, which it described as being "reasonably read as interpreting Title VII to include a practical significance requirement."[117]

Despite its willingness to concede these logical points brought up by the Department, the court forcefully rejected a practical significance requirement.[118] The court's first concern with such a requirement was a lack of any objective measure of practical significance, expressing apprehension at the idea of courts applying an "elusive, know-it-when-you-see-it standard."[119] The First Circuit next took the four-fifths rule to task, pointing out several flaws with using the rule as a measure of practical signifi-

negative in a later year, as compared to the probability of a negative result for someone first tested in the later year").

114. *See id.* at 48 (rejecting Department's central argument). The district court had adopted the Department's argument, holding that the plaintiffs could not establish a prima facie case based solely on a showing of two or more standard deviations in failure rates. *See* Jones v. City of Boston, No. 05-11832-Gao, 2012 WL 4530594, at *3 (D. Mass. Sept. 28, 2012) ("For these reasons, the defendants are correct that the plaintiffs have not satisfied the first step in making out a prima facie case.").

115. *See Jones*, 752 F.3d at 48–49 (discussing problematic connection between statistical significance and large sample sizes). The court provided a useful example regarding large sample size and statistical significance:

> 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.

Id. at 49.

116. *See id.* ("[S]tatisticians acknowledge that not all statistically significant results are practically significant, meaning 'practically meaningful or important.'" (quoting Xitao Fan, *Statistical Significance and Effect Size in Education Research: Two Sides of a Coin*, 94 J. EDUC. RESEARCH 275, 277 (2001))).

117. *See id.* at 50. The court noted that the federal regulation that created the four-fifths rule considers both practical and statistical significance as relevant in determining whether disparate impact exists. *See id.* (discussing 29 C.F.R. § 1607.4(D)).

118. *See id.* at 50–53 (rejecting Department's main argument).

119. *See id.* at 50–51 (discussing lack of objective measure of practical significance). The court described the concept of practical significance as "impossible to define in even a remotely precise manner." *Id.* at 50.

cance.[120] Although other courts often use the four-fifths rule as a "rule of thumb," the court noted that the Supreme Court and the EEOC have questioned the rule as a decisive tool.[121] Further, the court found the rule to be flawed because it can lead to inconsistent results depending on how it is interpreted.[122] Although the court recognized that the four-fifths rule may serve important goals in contexts outside of the courtroom, it dismissed the Department's reliance on the rule as a measure of practical significance.[123]

Similarly, the court recognized that, although there may be some "theoretical benefits" to requiring a showing of practical significance, it would simply be too difficult to define or enforce such a standard in any "principled and predictable manner."[124] Because it could not find an acceptably accurate measure of practical significance, the court chose to simply discard the requirement altogether.[125] The court concluded its analysis by suggesting that plaintiffs must fulfill other requirements in proving a disparate impact case that indirectly satisfy the benefits a practical significance standard would yield.[126]

---

120. *See id.* at 51–53 (describing flaws of four-fifths rule).

121. *See id.* at 51 (citing Watson v. Fort Worth Bank & Trust, 487 U.S. 977, 995 (1988) (plurality opinion)). *But see id.* (citing Adoption of Questions and Answers to Clarify and Provide a Common Interpretation of the Uniform Guidelines on Employee Selection Procedures, 44 Fed. Reg. 11996, 11998 (Mar. 2, 1979)) (noting that four-fifths rule was not meant to be legal definition or controlling in all circumstances). The First Circuit also noted its own reluctance to rely on the rule in a past disparate impact case. *See id.* (citing Fudge v. City of Providence Fire Dep't, 766 F.2d 650, 658 n.10 (1st Cir. 1985)) (describing four-fifths rule as inaccurate test of discriminatory impact). In *Fudge*, the court noted that the 4% passage rate of black applicants, as compared to the 13% passage rate of white applicants, fell well short of the four-fifths rule's requirements, but it rejected the rule as a measure of impact because the sample size was too small. *See Fudge*, 766 F.2d at 656, 658 n.10 (noting that sample size of 248 applicants was not large enough to justify use of four-fifths rule).

122. *See Jones*, 752 F.3d at 51–52 (highlighting several flaws with four-fifths rule). The *Jones* court pointed out that a given sample may be deemed to meet the four-fifths requirement depending on which "selection rate" is chosen by the statistician; choosing a selection rate at which employees pass a test, as opposed to the rate at which employees fail a test, may lead to an entirely different conclusion. *See id.* at 52. Further, the court took issue with the rule's strict 80% requirement, explaining that a court implementing the rule would put no greater weight on a policy that leads to termination of 100 in 900 black employees than it would to a policy that leads to the termination of one in nine black employees. *See id.*

123. *See id.* (rejecting application of four-fifths rule to determine practical significance of statistically significant sample).

124. *Id.* at 53. Having rejected the four-fifths rule, the court decided that there was no other mathematical measure of practical significance available for its consideration. *See id.* ("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.").

125. *See id.*

126. *See id.* (describing existing requirements imposed by Title VII on plaintiffs in disparate impact cases). The court construed the need for plaintiffs to show statistical significance as a filter that would preclude "small impacts" from making

IV. SWING AND A MISS: THE FIRST CIRCUIT FAILED TO EXAMINE
PRACTICAL SIGNIFICANCE

In *Jones*, the First Circuit had the opportunity to set forth a new standard in disparate impact claims, but it instead fell into the same trap that courts have fallen into over the past thirty years: placing too much emphasis on a finding of statistical significance, while overlooking or ignoring the statistics' practical meaning.[127]

The first misstep the court made was disregarding the potential influence of other circuit courts and scholarship on the subject, as the court dismissed case law from other circuits as "sparse" and unclear.[128] The court's claim is simply incorrect.[129] Several courts have adopted a requirement of showing practical significance, which is typically added as a second tier of analysis in conjunction with statistical evidence.[130] Further, it is well-established in the study of statistics that a showing of statistical significance is not dispositive, particularly when dealing with large sample sizes, as was the case in *Jones*.[131] Moreover, the First Circuit implied that the concepts of statistical significance and practical significance are mutually exclusive, when, in fact, they are intertwined and most beneficial when used in tandem.[132]

The First Circuit also erred in misinterpreting its own precedent on practical significance.[133] Although the court has not explicitly addressed

it past the prima facie stage, because proving statistical significance with a small impact requires attaining often-unavailable large sample sizes. *See id.* Further, the court felt confident that the consequent steps in a disparate impact case would give a defendant ample opportunity to weed out insignificant or otherwise justifiable claims of impact. *See id.* ("[E]ven 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."). Having rejected the Department's proposal for a practical significance requirement, the First Circuit reversed the district court's decision. *See id.* at 60 ("The plaintiffs have proven beyond reasonable dispute a prima facie case of disparate impact under Title VII . . . .").

127. For an in-depth discussion of the court's refusal to adopt a practical significance standard, see *infra* notes 128–43 and accompanying text.

128. *See Jones*, 752 F.3d at 50 (labeling case law from other federal circuit courts as "sparse" and unable to offer answers as to practical significance standard).

129. For a discussion of case law from other courts examining practical significance, see *supra* note 75 and accompanying text.

130. *See, e.g.*, Waisome v. Port Auth. of N.Y. & N.J., 948 F.2d 1370, 1376 (2d Cir. 1991) (holding that case-by-case analysis of all facts and circumstances surrounding statistical disparities is appropriate consideration).

131. For a discussion of scholarship advocating for a practical significance standard, see *supra* note 88.

132. *See Jones*, 752 F.3d at 53 (holding that plaintiffs' evidence of statistical significance was sufficient to establish prima facie case where no demonstration of practical significance was made). The court also spoke of practical significance as "trump[ing] a showing of statistical significance," implying that the court viewed the two concepts as competing, not complimentary, considerations. *Id.* at 52.

133. For a discussion on the First Circuit's misinterpretation of its own precedent, see *infra* notes 134–37 and accompanying text.

"practical significance" as a standard, prior First Circuit cases can be read as having already deployed such an analysis in the disparate impact context.[134] In *Beecher*, the court determined that the plaintiffs' use of "other evidence" in support of a disparate impact claim may be highly relevant.[135] The First Circuit later expanded on the *Beecher* ruling in *Fudge*, where it held that the appropriate analysis in disparate impact claims includes consideration of both statistical and practical evidence.[136] District courts within the First Circuit have also applied a similar analysis.[137]

The court also misinterpreted the four-fifths rule as being the only viable measure of practical significance.[138] Though the court noted that there are no other *mathematical* measures of practical significance, it failed to recognize that not all evidence in the employment discrimination context must be mathematical in nature.[139] Although the four-fifths rule has been widely criticized, it is not the only measure of practical significance; as illustrated above, there are several suggestions on how to consistently calculate practical significance.[140] In *Jones*, the defendant's expert witness offered such a suggestion to determine practical significance, but the First Circuit dismissed the recommendation without addressing it.[141]

---

134. For a discussion of prior First Circuit cases considering practical significance, see *supra* notes 76–79 and accompanying text.

135. *See* Boston Chapter, NAACP, Inc. v. Beecher, 504 F.2d 1017, 1021 n.7 (1st Cir. 1974) (holding that other evidence may be probative where proffered sample size is small).

136. *See* Fudge v. City of Providence Fire Dep't, 766 F.2d 650, 657 (1st Cir. 1985) (rejecting four-fifths rule as proper analytical tool where plaintiff offered no statistical evidence as to alleged disparity). In an important passage, the court held that: "In a case involving a claim . . . [of] disparate and adverse impact . . . the initial inquiry must be whether there is a discrepancy in the rate of passage . . . . If so, an *intuitive judicial judgment* must next be made whether the discrepancy is *substantial.*" *Id.* (emphasis added).

137. *See, e.g.*, Bradley v. City of Lynn, 443 F. Supp. 2d 145, 164 (D. Mass. 2006) (examining statewide aggregated data against localized data to determine practical significance of varying outcomes).

138. *See* Jones v. City of Boston, 752 F.3d 38, 52 (1st Cir. 2014) ("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.").

139. *See, e.g.*, EEOC v. Sears, Roebuck & Co., 628 F. Supp. 1264, 1353 (N.D. Ill. 1986) (holding that statistical disparities in hiring between men and women were attributed externally to women's lack of interest rather than to internal discriminatory motive). In systemic disparate treatment cases, courts frequently take other, non-mathematical considerations into account when determining liability. *See* Tristin K. Green, *The Future of Systemic Disparate Treatment Law*, 32 Berkeley J. Emp. & Lab. L. 395, 417 (2011) (commenting that statistical evidence of disparities is probative, but not dispositive, in employment discrimination context).

140. For a discussion of proposed methods for determining practical significance, see *supra* notes 85–88 and accompanying text.

141. *See* Expert Report of Dr. Kathleen K. Lundquist at 4, Jones v. City of Boston, No. 05-11832-Gao, 2012 WL 4530594 (D. Mass. Mar. 13, 2009), 2009 WL 8754866 (relying on EEOC Uniform Guidelines and applying statistical and practical significance tests). Lundquist, the defendant's expert witness, demonstrated in

Finally, the court's reasoning is overly broad in that it rejects the need for a showing of practical significance altogether.[142] This holding flies in the face of supporting academic literature, and it defies sound logic proposed by other courts that the First Circuit seemed to simply ignore.[143]

V. Instant Replay: Proposal of a Two-Part Significance Framework and How Jones Should Have Played Out

Jones illustrates the need for a new framework in disparate impact analysis that adequately accounts for statistical and practical concerns.[144] Such a framework would more accurately address the concerns of Title VII by screening out employment practices that actually have a discriminatory effect on a given class.[145] Had the First Circuit used a framework that incorporated concerns of practical effects, the outcome in *Jones* would have been different.[146]

A. *Let's Play Two!: A New Approach to Disparate Impact*

Although the four-fifths rule may be a poor method of measuring disparate impact, the EEOC did accurately assess the need for *some* consideration of practical significance.[147] Statistical significance and practical significance separately are useful tools to determine the existence of disparate impact.[148] However useful they may be, the two standards share a weakness: they both, individually, only measure one of two important in-

---

her report that in an analysis using a sample size as large as the entire Boston Police Department, statistical significance tests may have no practical significance. *See id.* at 3–4. By way of example, Lundquist provided the following hypothetical situation: "For instance, if 100% of white officers passed the Hair Drug Test in 2000, a pass rate of 99.4% [of black officers] would have been identified as statistically significant." *Id.*

142. For a discussion of the need for a practical significance standard in disparate impact cases, see *supra* notes 86–88 and accompanying text.

143. For a discussion of academic support for practical significance and other courts' analyses of the standard, see *supra* notes 73–88 and accompanying text.

144. For a critical analysis of *Jones*, see *supra* notes 127–43 and accompanying text.

145. *See* Peresie, *supra* note 40, at 792 (arguing that two-prong test at prima facie stage would more accurately measure existence of impact). For a discussion of a proposed two-part framework incorporating practical significance at the prima facie stage, see *infra* notes 147–65 and accompanying text.

146. For a discussion of the proposed two-part framework as applied to *Jones*, see *infra* notes 166–76 and accompanying text.

147. *See* Murphy & Jacobs, *supra* note 88, at 486 (advocating for federal enforcement agencies to define disparate-impact theory with both statistical significance and practical significance in mind); George Rutherglen, *Disparate Impact Under Title VII: An Objective Theory of Discrimination*, 73 Va. L. Rev. 1297, 1323 (1987) (noting that plaintiffs' burden in demonstrating disparate impact should require more than statistically significant analysis).

148. *See* Dunleavy, *supra* note 67, at 4–8 (discussing benefits of using statistical significance and practical significance in conjunction).

quiries in disparate impact cases.[149] On the one hand, statistical signifi-
cance allows plaintiffs to demonstrate that a particular practice causes
some disparity between classes (the "disparate" prong of the inquiry); on
the other, practical significance determines if that disparity is large
enough to have real-world implications (the "impact" prong of the in-
quiry).[150] Practices that do in fact create a noticeable disparate impact
would implicate both of these considerations.[151]

The first step that courts should take under this new framework
would be very familiar; as to statistical significance, courts should continue
to employ the standard deviation analysis as a tool to determine if a dispar-
ity actually exists.[152] As part of the first step, plaintiffs would be required
to demonstrate that a disparity exists that is at least two standard deviations
from the outcome that assumes equal opportunity.[153] This disparity is
large enough to give a sample probative value because, as described above,
two standard deviations is equivalent to a showing that the demonstrated
disparity would occur as a result of random chance only one in twenty
times.[154] Defendants would be able to continue to attack the statistical
analysis and question the methods used or the sample size analyzed.[155]

---

149. *See* Peresie, *supra* note 40, at 790 (describing tests as problematic when
used individually).

150. *See id.* at 790–91.

151. *See* Rutherglen, *supra* note 147, at 1324 (noting that employment prac-
tices that create great adverse impact likely serve as pretexts for discrimination and
deny equality of opportunity).

152. *See* Peresie, *supra* note 40, at 792–93 (arguing that courts should con-
tinue to utilize standard deviation as measure of disparity). Retaining the same
analytical tool carries several advantages. *See id.* The standard deviation analysis is,
compared to other statistical analyses, relatively easy to understand. *See* Kaye, *supra*
note 11, at 1337 (noting that concept of statistical significance is "easily grasped").
Further, it is well-engrained in disparate impact case law nationwide; courts would
have no difficulty keeping a standard in place that they have already used for more
than forty years. *See supra* notes 41–61 and accompanying text.

153. *See* Kadas v. MCI Systemhouse Corp., 255 F.3d 359, 362–63 (7th Cir.
2001) (describing problems with using lower or higher *p*-value to determine signif-
icance). Judge Posner, writing for the Seventh Circuit, observed:

A lower significance level may show that the correlation is spurious, but
may also be a result of "noise" in the data or collinearity . . . and such
evidence, when corroborated by other evidence, need not be deemed
worthless. Conversely, a high significance level may be a misleading arti-
fact of the study's design; and there is always the risk that . . . in the
circumstances, it was a chance result with no actual evidentiary
significance.

Id. at 362.

154. For a discussion on the standard deviation analysis and levels of signifi-
cance, see *supra* notes 41–43 and accompanying text.

155. *See* Peresie, *supra* note 40, at 778 (describing shift of burden of proof to
employer). If a plaintiff can offer a statistically significant disparity, the defendant
can then attempt to rebut the plaintiff's statistical evidence or demonstrate that
the questioned employment practice is job-related and consistent with business
necessity. *See id.*

If plaintiffs are able to demonstrate that a statistically significant disparity exists as a result of an employment practice, they should then be required to show that this disparity is practically significant.[156] As the court in *Jones* observed, practical significance is a difficult concept to define in any concrete manner.[157] However, various methods exist for determining practical significance that plaintiffs and defendants alike could use to persuade a court, much in the same way the four-fifths rule has been used in courts in the past.[158] Courts should nevertheless discredit the four-fifths rule as a viable measure of disparate impact altogether due to its well-established history of inconsistent results.[159] Although a "know-it-when-you-see-it" approach should be avoided, the particular nature of statistical analysis may lend itself to a "case-by-case" analysis or an "intuitive judicial judgment."[160] Until Congress or the Supreme Court adopts a new standard for practical significance, or until the EEOC provides a more coherent and consistently-applicable measure, courts may have to develop the standard through the common law.[161] Used in conjunction, the two tests will more accurately weed out those employment practices that are causing actual discrimination, and they will also prevent unjust findings of

---

156. *See id.* at 802 (advocating for adoption of two-part test for significance at prima facie stage in disparate impact claims).

157. For a discussion on the difficulty in defining practical significance precisely, see *supra* note 82 and accompanying text.

158. *See* Peresie, *supra* note 40, at 799–800 (proposing three methods for establishing practical significance). Courts might utilize a method similar to the four-fifths rule, but use the fail rate instead of the pass rate in comparing selection ratios. *See id.* at 799 ("One way to minimize this problem is to use the fail ratio instead of the pass ratio when the selection rates are very low."). Or, plaintiffs could be required to show that the difference in selection rates, not the ratio between the rates, establishes practical significance. *See id.* (noting that courts "could require plaintiffs to show that the difference was equal to, or greater than, a specified number of percentage points in the selection rate to establish practical significance"). Alternatively, defendants could demonstrate lack of practical significance by showing that subtracting a specified number of people from the disfavored group and adding them to the favored group discredits the significance of a statistical analysis. *See id.* at 800 ("If the addition of a certain number of persons to the pass group of the minority eliminates any statistically significant disparity, then the disparity is not practically significant.").

159. For a description of problems with the four-fifths rule, see *supra* notes 69–72 and accompanying text.

160. *See* Waisome v. Port Auth. of N.Y. & N.J., 948 F.2d 1370, 1379 (2d Cir. 1991) (citing Ottaviani v. State Univ. of N.Y. at New Paltz, 875 F.2d 365, 371 (2d Cir. 1989)) (noting that courts should take case-by-case approach that considers both statistical analyses and surrounding facts and circumstances in determining if disparate impact occurred); Fudge v. City of Providence Fire Dep't, 766 F.2d 650, 657 (1st Cir. 1985) (applying "intuitive judicial judgment" as proper mechanism for analyzing substantiality of established statistical discrepancy).

161. *See* Peresie, *supra* note 40, at 802 (recognizing necessity for practical and statistical significance tests and noting that choosing proper standard is "policy-laden decision"). *But see* Rutherglen, *supra* note 147, at 1330 (advocating that courts should establish modest threshold that includes consideration of practical significance).

discrimination against employers whose tests or other employment prac-
tices have no practically-meaningful impact on different classes.

Though this two-tiered approach has several benefits, it does have
some potential weaknesses as well.[162] First, it very obviously increases the
plaintiff's burden in establishing a prima facie case; plaintiffs very rarely
win disparate impact claims as it is, and this added burden would surely
screen out previously-acceptable claims.[163] Secondly, there is scholarship
that indicates that the standard deviation analysis may not, in fact, be the
most accurate statistical measure of impact.[164] However, until some mea-
sure of significance is suggested by the Supreme Court or is codified by
Congress, the approach proposed above would still allow suits that actually
identify discriminatory practices and policies to proceed.[165]

B. *Double Play: New Standard as Applied to* Jones

The First Circuit erred in *Jones v. City of Boston* by not requiring the
plaintiffs to show that their statistically significant sample was also practi-
cally significant.[166] But, what if the court *had* embraced practical signifi-
cance as a measure of impact—would the outcome have changed at all?[167]

---

162. For a discussion of weaknesses of the two-tiered approach, see *infra* notes
163–65 and accompanying text.

163. *See* Selmi, *supra* note 6, at 739–40 (reporting that plaintiffs win about
25.1% of disparate impact cases at trial in federal district courts); *see also* Kevin M.
Clermont & Stewart J. Schwab, *Employment Discrimination Plaintiffs in Federal Court:
From Bad to Worse?*, 3 Harv. L. & Pol'y Rev. 103, 103–04 (2009) (describing how
plaintiffs win low proportion of cases at trial). Plaintiffs in federal employment
discrimination cases also generally have a difficult time resolving cases pretrial. *See
id.* at 111 (noting lack of success for plaintiffs pretrial). They also win a low per-
centage of cases at trial and, when they do, are often overturned on appeal. *See id.*
at 109–11 (providing statistics for plaintiffs' success rates at trial).

164. *See, e.g.*, Murphy & Jacobs, *supra* note 88, at 486–87 (advocating for sev-
eral different methods of determining impact). One such method would be to
evaluate the difference between groups in terms of pooled deviations. *See id.* at
487 (noting that pooled deviations would "express[ ] the difference between
groups in terms of the number of standard deviations that separate [the groups']
average scores"). This method would produce a *d* value, which would range on the
low end from zero (no impact) to one (large impact). *See id.* Another potential
method would be to evaluate the percentage of variance explained by differences
between the groups being compared:

> If, for example, differences between men and women explain less than
> 1% of the variance in selection decisions, that figure will have the same
> meaning and the same implications regardless of the sample size. . . . PV
> values of .01, .05, .15, .20, respectively, are widely used to describe small,
> medium, moderately large, and large effects in the social and behavioral
> sciences.

Id.

165. *See* Peresie, *supra* note 40, at 792 (arguing that two-prong test at prima
facie stage would more accurately measure existence of impact).

166. For an analysis of the First Circuit's holding in *Jones*, see *supra* notes
100–26 and accompanying text.

167. For an application of the two-part test to the facts of *Jones*, see *infra* notes
168–76 and accompanying text.

Applying the proposed two-part test to the facts of *Jones* reveals several holes in the court's reasoning.[168] Under this new framework, the first part of the plaintiffs' prima facie case—offering statistical evidence at or above the two standard deviation threshold—would be satisfied.[169]

The second step in the proposed analysis, that would require plaintiffs to demonstrate practical significance, is where the most obvious divergences from the court's holding are evident.[170] Had the court taken practical significance into consideration, it would have had to more thoroughly examine the aggregation issue raised by the defendants.[171] Although that specific argument was deemed not to have been properly preserved upon appeal to the First Circuit, the Department could have also attacked the plaintiffs' aggregated data on the basis that the impact rose to an actionable level in only three of the eight years included in the sample.[172] It was only in the aggregate that the plaintiffs' sample seemed to shock the court into finding evidence of disparate impact.[173] In the same vein, had the court taken into account the sheer size of the sample

---

168. For a discussion of the flaws in the *Jones* court's reasoning, see *infra* notes 169–76 and accompanying text.

169. *See* Jones v. City of Boston, 752 F.3d 38, 44 (1st Cir. 2014) (displaying table of plaintiffs' statistical analysis). In *Jones*, the plaintiffs' proffered statistical sample proved, in the aggregate, that the actual result was 7.14 standard deviations away from the expected norm, well above the necessary showing. *See id.* at 44 n.8 (providing sample's "standard deviation of 7.14 for all years combined is far greater than the average of the standard deviations in the individual years").

170. For an application of the second part of the test to the facts of *Jones*, see *infra* notes 171–76 and accompanying text.

171. *See Jones*, 752 F.3d at 48 (detailing defendant's argument regarding aggregation of eight years of data); *see also supra* note 113 and accompanying text.

172. *See Jones*, 752 F.3d at 44–45 (presenting plaintiffs' proffered statistical data). In 1999, 2002, and 2003, the standard deviations between the actual and expected outcomes were 3.43, 4.41, and 2.01, respectively, which would be large enough disparities to be actionable on their own. *See id.* at 44. However, in the other five years in the sample, the standard deviation analysis did not produce statistically significant results, although some disparity did exist in all those years. *See id.*

173. *See id.* at 45 (providing court's stance on plaintiffs' statistical data). The First Circuit acknowledged its uncertainty in dealing with the aggregated data, but still accepted the analysis as appropriate:

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

Id.

Although the Code of Federal Regulations supports aggregation, it clarifies that it should only be utilized where the "numbers . . . are too small to be reliable . . . ." 29 C.F.R. § 1607.4(D) (1981) ("Where . . . evidence concerning the impact of a selection procedure indicates adverse impact but is based upon numbers which are too small to be reliable, evidence concerning the impact of the procedure over a longer period of time . . . may be considered in determining adverse impact.").

provided, it may have seen the necessity for a practical significance analysis.[174] Finally, the court should have taken into consideration the small difference in *actual* failed drugs tests between the races.[175] Had it examined the practical, real-world consequences of the disparity, the court would likely have discovered that these considerations rendered the plaintiffs' statistics, at the very least, to be much less probative than they originally understood.[176]

VI. ROUNDING THE BASES: LOOKING FORWARD AFTER *JONES*

The First Circuit's decision in *Jones v. City of Boston* is a prime example of how a consideration of practical significance can effectively and more justly change disparate impact law.[177] However, as the law currently stands, the *Jones* decision may have some lasting implications in both the First Circuit and in other courts nationwide.[178] Some commentators have suggested that the First Circuit's holding in *Jones* will make it easier for plaintiffs in the First Circuit to demonstrate a prima facie case of disparate impact.[179] The case may also be a point of reference for other courts going forward, as the use of statistics in disparate impact cases continues to challenge both courts and practitioners alike.[180] From a practitioner's

---

174. *See Jones*, 752 F.3d at 44 (displaying plaintiffs' sample). The aggregated eight-year sample included over 15,000 individuals tested. *See id.* For a discussion on how large sample sizes affect findings of statistical significance, see *supra* notes 56–58 and accompanying text.

175. *See Jones*, 752 F.3d at 44 (providing statistics). In total, fifty-five black officers and recruits tested positive for cocaine, compared to thirty white officers and recruits. *See id.* at 41. The total numbers reveal that 0.3% of white officers (30/10,835) and 1.3% of their black co-workers (55/4,222) failed the drug test. *See id.* at 44 (comparing cocaine use between black and white officers). The court admitted that the difference between these percentages is "very small." *Id.* at 42.

176. *See id.* at 45 (acknowledging that plaintiffs' statistical sample "does not establish that the differences in outcomes were large").

177. *See supra* notes 166–76 and accompanying text.

178. For a discussion on the impact of the *Jones* decision, see *infra* notes 179–83 and accompanying text.

179. *See* Gerald L. Maatman, Jr. & Jennifer A. Riley, *First Circuit Creates Potential Circuit Split by Rejecting "Practical Significance" Requirement and Finding that Hair-Based Drug Test Has Disparate Impact*, SEYFARTH SHAW LLP (May 14, 2014), http://www.workplaceclassaction.com/2014/05/first-circuit-creates-potential-circuit-split-by-rejecting-practical-significance-requirement-and-finding-that-hair-based-drug-test-has-disparate-impact/ ("If the decision stands, it smooths the road for plaintiffs looking to establish disparate impact in the First Circuit."); *see also* Amici Curiae Brief of Massachusetts Employment Lawyers Association et al. at 4, Jones v. City of Boston, 752 F.3d 38 (1st Cir. 2014) (No. 12-2280), 2013 WL 2149405 (describing plaintiffs' prima facie burden as mere threshold showing not intended to be difficult).

180. *See* Alan H. Schorr, *Alan Schorr's Employment Case of the Week Ending 5/9/14:* Jones v. City of Boston, ALAN H. SCHORR & ASSOCS., P.C. (May 12, 2014, 2:07 PM), http://www.schorrlaw.com/blog/154-week-ending-5-9-jones.html ("[T]he use of statistics in disparate impact cases continues to make this one of the most interesting and challenging aspects of employment law.").

standpoint, *Jones* will make it more difficult to defend against a statistically significant sample, despite any other factors that may influence that outcome.[181] Further, this case may be persuasive in light of the Supreme Court potentially deciding a disparate impact case in the context of the Fair Housing Act in the near future.[182] Finally, employers may take the *Jones* ruling into consideration when implementing new employment tests or analyzing existing tests by remaining cognizant that they may now be liable for any statistically significant disparity, regardless of its real-world implications.[183]

---

181. *See* Lydell C. Bridgeford, *Q&A: Statistical Proof of Discrimination Isn't Static*, BLOOMBERG BNA LAB. & EMP. BLOG (June 20, 2014), http://www.bna.com/qa-statistical-proof-b17179891425/ (noting that *Jones* eliminates employers' argument for consideration of practical significance of small differences in selection rates); Maatman & Riley, *supra* note 179 (asserting that plaintiffs are likely to cite *Jones* as disposing of practical considerations and will pursue "practically meaningless" statistical disparities in court).

182. *See* Sam Hananel, *Supreme Court May Get 'Disparate Impact' Housing Case*, INS. J. (Sept. 8, 2014), http://www.insurancejournal.com/news/national/2014/09/08/339951. htm (reporting that several cases challenging Department of Housing and Urban Development's regulations based on disparate impact claims may be heard by Supreme Court in near future).

183. *See* Maatman & Riley, *supra* note 179 (noting that practically insignificant disparities are likely to be pursued by plaintiffs in First Circuit).