

Oklahoma Law Review

Volume 68 | Number 3

2016

Lies, Damned Lies, and Statistics: The Case to Require “Practical Significance” to Establish a Prima Facie Case of Disparate Impact Discrimination

Katie Eissenstat

Follow this and additional works at: <https://digitalcommons.law.ou.edu/olr>



Part of the [Labor and Employment Law Commons](#)

Recommended Citation

Katie Eissenstat, *Lies, Damned Lies, and Statistics: The Case to Require “Practical Significance” to Establish a Prima Facie Case of Disparate Impact Discrimination*, 68 OKLA. L. REV. 641 (2016), <https://digitalcommons.law.ou.edu/olr/vol68/iss3/6>

This Note is brought to you for free and open access by University of Oklahoma College of Law Digital Commons. It has been accepted for inclusion in Oklahoma Law Review by an authorized editor of University of Oklahoma College of Law Digital Commons. For more information, please contact darinfox@ou.edu.

Lies, Damned Lies, and Statistics: The Case to Require “Practical Significance” to Establish a Prima Facie Case of Disparate Impact Discrimination

Introduction

From advertisements, to advice, to policy discussions, statistics surround us daily.¹ Four out of five dentists recommend Trident gum. The president’s approval rating has fallen by 5%. Smoking shortens a person’s life span by at least ten years. A total of 5.7% of Americans are currently unemployed. The list goes on and on.² “In the problem of . . . discrimination, statistics often tell much, and Courts listen.”³ Because these mathematical calculations provide a seemingly clear objective standard, statistics can be a powerful tool for litigants in discrimination cases. Their persuasive capabilities instill a sense of unbiased security that subjective judgment calls cannot provide.⁴ It is widely believed this unbiased standardized determination will bolster justice and pinpoint when discrimination has occurred.⁵ While—as the popular cliché goes—the road to hell is paved with good intentions, the differing approaches to law and the science of statistics unfortunately often lead to confusion and misunderstandings in application.⁶

1. Mikki Hebl, *Importance of Statistics*, in INTRODUCTION TO STATISTICS 13, 13 (David M. Lane ed., n.d.), http://onlinestatbook.com/Online_Statistics_Education.pdf.

2. *Id.* (providing examples of common claims that are statistical in character).

3. *Alabama v. United States*, 304 F.2d 583, 586 (5th Cir. 1962) (discussing the usefulness of statistics in the context of racial discrimination cases); see Kingsley R. Browne, *Statistical Proof of Discrimination: Beyond “Damned Lies”*, 68 WASH. L. REV. 477, 477-79 (1993) [hereinafter Browne, *Statistical Proof*].

4. See RAMONA L. PAETZOLD ET AL., THE STATISTICS OF DISCRIMINATION: USING STATISTICAL EVIDENCE IN DISCRIMINATION CASES § 2.04, at 2-14 (2006); Kent Spriggs, *Probative Value of Statistical Proof*, EMPLOYMENT LAW: THE BIG CASE 505, 512-13 (ALI-ABA Course of Study, Oct. 31 - Nov. 2, 1996); Note, *Beyond the Prima Facie Case in Employment Discrimination Law: Statistical Proof and Rebuttal*, 89 HARV. L. REV. 387 (1975). *But see* Browne, *Statistical Proof*, *supra* note 3, at 477 (noting the courts misplaced reliance in statistics, and that statistical proof is “based upon faulty statistical and factual assumptions”).

5. See *Watson v. Fort Worth Bank & Tr.*, 487 U.S. 977 (1988); *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299 (1977); *Int’l Bhd. of Teamsters v. United States*, 431 U.S. 324 (1977); *Isabel v. City of Memphis*, 404 F.3d 404 (6th Cir. 2005).

6. See PHILLIP DAWID, PROBABILITY AND STATISTICS IN THE LAW 1-2 (2006); see also Thomas J. Campbell, *Regression Analysis in Title VII Cases: Minimum Standards, Comparable Worth, and Other Issues Where Law and Statistics Meet*, 36 STAN. L. REV. 1299 (1984) (discussing the uneasy fit between the science of statistics and the law); D.H.

Specifically, plaintiffs establishing prima facie claims of disparate impact discrimination rely on statistical disparities as key evidence.⁷ Unlike disparate treatment, disparate impact theory is a method of proving discrimination based on a discriminatory policy or practice rather than proof of discriminatory motive. Because no clear test exists to determine when disparate impact has occurred, courts choose between two predominant methods: the four-fifths rule and statistical significance tests.⁸ Unfortunately, these methods often produce opposite results.⁹ The conflicting nature of the tests allows judges to choose whichever test supports their subjective opinion and equalizing view of the claim—leading to a skewed sense of justice and confusion among the circuits.¹⁰

The First Circuit's decision in *Jones v. City of Boston* creates further confusion in the already murky waters of disparate impact discrimination by rejecting federal employment agency guidelines. These guidelines require disparities sufficient to establish a prima facie case of disparate impact discrimination to be “practically significant.”¹¹ If “practically

Kaye, *Is Proof of Statistical Significance Relevant?*, 61 WASH. L. REV. 1333 (1986) (discussing the varying statistical tests used and the ambiguity of results); Ramona L. Paetzold, *Problems with Statistical Significance in Employment Discrimination Litigation*, 26 NEW ENG. L. REV. 395 (1991) (discussing the vulnerability of hypothesis testing to misuse and misinterpretation); Laurence H. Tribe, *Trial By Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329 (1971) (discussing the dangers of using mathematical methods in the legal process).

7. *Watson*, 487 U.S. at 991-92; *Teamsters*, 431 U.S. at 339; *Hazelwood*, 433 U.S. at 307-09.

8. See 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 (2008) (comparing the two methods and the varying outcomes based on which test is used); Jennifer L. Peresie, *Toward a Coherent Test for Disparate Impact Discrimination*, 84 IND. L.J. 773, 780-87 (2009) (explaining the four-fifths test and statistical significance tests).

9. Scott B. Morris, *Significance Tests and Confidence Intervals for the Adverse Impact Ratio*, ILL. INST. OF TECH. (Feb. 8, 2015, 4:15 PM), <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.218.7150&rep=rep1&type=pdf> (discussing the different operational definitions of the two methods which leads to their conflicting results); see also McKinley, *supra* note 8, at 171; Peresie, *supra* note 8, at 780-87.

10. Richard A. Primus, *Equal Protection and Disparate Impact: Round Three*, 117 HARV. L. REV. 494, 518-19 (2003) (discussing the different equalizing views of judges in evaluating impact cases and how such views are decisive in which test they choose to utilize); see PAETZOLD ET AL., *supra* note 4, at 2-13; Peresie, *supra* note 8, at 779.

11. *Jones v. City of Boston*, 752 F.3d 38, 48 (1st Cir. 2014) (finding that “Title VII does not require plaintiffs to prove that the observed differential is ‘practically significant’ in order to establish a prima facie case of disparate impact”).

significant,” the disparity is “sufficiently important substantively for the court to be concerned.”¹² The court in *Jones* held a 1% difference in hair follicle drug testing between white and black police officers sufficiently established a prima facie case of disparate impact discrimination.¹³ This decision—in conjunction with recent employment discrimination cases—detrimentally impacts employers. Companies may now have to invest extensive time and resources to determine whether their employment practices and procedures create even small deviations among hiring practices.

This Note argues for courts to hold “practical significance” as an essential element for plaintiffs establishing a prima facie case of disparate impact discrimination. Part I of this Note analyzes the purpose and history of disparate impact, the two tests most utilized to establish a prima facie case of disparate impact discrimination, and several courts’ recent trend of relying on statistical regression analysis to evaluate disparities. Part II explores the facts, holding, and rationale for the First Circuit’s rejection of “practical significance” in disparate impact cases. Part III argues the court’s decision was incorrect because it (1) misinterprets scientific language, (2) ignores the purpose of Title VII of the Civil Rights Act of 1964, (3) leads to scattered results, (4) fails to focus on the utility of maintaining a uniform, practical significance test, and (5) ignores the need for legislative action and clear, uniform judicial review on the matter. Part III additionally advocates for stronger deference toward EEOC guidelines and offers the 10% rule as an alternative method for courts to uniformly evaluate “practical significance.”

I. Law Before Jones v. City of Boston

A. Title VII Prohibition: Disparate Impact in a Nutshell

Disparate impact theory¹⁴ is one of the most important and controversial developments in employment discrimination law.¹⁵ Whereas “disparate

12. Daniel L. Rubinfeld, *Reference Guide on Multiple Regression*, in 1 MOD. SCI. EVIDENCE § 7.11 (2d ed. 2000). The EEOC actually suggests “practical significance” as a potential third alternative to statistical significance tests and the four-fifths rule “under which the court evaluates whether findings of statistical significance are ‘practically’ sound, rather than just ‘barely significant.’” *Id.*

13. *Jones*, 752 F.3d at 53.

14. Civil Rights Act of 1964, 42 U.S.C. § 2000e-1 to -15 (2012). Before 1971, it was widely believed that Title VII of the Civil Rights Act only applied to direct acts of discrimination. Peresie, *supra* note 8, at 776. For example, discrimination is clearly evident and actionable against employers who refuse to hire women or specific ethnic minority

treatment” is an intentional decision to treat a class of people differently based on race, age, or sex,¹⁶ “disparate impact” occurs when the treatment is unintentional, or there is a hidden intent.¹⁷ Disparate impact theory provides a way to prove employment discrimination based on the effects of an employment policy or practice rather than the motivation behind it.¹⁸ Such unintentional discrimination may result from a subtle, subconscious

groups. *Id.* The Supreme Court first established employer liability under Title VII in *Griggs v. Duke Power Co.*, finding defendant’s facially neutral policies and practices had an adverse impact on members of a protected class. 401 U.S. 424 (1971). Under the doctrine of disparate impact, a facially neutral employment practice is one that does not appear to be discriminatory on its face; rather it is one that is discriminatory in its application or effect. *See* Uniform Guideline on Employee Selection Procedures, 29 C.F.R. § 1607 (2008) (codifying the *Griggs* decision in the 1991 amendments).

15. Michael Selmi, *Was the Disparate Impact Theory a Mistake?*, 53 UCLA L. REV. 701, 708 (2006) (explaining the controversy surrounding disparate impact theory in anti-discrimination law). The doctrine is an important tool in proving discrimination where there is no evidence of overt bias or animus. *See* Mark S. Brodin, *Costs, Profits, and Equal Employment Opportunity*, 62 NOTRE DAME L. REV. 318, 358 (1987) (suggesting the Supreme Court established disparate impact theory in part based on the difficulty of proving intent); *see also* George Rutherglen, *Disparate Impact Under Title VII: An Objective Theory of Discrimination*, 73 VA. L. REV. 1297 (1987) (arguing that one justification for disparate impact theory is the difficulty of proving intent under disparate treatment models). It has been widely criticized, however, for its lack of overall impact in effecting change, and for the heavy burdens it can place on employers. Peresie, *supra* note 8, at 775. Under the disparate impact doctrine, courts have invalidated numerous employment practices, including written tests, physical tests, height and weight requirements, and subjective evaluation processes for having a disparate impact on a protected class without a business justification. *Id.* at 777 (citing *Fickling v. N.Y. State Dep’t of Civil Serv.*, 909 F. Supp. 185, 193 (S.D.N.Y. 1995) (concluding that written examination for welfare eligibility examiners had a racially disparate impact); *Brunet v. City of Columbus*, 642 F. Supp. 1214 (S.D. Ohio 1986); *Dothard v. Rawlinson*, 433 U.S. 321 (1977) (holding that a height and weight requirement for prison guards had a disparate impact on female applicants); *Stender v. Lucky Stores, Inc.*, 803 F. Supp. 259, 335-36 (N.D. Cal. 1992) (concluding that the employer’s “standard policy of discretionary, subjective and frequently unreviewed decision making with respect to initial placement, promotion and training” had a disparate impact on women).

16. 42 U.S.C. § 2000e-1 to -15. The statute read in full prohibits employment discrimination based on race, color, religion, sex, national origin, age, or disability.

17. 3 DAN BIDDLE, *ADVERSE IMPACT AND TEST VALIDATION: A PRACTITIONER’S GUIDE TO VALID AND DEFENSIBLE EMPLOYMENT TESTING* 2-5 (2d ed. 2006); *see also* Rutherglen, *supra* note 15, at 1300 (explaining how indirect acts are implied even though the statute nowhere says anything about indirect acts, only intentional).

18. BIDDLE, *supra* note 17, at 2; *see Griggs*, 401 U.S. at 429-36.

bias, or merely an ignorance of the effects one's employment practices have on workers from disadvantaged social origins.¹⁹

Disparate impact discrimination becomes actionable when employers apply identical standards or procedures uniformly despite substantial differences in employment outcomes for members of a particular minority group.²⁰ As in all Title VII cases, the burden of proof initially lies with the plaintiff to establish a prima facie case of disparate impact discrimination.²¹ To establish disparate impact discrimination, the plaintiff must (1) identify and isolate an employment practice implemented by the defendant, and (2) demonstrate the isolated employment practice or procedure "causes a disparate impact on the basis of race, color, religion, sex, or national origin"²² If plaintiff can meet these two requirements, the burden shifts to the employer to prove the disputed employment practice is a "business necessity."²³ If proven, the plaintiff may still prevail by offering an

19. Selmi, *supra* note 15, at 709.

20. *See id.* In other words, plaintiffs can prove disparate impact by demonstrating that a facially neutral employment practice or procedure has a disproportionately adverse effect on a protected group or class of people compared to other groups.

21. *See* 42 U.S.C. § 2000e-2(k)(1)(A).

22. *Id.* (with respect to the number of protected class members in the workforce compared to the protected class members in the relevant labor market). To establish causation, courts would ideally look at whether the challenged practice would create a disparate impact if exercised on the overall relevant labor market. *Watson v. Fort Worth Bank & Tr.*, 487 U.S. 977, 997 (1988) ("[P]roper comparison was between the racial composition of [the employer's] teaching staff and the racial composition of the qualified public school teacher population in the relevant labor market.") (quoting *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 308 (1977)). Because this data is usually unavailable, however, plaintiffs generally must rely on statistical disparities resulting from applicant data as a representation of the relevant population. *See Peresie, supra* note 8, at 775. This reliance on applicant data often leads to problems when there is not enough data to represent the relevant population, thereby deemphasizing apparent discrimination. *See Browne, Statistical Proof, supra* note 3, at 551. Conversely, it can lead to problems when there is too much data to represent the relevant population, overemphasizing even very small disparities. *See id.* at 550-51 (advocating the four-fifths rule as evaluating substantiality of the disparity because statistical significance tests can magnify even miniscule disparities); *see also* Arthur B. Smith, Jr. & Thomas G. Abram, *Quantitative Analysis and Proof of Employment Discrimination*, 1981 U. ILL. L. REV. 33, 53 ("The Court's emphasis in *Teamsters* and in *Hazelwood* on evidence of 'long-lasting and gross disparities' suggests that more is required in the evaluation of statistical proof of disparate impact than statistically significant disparities.").

23. 42 U.S.C. § 2000e-2(k)(1)(A); *see also* *Griggs v. Duke Power Co.*, 401 U.S. 424, 431-32 (1971); *see* Ernest F. Lidge III, *Financial Costs as a Defense to an Employment Discrimination Claim*, 58 ARK. L. REV. 1, 27-30 (2005).

alternative employment practice that accomplishes the business necessity but creates a lesser discriminatory effect than the current practice.²⁴

B. Which Test to Use?: How Courts Determine Disparities

The bare text of Title VII fails to indicate at what point a disparity becomes actionable.²⁵ The Supreme Court simply states a plaintiff must show “that the tests in question select applicants for hire or promotion in a . . . pattern significantly different from that of the pool of applicants.”²⁶ In *Watson v. Forth Worth Bank & Trust*, Justice O’Connor wrote, “Our formulations . . . have consistently stressed that statistical disparities must be *sufficiently substantial* that they raise such an inference of causation.”²⁷ Similarly, in *Hazelwood School District v. United States*, the Court held, “Where *gross statistical disparities* can be shown, they alone may in a proper case constitute prima facie proof of a pattern or practice of discrimination.”²⁸ Aside from these vague guidelines, no clear answer exists as to when a disparity becomes actionable.²⁹ Therefore, courts evaluate whether an employment practice causes a disparity using two primary methods: the four-fifths rule and statistical significance tests.³⁰

24. 42 U.S.C. § 2000e-2(k)(1)(A)-(C); see also Joel P. Rudin & Kathryn L. Glover, *Alternative Employment Practices: A Call to Arms*, 58 LAB. L.J. 39, 42, 45 (2007) (noting that there appears to be no basis in the law for employers to adopt these offered alternatives if proven acceptable); Peter E. Mahoney, *The End(s) of Disparate Impact: Doctrinal Reconstruction, Fair Housing and Lending Law, and the Antidiscrimination Principle*, 47 EMORY L.J. 409, 483-96 (1998) (discussing the lack of burden on the defendant to show that there is no less discriminatory alternative practice).

25. 42 U.S.C. § 2000e-2(k)(1)(A). The other inherent problem with the disparate impact doctrine is Congress’s failure to define “disparity.” There is no indication of what statistical showing is required to establish actionable disparate impact. *Jones v. City of Boston*, 752 F.3d 38, 49-50 (1st Cir. 2014).

26. *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 425 (1975).

27. 487 U.S. 977, 994-95 (emphasis added); see *id.* at 995 (rejecting a “rigid mathematical formula” for disparate impact).

28. 433 U.S. 299, 307-08 (1977) (emphasis added) (citing *Int’l Bhd. of Teamsters v. United States*, 431 U.S. 324, 339-40 n.20 (1977)); see also Browne, *Statistical Proof*, *supra* note 3, at 549-50 (discussing the *Hazelwood* holding and the need for gross disparities).

29. See Peresie, *supra* note 8, at 780-84 (noting the difficulties in proving causation).

30. Peresie, *supra* note 8, at 774. While these are the two primary methods, courts occasionally use other tests as well. *Richardson v. Lamar Cty. Bd. of Educ.*, 729 F. Supp. 806 (M.D. Ala. 1989). In *Richardson*, the court applied three different formulas in assessing the impact of early childhood education and elementary education tests. *Id.* at 816. In addition to the four-fifths rule and standard deviation analysis, the court applied the Shoben formula. *Id.* This formula measures the difference between independent proportions, referred

All of the federal employment agencies—including the U.S. Equal Employment Opportunity Commission (EEOC), the Office of Federal Contract Compliance Programs (OFCCP), the Department of Labor, and the Department of Justice in Title VII enforcement—have adopted the Uniform Guidelines of Employee Selection Procedures (UGESP) to provide a uniform set of principles governing use of employment practices.³¹ Recognizing the unlikelihood of a particular employment practice causing merely isolated small disparities, the UGESP advocates for a “practical significance” requirement in determining whether a plaintiff can establish a prima facie case of disparate impact discrimination.³² The Guidelines further suggest the four-fifths rule as clear evidence of disparate impact discrimination, while still recognizing smaller disparities may nevertheless constitute a disparate impact if both statistically and practically significant.³³

1. The Four-Fifths Rule

Historically, courts relied on the four-fifths rule as the most popular test for evaluating disparate impact.³⁴ The four-fifths rule finds a disparity actionable when one group’s selection rate—a group’s ability to successfully meet the criteria of the hiring or employment procedure—is

to as the *Z* value. *Id.* A *Z* value of 1.96 is considered the threshold for legal significance, making statistical and legal significance interchangeable under this approach. *Id.*

31. 29 C.F.R. § 1607.1(A) (2015). The UGESP, adopted by the EEOC and OFCCP, are a set of guidelines that incorporate a single set of principles which are designed to assist employers, labor organizations, employment agencies, and licensing and certification boards to comply with requirements of federal law prohibiting employment practices which discriminate on the grounds of race, color, religion, sex, and national origin. *See id.* § 1607.1. The UGESP is designed to assist employers, labor organizations, employment agencies, and licensing and certification boards to comply with requirements of federal employment discrimination law. *Id.* § 1607.1(B). It was created to be consistent with applicable legal standards and validation standards generally accepted by the psychological profession. *Id.* § 1607.1(C).

32. *Id.* § 1607.4(D).

33. *Id.*

34. BIDDLE, *supra* note 17, at 2-5. The test was originally framed by a panel of professionals called the Technical Advisory Committee on Testing and was later codified in the 1978 UGESP. *Id.* The EEOC—the federal agency charged with enforcing federal civil rights laws—adopted the four-fifths rule in the 1970s. *Id.* Since then, the Department of Labor, the Department of Justice, and the Office of Personnel Management (formerly known as the Civil Service Commission), have also relied on the UGESP and adopted the four-fifths rule for measuring disparate impact. *Id.*

less than four-fifths, or 80%, of another group's selection rate.³⁵ This "impact" of 80% or less demonstrates the existence of a "sufficiently substantial" disparity, meeting the requirement of "practical significance" outlined in the UGESP.³⁶

To demonstrate, suppose female applicants allege a physical examination requiring all firefighters to successfully perform one hundred consecutive pushups before hire creates a disparity between genders. 80% of men in the applicant pool meet this requirement compared to only 60% of females. Dividing the female selection rate by the male selection rate (0.6/0.8), we find the selection ratio of females to males is 75%. Therefore, the impact is less than 80%—making the gender disparity claim actionable under the four-fifths rule.³⁷

The advantages of the four-fifths rule are clear: it is an easy-to-calculate, simple test that puts responsible parties on notice of the relative balance an employer must achieve in its workforce to avoid liability.³⁸ It evaluates the impact of the disparity and ensures substantiality.³⁹ Several courts, however, criticize the four-fifths rule for creating mixed results.⁴⁰ Rather than directly addressing causation, many courts believe the four-fifths rule sets a seemingly arbitrary and unachievable evaluation of impact difficult for plaintiffs to meet.⁴¹ This heavy critique indicates employers may no longer rely on the four-fifths rule when evaluating disparities in their hiring practices. Instead, many courts now look to statistical significance tests.⁴²

35. 29 C.F.R. § 1607.4(D); see Peresie, *supra* note 8, at 781-84.

36. 29 C.F.R. § 1607.4(D). The committee chose the 80% rule as a guideline of convenience. If an employer hires a minority group at a rate lower than four-fifths of a majority group, the employer should be concerned that it was the employment procedure utilized causing a significant impact of more than 20%. See *id.*

37. See Peresie, *supra* note 8, at 775 (offering a similar illustration).

38. Paul Meier et al., *What Happened in Hazelwood: Statistics, Employment Discrimination, and the 80% Rule*, 1984 AM. B. FOUND. RES. J. 139, 166-70 (praising the four-fifths rule).

39. *Watson v. Fort Worth Bank & Tr.*, 487 U.S. 977, 979 (1988) (providing the "sufficiently substantial" language).

40. McKinley, *supra* note 8, at 182-85 (discussing the circuit split).

41. See McKinley, *supra* note 8, at 188-99; Peresie, *supra* note 8, at 782; see, e.g., Anthony E. Boardman, *Another Analysis of the EEOC "Four-Fifths" Rule*, 25 MGMT. SCI. 770, 770-76 (1979); Elaine W. Shoben, *Differential Pass-Fail Rates in Employment Testing: Statistical Proof Under Title VII*, 91 HARV. L. REV. 793, 805-12 (1978).

42. Coupled with the EEOC's recognition that statistical significance tests are a valuable alternative to the four-fifths rule when the data used is either too small or too large, the *Hazelwood* decision ignited a wave of courts disfavoring the four-fifths rule to look to statistical significance test as a more precise means of evaluating disparities. 433 U.S. 299

2. Statistical Significance Tests

While a wide variety of statistical significance tests exist, each serves the purpose of calculating the overall likelihood an observed disparity occurred due to random chance.⁴³ Using this method, a disparity becomes actionable when one can be confident at a designated level—typically 95%—the disparity observed is not attributable to random chance.⁴⁴ In other words, a plaintiff may establish a prima facie case of disparate impact discrimination if the test indicates at least a 5% disparity—approximately two standard deviations—in the relevant labor market, meaning a roughly one-in-twenty chance the disparity occurred at random.⁴⁵

Like the four-fifths rule, several courts heavily criticize statistical significance tests and the use of statistical regression analysis for similar varied and ambiguous results.⁴⁶ First, courts disagree over whether statistical significance tests directly evaluate causation. Many experts contend evaluation of chance—statistical significance—relies just as much on inference as evaluation of impact—the four-fifths rule.⁴⁷ Second,

(1977); Peresie, *supra* note 8, at 786; *see also* Jones v. City of Boston, 752 F.3d 38, 46-47 (1st Cir. 2014) (accepting 1.96 standard deviations as sufficient to establish a prima facie case); Chin v. Port Auth. of N.Y. & N.J., 685 F.3d 135, 145, 153-54 (2d Cir. 2012) (“[S]tatistical significance at the five-percent level is generally sufficient”) (internal quotations omitted) (citing Smith v. Xerox Corp., 196 F.3d 358, 366 (2d Cir 1999)); Stagi v. Nat’l R.R. Passenger Corp., 391 F. App’x 133, 140 (3d Cir. 2010) (holding that the threshold is generally a “probability level at or below 0.05, or at [two] to [three] standard deviations or greater”); Paige v. California, 233 F. App’x 646, 648 (9th Cir. 2007) (accepting 1.96 standard deviations as the threshold for statistical significance).

43. Peresie, *supra* note 8, at 785 (“Statistical significance tests come in various technical forms, including multiple regressions, t-tests, Z-tests, the chi-square test, and the Fisher exact test,” but they all serve the same function.)

44. *Id.* at 774; *see* PAETZOLD ET AL., *supra* note 4, § 2.4 (this percentage varies).

45. Peresie, *supra* note 8, at 786; *see also* McKinley, *supra* note 8, at 188-96 (explaining the science behind the statistical significance tests).

46. *See* Kingsley R. Browne, *The Strangely Persistent “Transposition Fallacy”: Why “Statistically Significant” Evidence of Discrimination May Not Be Significant*, 14 LAB. LAW. 437 (1998) [hereinafter Browne, “*Transposition Fallacy*”]; *see also* Browne, *Statistical Proof*, *supra* note 3, at 556-58 (discussing the weaknesses of statistical significance in evaluating disparate impact); Kaye, *supra* note 6 (criticizing the wide variety of statistical significance tests available); Paetzold, *supra* note 6, at 411 (criticizing the lack of awareness in courtrooms as to how statistical significance works).

47. Browne, “*Transposition Fallacy*”, *supra* note 46, at 449 (“[E]stablishing a significance level of .05 does not mean the law is demanding a 95% certainty of discrimination. . . . ‘The court’s assumption . . . that when the “probability of statistical error is less than 5%,” the “scientific fact is at least 95% certain” exemplifies a common misunderstanding of the role of statistical tests in statistical inference.’”) (quoting David H.

statistical significance tests vary wildly from one another.⁴⁸ Furthering the problem, there is no established threshold of statistical significance.⁴⁹ While most courts choose a confidence level of 95%, nothing restricts courts to this level, making it an imprecise standard.⁵⁰ Third, judges are not equipped with the tools to evaluate the scientific nature of such tests.⁵¹ The variety of tests available allows litigants to act as amateur statisticians and manipulate numbers.⁵² Consequently, judges often subconsciously favor whichever test matches their subjective opinions.⁵³ Finally, these tests are extremely sensitive to sample size; the larger the number of applicants, the greater chance the data will magnify even miniscule disparities.⁵⁴

C. Consequences of Multiple Tests: The Circuit Split

Because the four-fifths rule and statistical significance tests have different operative functions—one evaluating impact, the other evaluating chance—they often lead to conflicting results.⁵⁵ With no direction other than ambiguous guidance provided by the Supreme Court that actionable disparities be “gross”⁵⁶ or “sufficiently substantial,” lower courts freely choose whichever test allows their preferred party to prevail, often ignoring

Kaye, *Statistical Significance and the Burden of Persuasion*, 46 LAW & CONTEMP. PROBS., Autumn 1983, at 13, 21-22).

48. See *Richardson v. Lamar Cty. Bd. Of Educ.*, 729 F. Supp. 806, 816 (1989); *supra* note 43 and accompanying text.

49. Louis J. Braun, *Statistics and the Law: Hypothesis Testing and Its Application to Title VII Cases*, 32 HASTINGS L.J. 59, 69-70 (1980) (examining error that inevitably results when using statistical significance tests); John M. Dawson, *Are Statisticians Being Fair to Employment Discrimination Plaintiffs?*, 21 JURIMETRICS J. 1, 2 (1980).

50. *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 311 n.17 (1977) (noting that a disparity is statistically significant where it is more than two or three standard deviations from the expected rates, but this standard is far from precise).

51. Browne, “*Transposition Fallacy*”, *supra* note 46, at 451-55.

52. See Peresie, *supra* note 8, at 778.

53. See Primus, *supra* note 10, at 518-19; see DAWID, *supra* note 6, at 89-90.

54. See Paetzold, *supra* note 6, at 402 (explaining how large sample sizes have the “effect . . . of increasing the power of a magnifying glass being used to detect the difference or deviation from the hypothesis that is assumed to be true. For a large enough sample, then, even trivial deviations . . . will tend to appear ‘significant.’”). *But see* *Jones v. City of Boston*, 752 F.3d 38, 50 (1st Cir. 2014); *Stagi v. Nat’l R.R. Passenger Corp.*, 391 F. App’x 133, 144 (3d Cir. 2010) (advocating statistical significance testing).

55. See Morris, *supra* note 9, at 89-90.

56. See Meier et al., *supra* note 38, at 140.

the UGESP standards employers, labor organizations, and many others rely upon.⁵⁷

Some circuits continue to back the four-fifths rule as the proper method for evaluating causation under disparate impact theory, finding it a reliable indicator of “practical significance.”⁵⁸ In 2012, the Tenth Circuit held that small differences in selection rates were not practically significant, even though, because of the large number of selections, a disparity in excess of four-and-a-half standard deviations occurred—more than twice the threshold needed for statistical significance.⁵⁹ Likewise, the Seventh Circuit continues its reliance on the four-fifths rule, holding the 5% standard utilized by the majority of statistical significance tests “arbitrary.”⁶⁰ According to the Seventh Circuit, it is the judge's role—relying on evidence examined by a trained statistician—to determine whether a particular significance level is too low to make the study worthy of consideration by the fact finder.⁶¹

The Supreme Court's decisions in *Castaneda v. Partida*⁶² and *Hazelwood School District v. United States* prompted a shift toward judicial reliance on statistical significance tests when evaluating causation under disparate impact theory.⁶³ Because such tests create “[a] precise method of measuring the significance of . . . disparities,”⁶⁴ the Court held use of these seemingly decisive assessments proper. When examining large sample sizes specifically, the Court noted an inference of discrimination generally arises

57. See Peresie, *supra* note 8, at 777-78, 782-83; Primus, *supra* note 10, at 518; Jones, 752 F.3d at 50.

58. Meier et al., *supra* note 38, at 168.

59. *Apsley v. Boeing Co.*, 691 F.3d 1184, 1196, 1206-07 (10th Cir. 2012). *But see Tabor v. Hilti, Inc.*, 703 F.3d 1206, 1223, 1226 (10th Cir. 2013) (describing a statistical significance threshold of “two or three standard deviations”).

60. *Kadas v. MCI Systemhouse Corp.*, 255 F.3d 359, 361-62 (7th Cir. 2001) (noting that the 5% threshold is influenced by “the fact that scholarly publishers have limited space and don't want to clog up their journals and books with statistical findings that have a substantial probability of being a product of chance rather than of some interesting underlying relation between the variables of concern”).

61. *Id.* at 362-63.

62. *Castaneda v. Partida*, 430 U.S. 482, 497 n.17 (1977) (adopting a two or three standard deviation standard for a case involving racial discrimination in jury selection).

63. *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 308-09 n.14 (1977).

64. *Id.* at 308-09 n.14 (referencing *Castaneda v. Partida*, 430 U.S. 482, 496-97 n.17 (1977)); see also Peresie, *supra* note 8, at 786 (describing the shift toward statistical significance).

when the disparity is “greater than two or three standard deviations.”⁶⁵ Coupled with the EEOC’s recognition of the value of statistical significance tests over the four-fifths rule when dealing with large or small amounts of data, the *Castaneda* and *Hazelwood* decisions struck a chord with many members of the judiciary. Now, several courts shun the four-fifths rule—looking instead to statistical significance tests as a more exact means of evaluating disparities.⁶⁶

Despite the Court’s disclaimer that it did not intend computations of “two or three standard deviations” required as strict procedure, many lower courts have done just so.⁶⁷ In *Moultrie v. Martin*, the Fourth Circuit held, “[I]n all cases involving racial discrimination, the courts of this circuit must apply a standard deviation analysis such as that approved by the Supreme Court in *Hazelwood* before drawing conclusions from statistical comparisons.”⁶⁸ While other circuits have declined to follow such an extreme conclusion, many plaintiffs successfully rely on these decisions to support utilization of “statistical significance” and statistical regression analysis as primary methods for establishing or disestablishing disparities.⁶⁹

65. *Castaneda*, 430 U.S. at 496-97 n.17 (noting that finding a disparity “greater than two or three standard deviations” random would be “suspect to a social scientist”); *Hazelwood*, 433 U.S. at 308-09 n.14 (relying on *Castaneda* to support finding that a disparity “greater than two or three standard deviations” was evidence of racial disparate impact discrimination); see also Kaye, *supra* note 6, at 1335.

66. Peresie, *supra* note 8, at 786; see Adoption of Questions and Answers to Clarify and Provide a Common Interpretation of the Uniform Guidelines on Employee Selection Procedures, 44 Fed. Reg. 11,996, 11,998-11,999 (Mar. 2, 1979) (acknowledging the use of statistical significance tests in large selection numbers and also noting “[w]here the sample of persons selected is not large, even a large real difference between groups is likely not to be confirmed by a test of statistical significance”).

67. Kaye, *supra* note 6, at 1335. *But see* Jones v. City of Boston, No. 05-11832-Gao, 2012 WL 4530594, at *2 (D. Mass. Sept. 28, 2012) (“The passage does not mean what the plaintiffs try to make of it. The Court was merely acknowledging the general convention used by social scientists. In its holding the Court then affirmed the finding of a prima facie case of discrimination based on evidence *twelve* standard deviations, not two or three.”), *rev’d*, Jones v. City of Boston, 752 F.3d 38 (1st Cir. 2014).

68. 690 F.2d 1078, 1082 (4th Cir. 1982) (“When a litigant seeks to prove his point exclusively through the use of statistics, he is borrowing the principles of another discipline, mathematics . . . [He] cannot be selective in which principles are applied. He must employ a standard mathematical analysis. Any other requirement defies logic to the point of being unjust. Statisticians do not simply look at two statistics . . . and make a subjective conclusion that the statistics are significantly different. Rather, statisticians compare figures through an objective process known as hypothesis testing.”); see also Kaye, *supra* note 6, at 1335.

69. See Kaye, *supra* note 6, at 1335; *supra* note 43 and accompanying text.

The Third Circuit explicitly rejected the four-fifths rule in 2010.⁷⁰ Regardless of the size of the difference between groups' outcomes, the Third Circuit found demonstrations of statistically significant differences sufficient to establish disparate impact.⁷¹ Like the First Circuit in *Jones*,⁷² the Third Circuit held a plaintiff's demonstration of statistically significant disparity sufficient to establish a prima facie case of disparate impact discrimination.⁷³ In addition, many other circuits similarly hold statistical significance tests sufficient for demonstrating a legally significant adverse impact, even when the data fails the four-fifths guideline.⁷⁴

The conflicting results of these two methods allow well-meaning judges to decide claims based on preference rather than any real, objective standard—often leading to varied and conflicting precedents among the circuits.⁷⁵ The First Circuit in *Jones* further complicates this problem not only by outright rejecting the four-fifths rule as a valid alternative to the statistical significance tests but also by rejecting the “practical significance” requirement on the speculative premise the difficulty in its application outweighs any of its benefits.⁷⁶

II. Statement of the Case

A. Facts

Ten black plaintiffs—all who alleged to have suffered adverse employment actions after testing positive for cocaine via a drug hair test—

70. *Stagi v. Nat'l R.R. Passenger Corp.*, 391 F. App'x 133, 144-45 (3d Cir. 2010).

71. *Id.*

72. *Jones v. City of Boston*, 752 F.3d 38, 53 (1st Cir. 2014).

73. *Stagi*, 391 F. App'x at 144-45.

74. *See id.* at 144; *see also Adams v. Ameritech Servs., Inc.*, 231 F.3d 414, 424 (7th Cir. 2000) (“Two standard deviations is normally enough to show that it is extremely unlikely (that is, there is less than a 5% probability) that the disparity is due to chance, giving rise to a reasonable inference that the hiring was not race-neutral”); *Anderson v. Zubieta*, 180 F.3d 329, 33-41 (D.C. Cir. 1999) (indicating that “disparities . . . exceed[ing] 1.96 standard deviations under a two-tailed test of statistical significance” are sufficient to establish a prima facie case of disparate impact). *But see Kaye, supra* note 6, at 1335 n.21 (criticizing the misplaced reliance in standard deviation analysis).

75. *See Peresie, supra* note 8, at 779 (noting “that the rule should be used only ‘to the extent that [it is] useful . . . for advancing the basic purposes of Title VII,’” demonstrating that under this ambiguous standard judges may be using the statistical tests to further their own views depending on what they believe advances the “purposes of Title VII” entail) (quoting *Isabel v. City of Memphis*, No. 01-2533 ML/BRE, 2003 WL 23849732, at *3 n.5 (W.D. Tenn. Feb. 21, 2003)), *aff'd*, 404 F.3d 404 (6th Cir. 2005)).

76. *Jones v. City of Boston*, 752 F.3d 38, 50-53 (1st Cir. 2014).

came forward to challenge the Boston Police Department's (Department) drug testing program.⁷⁷ "[E]ven though over two-thirds of the officers and cadets tested were white," a higher number of black employees than white employees tested positive for cocaine.⁷⁸ The plaintiffs denied any cocaine use, asserting "the hair test employed by the department generated false-positive results in processing the type of hair common to many black individuals."⁷⁹ While the plaintiffs did not object to the business need for drug screenings,⁸⁰ they objected to the use of hair samples as the method of screening.⁸¹ The plaintiffs claimed this employment practice had a disparate impact on black officers in violation of Title VII.⁸²

B. Procedural History, Issue, and Holding

At the district court level, the Department filed a motion for summary judgment regarding the Title VII claim, arguing that a statistical disparity of 1% was insufficient to establish a prima facie case of disparate impact discrimination.⁸³ The district court agreed and entered summary judgment for the defendant on all claims.⁸⁴ The court held a "statistically significant imbalance does not automatically constitute disparate impact where practical significance is lacking."⁸⁵ In the aggregate and in each individual disputed year, blacks consistently passed the test at a rate of 97% or higher

77. *Id.* at 41.

78. *Id.* at 42-44. "During the eight years for which the plaintiffs present data, black officers and cadets tested positive for cocaine approximately 1.3% of the time." *Id.* at 41. Conversely, "white officers and cadets tested positive just under 0.3% of the time." *Id.*

79. *Id.* Plaintiffs asserted that cocaine easily binds to the higher amounts of melanin present in black individuals' hair, generating higher amounts of false positives than among white officers. *Id.*

80. *Id.* at 54.

81. *Id.* at 45 (arguing that the federal government does not use hair tests in drug screenings and that these tests are unreliable). Notably, the plaintiffs relied "on evidence that showed differences in the rates at which African Americans *failed*, rather than *passed* the hair test, were statistically significant to the extent of between two to four standard deviations." *Jones v. City of Boston*, No. 05-11832-Gao, 2012 WL 4530594, at *2 (D. Mass. Sept 28, 2012).

82. *See Jones*, 752 F.3d at 41; Civil Rights Act of 1964, 42 U.S.C. § 2000e-1 to -15 (2012).

83. *Jones v. City of Boston*, No. 05-11-832-Gao, 2012 WL 4530594, at *1-2 (D. Mass. Sept. 28, 2012) (noting that "the passing rate for African Americans was at least 97% of the passing rate for whites").

84. *Id.* at *6.

85. *Jones*, 752 F.3d at 49 (discussing the district court's reliance on the four-fifths rule); *see Jones*, 2012 WL 4530594, at *1-3.

compared to whites.⁸⁶ Because the statistical deviation did not meet the four-fifths “rule of thumb” relied on by employers, the court found the disparity not “practically significant.”⁸⁷ Therefore, the plaintiffs failed to offer sufficient evidence establishing a prima facie case of disparate impact discrimination.⁸⁸

On appeal, the First Circuit took a radically different approach in addressing the Title VII claim.⁸⁹ Examining the case de novo, the court chose to highlight the flaws of the four-fifths rule and ultimately rejected the rule as a proper method to establish a prima facie case of disparate impact discrimination.⁹⁰ With no other means of determining or predictably applying “practical significance,” the court then rejected this EEOC-advocated requirement as well.⁹¹

C. Decision

Beginning its analysis with an overview of the disparate impact doctrine, the First Circuit emphasized the steps necessary to establish a prima facie case of disparate impact discrimination.⁹² First, a plaintiff “isolat[es] and identif[ies]” the challenged employment practice.⁹³ Second, a plaintiff must show the identified practice “causes a disparate impact on the basis of race.”⁹⁴ The Department did not dispute hair testing as a “particular employment practice.”⁹⁵ Additionally, relying on the EEOC-advocated “practical significance” requirement,⁹⁶ the Department believed these disparities too small in size to be disconcerting. Therefore, it did not willfully challenge the raw math of the calculations indicating the

86. *Jones*, 752 F.3d 44-45 (showing the data calculations for each year); *Jones*, 2012 WL 4530594, at *2.

87. *Jones*, 752 F.3d at 48-49; see *Jones*, 2012 WL 4530594, at *2-3.

88. *Jones*, 2012 WL 4530594, at *3.

89. *Jones*, 752 F.3d at 46-55.

90. *Id.* at 46, 49-53.

91. See *id.* at 52 (finding that “as a matter of theory . . . [we cannot] expect to find any single measure of the size of the impact to determine . . . practical significance”); see also *id.* at 50 (acknowledging an argument against adopting a practical significance requirement because “the concept of practical significance is impossible to define in even a remotely precise manner”).

92. *Id.* at 46.

93. *Id.* (quoting *Watson v. Fort Worth Bank & Tr.*, 487 U.S. 977, 944 (1988)).

94. *Id.* (quoting 42 U.S.C. § 2000e-2 (k)(1)(A)(i) (2012)).

95. *Id.* (quoting 42 U.S.C. § 2000e-2 (k)(1)(A)(i)).

96. *Jones*, 752 F.3d at 48-49.

employment practice resulted in a statistically significant correlation with race.⁹⁷

The court attacked the Department's assertion "that even a statistically significant racial skew in outcomes does not constitute a disparate impact unless the racial differential is also sufficiently large, or 'practically significant.'"⁹⁸ The First Circuit stated, "Title VII does not require plaintiffs to prove that the observed differential [disparity] is 'practically significant' in order to establish a prima facie case of disparate impact."⁹⁹ To support this decree, the court focused on three main points: (1) the inherent flaws of the four-fifths test as a measure of "practical significance,"¹⁰⁰ (2) the ambiguity of the term "disparity" and the difficulty in defining and applying "practical significance,"¹⁰¹ and (3) the adequacy of subsequent requirements to prove disparate impact as safeguards to avoid impractical claims.¹⁰²

First, the court dismissed the Department's reliance on the four-fifths rule as a proper measure for determining disparities.¹⁰³ Although the court recognized the important needs served by the four-fifths rule—including its use in guiding exercise of agency discretion and as a crucial guideline for employers wanting to avoid liability—it found this utility did not justify plaintiffs' use of the rule to establish a prima facie case of disparate impact

97. *See id.* at 47-53. "As their threshold for statistical significance, the plaintiffs chose a p-value of five percent, or 1.96 standard deviations, the threshold most commonly used by social scientists" and other federal courts utilizing statistical significance tests. *Id.* at 46-47. Using this threshold selected by the plaintiffs, the court found that in three of the eight years of the relevant period, the difference in pass rates for black and white employees was statistically significant. *Id.* at 47. The court also found that when aggregating together the eight years in dispute, the disparity resulted in "more than seven standard deviations from the expected norm." *Id.* For an explanation of p-values and other statistical significance tests, see generally David H. Kaye & David A. Freedman, *Reference Guide on Statistics*, in FED. JUDICIAL CTR., REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 83 (2d ed. 2000), [http://www.fjc.gov/public/pdf.nsf/lookup/sciman00.pdf/\\$file/sciman00.pdf](http://www.fjc.gov/public/pdf.nsf/lookup/sciman00.pdf/$file/sciman00.pdf).

98. *Id.* at 48-53.

99. *Id.*

100. *Id.* at 51-52.

101. *Id.*

102. *Id.* at 53 ("Our confidence in rejecting a practical significance requirement is bolstered by the . . . two other requirements to be met by a plaintiff in a Title VII disparate impact case [T]he statute as designed by Congress effectively assigns case-specific practical significance to the size of the impact").

103. *Id.* at 52.

discrimination.¹⁰⁴ The court supported its rejection of the four-fifths rule by relying on a previous First Circuit decision to reject the four-fifths rule where a small sample size precluded a showing of statistical significance.¹⁰⁵ Comparing the analysis of small sample size to large sample size, the First Circuit found similar analysis applied and also precluded a showing of statistical significance.¹⁰⁶

Second, having rejected the four-fifths rule and without any other standard to evaluate “practical significance” in an objective or precise manner, the court refused to adopt the UGESP’s “practical significance” requirement.¹⁰⁷ The court found the “practical significance” label lacked objectivity and allowed the person applying the requirement to subjectively determine whether the disparity is substantial enough to create liability.¹⁰⁸ “Courts would find it difficult to apply such an elusive, know-it-when-you-see-it standard, let alone instruct a jury on how to do so, and parties may find it impossible to predict results.”¹⁰⁹ Noting the text of Title VII’s failure to define “disparate,” the court concluded the term could simply mean nonrandom, or “different,” and rejected the “practical significance” requirement.¹¹⁰

Third, the First Circuit looked to the shifting burden of proof to support its rejection of “practical significance” and the four-fifths rule.¹¹¹ According to the court, these subsequent steps act as additional safeguards to “indirectly secure most of the advantages that might be gained were it possible to fashion a principled and predictable direct test of practical significance.”¹¹² Congress, the court noted, designed these safeguards to

104. *Id.* (“The rule itself has some practical utility. There is simply nothing in that utility, however, to justify affording decisive weight to the rule to negate or establish proof of disparate impact in a Title VII case.”).

105. *Id.* at 51. The court relied on *Fudge v. City of Providence Fire Department* to support its rejection of the four-fifths rule. 766 F.2d 650 (1st Cir. 1986). In *Fudge*, the First Circuit found that even though the acceptance rates of blacks was “well below the ‘four-fifths’ rate established by the EEOC . . . [w]here the sample size is small . . . the ‘four-fifths rule’ is not an accurate test of discriminatory impact.” 766 F.2d at 659 n.10; see Shoben, *supra* note 41, at 806-10 (finding that the four-fifths rule creates problems in both small sample sizes and large sample sizes).

106. *Jones*, 752 F.3d at 52.

107. *Id.* at 52-53.

108. *Id.* at 50-51.

109. *Id.*

110. *Id.* at 49-50 (citing WEBSTER’S NEW COLLEGIATE DICTIONARY 329 (8th ed. 1977); MERRIAM-WEBSTER’S COLLEGIATE DICTIONARY 360 (11th ed. 2003)).

111. *Id.* at 53.

112. *Id.*

effectively assign case-specific practical significance as to the size of the impact.¹¹³ Extensive data is required to prove an alternative practice would not have a similar effect to the current employment practice, and “as the size of the impact increases, so too does the ease of demonstrating an alternative practice that reduces the impact.”¹¹⁴ Furthermore, the court noted that if the defendant cannot establish the practice as a “business necessity,” there is no need to retain the practice merely because the size of those affected is small.¹¹⁵

Having rejected the Department’s objections to the plaintiff’s use of a statistical significance test to demonstrate a statistical disparity, the First Circuit overturned summary judgment on the matter.¹¹⁶ The court declined to consider whether the Department could establish the employment practice as a “business necessity” and remanded the case back to the district court for further proceedings.¹¹⁷

III. Analysis

The First Circuit’s decision in *Jones* is incorrect for several reasons. First, the court placed an unfound reliance in statistics by obscuring the meaning of the phrase “statistically significant.” Second, by illustrating ways the test clearly does not accomplish the goals of combatting unintentional discrimination, the court wrongly rejected the practical significance test because, like the four-fifths rule, there will always be instances of arbitrary application if applied uniformly. The court placed too much weight on the equivocal results that can occur when applying the four-fifths rule instead of focusing on the utility of maintaining a uniform practical significance test. This decision to boldly reject the EEOC’s adopted guidelines creates further confusion among circuits and employers in an already ambiguous area of law. Third, the *Jones* decision emphasizes a growing problem with the arbitrary use of statistical significance tests.

113. *Id.*

114. *Id.* Cases involving small impacts require large sample sizes. Because this data is usually unavailable, there will generally not be enough data to prove a statistical significance.

115. *Id.* In cases where data is available, the employer may justify their employment practice by proving it is a “business necessity.” Once demonstrated to be a business necessity, a plaintiff can only prevail if they prove the existence of an alternative practice that satisfies the department’s legitimate business needs and does not have a similarly “undesirable racial effect.”

116. *Id.*

117. *Id.* at 54-55, 60.

Courts should defer to the EEOC when reasonable and require “practical significance” as a more coherent method of evaluating disparities.

A. Mixing Science and the Law

It is a frequently voiced suspicion that statistics can prove anything;¹¹⁸ as lamented by Mark Twain, “Facts are stubborn, but statistics are more pliable.”¹¹⁹ The First Circuit’s decision to reject a “practical significance” requirement in favor of “statistical significance” places far too much faith in the accuracy of statistical regression analysis.¹²⁰ The science of statistics does not mesh well with the application of law, and courts routinely make substantial errors when interpreting statistical evidence.¹²¹ This misinterpretation leads to varying legal precedents and confusion for litigants.¹²²

1. One Term Fits All? The Ill Fit Between Scientific Terms and Legal Jargon

From words like “average”¹²³ to “proof” to “theory,” scientists grit their teeth at the way many ideas have left the world of science and infected the

118. Kaye, *supra* note 6, at 1334; *see, e.g.*, EEOC v. Fed. Reserve Bank, 698 F.2d 633, 645-46 (4th Cir. 1983), *rev’d on other grounds sub nom.*, Cooper v. Fed. Reserve Bank, 467 U.S. 867 (1984).

119. Mark Twain, *Mark Twain Quotes*, BRAINY QUOTE, <http://www.brainyquote.com/quotes/quotes/m/marktwain163414.html> (last visited Sept. 7, 2015) (discussing customer analytics); *see also* 1 SAMUEL LANGHORNE CLEMENS, MARK TWAIN’S AUTOBIOGRAPHY 246 (1924) (recounting Mark Twain’s disdain for statistics).

120. *See supra* note 6 and accompanying text.

121. *See supra* note 47 and accompanying text.

122. *See Meier et al., supra* note 38, at 185; Peresie, *supra* note 8, at 786-87.

123. For example, the colloquial “average” means a “typical amount.” The mathematical term “average” however refers to the mathematical mean: the number you could use in place of each of these values, and still have the same sum. You find the mean by adding the numbers and dividing them by how many there are. Deb Russell, *The Mean, the Median, and the Mode*, ABOUT EDUCATION, <http://math.about.com/od/statistics/a/MeanMedian.htm> (last visited Oct. 17, 2015). Imagine a small town has a population of roughly 10,000 people. The average annual income for this town is \$100,000. Based on this statistic alone, many would believe this to be an overall financially stable community, because in everyday language, the term “average” means “most people.” When one actually examines the numbers however, it reveals 2000 centrally located individuals receive a generous annual income of \$420,000 each per year, while the remaining 8000 individuals receive an annual income of just \$20,000 a year. In this over-simplified scenario, 80% of employees earn only one-fifth of the “average” income, demonstrating the problems and misconceptions that can arise when we begin to mix the popular use of the term “average” with its mathematical term. *See also* James E. Owers, *Court Determines Average Weekly Wage Calculation*, N.H.

everyday language of a public who usually gets their meaning wrong.¹²⁴ “Statistically significant” is one of those phrases scientists would love to take back and rename.¹²⁵ Using our knowledge of everyday language, most people would say a “significant” difference indicates an important difference. In scientific terms, however, a difference is “statistically significant” if it is unlikely to have occurred at random.¹²⁶ While the colloquial “significance” suggests importance, the scientific test of “statistical significance” does not measure the value of such a difference; it only measures whether it can be distinguished, using the most advanced statistical tools, from a perfect, imaginary model.¹²⁷ While seemingly important, scientists often determine “statistically significant” differences to be unimportant and factually insignificant.¹²⁸

While the First Circuit properly defines “statistical significance” as a non-random disparity, it improperly attributes *legal* significance to the term—a significance scientists and statisticians did not intend it to have.¹²⁹ Just because it is highly unlikely a disparity occurred at random does not mean the employment practice caused the disparity or that it is one courts should be concerned about.¹³⁰ As stated in the Federal Judicial Center (FJC) Manual of Statistics, “[Statistically] significant differences are evidence that something besides random error is at work, but they are not evidence that this ‘something’ is legally or practically important.”¹³¹ Despite the court’s attempts to attach an improper meaning to the term, it is clear

EMP. L. LETTER, Jan. 1, 2001, at 2 (discussing the problems that can arise when the wrong data is used to calculate “average” weekly compensation).

124. See Tia Ghose, *‘Just a Theory’: 7 Misused Science Words*, LIVE SCI. (Apr. 1, 2013, 5:52 PM), <http://www.livescience.com/28347-most-misused-science-words.html>.

125. See JORDAN ELLENBERG, *HOW NOT TO BE WRONG: THE POWER OF MATHEMATICAL THINKING* 21-30 (2014).

126. *Id.*

127. *Id.*

128. *Id.*; see also Kaye, *supra* note 6, at 1345-49.

129. See Kaye & Freedman, *supra* note 96, at 124; see also Allan G. King, “*Gross Statistical Disparities*” as Evidence of a Pattern and Practice of Discrimination: *Statistical Versus Legal Significance*, 22 LAB. LAW. 271 (2007) (noting that statistical significance does not indicate legal significance).

130. See Browne, *Statistical Proof*, *supra* note 3; Browne, “*Transposition Fallacy*”, *supra* note 46; Campbell, *supra* note 6, at 1334; Kaye, *supra* note 6, at 1334; Tribe, *supra* note 6, at 1329.

131. See Kaye & Freedman, *supra* note 96, at 124.

statisticians agree that “[w]hen practical significance is lacking . . . there is no reason to worry about statistical significance.”¹³²

2. *What Happens When You Assume: Problems with Scientific Assumption*

Because a statistically significant disparity is not meaningful on its own, it is disconcertingly incapable of accurately depicting legal burdens of proof.¹³³ Statistical models are not perfect pictures of reality; they are approximations, nothing more.¹³⁴ Statisticians—who rely on a series of assumptions that would concern a lawyer if he were more keenly aware of them—do not consider in their calculations the obligation of attorneys to prove an assertion by a preponderance of the evidence. Conversely, lawyers—who do not have to deal with the statistical concept of a “null hypothesis”—often phrase inappropriate questions to statisticians without fully considering this reality.¹³⁵

Statistics derived from inappropriate models provide useless answers.¹³⁶ By holding any non-random disparity—no matter how insignificant—to be an actionable disparity, the First Circuit ignores the functional reality of statistics.¹³⁷ As stated by one exasperated statistician, “No statistician or other scientist should ever put himself/herself in a position of trying to prove or disprove discrimination.”¹³⁸ The trier of fact is to resolve disputed factual questions as best as it can, and the decision-making task should not be delegated to statisticians and other experts by trusting “superficially impressive methods whose seeming objectivity does not withstand analysis.”¹³⁹ Judges and juries often will not recognize that an expert’s reference to statistical proof as “highly significant” does not necessarily indicate that a substantial effect exists, which leads to a dangerous and misplaced reliance that outweighs the limited value of this testimony.¹⁴⁰

132. *Id.*; *Jones v. City of Boston*, 752 F.3d 38, 48 (1st Cir. 2014) (finding if not sufficiently important, statisticians agree there is no reason to worry about statistical significance).

133. DAWID, *supra* note 6, at 89-90; *see also* PAETZOLD ET AL., *supra* note 4.

134. *See* Campbell, *supra* note 6, at 1299.

135. *Id.* at 1302.

136. *See* Paetzold, *supra* note 6, at 397.

137. *See* Campbell, *supra* note 6, at 1323-24.

138. *Id.* at 1324.

139. Kaye, *supra* note 6, at 1337. A statistician may find a disparity “significant,” but because this significance is not a direct measure of the magnitude of an observed disparity, they may ultimately conclude any observed difference is meaningless.

140. *Id.*; *see also supra* note 46 and accompanying text.

Furthering the problem, litigants often manipulate statistics to support a variety of positions.¹⁴¹ Therefore, these seemingly objective tests often serve as clever disguises for parties to mask their very subjective views.¹⁴² Professor Richard Primus explains two different schools of thought among judges which lead to the decision to choose one test over the other.¹⁴³ One view sees disparate impact as an “evidentiary dragnet designed to discover hidden instances of intentional discrimination.”¹⁴⁴ The other view sees disparate impact as a “more aggressive attempt to dismantle racial and [other] hierarchies.”¹⁴⁵ Where there is no proof of ill-intent on behalf of the employer, judges who view disparate impact as a means of disclosing employers who secretly harbor loathsome feelings toward a protected class are much more likely to choose whichever statistical test favors the defendant (Primus’s first view).¹⁴⁶ Judges who view disparate impact as a “grand way of leveling the playing field between different groups of people” are more likely to penalize the employer and find a statistical disparity if the plaintiff can satisfy either test (Primus’s second view).¹⁴⁷ Therefore, strong evidence exists supporting the notion judges may use statistical tests to further their equalizing opinions of disparate impact theory.¹⁴⁸

Several other areas of law criticize the utilization of statistics when interpreting evidence, specifically its use in criminal jury trials, class-action litigation, environmental policy, and more.¹⁴⁹ While statistics are seemingly

141. ELLENBERG, *supra* note 125, at 21-30.

142. Peresie, *supra* note 8, at 786; Primus, *supra* note 10, at 518; *see also* Michael Perry, *A Brief Comment on Motivation and Impact*, 15 SAN DIEGO L. REV. 1173, 1178-81 (1978).

143. Peresie, *supra* note 8, at 779; Primus, *supra* note 10, at 518-32.

144. Peresie, *supra* note 8, at 779; Primus, *supra* note 10, at 518.

145. *See supra* note 144 and accompanying text.

146. Peresie, *supra* note 8, at 779.

147. *Id.*

148. Primus, *supra* note 10, at 518-32; *see* Meier et al., *supra* note 38, at 142-43; Peresie, *supra* note 8, at 779; *see also* McKinley, *supra* note 8, at 171; *Isabel v. City of Memphis*, 404 F.3d 404 (6th Cir. 2005) (demonstrating the dangerous results).

149. *See* David E. Adelman, *Scientific Activism and Restraint: The Interplay of Statistics, Judgment, and Procedure in Environmental Law*, 79 NOTRE DAME L. REV. 497 (2004) (discussing the use of statistics in environmental law); Bruce Brown & Lirieka Meintjes-van der Walt, *The Use and Misuse of Statistical Evidence in Criminal Proceedings*, 32 J. JURID. SCI. 1 (2007) (criticizing statistics in criminal trials); Saby Ghoshray, *Hijacked by Statistics, Rescued by Wal-Mart v. Dukes: Probing Commonality and Due Process Concerns in Modern Class Action Litigation*, 44 LOY. U. CHI. L.J. 467, 468-69 (2012) (criticizing statistics in class-action litigation); D.H. Kaye, *The Dynamics of Daubert: Methodology, Conclusions, and Fits in Statistical and Econometric Studies*, 87 VA. L. REV. 1933 (2001)

precise, innocuous methods for evaluating data, misunderstandings of how the science operates lead to harmful conclusions. By rejecting a “practical significance” requirement, the First Circuit places a heavy burden on employers who must now protect against even insignificant disparities that may result from isolated employment practices or procedures.¹⁵⁰

B. Practical Significance Requirement: Arbitrary or Efficient?

Strong evidence supports the EEOC’s interpretation of Title VII mandating a “practical significance” requirement. By hyper-focusing on potential ambiguities, the court failed to consider the functional benefits of “practical significance.” The court’s decision: (1) neglects to consider the utility of a “practical significance” requirement in accomplishing the goals of Title VII and (2) fails to consider the importance of practical significance when evaluating large and small sample sizes.

1. Honoring Title VII

The First Circuit’s decision to ignore employment agency standards and reject “practical significance”¹⁵¹ as a requirement for actionable disparities ignores the objectives of Title VII.¹⁵² Title VII’s disparate impact doctrine

(criticizing the use of statistics in evaluating the admissibility of evidence); James T. McKeown, *Statistics for Wage Discrimination Cases: Why the Statistical Models Used Cannot Prove or Disprove Sex Discrimination*, 67 IND. L.J. 633 (1992) (criticizing the use of statistics in evaluating wage claims and discrimination).

150. Notably, while employers struggle to comply with the varied case law regarding disparate impact, plaintiffs still continuously struggle to establish a prima facie case in the first sense due to the lack of data necessary to establish connections between employment practices and disparities. In the end, neither party benefits from the use of “statistical significance” as the primary test for establishing actionable disparity. See Marcel C. Garaud, *Legal Standards and Statistical Proof in Title VII Litigation: In Search of a Coherent Disparate Impact Model*, 139 U. PA. L. REV. 455, 466 n.53 (1990) (discussing the heavy burden placed on employers by statistical analyses). The 5% threshold—roughly two standard deviations—equates to roughly a one-in-twenty possibility that the correlation between selection rates and an employment practice occurred by chance. For comparison, three standard deviations equates to roughly a one-in-two-hundred possibility. *Id.*

151. Peresie, *supra* note 8, at 790 n.96 (noting that the EEOC actually suggests “practical significance” as a potential third alternative to statistical significance tests and the four-fifths rule “under which the court evaluates whether findings of statistical significance are ‘practically’ sound, rather than just ‘barely significant’”). The suggestion of practical significance as a third alternative overlooks the fact that the four-fifths rule is actually a measure of “practical significance” and provides a means of assessing whether findings are practically sound.

152. 42 U.S.C. § 2000e-2(a) (2012); see Lindsay Roshkind, *Employment Law: An Adverse Action Against Employers: The Supreme Court’s Expansion of Title VII’s Anti-*

seeks to compensate individuals who have suffered from use of an unlawful employment practice.¹⁵³ Congress did not design the doctrine to dispute inevitable differences in diversity that occur in the natural course of employment procedures.¹⁵⁴ The “practical significance” requirement allows courts to focus their efforts on invalidating employment procedures actually causing perceptible and troublesome disparities.¹⁵⁵ If flipped twice, a perfectly fair coin will come up two heads 25% of the time. Flipped three times, the fair coin will yield three heads 12.5% of the time.¹⁵⁶ Similarly, “a

Retaliation Provision, 59 FLA. L. REV. 707, 718 (2007); see also *Faragher v. City of Boca Raton*, 524 U.S. 775, 805-06 (1998) (citing *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 418 (1975)) (noting that Title VII is not a “general civility”); Harry E. Groves & Albert Broderick, *Affirmative Action Goals Under Title VII: Statute, Legislative History, and Policy*, 11 T. MARSHALL L. REV. 327, 335 (1986) (noting that the legislative history tells us “race-conscious goals as relief for egregious discrimination . . . violate both Title VII and the equal protection clause of the 14th amendment [T]he statutory and constitutional notion [is] that we live in a ‘color-blind’ society.”).

153. See 42 U.S.C. § 2000e-2(a); *supra* note 152 and accompanying text.

154. See Groves & Broderick, *supra* note 152, at 336 (“Title VII was ‘a law triggered by a Nation’s concern over centuries of racial injustice and intended to improve the lot of those who had “been excluded from the American dream for so long,” and ‘to break down old patterns of racial segregation and hierarchy.’”) (citing *United Steelworkers of Am. v. Weber*, 443 U.S. 193, 208 (1979)); *id.* at 329 n.12 (noting the clear unconstitutionality of quotas). The well-accepted illegality of quotas demonstrates that correlatively employers should not promote minority groups over majority groups for the sole purpose of creating diversity and avoiding disparate impact liability. Michael Evan Gold, *Grigg’s Folly: An Essay on the Theory, Problems and Origin of the Adverse Impact Definition of Employment Discrimination and a Recommendation for Reform*, 7 INDUS. REL. L.J. 429 (1985) (criticizing the potentiality of disparate impact to incentivize quotas which clearly goes against the goals of Title VII and noting the legislative history of Title VII does not support disparate impact theory at all and calls for it to be dismantled); see also *Ricci v. DeStefano*, 557 U.S. 557 (2009) (finding fear of disparate impact liability insufficient to validate favoring minority groups). *But cf.*, e.g., Robert Belton, *The Dismantling of Griggs Disparate Impact Theory and the Future of Title VII: The Need for a Third Reconstruction*, 8 YALE L. & POL’Y REV. 223 (1990) (noting Congress has now made clear that the goal of Title VII is to achieve workplace equality); Alfred W. Blumrosen, *The Legacy of Griggs: Social Progress and Subjective Judgments*, 63 CHI.-KENT L. REV. 1, 14-16 (1987) (finding the legislative history inconclusive).

155. See *Apsley v. Boeing, Co.*, 691 F.3d 1184 (10th Cir. 2012) (finding that the “statistically significant” disparities were too small to be “practically significant” and therefore not actionable); Kaye & Freedman, *supra* note 96, at 124 (noting that when “practical significance” is lacking, there is no reason to be concerned about statistical significance); see also King, *supra* note 129, at 271 (highlighting the problems that can occur in disregarding “practical significance”).

156. See Kaye & Freedman, *supra* note 96, at 151.

perfectly fair test given to a pool of blacks and whites will not always produce results that precisely mirror the racial percentages in the pool.”¹⁵⁷

Congress did not design the disparate impact doctrine to invalidate innocuous employment procedures that invariably result in diversity differences—as any equitable test might—but to protect historically burdened minorities from unfair employment practices.¹⁵⁸ The Supreme Court’s consistent interpretations further support the existence of an implied “practical significance” requirement within Title VII, continuously stating disparities must be “sufficiently substantial”¹⁵⁹ to be actionable.¹⁶⁰ By merely requiring disparities be statistically significant and not “gross,”¹⁶¹ the First Circuit puts too much faith in statistical models, creating an unjustifiable risk of improper findings of liability.¹⁶²

2. *The Problem with Sample Size and Significance*

Rejecting the “practical significance” requirement and ignoring the need for disparities to be “sufficiently substantial,” the First Circuit disconcertingly found any demonstration of statistical significance sufficient for a plaintiff to establish an actionable disparity.¹⁶³ As demonstrated in *Jones*, using large amounts of data to establish statistical disparities can magnify even miniscule impacts.¹⁶⁴ Because with large samples even trivial differences in selection rates may be statistically significant, “the fact that statistical models are only approximations of the selection process assumes greater importance in cases involving large numbers of employment decisions.”¹⁶⁵ To illustrate, imagine you flip a coin one million times. The coin lands on tails exactly 50.1% of the time. The deviation from the expected result of 50% tails and 50% heads would be statistically significant, despite that it amounts to just one flip per thousand.¹⁶⁶

157. *Fudge v. City of Providence Fire Dep’t*, 766 F.2d 650, 659 (1st Cir. 1985).

158. See *supra* note 154 and accompanying text.

159. *Watson v. Fort Worth Bank & Tr.*, 487 U.S. 977, 995 (1988).

160. See *Browne, Statistical Proof, supra* note 3, at 483 (quoting *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299 (1977)); *supra* notes 27-29 and accompanying text.

161. *Hazelwood Sch. Dist.*, 433 U.S. at 301.

162. See *Browne, Statistical Proof, supra* note 3, at 483-84; *supra* note 47.

163. *Jones v. City of Boston*, 752 F.3d 38, 48-53 (1st Cir. 2014).

164. See *Browne, Statistical Proof, supra* note 3, at 484; *Garaud, supra* note 150, at 466; *King, supra* note 129, at 271.

165. *Browne, Statistical Proof, supra* note 3, at 550 (quoting *Hazelwood Sch. Dist.*, 433 U.S. 299).

166. See *Kaye & Freedman, supra* note 96, at 123-25.

It is highly unlikely such small impacts are the product of discrimination.¹⁶⁷ While proof of intentional discrimination is unnecessary to sustain a disparate impact claim, “[t]he disparate impact theory nevertheless serves, in part, to root out hidden intentional discrimination.”¹⁶⁸ When the size of a disparity is not “practically significant,” or when a racial disparity is so small it cannot be determined without detailed statistical analysis, there is a correspondingly small chance the disparity reveals a hidden intent to discriminate.¹⁶⁹ Furthermore, where an employer can establish an employment practice constitutes a “business necessity,” it unnecessarily extends litigation of a case likely to fail because of the difficulty in concluding with confidence an alternative practice will truly lessen this already small effect.¹⁷⁰

The First Circuit’s decision to reject the EEOC’s interpretation of a “practical significance” requirement because of an alleged lack of functional application ignores its value as an important mechanism to avoid needless litigation and accomplish the purpose of Title VII.¹⁷¹ Without practical significance, there is little indication an employment practice has *caused* a statistical disparity—a requirement explicitly mandated by the text of Title VII.¹⁷² A statistically significant racial skew in outcomes does not constitute a disparate impact unless the minority differential is also sufficiently large.¹⁷³ The “practical significance” requirement ensures the strictly scrutinized disparate impact doctrine does not wrongly focus on insignificant, unimportant differences inevitably resulting in hiring practices but retains credibility and achieves its objectives in working to eliminate discrimination.¹⁷⁴

167. Primus, *supra* note 10, at 498-99, 520-21.

168. *Id.*; Jones, 752 F.3d at 50-51.

169. See Browne, *Statistical Proof*, *supra* note 3, at 482; Campbell, *supra* note 6, at 1299.

170. Jones, 752 F.3d at 50; see Primus, *supra* note 10, at 520-22.

171. See *supra* note 157-165 and accompanying text.

172. 42 U.S.C. § 2000e-1 to -15 (2012).

173. See *supra* note 26-29 and accompanying text. See, e.g., Xitao Fan, *Statistical Significance and Effect Size in Education Research: Two Sides of a Coin*, 94 J. EDUC. RES. 275, 277 (2001) (lack of practical significance indicates a lack of legal importance).

174. Critics have doubted the utility of the disparate impact doctrine since its inception. As we move closer toward an equal-opportunity society, many view the doctrine as an unnecessary hindrance rather than as a method of uprooting biased employment practices. See Katie R. Kormanyos, Ricci v. DeStefano: *How the Supreme Court Muddled Employment Discrimination Law and Doomed Employers to Costly Litigation*, 41 U. TOL. L. REV. 975 (2010). Despite this criticism, several civil rights advocates see continuing importance in retaining the doctrine. See Susan D. Carle, *A Social Movement History of Title*

C. *The Problem with Statistical Significance*

The First Circuit sharply criticized the four-fifths rule and the “practical significance” requirement, ignoring the similar anomalous results that can result from application of statistical significance.¹⁷⁵ Consider the implications of the following problem resulting from the First Circuit’s decision that a 1% difference is an actionable disparity:

An employer considers 100 male applicants and 100 female applicants. It hires 99 males and 98 females. The selection rate is .99 for males and .98 for females, only a .01 difference, with a statistical significance of only .058 [standard deviations]. This would seem to give the employer confidence that its selection procedure had no adverse impact on females and it is safe to continue using the procedure. However, if the employer continues hiring men and women at the same rate, and considers 1,200 males and 1,200 females, hiring 1,188 males (99 percent of male applicants) and 1,176 females (98 percent of female applicants), the statistical significance reaches 2.01 standard deviations adverse to female applicants.¹⁷⁶

The apparently insignificant .01 difference between male and female selection rates did not change. The difference does not remotely approach the four-fifths rule, as the female rate is 99% of male hiring rate. However, this small .01 difference is now actionable disparate impact discrimination under *Jones*.¹⁷⁷ The disparate impact continues to grow with the number of selections, even when the same .01 selection rate difference remains constant. In this example, for instance, by the time 9900 males and 9800 females are hired, the disparate impact on female applicant is 5.82 standard deviations.¹⁷⁸

VII Disparate Impact Analysis, 63 FLA. L. REV. 251, 300 (2011). Disparate impact analysis ensures employers and their agents not only avoid engaging in direct acts of prejudice decision-making but that they also attend to the effects of their employment practices. It seeks to invalidate those practices implemented with a hidden malicious intent as well as those without a discriminatory motive. Such unintentional discrimination may result from a subtle, subconscious bias, or merely an ignorance of the effects one’s employment practices may have on a worker from disadvantaged social origins. *Id.*

175. *Jones*, 752 F.3d at 53.

176. Lydell C. Bridgeford, *Q&A: Statistical Proof of Discrimination Isn’t Static*, BLOOMBERG BNA (Feb. 8, 2015, 9:15 PM), <http://www.bna.com/qa-statistical-proof-b17179891425/> (illustration provided by David Cohen of DCI Consulting Inc.).

177. *Jones*, 752 F.3d at 50.

178. *Id.*

Ignoring the manner in which large amounts of data magnify insignificant disparities,¹⁷⁹ the First Circuit held that viewing the data in the aggregate demonstrated a near certainty that the difference in outcomes could not be due to chance alone.¹⁸⁰ However, aggregation of data to achieve statistical significance may actually “lessen the probative force of the data.”¹⁸¹ Just as large sample size magnifies small disparities, aggregation improperly inflates statistical significance.¹⁸² By relying on statistical significance as the threshold, aggregating data, and relying on failure rates instead of the more accepted survival rates, it is clear the First Circuit relied on manipulated numbers to support its decision to deny summary judgment to defendants. As a result, the *Jones* decision allows no room for employers to argue small selection rate differences are insufficient to create a *prima facie* case of disparate impact discrimination. It warns employers even very small, apparently insignificant differences in selection rates may become actionable when there are a large number of selections involved.

179. See *Apsley v. Boeing Co.*, 691 F.3d 1184, 1186 (10th Cir. 2012); see also PAETZOLD ET AL., *supra* note 4, at 2-13 (“Statistical significance is affected by the number of observations, so that for large samples, spurious significance can result.”); Rubinfeld, *supra* note 12, § 7.11 (stating that the reason that small disparities can be statistically significant is that “statistical significance is determined, in part, by the number of observations in the data set”).

180. *Jones*, 752 F.3d at 45. Compare *id.* with *Apsley*, 691 F.3d at 1186 (rejecting aggregation of data under similar circumstances). The First Circuit allowed examination of the data in *Jones* in the aggregate, demonstrating a disparity of 7.14 standard deviations over eight years. When examined individually however, disparities greater than 1.96 standard deviations existed in only three out of the eight aggregated years. In five out of eight years, the disparity would notably not be statistically significant even under the stringent *Jones* standard.

181. *United States v. City of Yonkers*, 609 F. Supp. 1281, 1288 (S.D.N.Y. 1984).

182. *Jones v. City of Boston*, No. 05-11832-Gao, 2012 WL 4530594 (D. Mass. Sept. 28, 2012) (district court). Because the department conducted the drug test annually, several of the same employees were included in the results from 1999-2006. Accordingly, one cannot observe such results as independent. In addition, viewed separately the standard deviations here are relatively modest—plaintiffs’ expert’s calculations reflect 0.33 to 3.99 standard deviations. Many courts have found a failure to establish a *prima facie* case where there were comparable standard deviations. See, e.g., *Waisome v. Port Auth. of N.Y. & N.J.*, 948 F.2d 1370 (2d Cir. 1991) (2.68 standard deviations); *Apsley v. Boeing Co.*, 722 F. Supp. 2d 1218 (D. Kan. 2010) (four to greater than five standard deviations).

D. Respecting Deference

1. Look to the EEOC: Honoring Agency Deference

In addition to its misplaced reliance and flawed application, the First Circuit's decision to completely ignore the employment-agency-advocated UGESP denies the EEOC proper deference and cheapens the agency's authority.¹⁸³ Many employers look to the UGESP for compliance with employment discrimination laws, relying on the four-fifths rule and "practical significance" as reasonable guidelines for evaluating disparities.¹⁸⁴ This complete departure from the UGESP detrimentally leaves employers with virtually no reliable reference for employment discrimination compliance in an already murky area of law.

While the Guidelines are not binding law, the Supreme Court has found they are entitled to great deference.¹⁸⁵ In *EEOC v. Commercial Office Products Co.*, the Court held the EEOC is entitled to deference in interpreting ambiguous language "where it is reasonable."¹⁸⁶ In light of Supreme Court precedent that disparities be "gross" and "sufficiently substantial," the EEOC's interpretation of a "practical significance" requirement is undoubtedly reasonable.¹⁸⁷ Therefore, in order to promote

183. *Waisome*, 948 F.2d at 1376 (finding no disparate impact where, "though the disparity was found to be statistically significant, it was of limited magnitude"). In this case, over the range of eight years in which the hair test was required, it is undisputed that white officers passed the test at rates of 99% to 100%, and African Americans passed the test at rates between 97% and 99%. Thus, the passing rate for African Americans was repeatedly at least 97% of the passing rate for whites. Under the four-fifths rule, the question is not even close; the EEOC would not regard the results to amount to adverse impact.

184. See 42 U.S.C. § 2000e-1 to -15 (2012); Peresie, *supra* note 8, at 786.

185. *Kormanyos*, *supra* note 174, at 982 (citing *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 430-31 (1975)); see also *Griggs v. Duke Power Co.*, 401 U.S. 424, 433-34 (1971) ("The administrative interpretation of the Act by the enforcing agency is entitled to great deference.").

186. *EEOC v. Commercial Office Prods. Co.*, 486 U.S. 107, 108 (1988) ("The reasonableness of the EEOC's interpretation of 'terminate' in its statutory context is more than amply supported by the legislative history of Title VII's deferral provisions, the purposes of those provisions, and the language of other sections of the Act."); see also *Skidmore v. Swift & Co.*, 323 U.S. 134, 139-40 (1944) (noting that *Commercial Office Products* has not been overruled).

187. See *supra* notes 26-28 and accompanying text.

efficiency and clarity, courts should defer to the EEOC's judgment in interpreting a "practical significance" requirement.¹⁸⁸

The First Circuit chose to reject the "practical significance" requirement based on its analyses of the bare text of Title VII.¹⁸⁹ Noting the Supreme Court's minimal guidance on the matter, the court found the term "disparate" highly ambiguous.¹⁹⁰ Because it did not find the UGESP's interpretation persuasive, the First Circuit looked to Merriam Webster's dictionary for reference.¹⁹¹ The dictionary defines "disparate" as meaning "fundamentally different" or "markedly distinct."¹⁹² While these two definitions support the idea that statistical deviations must be consequential in order to constitute a disparity, the additional listing of "different" as a synonym led the First Circuit to conclude "disparate" could simply mean "non-random" rather than "sufficiently large."¹⁹³ This conclusion makes little sense, and the court cites no persuasive authority to support its decision.

By ignoring the UGESP, the First Circuit denied the EEOC proper deference.¹⁹⁴ The EEOC adopted the Guidelines to "constitute a body of experience and informed judgment to which courts and litigants may properly resort for guidance."¹⁹⁵ With no clear guidance from case law or the text of Title VII, employers and labor organizations systematically rely on the UGESP to comply with federal employment law.¹⁹⁶ The First Circuit's rejection of requirements advocated by the EEOC—the agency with primary enforcement responsibility in combatting employment discrimination—in favor of a synonym listed in Merriam Webster's dictionary completely undermines the agency, dismantling guidelines relied on by many.

The Supreme Court has most recently described a *prima facie* case of disparate impact discrimination as "essentially, a threshold showing of a

188. See Rutherglen, *supra* note 15, at 1330 (advocating for a "modest threshold for the plaintiff's *prima facie* case that emphasizes the practical significance of the evidence of adverse impact").

189. Jones v. City of Boston, 752 F.3d 38, 49 (1st Cir. 2014).

190. *Id.* at 50.

191. *Id.*

192. *Id.* (citing WEBSTER'S NEW COLLEGIATE DICTIONARY 329 (8th ed. 1977)); MERRIAM-WEBSTER'S COLLEGIATE DICTIONARY 360 (11th ed. 2003) (offering the same definitions and synonyms)).

193. *Id.* at 52.

194. See *supra* notes 185-86 and accompanying text.

195. *Skidmore*, 323 U.S. at 140; see 29 C.F.R. § 1607.1(B).

196. Peresie, *supra* note 8, at 775.

significant statistical disparity . . . nothing more.”¹⁹⁷ It is clear statistical significance does not indicate legal significance.¹⁹⁸ Therefore, in order to establish a *prima facie* case, it is reasonable to conclude a plaintiff need only identify (1) an employment practice or procedure and (2) a “practically significant” disparity. Although heavily criticized, the four-fifths rule advocated by the EEOC is the best available test to meet this standard.¹⁹⁹ While critics condemn the test for evaluating impact—rather than causation—many overlook the four-fifths’ initial *assumption* of causation. The test then performs the important function of evaluating the overall negative consequences of the disparity. In other words, it determines whether the *impact* of the disparity is sufficiently enough large to concern the court.²⁰⁰

It makes perfect sense that when evaluating disparities in the context of the disparate *impact* doctrine, one would look to the impact of the disparity.²⁰¹ Because causation is nearly impossible to prove, and because small disparities are unlikely the result of discrimination, the evaluation of impact allows the court to ensure there is a “gross” disparity before shifting the burden to the defendant to prove business necessity.²⁰² Because subsequent steps in the burden-shifting process further protect any risk of insufficient causation, courts should defer to the judgment of the EEOC and utilize the four-fifths “rule of thumb” and “practical significance” as the

197. *Id.* (quoting *Ricci v. DeStefano*, 557 U.S. 557, 559 (2009)). It is noteworthy that the First Circuit similarly tried to rely on this statement to support its decision to hold any statistically significant disparity actionable. *Jones*, 752 F.3d at 46. The Supreme Court, however, never actually mentioned “statistical significance” anywhere in *Ricci* but instead pointed to the four-fifths rule to show that where the impact ratio was approximately 50%, the plaintiffs undisputedly established a *prima facie* case. 557 U.S. at 586-87. Thus, in discussing “statistical disparity,” the Court was addressing the degree of disparate impact rather than the technical term “statistical *significance*.” In sum, the First Circuit has “no authority for the proposition that a showing of two or more standard deviations in failure rates alone is sufficient to establish a *prima facie* case. This is especially true here, where the sample size is so large, because statistical significance is highly sensitive to sample size.” *Jones v. City of Boston*, No. 05-11832-Gao, 2012 WL 4530594, at *3 (D. Mass. Sept. 28, 2012).

198. *See supra* Section III.A.

199. *See* WALTER B. CONNOLLY, JR. ET AL., *USE OF STATISTICS IN EQUAL EMPLOYMENT OPPORTUNITY LITIGATION* § 2.01[1] (1992); Meier et al., *supra* note 38, at 1294 (advocating for the four-fifths rule as the proper method for measuring disparate impact).

200. *See* CONNOLLY, JR. ET AL., *supra* note 199.

201. *See id.* (“Disparate *impact* is, by definition, established by statistics since impact is described by quantitative patterns.”).

202. *See* 42 U.S.C. § 2000e-1 to -15 (2012).

proper methods for establishing disparate impact.²⁰³ The benefit of maintaining a uniform, reliable standard outweighs the risk of ambiguity. Furthermore, it relieves both employers and plaintiffs the burden of expense in examining complicated statistical significance tests that do not indicate any stated legal significance.

2. A Novel Proposal: Focusing on Practical Significance

Acknowledging the resistance of courts to adopt the four-fifths rule, this Note offers an alternate bright-line test to evaluate practical significance: the 10% rule. The 10% rule is an extremely simple test proposed for its efficiency, clarity, and ease of application. Under this approach—provided the sample size is not too small—courts would simply compare the percentage of selection rates of the minority group to the percentage of selection rates of the majority group.²⁰⁴ Next, the court would evaluate the percentage of the resulting reduction in workforce of the minority group.²⁰⁵ If the difference between the two survival rate percentages is 10% or greater, and the reduction in the minority workforce is 10% or greater, it would imply “practical significance” and establish a rebuttable prima facie case of disparate impact discrimination.

To illustrate, imagine 100 men and 100 women undergo an examination. A total of 88% of men pass compared with only 76% of women. The difference in pass rates is: $88\% - 76\% =$ twelve percentage points. The reduction in the female workforce is: $100\% - 76\% =$ twenty-four percentage points. Under the 10% rule, this would demonstrate a “practically significant” disparity between men and women, thereby establishing a prima facie case of disparate impact discrimination. If the plaintiff fails to meet this threshold, she may still be able to establish a prima facie case via an examination of the facts and surrounding circumstances of the claim as currently provided by the law.²⁰⁶

There are several policy reasons to support use of the 10% rule. First, it is incredibly easy to apply—anyone with a grade-school education could

203. See *Jones v. City of Boston*, 752 F.3d 38, 52 (1st Cir. 2014) (discussing the safeguards of the burden-shifting framework).

204. If the sample size is too small, this would indicate the disparity is not one courts should be concerned about. Just as Title VII denies claims when the employer employs less than fifteen people, Title VII should similarly deny claims when the “selection rates” of the minority group is less than roughly 15 people. See 42 U.S.C. § 2000e-1 to -15.

205. See *Apsley v. Boeing Co.*, 691 F.3d 1184, 1201 (10th Cir. 2012) (noting the importance in the reduction of the workforce in evaluating disparities).

206. *Id.* at 1195.

perform the test with knowledge of the hiring and termination percentages. This lifts the burden on both plaintiffs and employers to spend extensive time and money evaluating a wide variety of statistical tests that—viewed in their current state—do not indicate with any more certainty a “gross” disparity has occurred. Second, it puts employers on clear notice of when their employment practices create an actionable disparity. This clarity would alert them to any employment practices causing unintentional discriminatory effects and allow them to revisit the business necessity of their procedures, thus accomplishing the major goals of the disparate impact doctrine.²⁰⁷

Employers will likely argue the 10% rule ignores the chance for random error within limited data. While this may be true, the First Circuit correctly found subsequent safeguards of the burden-shifting mechanism capable of diminishing any risk of undue burden to the defendant.²⁰⁸ If the defendant can offer a business necessity for the practice, the burden shifts back to the plaintiff.²⁰⁹ If a defendant cannot offer a business necessity for the employment practice, there is no reason to retain a potentially discriminatory practice just because the causal connection is unclear. If a defendant can offer a business necessity, but the plaintiff can offer an alternative employment practice accomplishing the necessity to the same degree in a less discriminatory manner,²¹⁰ there is also no need to retain it when another viable option exists. This process reduces any risk of unwarranted liability from a truly random occurrence triggering the 10% rule.

While not a perfect test, uniform application of the 10% rule is preferable to uniform application of either of the current methods. The four-fifths standard is extremely difficult for plaintiffs to meet, denying victims of apparent disparities their day in court. On the other hand, requiring merely a showing of statistical significance places an inefficiently low burden on plaintiffs when establishing a *prima facie* case of disparate impact discrimination. Despite obvious manipulation, any demonstration of a

207. See *supra* Section III.B. It further simplifies things by looking at the employment data in question rather than the relevant labor market. While the relevant labor market would undoubtedly present a more reliable standard of comparison, this data is usually not available, creating an insurmountable burden on plaintiffs who have clearly suffered a wrong but are without access to such extensive data to prove it.

208. See *Jones*, 752 F.3d at 53 (also discussing how the subsequent steps required of the plaintiff add in a “case-specific” practical significance requirement).

209. See 42 U.S.C. § 2000e-1 to -15.

210. *Id.*

“difference” in numbers meets this burden, thereby forcing employers to defend even frivolously small impacts.

The 10% rule creates a middle ground. Imagine now that 100% of men survived termination and only 82% of women survived termination under use of an employment procedure. Despite clear discrimination, the four-fifths rule would deny these women a prima facie case of disparate impact discrimination. Using the 10% rule however, both the discrepancy between the two groups and the reduction in the female workforce is eighteen percentage points, undoubtedly establishing a prima facie case. In addition, recall the hypothetical in which 98% of women survived use of an employment procedure compared with 99% of men. Despite a difference in data that is almost impossible to recognize without extensive statistical analysis, the statistical-significance test found this to be an actionable disparity. The 10% rule would instead reject this claim because a 1% disparity is obviously not large enough to concern the court, nor is the 2% overall reduction in workforce of the minority group overwhelmingly disconcerting.

This Note offers the 10% threshold as a rule of convenience—it is easy to spot and sufficiently large to convey “practical significance” without placing impossible burdens on the plaintiff, evaluating both the disparity between the minority and the majority group and the disparity within the minority group itself. Its tendency to lead to arbitrary results is no greater than the current methods utilized, and its simplicity benefits both plaintiffs and employers alike. While certain application of such a bright-line rule may inevitably deny certain plaintiffs their day in court, that is the balance struck by the utility of the test and the ease and efficiency with which it is applied.²¹¹

211. *But see* Morris, *supra* note 9, at 12 (offering an extremely complicated but overall more efficient statistical test to evaluate disparities).

The proposed method, which includes estimates of the impact ratio along with confidence intervals, provides a more informative and more consistent framework for assessing adverse impact. While this procedure is more complex than current methods for assessing adverse impact, the complexity is offset by several advantages. First, it provides an indication of adverse impact in a common metric (i.e., the impact ratio), with a generally accepted criterion for practical significance ($IR < 0.8$). Thus, the recommended procedure does not require a fundamental change in how organizations and the courts define adverse impact. Second, it provides sufficient information so that both both practical and statistical significance can be evaluated, through the inclusion of the standard error. Third, the width of the confidence interval should provide practitioners with a better sense of the general lack of power of these tests, and

IV. Conclusion

As Ernest Rutherford cheekily noted, “If your experiment needs a statistician, you need a better experiment.”²¹² Courts should not burden employers and plaintiffs with resorting to complicated statistical analyses and instead should look to “practical significance” to evaluate disparities. While the four-fifths rule is the best available standard to evaluate “practical significance,” the 10% rule would similarly promote the objectives of Title VII and provide a clear, reliable standard for both plaintiffs and employers. In addition, it would accomplish this without placing an extreme burden on unlucky plaintiffs who fail the four-fifths standard.

If courts continue to reject the four-fifths rule and the “practical significance” requirement, legislative action and clear judicial review is necessary to provide clarity for courts and employers trying to avoid accusations of disparate impact discrimination. Because courts often do not pay the EEOC proper deference and it therefore cannot remedy the issues resulting in present-day Title VII litigation by itself, legislative action or judicial review is necessary for adopting a coherent test. Without a clear test, the uncertainty of whether utilized employment procedures are creating small, insignificant deviations that may become actionable lawsuits under the *Jones* decision will continue to burden employers. This forces them to either invest extensive time and resources to determine whether their chosen procedures create even small deviations in selection rates over large amounts of data or risk facing costly litigation.

Katie Eissenstat

the potential for inaccurate decisions based on sampling error.

Id.

212. *Ernest Rutherford Quotes*, GOODREADS, http://www.goodreads.com/author/quotes/437411.Ernest_Rutherford (last visited Nov. 15, 2015).