

2016

Big Enough to Matter: Whether Statistical Significance or Practical Significance Should Be the Test for Title VII Disparate Impact Claims

Elliot Ko

Follow this and additional works at: <https://scholarship.law.umn.edu/mlr>



Part of the [Law Commons](#)

Recommended Citation

Ko, Elliot, "Big Enough to Matter: Whether Statistical Significance or Practical Significance Should Be the Test for Title VII Disparate Impact Claims" (2016). *Minnesota Law Review*. 150.
<https://scholarship.law.umn.edu/mlr/150>

This Article is brought to you for free and open access by the University of Minnesota Law School. It has been accepted for inclusion in Minnesota Law Review collection by an authorized administrator of the Scholarship Repository. For more information, please contact lenzx009@umn.edu.

Note

Big Enough To Matter: Whether Statistical Significance or Practical Significance Should Be the Test for Title VII Disparate Impact Claims

*Elliot Ko**

Seventeen years ago, the Boston Police Department began testing its officers and cadets for illegal drug use.¹ Less than one percent of the force's officers and cadets tested positive for illegal drug use.² But African-American officers and cadets tested positive for illegal drug use almost five times more frequently than white officers and cadets.³ Ten African-American officers and cadets sued the Boston Police Department, arguing that the department's drug testing program violated Title VII of the Civil Rights Act of 1964.⁴ The plaintiffs probably expected the police department to defend its drug testing program on the grounds that it was a matter of "business necessity."⁵ Instead, this case addressed the unearthed and often-neglected legal question: What does "disparate impact" mean?⁶

Title VII broadly proscribes the use of any employment practice that "causes a *disparate impact* on the basis of race, color, religion, sex, or national origin."⁷ But what does "disparate impact" mean? Does it mean any disparity that is statisti-

* J.D., 2016, University of Minnesota Law School. Special thanks to Professor Jessica Clarke, Adam Beaupre, Kusha Mohammadi, Gregory Escobar, Danielle Bulta, and the *Minnesota Law Review* staff for their input and help with preparing this Note. Copyright © 2016 by Elliot Ko.

1. *Jones v. City of Boston*, 752 F.3d 38, 42 (1st Cir. 2014).
2. *Id.* at 42–43, 44.
3. *Id.* at 43.
4. *Id.* at 41.
5. *Cf. id.* at 54–55 (leaving this question "open for further consideration").
6. See Michael Stenger, *The First Circuit Strikes Out in Jones v. City of Boston: A Pitch for Practical Significance in Disparate Impact Cases*, 60 VILL. L. REV. 411, 413–14 (2015).
7. 42 U.S.C. § 2000e–2(k)(1)(A)(i) (2012) (emphasis added).

cally significant, no matter how small that disparity may be?⁸ Or does it refer only to a disparity that is both statistically *and* practically significant?⁹ Why does this matter? Statistical significance measures the likelihood that a certain disparity is due to random chance instead of some other factor.¹⁰ Practical significance, on the other hand, asks whether this disparity is large enough to matter in practical terms.¹¹ In many cases, a disparity that is statistically significant will also be practically significant. But not always.¹² In some cases, statistically significant disparities may “have little or no real-world importance.”¹³ Thus, the courts’ answer to the question posed above can either doom or save claims with a high level of statistical significance but a low level of practical significance.

Neither Congress nor the Supreme Court has provided much guidance on this question.¹⁴ Title VII itself prohibits the use of any employment practice that would “adversely affect” or otherwise have a “disparate impact” on minorities.¹⁵ Neither term, however, is defined in Title VII.¹⁶ As for the Supreme Court, it has alternately spoken of employment practices that disqualify minorities at “a substantially higher rate” than non-minority applicants,¹⁷ “significant statistical disparit[ies]”,¹⁸ and statistical disparities that have a “significantly different,”¹⁹ “significantly greater,”²⁰ or “significantly discriminatory im-

8. See discussion and cases cited *infra* Part I.C.1.

9. See discussion and cases cited *infra* Part I.C.2.

10. See *infra* Part I.B.1.

11. See *infra* Part I.B.2.

12. See *infra* Part I.B.3.

13. See *infra* Part I.B.3.

14. BARBARA T. LINDEMANN & PAUL GROSSMAN, 1 EMPLOYMENT DISCRIMINATION LAW 124 (4th ed. 2007) (“The Supreme Court has variously spoken of ‘statistical disparities . . . sufficiently substantial [to] raise an inference of causation,’ a ‘significantly different’ selection rate, and a ‘substantially disproportionate’ disqualification rate as constituting evidence of adverse impact, but it has given no definitive guidance on ‘just what threshold mathematical showing of variance . . . suffices as a “substantial disproportionate impact.” . . . [A]nd the text of Title VII, as amended by the 1991 Act, provides no definitive answer.” (citations omitted)).

15. See 42 U.S.C. § 2000e-2(a)(2), (k) (2012).

16. See, e.g., *id.* § 2000e (defining many terms, but not “adverse affect” or “disparate impact”).

17. *Griggs v. Duke Power Co.*, 401 U.S. 424, 426 (1971).

18. *Ricci v. DeStefano*, 557 U.S. 557, 587 (2009); *Watson v. Fort Worth Bank & Tr.*, 487 U.S. 977, 1004 (1988) (Blackmun, J., concurring).

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

20. *N.Y. Transit Auth. v. Beazer*, 440 U.S. 568, 586 (1979).

pact”²¹ on minorities. At other times, the Court has spoken of statistical disparities that are large enough to have sufficient “probative value”²² or that are “sufficiently substantial” to “raise . . . an inference of causation.”²³ In still other instances, the Supreme Court has cited the Equal Employment Opportunity Commission’s (EEOC) four-fifths rule (a rough proxy for both practical and statistical significance) and standard deviation analysis (a test for statistical significance) as possible tests for disparate impact claims.²⁴ What the Supreme Court has not done is clarify whether a “significant” disparity is one that is statistically significant or practically significant.²⁵ This has led to a split in the lower courts on this question.²⁶

This Note catalogs the circuit split on this question. It also analyzes the text, legislative history, and judicial interpretation of Title VII to determine whether practical significance should be a necessary element of every disparate impact claim. Part I briefly summarizes the history of disparate impact claims, introduces the difference between statistical and practical significance, and examines each circuit’s case law on this issue. Part II argues that the text of Title VII, its legislative history, and applicable Supreme Court precedent should not require plaintiffs to prove practical significance as part of their *prima facie* cases. Part III makes two proposals. First, courts should abolish the practical significance requirement as part of the plaintiff’s *prima facie* case and adopt a rebuttable presumption that any statistically significant disparity is a “disparate impact” for the purposes of Title VII disparate impact claims. Second, to the extent that courts still wish to retain practical significance as a way for a defendant to rebut the presumption that a statistically significant disparity is actionable under Title VII, courts should develop a more concrete test for measuring practical significance.

21. *Connecticut v. Teal*, 457 U.S. 440, 446 (1982).

22. *Id.* at 463 n.7.

23. *Watson*, 487 U.S. at 995.

24. *Id.* at 995 n.3.

25. *See, e.g., id.* at 995.

26. *See infra* Part I.C.

I. THE HISTORY OF DISPARATE IMPACT CLAIMS AND
UNDERSTANDING THE DIFFERENCE BETWEEN
STATISTICAL AND PRACTICAL SIGNIFICANCE

This Part explores the history of disparate impact claims, explains the difference between statistical and practical significance, and surveys the existing case law on this issue. Its purpose is to summarize the development of disparate impact liability from 1971 to the current day. Its purpose is also to explain why the distinction between statistical and practical significance matters in the real world and summarize the existing case law on whether practical significance is required for disparate impact claims. To that end, Section A briefly lays out the history of disparate impact theory. Section B explains the difference between statistical and practical significance. Finally, Section C examines each circuit's case law on this question.

A. A BRIEF HISTORY OF DISPARATE IMPACT THEORY

Title VII, as it was originally enacted, did not contain the words “disparate impact.”²⁷ It did, however, prohibit employers from “discriminat[ing] against any individual . . . because of such individual’s race, color, religion, sex, or national origin”²⁸ or acting in any other way “which would deprive or tend to deprive any individual of employment opportunities or otherwise adversely affect his status as an employee.”²⁹ Seven years after Title VII’s enactment, in *Griggs v. Duke Power Co.*, the Supreme Court held that this language prohibited not only intentional discrimination “because” of a protected characteristic, but also facially neutral policies that had the *effect* of disproportionately harming minorities.³⁰ In other words, Title VII prohibits “not only overt discrimination but also practices that are fair in form but discriminatory in practice[],” at least if such practices are not justified by “business necessity.”³¹ The

27. See *Lewis v. City of Chicago*, 560 U.S. 205, 211 (2010); see also 42 U.S.C. § 2000e-2 (1970) (containing no subsection (k), six years after Title VII was originally enacted).

28. 42 U.S.C. § 2000e-2(a)(1).

29. *Id.* § 2000e-2(a)(2).

30. *Griggs v. Duke Power Co.*, 401 U.S. 424, 430 (1971).

31. *Id.* at 431. If the challenged employment practice is “absolutely essential to the operation” of the business, the employer may assert this as an affirmative defense to Title VII liability. Andrew C. Spiropoulos, *Defining the Business Necessity Defense to the Disparate Impact Cause of Action: Finding the Golden Mean*, 74 N.C. L. REV. 1479, 1483 (1996).

first type of claim came to be known as a “disparate treatment” claim, whereas the second was called a “disparate impact” claim.³² A disparate treatment claim is appropriate when an employer intentionally discriminates against an applicant or an employee because of her race, sex, or religion.³³ A disparate impact claim, on the other hand, is appropriate when an employer adopts a facially neutral practice that disproportionately harms members of a protected class, even if the employer had no ill intent or animus towards members of that class.³⁴

Seventeen years after *Griggs*, in *Wards Cove Packing Co. v. Atonio*,³⁵ the Supreme Court “significantly” cut back the scope of disparate impact claims in three ways.³⁶ First, the Court imposed a “causation requirement,” requiring the plaintiff to show as part of his prima facie case that the statistical disparity complained of was the “result” of one or more specific employment practices.³⁷ Second, the Court lowered the standard for the employer’s affirmative defense, allowing employers to rebut a disparate impact claim by simply pointing to a “legitimate business justification” for the challenged employment practice.³⁸ Third, the Court held that the employer only bore the burden of production, and not the burden of persuasion, on its business justification defense.³⁹

This regime only lasted two years. In 1991, Congress passed the Civil Rights Act of 1991.⁴⁰ This Act did three things. First, the Act implicitly ratified the viability of disparate impact theory by stating that an unlawful employment practice

32. See, e.g., *Ricci v. DeStefano*, 557 U.S. 557, 577 (2009) (explaining the difference between these two types of claims).

33. *Id.*

34. *Id.*

35. *Wards Cove Packing Co. v. Atonio*, 490 U.S. 642 (1989), *superseded by statute*, Civil Rights Act of 1991, Pub. L. No. 102-166, 105 Stat. 1074, *as recognized in* *Tex. Dep’t of Hous. & Cmty. Affairs v. Inclusive Cmty. Project, Inc.*, 135 S. Ct. 2507 (2015).

36. *Ricci*, 557 U.S. at 623 (Ginsburg, J., dissenting) (describing the impact of *Wards Cove*).

37. *Wards Cove*, 490 U.S. at 657.

38. See *id.* at 659 (“[A]t the justification stage of . . . a disparate-impact case, the dispositive issue is whether a challenged practice serves, in a significant way, the legitimate employment goals of the employer [T]here is no requirement that the challenged practice be ‘essential’ or ‘indispensable’ to the employer’s business for it to pass muster”).

39. See *id.* at 659–60.

40. Civil Rights Act of 1991, Pub. L. No. 102-166, 105 Stat. 1071 (1991).

could be “based on disparate impact.”⁴¹ Second, the Act partially codified *Wards Cove*’s first holding by requiring the plaintiff to prove that “a particular employment practice . . . cause[d] a disparate impact,”⁴² unless “the elements of a respondent’s decision making process are not capable of separation for analysis.”⁴³ Third, the Act abrogated the last two holdings of *Wards Cove* and restored pre-*Wards Cove* case law on the business necessity defense by putting the burden of proof on the employer to prove that a challenged employment practice was “consistent with business necessity.”⁴⁴

More recently, the Supreme Court has extended disparate impact theory to two other anti-discrimination statutes: the Age Discrimination in Employment Act of 1967⁴⁵ and the Fair Housing Act.⁴⁶

B. STATISTICAL VS. PRACTICAL SIGNIFICANCE: TWO DIFFERENT THINGS

This Section attempts to explain the difference between statistical and practical significance. Even though statistical and practical significance are two different things, many courts do not analyze these concepts separately.⁴⁷ To make matters worse, there is little consensus on how to measure these two different concepts, if they are required in the first place.⁴⁸ Subsection 1 explains what statistical significance means and how to measure it. Subsection 2 does the same for practical significance. Subsection 3 discusses when the difference between sta-

41. 42 U.S.C. § 2000e-2(k)(1)(A)(i) (2012); see also *Ricci*, 557 U.S. at 624 (describing the Civil Rights Act of 1991 as “formally codify[ing] the disparate-impact component of Title VII”).

42. 42 U.S.C. § 2000e-2(k)(1)(A)(i).

43. *Id.* § 2000e-2(k)(1)(B)(i); see also *Tex. Dep’t of Hous. & Cmty. Affairs v. Inclusive Cmty. Project, Inc.*, 135 S. Ct. 2507, 2523 (2015) (reiterating *Wards Cove*’s causation requirement and describing *Wards Cove* as “superseded by statute on other grounds”).

44. 42 U.S.C. § 2000e-2(k)(1)(A)(i).

45. See *Smith v. City of Jackson*, 544 U.S. 228, 235–38 (2005).

46. See *Cmty. Affairs*, 135 S. Ct. at 2525.

47. See cases cited *infra* Part I.C. This imprecision is also common in the academic literature. See, e.g., 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) (discussing both the four-fifths rule and various ways to measure statistical significance without distinguishing between statistical and practical significance).

48. See discussion *infra* Parts I.B.1, I.B.2.

tistical and practical significance might actually matter in a real-world case.

1. Statistical Significance

Broadly speaking, statistical significance attempts to measure the likelihood that a statistical disparity is attributable to something more than chance.⁴⁹ To ask whether a certain employment practice adversely affects a protected class in a statistically significant way is another way of asking whether there is some “relationship” between “the challenged employment practice” and membership in a protected class.⁵⁰ Some courts have referred to statistical significance as an “indication of causation.”⁵¹ Statistical analysis, however, can never “conclusively establish the cause . . . for an observed disparity.”⁵² At best, it can suggest that it is highly unlikely that a certain adverse impact on a protected class is due solely to chance.⁵³

Statisticians use a variety of tests to measure statistical significance.⁵⁴ The relative merits of these tests is beyond the scope of this Note,⁵⁵ as are the precise details of how to conduct these tests.⁵⁶ That said, a basic understanding of some of the terms used in this field may be helpful for understanding what courts mean when they use certain statistical terms. Three terms bear special mention here: (a) p-values; (b) confidence intervals; and (c) standard deviations.

49. *Jones v. City of Boston*, 752 F.3d 38, 43 (1st Cir. 2014); STEPHANIE R. THOMAS, *STATISTICAL ANALYSIS OF ADVERSE IMPACT: A PRACTITIONER'S GUIDE* 25 (2011).

50. THOMAS, *supra* note 49, at 44–45.

51. *Matrixx Initiatives, Inc. v. Siracusano*, 563 U.S. 27, 40 (2011); *see also* *Watson v. Fort Worth Bank & Tr.*, 487 U.S. 977, 994–95 (1988) (correlating statistical testing with the causation requirement of Title VII).

52. THOMAS, *supra* note 49, at 63.

53. *Id.* at 25.

54. *See, e.g., id.* at 41–61 (describing binomial distribution analysis, hypergeometric distribution analysis, the Chi Square Test, Fisher's Exact Test, the Mantel-Haenszel Test, and logistic regression analysis as just a few of the most commonly used statistical tools for examining adverse impact). The EEOC's “four-fifths rule” is also sometimes used as a rough proxy for measuring both statistical and practical significance.

55. *See id.* at 46 (stating that “[c]hoosing the appropriate test depends upon the nature of the selection process,” and, among other things, whether there are a “variable number of selections” or a “fixed number of selections”).

56. *See generally* DAVID COPE, *FUNDAMENTALS OF STATISTICAL ANALYSIS* (2005) (providing an introduction to statistical analysis for lawyers).

a. *P-Values*

Statisticians can “calculate the probability of observing a difference equal to or greater than that which actually occurred, assuming equal opportunity.”⁵⁷ This probability is called the “p-value,” and by convention, studies with “p-values” of less than 5% are generally considered “statistically significant.”⁵⁸ In other words, if the likelihood of a certain disparity occurring by chance (i.e., the “null hypothesis”) is less than 5%, then a “majority of scientific journals (and courts)” will deem that disparity to be statistically significant.⁵⁹

That said, the selection of the “alpha level” (the number that the “p-value” has to equal or less for the study to be considered statistically significant) is somewhat arbitrary.⁶⁰ Some researchers have therefore moved towards reporting the “specific probability (p) of getting a particular result.”⁶¹ Instead of saying that a certain disparity is “significant at the 0.05 level,” these researchers might specify that “ $p = 0.016$,” allowing “the consumer of the data to decide the significance” of the data in question.⁶²

b. *Confidence Intervals*

Alternatively, statisticians can measure the probability that the “true value” of a parameter lies within a certain “range of values.”⁶³ This probability is usually referred to as a “confidence interval,” the “margin of error,” or an “interval estimate.”⁶⁴ Confidence intervals are “often used” instead of (or in addition to) p-values to measure statistical significance.⁶⁵ A

57. *Jones v. City of Boston*, 752 F.3d 38, 43 (1st Cir. 2014).

58. *Id.*

59. COPE, *supra* note 56, at 40.

60. *See id.* (“[N]othing in the definition of statistical significance singles out 0.05 as *the* level that must be met for the null hypothesis to be rejected.”). Multiple regression analysis also requires analysts to choose which variables to include in their equations. This can add another element of subjectivity into the process. *See* D. James Greiner, *Causal Inference in Civil Rights Litigation*, 122 HARV. L. REV. 533, 545 (2008).

61. COPE, *supra* note 56, at 41 n.10.

62. *Id.*

63. *Id.* at 48.

64. *Cf. id.* (using the term “interval estimate”).

65. Wayne W. LaMorte, *Random Error: Confidence Intervals and p-Values*, BOS. SCH. PUB. HEALTH, http://sphweb.bumc.bu.edu/otlt/MPH-Modules/EP/EP713_RandomError/EP713_RandomError6.html (last modified June 6, 2016).

95% cutoff point for statistical significance is common for confidence intervals.⁶⁶ Expressed in statistical terms, “if the null value is not contained within the 95% confidence interval, then the probability that the null is the true value is less than 5%.”⁶⁷ This means the corresponding p-value “must be [less than] 0.05,” and the documented disparity is likely statistically significant.⁶⁸

c. Standard Deviations

Standard deviations are a measure of how much a certain result differs from the mean of the relevant data set. For example, if a 5’6” height requirement for firefighter applicants would exclude 90% of female applicants but only 10% of all applicants, the distance between those two numbers (90% and 10%) could be measured in terms of standard deviations. There is a rough correlation between standard deviations and confidence intervals, with two standard deviations “roughly correspond[ing] to a [95%] confidence interval, meaning that there is a [5%] chance that the disparity is random,” whereas “three standard deviations corresponds to roughly a 99.75% confidence interval, meaning that there is a 0.25% chance that the disparity is random.”⁶⁹ In the Title VII context, a disparity may usually be considered “statistically significant when it is more than two or three standard deviations from the expected rates.”⁷⁰

Again, the details and respective merits of different tests for measuring statistical significance are beyond the scope of this Note. It is enough to observe that each of these tests seeks to calculate the likelihood that a certain disparity is due to something more than chance. At the same time, each of these tests requires the arbitrary selection of a certain value. Statistical significance is really not a yes/no proposition so much as it is a range in which statisticians can calculate their respective confidence that a certain result is due to more than chance.⁷¹ To

66. *Id.*

67. *Id.*

68. *Id.*

69. See, e.g., Jennifer L. Peresie, *Toward a Coherent Test for Disparate Impact Discrimination*, 84 IND. L.J. 773, 786 (2009).

70. *Id.* (citing *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 308 n.14 (1977)); see also *Brown v. Nucor Corp.*, 785 F.3d 895, 908 (4th Cir. 2015) (noting that *Hazelwood’s* standard deviation test is commonly used for disparate impact claims).

71. See *supra* notes 60–62 and accompanying text.

express that certainty as a yes/no proposition requires the arbitrary selection of a certain value, below or above which the result can be deemed “statistically significant.”

2. Practical Significance

Unlike statistical significance, practical significance measures the “real-world importance” of a statistical disparity.⁷² In other words, practical significance asks whether a certain disparity is large enough to matter.⁷³ If the size of the disparity is “negligible,” “practical significance is lacking.”⁷⁴ Practical significance, like statistical significance, can be measured in many ways. Two bear mention here: (a) the EEOC’s four-fifths rule and (b) the “case-by-case” analysis method.

a. *The EEOC’s Four-Fifths Rule*

One method⁷⁵ for measuring practical significance is the EEOC’s “four-fifths rule,” which states that “[a] selection rate for any race, sex, or ethnic group which is less than four-fifths . . . of the rate for the group with the highest rate will generally be regarded by the federal enforcement agencies as evidence of adverse impact.”⁷⁶ Unfortunately, the four-fifths rule is not clear on whether its reference to the “selection rate” for any group refers to the firing rate or retention rate for that group.⁷⁷

72. COPE, *supra* note 56, at 52–54.

73. *Jones v. City of Boston*, 752 F.3d 38, 49 (1st Cir. 2014) (describing the practical significance inquiry as asking whether a disparity “is sufficiently large” instead of asking whether that disparity “is nonrandom”); *see also* THOMAS, *supra* note 49, at 25 (describing the practical significance inquiry as an “additional question” that asks whether an “observed difference[]” is “big enough to be important from a practical perspective”).

74. FED. JUDICIAL CTR., REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 252 (3d ed. 2011). That said, in some fields, even “very small differences . . . can be of enormous significance” if the consequence of the disparity is significant, or if a very large sample size is involved, such that even a 0.1% difference might affect thousands of people. COPE, *supra* note 56, at 52–54.

75. As noted above, the four-fifths rule can be used as a rough measure of both statistical and practical significance. *See, e.g., McKinley, supra* note 47, at 185 (describing the four-fifths rule as one of several ways to show *statistical* significance).

76. 29 C.F.R. § 1607.4(D) (2015); *see also McKinley, supra* note 47, at 177–78 (giving an example of how the four-fifths rule can be applied in practice).

77. *See, e.g., Jones, 752 F.3d* at 51–52 (explaining how this ambiguity can lead to conflicting conclusions even with the same data); *Council 31, Am. Fed’n of State, Cnty., & Mun. Emps. v. Ward*, 978 F.2d 373, 376, 379 (7th Cir. 1992) (giving one example of this problem).

This “often leads to inconsistent application of the rule.”⁷⁸ Many circuits have rejected mechanical application of the EEOC’s four-fifths rule, but some still use it as a rough rule of thumb.⁷⁹

b. The Case-by-Case Approach

Those courts that have rejected mechanical application of the EEOC’s four-fifth rules often phrase the practical significance test as an analysis of whether a certain disparity is “significant” enough to matter.⁸⁰ These courts largely do not distinguish between statistical significance and practical significance.⁸¹ Phrased this way, the practical significance inquiry can take into account not just the quantitative size of the impact, but the “qualitative nature and weight” of its impact.⁸² For example, if a police department with 500 black officers and 1200 white officers “implements a policy leading to the termination of [ninety] black officers and no white officers,” the court might well determine that this disparity is practically significant, even if the retention rate of black officers (82%) is more than four-fifths of the retention rate of white officers (100%).⁸³ Similarly, a court might distinguish between an employment practice that leads to the firing of one of six black employees and a practice that leads to the firing of 100 of 600 black employees, even though these two situations would be treated identically under the EEOC’s four-fifths rule.⁸⁴ Those courts that have rejected mechanical application of the EEOC’s four-fifths rule often stress the flexible, fact-dependent nature of their alternative analysis.⁸⁵ Either way, practical significance is not a precise or purely mathematical concept.⁸⁶ Instead, it in-

78. Stenger, *supra* note 6, at 423 n.71.

79. See, e.g., McKinley, *supra* note 47, at 182–85 (describing the split among circuit courts regarding how much deference should be given to the four-fifths rule); Stenger, *supra* note 6, at 422 n.65 (collecting a few cases).

80. See cases cited *infra* Part I.C.1.

81. See cases cited *infra* Part I.C.1.

82. *Jones*, 752 F.3d at 52 (citing Steve Goodman, *The Dirty Dozen: Twelve P-Value Misconceptions*, 45 SEMINARS IN HEMATOLOGY 135, 136–37 (2008)).

83. *Id.* at 51.

84. *Id.* at 52.

85. See, e.g., *Waisome v. Port Auth. of N.Y. & N.J.*, 948 F.2d 1370, 1376 (2d Cir. 1991) (“Courts should take a ‘case-by-case approach’ in judging the significance or substantiality of disparities, one that considers not only statistics but also all the surrounding facts and circumstances.” (citing *Ottaviani v. State Univ. of N.Y. at New Paltz*, 875 F.2d 365, 372–73 (2d Cir. 1989))); see also cases cited *infra* Part I.C.1.

86. *Jones*, 752 F.3d at 50 (observing that “the concept of practical signifi-

vites social, legal, and common sense determinations about whether a certain disparity ought to be large enough to result in liability for a given employer.⁸⁷

3. Statistical vs. Practical Significance: Why This Distinction Matters

Statistical and practical significance often go together, but not always. As the American Statistical Association explains:

Statistical significance is not equivalent to scientific, human, or economic significance. Smaller *p*-values do not necessarily imply the presence of larger or more important effects, and larger *p*-values do not imply a lack of importance or even lack of effect. Any effect, no matter how tiny, can produce a small *p*-value if the sample size or measurement precision is high enough⁸⁸

In other words, with a good study, “even very small differences can be statistically significant.”⁸⁹ This is especially common in studies with “large sample sizes,” where the greater amount of available data decreases the likelihood that a certain disparity is simply due to chance.⁹⁰ For example, in *Jones v. City of Boston*, the challenged employment practice (a drug testing policy) adversely affected 1.3% of African-American officers and cadets and 0.3% of white officers and cadets.⁹¹ Over eight years, this employment practice disqualified 55 out of 4222 African-American officers and cadets but only 30 out of 10,835 white officers and cadets.⁹² This sample size was large enough to yield a statistical disparity that corresponded to a standard deviation of 7.14 from the expected mean.⁹³ This disparity was statistically significant.⁹⁴ The police department argued, however, that the difference between 0.3% and 1.3% was not practically significant because very few blacks or whites

cance is impossible to define in even a remotely precise manner,” as there is no generally accepted “objective measure of practical significance”).

87. *Id.*

88. Am. Statistical Assoc. Bd. of Dirs., *ASA Statement on Statistical Significance and P-Values*, AM. STATISTICIAN, June 9, 2016, at 132.

89. *Jones*, 752 F.3d at 49.

90. *Id.* at 53; *cf.* Stenger, *supra* note 6, at 420–21 (noting that courts will generally allow plaintiffs to “aggregate data from numerous years in order to make their proffered sample statistically significant” but may “question[] the viability of aggregation by scrutinizing its probative value in light of other considerations”).

91. *Jones*, 752 F.3d at 41.

92. *Id.* at 44.

93. *Id.* at 45.

94. *Id.* at 43–44.

failed the drug test policy.⁹⁵ The result in *Jones* thus turned on whether the plaintiffs had to prove both statistical and practical significance as part of their prima facie disparate impact case.⁹⁶

Jones is just one example of a case where statistical significance does not necessarily equate to practical significance. The next Section of this Note collects cases from several circuits that have required practical significance separately from statistical significance.

C. THE CIRCUIT SPLIT ON PRACTICAL SIGNIFICANCE

The courts are split on whether to define “disparate impact” in terms of statistical or practical significance. A few circuits have tackled this issue head-on and come down on opposite sides of the debate. Other circuits have used language that only inferentially suggests that they may or may not require a showing of practical significance separately from statistical significance. Other circuits do not appear to distinguish between statistical and practical significance at all. This Section summarizes the existing case law from all twelve circuits on this issue. Subsection 1 discusses precedents from six circuits that require something more than statistical significance for disparate impact claims. Subsection 2 discusses the three circuits that have reached the opposite conclusion. Subsection 3 discusses case law from three circuits that have yet to weigh in on this issue.

1. Circuits Requiring a Showing of Practical Significance

Six circuits—the Second, Fourth, Fifth, Sixth, Ninth, and Eleventh—take a holistic approach to disparate impact claims, considering both statistical and practical significance. All of these circuits have indicated that something more than statistical significance may be required for disparate impact claims in some cases.

95. See *id.* at 42 (“A very small percentage of officers and cadets, either white or black, tested positive for cocaine during the period covered by this lawsuit.”); *id.* at 48–49 (summarizing the defendant’s practical significance argument). More specifically, only 55 out of 4222 African-American officers and 30 out of 10,835 white officers tested positive for cocaine in the eight-year period that was covered by the lawsuit. *Id.* at 45.

96. Compare *id.* at 46–48 (concluding that there was no genuine dispute that the statistical disparity was statistically significant), with *id.* at 48–53 (declining to adopt an additional practical significance requirement).

The Second Circuit has concluded that “[c]ourts should take a ‘case-by-case approach’ in judging the significance or substantiality of disparities, one that considers not only statistics but also all the surrounding facts and circumstances.”⁹⁷ Even a statistically significant disparity may not be sufficient to establish a prima facie case of disparate impact if that disparity is of “limited magnitude” (meaning that the actual size of the disparity is relatively small).⁹⁸ A more recent case from the Second Circuit focuses on statistical significance but quotes approvingly the portion of *Waisome* that discusses using a “case-by-case approach” to statistics.⁹⁹

In the same way, the Fourth Circuit has also noted that statistical significance “is not always synonymous with legal significance.”¹⁰⁰ Statistical disparities that are greater than two standard deviations are generally considered statistically significant.¹⁰¹ But the usefulness of statistical evidence “depends on all of the surrounding facts and circumstances.”¹⁰² Thus, even though statistical significance is an important part of disparate impact theory,¹⁰³ it is not always a sufficient predicate for a disparate impact claim.

The Fifth Circuit has taken a similar approach. In one case, the Fifth Circuit affirmed a grant of summary judgment for the defendant, even where the difference between the ex-

97. *Waisome v. Port Auth. of N.Y. & N.J.*, 948 F.2d 1370, 1376 (2d Cir. 1991) (citing *Ottaviani v. State Univ. of N.Y. at New Paltz*, 875 F.2d 365, 372–73 (2d Cir. 1989)).

98. *Id.*

99. *Chin v. Port Auth. of N.Y. & N.J.*, 685 F.3d 135, 153 (2d Cir. 2012); see also *Burgis v. N.Y.C. Dep’t of Sanitation*, 798 F.3d 63, 69 (2d Cir. 2015) (concluding that a § 1981 or Equal Protection case could be “based on statistics alone, [but] the statistics must not only be statistically significant in the mathematical sense, but they must also be of a level that makes other plausible non-discriminatory explanations very unlikely”).

100. *Brown v. Nucor Corp.*, 785 F.3d 895, 908 (4th Cir. 2015) (noting that the two-standard deviation test is the most common statistical test for measuring disparate impact).

101. *Id.*

102. *Id.* (quoting *Int’l Bhd. of Teamsters v. United States*, 431 U.S. 324, 340 (1977)). *Teamsters*, however, was a systemic disparate treatment case, so courts should not assume that every disparate impact case will also require the same level of buttressing anecdotal evidence.

103. See *Lilly v. Harris-Teeter Supermarket*, 720 F.2d 326, 336 & n.17 (4th Cir. 1983); *EEOC v. United Va. Bank/Seaboard Nat’l*, 615 F.2d 147, 151–53 (4th Cir. 1980); *Vanguard Justice Soc., Inc. v. Hughes*, 471 F. Supp. 670, 698 (D. Md. 1979), *decision supplemented*, 592 F. Supp. 245 (D. Md. 1984) (referring to a “statistical showing of a significant discrepancy”).

pected and actual pass rates for a minority group equated to 3.93 standard deviations.¹⁰⁴ This difference would typically be considered statistically significant under the standard deviation test discussed above, but the Fifth Circuit concluded that the “7.1% selection differential between black and white applicants” was not large enough to support a disparate impact claim.¹⁰⁵ In another case, the Fifth Circuit suggested that it might not be enough for there to be “a statistically significant correlation between test scores and experimental ratings” if “that . . . correlation [was] of very low magnitude and lack[ed] practical significance.”¹⁰⁶ The Fifth Circuit also noted in dicta that the EEOC’s regulations required a showing of “[p]ractical or operational significance,”¹⁰⁷ not just statistical significance.

The Sixth Circuit has also suggested that a showing of practical significance may be required in some cases. In *Isabel v. City of Memphis*, the Sixth Circuit rejected strict adherence to the EEOC’s four-fifths rule and advocated for a case-by-case approach to statistical analysis.¹⁰⁸ In doing so, it quoted with approval the portion of the EEOC regulation that “[s]maller differences in selection rate may nonetheless constitute adverse impact, where they are significant in both statistical and practical terms”¹⁰⁹ Even the dissent in this case appeared to assume that any statistical disparity had to be significant in both statistical and practical terms.¹¹⁰

The Ninth Circuit’s position on this question is unclear, but two of its cases have looked to both statistical and practical significance to determine whether a disparate impact claim can

104. *Moore v. Sw. Bell Tel. Co.*, 593 F.2d 607, 608 (5th Cir. 1979) (per curiam). The 3.93 standard deviations figure is not from *Moore* itself, but from a district court opinion summarizing *Moore*. See *Groves v. Ala. State Bd. of Educ.*, 776 F. Supp. 1518, 1528 (M.D. Ala. 1991).

105. *Moore*, 593 F.2d at 608.

106. *Ensley Branch of NAACP v. Seibels*, 616 F.2d 812, 818 (5th Cir. 1980), cert. denied, 449 U.S. 1061 (1980); see also *Fisher v. Procter & Gamble Mfg. Co.*, 613 F.2d 527, 545 (5th Cir. 1980) (making a similar observation in another case).

107. *Ensley*, 616 F.2d at 818 n.15; cf. *Moore*, 593 F.2d at 608 & n.1 (concluding that a 7.1% selection differential was not large enough to qualify as a “substantially disproportionate impact,” as was required to make out an adverse impact claim).

108. *Isabel v. City of Memphis*, 404 F.3d 404 (6th Cir. 2005).

109. *Id.* at 412 (quoting 29 C.F.R. § 1607.4(D) (2015)).

110. *Id.* at 418 (Batchelder, J., dissenting) (“[T]he . . . statistical evidence . . . , while statistically significant, does not rise to the level of a Title VII infraction.”).

go forward.¹¹¹ Alternatively, the Ninth Circuit seems to have left open the possibility of some sort of hybrid test allowing courts to look at both statistical and practical significance, but without making either one a strict prerequisite for establishing a disparate impact claim.¹¹²

Since the Eleventh Circuit split off from the Fifth Circuit in 1981, Fifth Circuit cases decided before that year (including the *Moore* and *Ensley* decisions discussed above) are binding in the Eleventh Circuit.¹¹³ Later cases from the Eleventh Circuit have also drawn a distinction between statistical and practical significance and required proof of both for disparate impact claims.¹¹⁴

2. Circuits That Do Not Require a Showing of Practical Significance

Three circuits—the First, Third, and Tenth—have taken a contrary position. These circuits only require a showing of statistical significance, and not practical significance.

The First Circuit recently held that “a plaintiff’s failure to demonstrate practical significance cannot preclude that plaintiff from relying on competent evidence of statistical significance to establish a prima facie case of disparate impact.”¹¹⁵ It reached this decision after criticizing the four-fifths rule at some length and observing that, apart from the four-fifths rule, it knew of “no statute, regulation, or case law proposing any other mathematical measure of practical significance.”¹¹⁶

The Third Circuit has also definitively rejected the practical significance requirement. In *Stagi v. National Railroad*

111. See *Rudebusch v. Hughes*, 313 F.3d 506, 515–16, 516 n.1 (9th Cir. 2002); *Clady v. County of Los Angeles*, 770 F.2d 1421, 1428–29 (9th Cir. 1985); cf. *Penk v. Or. State Bd. of Higher Educ.*, No 80-436 FR, 1985 WL 25631, at *34 (D. Or. Feb. 13, 1985), *aff’d*, 816 F.2d 458 (9th Cir. 1987) (distinguishing between statistical and practical significance in evaluating a systemic disparate treatment claim).

112. Cf. *Rudebusch*, 313 F.3d at 528 (arguing that even if the defendant “didn’t understand statistical significance, he would have had to see from the practical effect of the raises that he was causing . . . discrimination”).

113. *Gee v. Boyd*, 471 U.S. 1058, 1059 n.3 (1985) (citing *Bonner v. City of Prichard*, 661 F.2d 1206 (11th Cir. 1981) (en banc)); *Hamer v. City of Atlanta*, 872 F.2d 1521, 1525 (11th Cir. 1989).

114. See, e.g., *Ensley Branch of NAACP v. Seibels*, 31 F.3d 1548, 1555 (11th Cir. 1994); *Groves v. Ala. State Bd. of Educ.*, 776 F. Supp. 1518, 1527–28 (M.D. Ala. 1991).

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

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

Passenger Corp., the Third Circuit asserted: “We can identify no Court of Appeals that has found ‘practical significance’ to be a requirement for a plaintiff’s prima facie case of disparate impact”¹¹⁷ Then, after dismissing the reference to practical significance in the EEOC’s four-fifths rule, the Third Circuit concluded that because “‘practical’ significance has not been adopted by our Court, and no other Court of Appeals requires a showing of practical significance, we decline to require such a showing as part of a plaintiff’s prima facie case.”¹¹⁸ Consistent with this rule, a more recent decision from the Third Circuit used only standard deviation analysis—a test for statistical significance—to determine whether a plaintiff had established a prima facie disparate impact claim.¹¹⁹

The Tenth Circuit has also declined to adopt a practical significance requirement, but only in a case involving the Age Discrimination in Employment Act (ADEA). In *Apsley v. Boeing Co.*, the Tenth Circuit observed that the defendant had “cite[d] no cases supporting a formal ‘practical significance’ requirement at the summary judgment stage.”¹²⁰ The Tenth Circuit conceded that even very small disparities could be statistically significant with large sample sizes but stressed that the “observed [employment] disparity persisted over the course of eight or nine thousand individual recommendations and offers.”¹²¹ Even though *Apsley* was an ADEA case, if anything, the permissible scope of disparate impact claims under Title VII is broader than the scope of similar claims under the ADEA.¹²²

117. *Stagi v. Nat’l R.R. Passenger Corp.*, 391 F. App’x 133, 139 (3d Cir. 2010). The accuracy of this statement is somewhat questionable in light of the cases discussed above, but perhaps the Third Circuit was only counting circuits that have *explicitly* adopted a practical significance requirement, without counting circuits that have adopted a more holistic approach that evaluates both statistical and practical significance.

118. *Id.* at 140.

119. *Meditz v. City of Newark*, 658 F.3d 364, 372 (3d Cir. 2011).

120. *Apsley v. Boeing Co.*, 691 F.3d 1184, 1199 (10th Cir. 2012).

121. *Id.* On the other hand, even though the Tenth Circuit may have disagreed with some of the district court’s reasoning, it affirmed the district court’s grant of summary judgment for the defendant because the plaintiffs failed to prove a “systemwide pattern or practice” of discrimination. *Id.* at 1200. The plaintiffs’ evidence indicated, at best, that the defendants might have engaged in “isolated or sporadic instances of . . . discrimination.” *Id.*

122. *See Smith v. City of Jackson*, 544 U.S. 228, 240 (2005).

3. Circuits Without Clear Precedent on the Need for Practical Significance

Three circuits—the D.C., Seventh, and Eighth Circuits—have not clearly indicated whether proof of practical significance should be required for disparate impact claims.

A dissenting D.C. Circuit judge once pointed out that “statistical significance is not the same as practical significance because in isolation . . . [statistical significance] tells nothing about the importance or magnitude of the differences.”¹²³ But this was a passing comment in a Federal Communications Commission case.¹²⁴ In the Title VII context, at least three district court opinions from the D.C. Circuit have focused on statistical significance, rather than practical significance, to evaluate the sufficiency of disparate impact claims.¹²⁵

The Seventh Circuit has repeatedly used standard deviation analysis to determine whether a certain disparity is statistically significant.¹²⁶ On the other hand, the Seventh Circuit has encouraged courts to avoid using the two-to-three standard deviations test as a bright line test, leaving open the application of some sort of practical significance test.¹²⁷ And at least one district court case from the Seventh Circuit has argued for a more thorough and holistic approach for evaluating *both* statistical *and* practical significance.¹²⁸

123. *Bilingual Bicultural Coal. on Mass Media, Inc. v. FCC*, 595 F.2d 621, 642 n.57 (D.C. Cir. 1978) (en banc) (Robinson, J., dissenting in part).

124. *Id.* The main question in this case was whether the FCC had a duty to give two minority associations a hearing to challenge the FCC’s renewal of the broadcast licenses for two radio stations that allegedly discriminated against racial minorities. *See id.* at 624. Furthermore, this D.C. Circuit judge made this comment only after noting that the FCC could “employ standards *different* from those utilized by the EEOC in carrying out its mandate” because the FCC’s main concern was with “intentional discrimination,” whereas the EEOC was “chartered to search out and remedy both discriminatory intent and discriminatory effect.” *Id.* at 642 (emphasis added).

125. *See Delgado v. Ashcroft*, No. 99-2311 (JR), 2003 WL 24051558, at *8 (D.D.C. May 29, 2003); *Hatcher-Capers v. Haley*, 786 F. Supp. 1054, 1063 (D.D.C. 1992); *Reynolds v. Sheet Metal Workers Local 102*, 498 F. Supp. 952, 966–67 (D.D.C. 1980), *aff’d*, 702 F.2d 221 (D.C. Cir. 1981). What these cases did not clarify is whether a showing of statistical significance is a sufficient, necessary, or just possible element of a disparate impact claim.

126. *See, e.g., Bew v. City of Chicago*, 252 F.3d 891, 893 (7th Cir. 2001); *Adams v. Ameritech Servs., Inc.*, 231 F.3d 414, 426–27 (7th Cir. 2000); *Coates v. Johnson & Johnson*, 756 F.2d 524, 536–40 (7th Cir. 1985).

127. *Coates*, 765 F.2d at 537 n.11, 547 n.22 (quoting *EEOC v. Am. Nat’l. Bank*, 652 F.2d 1176, 1192 (4th Cir. 1981)).

128. *See EEOC v. Sears, Roebuck & Co.*, 628 F. Supp. 1264, 1286–88 (N.D.

The Eighth Circuit has also used standard deviation analysis to evaluate the sufficiency of disparate impact claims.¹²⁹ It has stated that “[a] difference of two or three standard deviations is statistically significant at the [5%] significance level” and observed that a difference of 5.5 standard deviations (corresponding with a pass rate of 98% of whites but only 84% of blacks) is “highly unlikely” to be “due to chance.”¹³⁰ On the other hand, in another part of its opinion, it concluded that this difference in admission rates was “significant in both statistical and practical terms,” suggesting that it might be using standard deviation analysis in conjunction with the actual disparity of admitted students to establish both statistical and practical significance.¹³¹

Disparate impact claims require proof that a certain employment practice has a “disparate impact” on members of a protected class.¹³² Statistical significance is a test for determining whether a statistical disparity is due to something more than random chance.¹³³ Practical significance is a test for determining whether that disparity is large enough to have real-world importance.¹³⁴ Most disparities that are statistically significant will also be practically significant, but not always.¹³⁵ Neither Congress nor the Supreme Court has clarified whether Title VII requires plaintiffs to prove that a certain employment practice disproportionately affects members of a protected class in a way that is both statistically and practically significant, or just the former.¹³⁶ This has created a split among the circuits on this issue.¹³⁷

Ill. 1986), *aff'd*, 839 F.2d 302 (7th Cir. 1988).

129. *Hameed v. Int'l Ass'n of Bridge, Structural & Ornamental Iron Workers, Local Union No. 396*, 637 F.2d 506, 514 (8th Cir. 1980).

130. *Id.* at 513–14.

131. *Id.* at 514.

132. *See supra* Part I.A.

133. *See supra* Part I.B.1.

134. *See supra* Part I.B.2.

135. *See supra* Part I.B.3.

136. *See supra* Part I.C.

137. *See supra* Part I.C.

II. PRACTICAL PROBLEMS WITH THE PRACTICAL SIGNIFICANCE STANDARD

This Part argues that practical significance should not be required for disparate impact claims. It makes both legal and practical arguments against the adoption of the practical significance standard. Section A explains how the ordinary meaning of the words “disparate” and “impact” in Title VII can refer to any statistically significant disparity, no matter how small. Section B notes the inconclusiveness of the legislative history on this issue. Section C explains why past Supreme Court precedents need not be interpreted as requiring a practical significance requirement. Section D argues that the EEOC’s implicit endorsement of the practical significance requirement should not receive controlling deference from the courts. Section E suggests that practical significance should not be required for disparate impact claims because practical significance is basically a test for proximate causation—a requirement lacking from the actual text of Title VII. Finally, Section F argues that the current formulation of the practical significance test is problematic because it is so amorphous as to be essentially arbitrary.

A. THE TEXT DOES NOT SUPPORT A PRACTICAL SIGNIFICANCE STANDARD

Title VII does not define the term “disparate impact.”¹³⁸ Even so, the text of Title VII is instructive in several ways.

First, courts may look to the common dictionary definition of a statutory term when that term is not separately defined in the statute.¹³⁹ Here, courts must decide whether the term “disparate impact” refers to *any* difference in how an employment practice affects a protected class, or only a *large* difference in how the employment practice affects that class.

On one hand, several dictionaries define the word “disparate” as meaning “[f]undamentally distinct or different in kind,”¹⁴⁰ “entirely dissimilar,”¹⁴¹ “markedly distinct in quality or

138. *Cf.* 42 U.S.C. § 2000e (2012) (defining many other terms, but not “disparate impact”).

139. *Schwab v. Reilly*, 560 U.S. 770, 783 (2010) (“[W]e may look to dictionaries . . . to determine the meaning of words the Code does not define . . .”).

140. *Disparate*, THE AMERICAN HERITAGE DICTIONARY (5th ed. 2015).

141. *Id.*

character,”¹⁴² or “essentially different or diverse in kind.”¹⁴³ *Oxford’s English Dictionary* gives this example of how the word can be used: “As remote in their nature . . . as any two disparate things we can propose or conceive”¹⁴⁴ Arguably, then, even statistically significant differences may not be “disparate” impacts under Title VII if they are so small that they are not “markedly” “distinct in quality or character.”

On the other hand, *Oxford’s English Dictionary* also defines “disparate” as meaning “essentially different,” “dissimilar,” “unlike,” “distinct,” or “unequal.”¹⁴⁵ It also gives several examples where authors have referred to two things that are “utterly” or “very” disparate.¹⁴⁶ If the word “disparate,” standing alone, only encompassed large or extreme differences between two things, it would be redundant to qualify it with the adverbs “utterly” or “very.” Arguably, then, the word “disparate” refers to any statistically significant disparity of any size.

Second, courts generally should not read into a statute additional requirements that Congress has not itself placed in the statute.¹⁴⁷ Title VII does not require plaintiffs to prove that an employment practice had a “large” impact on a protected class.¹⁴⁸ Title VII just requires plaintiffs to prove that “a particular employment practice” had a disparate impact¹⁴⁹ on a protected class. Thus, courts should not read an additional magnitude requirement into Title VII’s disparate impact claims.

Title VII only requires proof of a “disparate impact,” not proof of a “very” disparate impact that is large enough to warrant societal or moral condemnation. Thus, any time a challenged employment practice impacts a protected class more harshly than a non-protected class, that should be enough to establish a Title VII violation, whether the challenged employment practice affects the protected class in a small or large way.

142. *Disparate*, MERRIAM-WEBSTER DICTIONARY, <http://www.merriam-webster.com/dictionary/disparate> (last visited Nov. 3, 2016).

143. *Disparate*, OXFORD ENGLISH DICTIONARY, <http://www.oed.com/view/Entry/54914?redirectedFrom=disparate#eid> (last visited Nov. 3, 2016).

144. *Id.* (quoting 2 THOMAS BURNET, *THE SACRED THEORY OF THE EARTH* 302 (1719)).

145. *Id.*

146. *Id.*

147. *See, e.g.*, *Desert Palace, Inc. v. Costa*, 539 U.S. 90, 98–99 (2003).

148. *See* 42 U.S.C. § 2000e-2(k)(1)(A) (2012).

149. *Id.* § 2000e-2(k)(1)(A)(i).

At this point, proponents of a practical significance requirement may argue that the preceding argument proves too much. To be consistent, they might say, the preceding argument would also require abolition of the statistical significance requirement. If the smallest disparity is enough to establish a Title VII violation, does that mean even a statistically *insignificant* disparity should be enough to support a disparate impact claim? Not necessarily, because the statistical significance requirement, unlike the practical significance requirement, is grounded in a different part of the text. In expounding a statute, courts must “look to the provisions of the whole law,”¹⁵⁰ keeping in mind that “a word is known by the company it keeps.”¹⁵¹ In this case, Title VII requires plaintiffs to prove that a particular employment practice “*cause[d]* a disparate impact” on the basis of a protected characteristic.¹⁵² Statistical significance, unlike practical significance, is at least inferentially related to causation.¹⁵³

For these reasons, although the text of Title VII is somewhat ambiguous, the better view is to read the text as requiring only a showing of statistical significance. Disparate impact plaintiffs should only have to prove that the challenged employment practice had “a” disparate impact on a protected class, however small.

B. THE LEGISLATIVE HISTORY DOES NOT SUPPORT A PRACTICAL SIGNIFICANCE STANDARD

The legislative history for the Civil Rights Act of 1964, the Civil Rights Act of 1990, and the Civil Rights Act of 1991 is inconclusive. But it contains three statements that suggest that Title VII’s disparate impact provisions were originally intended to only require a showing of statistical significance.

First, during the debate on the original Civil Rights Act of 1964, a senator read from an article which quoted an expert statistician as saying that a variance of 2.2% between the passage rate of minorities and non-minorities was “not statistically significant,” but a “variance of 6[%]. . . could be considered significant.”¹⁵⁴ Neither this senator nor any other senator made

150. *Massachusetts v. Morash*, 490 U.S. 107, 115 (1989) (quoting *Pilot Life Ins. Co. v. Dedeaux*, 481 U.S. 41, 51 (1987)).

151. *Gustafson v. Alloyd Co., Inc.*, 513 U.S. 561, 575 (1995).

152. See 42 U.S.C. § 2000e-2(k)(1)(B)(i) (emphasis added).

153. See *supra* Parts I.B.1, I.B.2.

154. 110 CONG. REC. 9041 (1964) (statement of Sen. John Tower).

any reference to any additional requirement of “practical significance.”

Second, in a Senate committee hearing for the Civil Rights Act of 1990 (an early precursor to the Civil Rights Act of 1991 that was ultimately enacted into law), a labor economist noted that “even a statistically significant disparity” between the employment rates of minorities and non-minorities might not “constitute[] conclusive evidence of race . . . discrimination by that employer.”¹⁵⁵ He explained that “statistical significance” only measured whether a disparity was “due to something other than pure ‘chance,’” but “even if the hypothetical disparity were found to be ‘statistically significant,’” this did “not mean that the disparity [was] necessarily due to discrimination.”¹⁵⁶ These statements apparently assumed that statistical significance was the relevant test for disparate impact claims.¹⁵⁷

Third, during a House subcommittee hearing for the Civil Rights Act of 1991, a lawyer and former commissioner of the EEOC suggested that the term “disparate impact” should be defined in terms of statistical significance.¹⁵⁸ She said the following:

Although . . . [the Civil Rights Act of 1991] does not contain a definition as to what “disparate impact” is, I would hope that the Committee would consider using the definitions previously used by the Supreme Court to limit the term by applying standards of statistical significance of two or three standard deviations¹⁵⁹

Again, as discussed above, standard deviations are a measure of statistical significance (i.e., statistical correlation), not practical significance.¹⁶⁰

All three of these statements referred specifically to statistical significance. Read together, they suggest that Congress understood the Civil Rights Act of 1991 to only require the plaintiff to prove that a certain employment practice had a sta-

155. *Hearing on S. 2104 Before the S. Comm. on Labor & Human Res.*, 101st Cong. 298 (1990) (statement of Paul J. Andrisani, Professor, Human Resource Management, Temple University).

156. *Id.* at 299.

157. *See id.* at 298.

158. *Hearings on H.R. 1 Before the H. Subcomm. on Civil & Constitutional Rights*, 102d Cong. 158–59 (statement of Cathie A. Shattuck, Partner, Epstein, Becker & Green, P.C.).

159. *Id.* Shattuck also made a passing reference to the EEOC’s four-fifths rule in her testimony. But that reference is not inconsistent with an implicit endorsement of a statistical significance standard because the EEOC’s four-fifths rule can be used as a rough proxy for statistical *or* practical significance.

160. *See supra* Part I.B.1.

tistically significant disparate impact on members of a protected class.

C. SUPREME COURT PRECEDENT DOES NOT SUPPORT A PRACTICAL SIGNIFICANCE STANDARD

The Supreme Court has never clearly specified “just what threshold mathematical showing of variance” is required for disparate impact claims.¹⁶¹ Its precedents, however, do not clearly support a practical significance standard requirement for two reasons.

First, *Griggs* itself does not require any showing of practical significance. Rather, it broadly prohibits the use of any employment practice “which operates to exclude” minorities.¹⁶² An employment practice that results in a statistically significant disparity “operates to exclude” minorities, whether it only excludes one or a million minority employees.¹⁶³ Thus, *Griggs* suggests that any statistically significant disparity should be sufficient to establish a prima facie disparate impact claim under Title VII.¹⁶⁴

Second, in *Hazelwood School District v. United States*, the Court stated that a “difference between the expected value and the observed number” that “is greater than two or three standard deviations” will generally be sufficient to infer some sort of violation of Title VII.¹⁶⁵ As explained above, standard deviations

161. LINDEMANN, *supra* note 14, at 124 (quoting *Moore v. Southwestern Bell Tel. Co.*, 593 F.2d 607, 608 (5th Cir. 1979) (per curiam)).

162. *Griggs v. Duke Power Co.*, 401 U.S. 424, 431 (1971).

163. It is true that *Griggs* struck down a high school degree requirement that disqualified African Americans at a “substantially higher rate than white applicants.” *Id.* at 426 (emphasis added). This, however, was the Court’s description of the effect of the employment practice in *Griggs*, not the actual rule from *Griggs*. The fact that a “substantially” discriminatory employment practice may be sufficient to establish a Title VII violation does not mean that a showing of a “substantial” disparity is a necessary part of every disparate impact claim. Further, even if a “substantial” disparity is a necessary part of every disparate impact claim, that still leaves unanswered the question of whether “substantiality” only requires a showing of statistical significance or a showing of both statistical and practical significance.

164. See *Griggs*, 401 U.S. 424. A Fifth Amendment case, *Washington v. Davis*, 426 U.S. 229 (1976), has described Title VII as prohibiting hiring practices that “disqualif[y] substantially disproportionate numbers of blacks.” *Id.* at 247 (emphasis added). *Davis* was a Fifth Amendment case, not a Title VII case, however, so its statements about the scope of Title VII are dicta. *Cf. id.* at 246–48 (declining to import Title VII standards wholesale into the Court’s Fifth Amendment jurisprudence).

165. *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 309 n.14 (1977)

are a measure of statistical significance, not practical significance.¹⁶⁶ *Hazelwood* was a systematic disparate treatment case, not a disparate impact case. However, following *Hazelwood's* lead, many courts have used the same standard deviation test to evaluate the statistical significance of disparate impact claims.¹⁶⁷

Proponents of a practical significance requirement may point out that the Court has described the prima facie test (in five cases decided between 1975 and 1988) as requiring the plaintiff to prove that the challenged employment practice resulted in a “significant statistical disparity”¹⁶⁸ or a “significant adverse effect[]” on a protected group¹⁶⁹ or otherwise harmed the plaintiff in a “significantly different,”¹⁷⁰ “significantly greater,”¹⁷¹ or “significantly discriminatory” way.¹⁷² But none of these cases specifically required a showing of practical significance. These references to a “significantly” different impact on minority employees can be read as requiring proof of statistical or practical significance.

Proponents of a practical significance requirement may also point out that a plurality of the Supreme Court stated in *Watson v. Fort Worth Bank* that the test for Title VII has “never been framed in terms of any rigid mathematical formula.”¹⁷³ The same plurality disclaimed exclusive reliance on standard deviation analysis and instead urged that “the ‘significance’ or ‘substantiality’ of numerical disparities” should be evaluated “on a case-by-case basis.”¹⁷⁴ To the extent that statistical significance tends to be more of a mathematical concept, and practical significance tends to be a more flexible approach, this statement tends to support a practical significance requirement for disparate impact claims. However, these two quotes come from Part II-D of Justice O’Connor’s opinion. This part of Jus-

(quoting *Castaneda v. Partida*, 430 U.S. 482, 497 n.17 (1977)).

166. See *supra* Parts I.B.1, I.B.2.

167. See, e.g., *Jones v. City of Boston*, 752 F.3d 38, 44 (1st Cir. 2014); *Meditz v. City of Newark*, 658 F.3d 364, 372 (3d Cir. 2011).

168. *Watson v. Fort Worth Bank & Tr.*, 487 U.S. 977, 1004 (1988) (Blackmun, J., concurring).

169. *Id.* at 986 (plurality opinion).

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

171. *N.Y. City Transit Auth. v. Beazer*, 440 U.S. 568, 585–87 (1979).

172. *Connecticut v. Teal*, 457 U.S. 440, 446 (1982); *Dothard v. Rawlinson*, 433 U.S. 321, 329 (1977).

173. *Watson*, 487 U.S. at 994–95.

174. *Id.* at 995 n.3.

tice O'Connor's opinion only commanded four votes¹⁷⁵ and thus is not controlling.¹⁷⁶

D. THE EEOC'S FOUR-FIFTHS RULE SHOULD NOT BE GIVEN CONTROLLING DEFERENCE

Proponents of the practical significance requirement sometimes cite to the EEOC's four-fifths rule to support their position. As discussed above, the EEOC's four-fifths rule is sometimes used as a rough proxy for practical significance.¹⁷⁷ The rule itself states that "[a] selection rate for any race, sex, or ethnic group which is less than four-fifths . . . of the rate for the group with the highest rate will generally be . . . evidence of adverse impact," but "[s]maller differences . . . may nevertheless constitute adverse impact, where they are significant in *both statistical and practical* terms."¹⁷⁸ In the EEOC's view, then, disparate impact claims require a showing of both statistical and practical significance.

This interpretation of Title VII, however, should not receive significant deference from the courts. If Congress has "delegated authority to the agency generally to make rules carrying the force of law," and "the agency interpretation . . . was promulgated in the exercise of that authority," courts must give *Chevron* deference to the agency's interpretation of the statute.¹⁷⁹ *Chevron* deference means that courts must defer to the agency's interpretation of the statute as long as the agency's interpretation is a permissible reading of an ambiguous statute.¹⁸⁰ This is a "highly deferential" standard that requires courts to defer even to many agency interpretations with which they disagree.¹⁸¹

175. See *id.* at 982 (recording that only three Justices joined Parts II.C and II.D of Justice O'Connor's opinion).

176. When no one opinion commands a majority of the Court, the controlling holding is the position taken by the Justices "who concurred in the judgment[] on the narrowest grounds." *Marks v. United States*, 430 U.S. 188, 193 (1977). In *Watson*, this would have been Justice Stevens' opinion, not Parts II.C and II.D of Justice O'Connor's opinion.

177. See *supra* Part I.B.2.

178. 29 C.F.R. § 1607.4(D) (2015) (emphasis added).

179. *United States v. Mead Corp.*, 533 U.S. 218, 226–27 (2001).

180. *Chevron, U.S.A., Inc. v. Nat. Res. Def. Council, Inc.*, 467 U.S. 837, 843–44 (1984).

181. See Elliot Greenfield, *A Lenity Exception to Chevron Deference*, 58 BAYLOR L. REV. 1, 51 & n.316 (2006) (noting that there are only a handful of cases where the Supreme Court has refused to defer to an agency interpretation at Step 2 of *Chevron*).

In this case, however, the EEOC does not have “rulemaking authority to interpret the substantive provisions of Title VII.”¹⁸² It only has congressional authority to issue “procedural regulations”¹⁸³ governing “the administrative stage of discrimination complaints.”¹⁸⁴ This means that the EEOC’s four-fifths rule is not entitled to *Chevron* deference from the courts.¹⁸⁵

Instead, the EEOC’s four-fifths rule is only entitled to *Skidmore* deference.¹⁸⁶ This means that it will be followed by the courts only in light of “the thoroughness evident in its consideration, the validity of its reasoning, its consistency with earlier and later pronouncements, and all those factors which give it power to persuade, if lacking power to control.”¹⁸⁷ The EEOC has provided little justification for its four-fifths rule, and both scholars and courts have often criticized this rule.¹⁸⁸ At various times, scholars have referred to this test as “sufficiently inflexible and statistically dubious”¹⁸⁹ or as an “inappropriate benchmark” in cases where not “all groups are equally well qualified on average.”¹⁹⁰ Courts have also criticized the mechanical application of the four-fifths rule to studies with small

182. *Edelman v. Lynchburg Coll.*, 535 U.S. 106, 122 (2002) (O’Connor, J., concurring); *see also* *EEOC v. Arabian Am. Oil Co.*, 499 U.S. 244, 257 (1991) (citing *Gen. Elec. Co. v. Gilbert*, 429 U.S. 125, 140–46 (1976)).

183. 42 U.S.C. § 2000e-12(a) (2012).

184. *Ebbert v. DaimlerChrysler Corp.*, 319 F.3d 103, 110 (3d Cir. 2003).

185. *See* *United States v. Mead Corp.*, 533 U.S. 218, 226–27 (2001); *Stagi v. Nat’l R.R. Passenger Corp.*, 391 F. App’x 133, 138 (3d Cir. 2010) (reaching the same conclusion specifically in regards to the four-fifths rule); Thomas W. Merrill, *Step Zero After City of Arlington*, 83 *FORDHAM L. REV.* 753, 760 (2014) (noting that the Supreme Court has consistently given only *Skidmore* deference to EEOC interpretations).

186. *See* *Arabian Am. Oil Co.*, 499 U.S. at 257.

187. *Skidmore v. Swift & Co.*, 323 U.S. 134, 140 (1944).

188. *See, e.g.*, *Watson v. Fort Worth Bank & Tr.*, 487 U.S. 977, 995 n.3 (1988) (observing that this rule “has been criticized on technical grounds”); *Jones v. City of Boston*, 752 F.3d 38, 51–52 (1st Cir. 2014) (listing three problems with the four-fifths rule and citing other circuit court cases that have “minimized the importance of the four-fifths rule” or “criticized it directly”).

189. Lawrence Rosenthal, *Saving Disparate Impact*, 34 *CARDOZO L. REV.* 2157, 2197 n.187 (2013); *see also* Joseph L. Gastwirth & Weiwen Miao, *Formal Statistical Analysis of the Data in Disparate Impact Cases Provides Sounder Inferences than the U.S. Government’s “Four-Fifths” Rule: An Examination of the Statistical Evidence in Ricci v. DeStefano*, 8 *L. PROB. & RISK* 171, 187–90 (2009) (comparing the four-fifths rule to other statistical methods and concluding that “formal statistical analysis of . . . data . . . in disparate impact [cases] yields sounder inferences than simply relying on the ‘four-fifths’ rule”).

190. Amy L. Wax, *Disparate Impact Realism*, 53 *WM. & MARY L. REV.* 621, 662 (2011).

sample sizes that “tend to produce inherently unreliable results.”¹⁹¹ This is not to say that the four-fifths rule cannot be a “helpful rule of thumb” for employers wishing to avoid potential Title VII liability without engaging in a more robust statistical analysis.¹⁹² The fact remains, however, that the four-fifths rule has not been persuasively justified by the EEOC or consistently followed by the courts.¹⁹³ As such, the four-fifths rule’s passing reference to practical significance should not be given much deference from the courts.

E. PRACTICAL SIGNIFICANCE IS ESSENTIALLY A TEST OF LEGAL/PROXIMATE CAUSATION, WHICH SHOULD NOT BE READ INTO THE TEXT OF TITLE VII

At their core, arguments in favor of a practical significance requirement may stem from the belief that the law should not impose liability on employers who adopt employment practices that unintentionally affect minorities in a small or remote way. As a policy matter, this is a reasonable position, as many other areas of law distinguish between intentional and unintentional wrongdoers and cut off the liability of unintentional wrongdoers for harm that is unforeseeable or otherwise too remote to have been proximately caused by the defendant’s behavior.

Title VII, however, does not include any sort of proximate cause requirement.¹⁹⁴ All that it requires is for the plaintiff to prove that the challenged employment practice “cause[d]” a disparate impact on the basis of a protected characteristic.¹⁹⁵ As discussed above, statistical significance is a better measure than practical significance of the correlation between a challenged employment practice and the alleged effect of that employment practice on a protected class.¹⁹⁶ Although it can never prove causation, it can at least measure the likelihood that a

191. EEOC v. Joint Apprenticeship Comm. of Joint Indus. Bd. of the Elec. Indus., 164 F.3d 89, 97 (2d Cir. 1998).

192. *Jones*, 752 F.3d at 51.

193. See cases cited *supra* note 188.

194. See generally Sandra F. Sperino, *Discrimination Statutes, the Common Law, and Proximate Cause*, 2013 U. ILL. L. REV. 1 (2013) (arguing that courts should not read a proximate cause element into most discrimination statutes); Sandra F. Sperino, *Statutory Proximate Cause*, 88 NOTRE DAME L. REV. 1199 (2013) (same).

195. 42 U.S.C. § 2000e-2(k)(1)(A)(i) (2012).

196. See *supra* Part I.B.1.

certain disparity is due to something more than random chance.¹⁹⁷

F. THERE IS NO OBJECTIVE WAY TO MEASURE PRACTICAL SIGNIFICANCE

The main argument against adopting a practical significance requirement is the difficulty in coming up with a workable and consistent way to test for practical significance, apart from the four-fifths rule, which has its own problems, as discussed above. The First Circuit has observed that “the concept of practical significance is impossible to define in even a remotely precise manner,” and there is no objective or mathematical way to measure practical significance, apart from the four-fifths rule.¹⁹⁸ As discussed above, those who have attempted to define practical significance without relying on the four-fifths rule have usually defined practical significance as meaning a difference that is “big enough to be important from a practical perspective.”¹⁹⁹ That formulation is inoffensive but useless. It does nothing more than to beg the question. In practice, then, despite the superficial attractions of a practical significance requirement, it is extremely difficult to apply it “in any principled and predictable manner.”²⁰⁰

To be fair, this problem is not unique to practical significance. Other areas of law also employ flexible, multi-factor tests. For example, *Skidmore* deference has been described as a “sliding scale” approach under which the deference owed to an agency’s interpretation of a statute depends on an ad hoc weighing of several factors.²⁰¹ Other subjects like torts,²⁰² contracts,²⁰³ federal jurisdiction,²⁰⁴ or constitutional law²⁰⁵ also reg-

197. See *supra* Part I.B.1.

198. *Jones v. City of Boston*, 752 F.3d 38, 50, 52 (1st Cir. 2014).

199. THOMAS, *supra* note 49, at 25; discussion *supra* Part I.B.2.

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

201. *United States v. Mead Corp.*, 533 U.S. 218, 250 (2001) (Scalia, J., dissenting). See also *Doe v. Leavitt*, 552 F.3d 75, 81 (1st Cir. 2009); Kristin E. Hickman & Matthew D. Krueger, *In Search of the Modern Skidmore Standard*, 107 COLUM. L. REV. 1235, 1259 (2007).

202. After all, tort law is the home of the “reasonable person” standard that terrorizes thousands of first-year law students every year. See RESTATEMENT (SECOND) OF TORTS §§ 282–283 (1965).

203. For example, some courts define “unconscionability” as meaning a contract which “is so grossly unreasonable or unconscionable in the light of the mores and business practices of the time and place as to be unenforceable [sic] according to its literal terms.” *Gillman v. Chase Manhattan Bank, N.A.*, 534 N.E.2d 824, 828 (N.Y. 1988) (quoting *Mandel v. Liebman*, 100 N.E.2d 149, 152

ularly require courts to exercise “reasoned judgment”²⁰⁶ on a case-by-case basis. Even statistical significance analysis requires statisticians to pick a certain “significance level” for the study.²⁰⁷

But the current test for practical significance is different. It does not just allow the court to engage in “reasoned judgment.”²⁰⁸ Courts have been doing that for years.²⁰⁹ Instead, the test for practical significance “begin[s] by begging the question.”²¹⁰ It gives the judge no factors to guide her in making her decision. It just asks the judge to decide whether a given disparity is practically significant based on the judge’s “common sense.” Because this formulation is practically impossible to apply “in any principled and predictable manner,”²¹¹ courts should not require practical significance in Title VII cases.

To recap, the text of Title VII only requires a showing of “disparate” impact, not a “large” or “practically significant” disparity.²¹² The legislative history for Title VII does not support a practical significance requirement.²¹³ Supreme Court precedent is unclear on this question.²¹⁴ The EEOC has endorsed a practical significance requirement, but courts need not defer to the EEOC’s interpretation of this portion of Title VII.²¹⁵ Statistical significance is a better measure of actual causation than practical significance.²¹⁶ And the courts have largely failed to come up with a workable definition of practical significance.²¹⁷

(N.Y. 1951)).

204. Even impersonal concepts like justiciability can require courts to make judgment calls on a case-by-case basis. *See, e.g.*, *Gunn v. Minton*, 133 S. Ct. 1059, 1066 (2013) (extending federal jurisdiction to state law claims that necessarily implicate “substantial” federal questions).

205. *See, e.g.*, *Planned Parenthood of Se. Pa. v. Casey*, 505 U.S. 833, 986 (1992) (Scalia, J., concurring and dissenting in part) (describing the joint opinion’s “undue burden” test as an “inherently manipulable” standard).

206. *Id.* at 849 (joint opinion).

207. *See* discussion *supra* Part I.B.1.

208. *Casey*, 505 U.S. at 849 (joint opinion).

209. *Id.*

210. *Id.* at 982 (Scalia, J., dissenting and concurring in part).

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

212. *See supra* Part II.A.

213. *See supra* Part II.B.

214. *See supra* Part II.C.

215. *See supra* Part II.D.

216. *See supra* Part II.E.

217. *See supra* Part II.F.

III. WHAT THEN? THE CASE FOR EVALUATING ONLY STATISTICAL SIGNIFICANCE OR DEVELOPING A MORE CONCRETE TEST FOR PRACTICAL SIGNIFICANCE

The courts are split on whether practical significance should be required separately from statistical significance. And those courts that require some showing of practical significance have little guidance to aid them in determining how to measure practical significance. This Part proposes two alternative solutions to this issue. Section A suggests that plaintiffs should not be required to prove practical significance as part of their prima facie cases. It suggests that the phrase “disparate impact” in Title VII should be read as referring broadly to any statistical disparity that is statistically significant, regardless of the size of that disparity. It also explains how courts can rely on other legal doctrines to address the policy concerns behind the practical significance requirement. Alternatively, Section B suggests that courts can keep the practical significance requirement but address the vagueness issue with that requirement by developing a more concrete test for measuring practical significance.

A. COURTS SHOULD ABOLISH THE PRACTICAL SIGNIFICANCE REQUIREMENT

This Section makes the case for abolishing the practical significance requirement as part of the plaintiff’s prima facie case. This means that disparate impact plaintiffs should be able to make out a prima facie case of disparate impact by simply showing that a challenged employment practice impacts minority employees in a statistically significant way. Subsection 1 argues that abolishing the practical significance requirement would further the purposes of Title VII. Subsection 2 explains how the policy concerns embodied in the practical significance can be addressed using other existing Title VII rules.

1. Abolishing the Practical Significance Requirement Would Further the Purposes of Title VII

This Subsection argues that abolishing the practical significance requirement would further the purposes of Title VII in three ways. First, this would help eliminate barriers to employment opportunity that keep minority applicants and employees from getting positions for which they are perfectly qualified. Second, to the extent that disparate impact claims are justified as a way to “smoke out” hidden discrimination, abol-

ishing the practical significance requirement would encourage hiring based on merit rather than unarticulated (or sometimes even unconscious) biases. Third, because statistical significance is a more objective measure than practical significance, relying solely on statistical significance would make it easier for employers to comply with Title VII ahead of time to avoid exposure to liability.²¹⁸

First, disparate impact theory can help eliminate barriers to employment opportunities for all applicants and employees. Even if an employment practice does not discriminate on the basis of race or gender, it can discourage employers from hiring or promoting the most qualified applicants or employees.²¹⁹ Consider the case of an employer who refuses to hire anyone with a criminal record. A blanket prohibition like this could adversely affect certain racial minorities. To address this problem, the EEOC issued policy statements in 1987, 1990, and 2012, encouraging employers to replace categorical bans on hiring convicted offenders with “a more nuanced approach, taking into account the nature of the offense, the nature of the position sought, and how much time has passed since the conviction.”²²⁰ If courts require plaintiffs to prove that a certain employment practice impacts a *large* number of people, employers will have less incentive to eliminate roadblocks to opportunity that only affect small numbers of their employees. Conversely, adopting a pure statistical significance standard for disparate impact claims would help eliminate small but statistically significant barriers to employment opportunity for all workers.

Second, disparate impact theory can help “smoke out” hidden discrimination, in both its intentional²²¹ and unintention-

218. This Note takes no position on whether Title VII’s primary purpose *should* be to eliminate barriers to employment opportunity, smoke out hidden discrimination, or make it easier for employers to voluntarily comply with its provisions. It simply argues that a statistical significance standard would further *all* of these policy goals.

219. See generally Joseph Fishkin, *The Anti-Bottleneck Principle in Employment Discrimination Law*, 91 WASH. U. L. REV. 1429 (2014) (describing how even facially non-discriminatory employment practices can prevent qualified workers from accessing employment opportunities).

220. *Id.* at 1460.

221. In the twenty-first century, employers who *want* to discriminate against minorities may use coded language or unnecessarily restrictive employment requirements to screen out applicants whom they do not wish to hire. Cf. Richard Dvorak, *Cracking the Code: “De-Coding” Colorblind Slurs During the Congressional Crack Cocaine Debates*, 5 MICH. J. RACE & L. 611, 613 (2000) (noting that modern racism is often expressed in “coded” form to

al²²² forms. Abolishing the practical significance requirement as part of the plaintiff's prima facie case will allow more disparate impact claims to survive a motion to dismiss, require more employers to justify challenged employment practices on the basis of business necessity, and root out more cases of hidden discrimination in the process. Hopefully, this process will also help employers be more aware of their intentional *and* unintentional biases.

Third, abolishing the practical significance requirement would help employers manage their liability under Title VII. Congress did not want Title VII to impose excessive compliance costs on employers.²²³ This was the primary impetus behind Congress's decision to exempt small businesses from Title VII's coverage.²²⁴ Practical significance is a slippery (and often circularly defined) concept, whereas statistical significance has a generally accepted definition in employment law circles. Therefore, a pure statistical significance standard would make it easier for employers to know exactly what Title VII requires.

2. The Policy Concerns Embodied in the Practical Significance Requirement Can Be Addressed Using Other Title VII Rules

The practical significance requirement does limit the scope of disparate impact liability. This can be a good thing insofar as it limits government intrusion into private business decisions,²²⁵ especially when "very small impacts are unlikely to be

avoid public censure). As Justice Scalia observed in his concurrence in *Ricci*, a disparate impact is "sometimes (though not always) a signal of" invidious discrimination. *Ricci v. DeStefano*, 557 U.S. 557, 595 (2009). Unsurprisingly, then, statistical disparities can sometimes be used as evidence of *intentional* discrimination to support a disparate *treatment* claim. *See id.*

222. Most people have hidden prejudices that even they may not recognize. For the seminal work on this problem, see Charles R. Lawrence III, *The I, the Ego, and Equal Protection: Reckoning with Unconscious Racism*, 39 STAN. L. REV. 317 (1987). Disparate impact theory can help address these implicit biases by striking at employer conduct that reflects these unintentional biases. *See* Eang L. Ngov, *War and Peace Between Title VII's Disparate Impact Provision and the Equal Protection Clause: Battling for a Compelling Interest*, 42 LOY. U. CHI. L.J. 1, 29, 41–43 (2010).

223. *See, e.g.*, 110 CONG. REC. 13092 (1964) (statement of Sen. Cotton); *id.* 13092–93 (statement of Sen. Morse).

224. *See id.* at 13088 (statement of Sen. Humphrey).

225. *See* Texas Dep't of Hous. & Cmty. Aff. v. Inclusive Cmty. Project, Inc., 135 S. Ct. 2507, 2518 (2015) ("[D]isparate-impact liability must be limited so employers and other regulated entities are able to make the practical business choices and profit-related decisions that sustain a vibrant and dynamic free-enterprise system."); *Waisome v. Port Auth. of N.Y. & N.J.*, 948 F.2d 1370,

the product of intentional discrimination.”²²⁶ But courts can limit the scope of disparate impact liability without requiring a showing of practical significance for disparate impact claims. Three other rules already limit the scope of disparate impact liability: (a) the statistical significance requirement; (b) the causation requirement; and (c) the business necessity defense.

a. *The First Safeguard—Statistical Significance*

Statistical and practical significance tend to go hand-in-hand. In most cases, small disparities that are of a “limited magnitude” will neither be statistically nor practically significant. To the extent that courts believe that small statistical disparities should not support disparate impact claims, the statistical significance requirement should be enough, standing alone, to weed out most “limited magnitude” cases.²²⁷ Not only so, the statistical significance requirement is easier to apply in an objective manner than the practical significance requirement.²²⁸ And the statistical significance requirement may actually be more effective than the four-fifths rule at weeding out small-impact cases in situations with small sample sizes.²²⁹

Any discrepancy between statistical and practical significance usually arises only with large sample sizes. In *Jones*, for example, the parties brought suit only after collecting eight years of data with “thousands of test results.”²³⁰ However, such large sample sizes “are often unavailable.”²³¹ Even when such large samples are available, the time and expense needed to comb through reams of such data should protect employers from being sued on a daily basis for every conceivable employment practice.²³² And *if* a plaintiff is willing to go through this trouble to challenge an employment practice that affects hundreds, perhaps even thousands, of similarly situated applicants

1376–77 (2d Cir. 1991) (assuming that disparities that are of a “limited magnitude” should not result in disparate impact liability).

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

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

228. See discussion *supra* Part III.A.

229. See *supra* Part II.C (pointing out the failings of the four-fifths rule in this situation).

230. *Jones*, 752 F.3d at 44.

231. *Id.* at 53.

232. *Cf. id.* (noting the need for extensive data in these cases).

or employees, the defendant most likely will be a larger business with more resources to defend against the lawsuit.²³³

b. The Second Safeguard—Causation

Disparate impact law requires the plaintiff to prove that a specific employment practice caused the alleged disparate impact.²³⁴ This requirement “protects defendants from being held liable for racial disparities they did not create” and prevents Title VII from morphing into a de facto racial quota system.²³⁵ It also keeps every employment practice from being actionable as a disparate impact violation.

c. The Third Safeguard—The Business Necessity Defense

Until this point, this Note has focused on the plaintiff’s prima facie disparate impact claim. It is worth noting, however, that the employer retains an affirmative defense of business necessity. Employment practices that have a disparate impact on members of a protected class may nevertheless be permitted if they are “consistent with business necessity.”²³⁶ This requires, in most cases, a case-by-case evaluation of all of the evidence. For example, education requirements and height requirements may be justified by business necessity in some cases but not in others.²³⁷ In *Jones*, the plaintiffs conceded that the police department had a legitimate purpose in ensuring that its police

233. This concern with the size of the business being sued reflects the special concern for small businesses that Congress codified when it explicitly exempted small businesses from the provisions of the Civil Rights Act of 1964. See 42 U.S.C. § 2000e(b) (2012) (defining the term “employer” to include only businesses with fifteen or more employees).

234. See *id.* § 2000e-2(k)(1)(B). There is a limited exception that allows all employment practices to be analyzed together as one general cause of disparate impact.

235. *Texas Dep’t of Hous. & Cmty. Aff. v. Inclusive Cmty. Project, Inc.*, 135 S. Ct. 2507, 2523 (2015) (citing *Wards Cove Packing Co. v. Atonio*, 490 U.S. 642, 653 (1989)).

236. 42 U.S.C. § 2000e-2(k)(1)(A)(i); see also *Griggs v. Duke Power Co.*, 401 U.S. 424, 431 (1971) (“The touchstone is business necessity. If an employment practice which operates to exclude Negroes cannot be shown to be related to job performance, the practice is prohibited.”).

237. See, e.g., *Dothard v. Rawlinson*, 433 U.S. 321, 331 (1977) (striking down a height and weight requirement for prison guards when the defendant failed to show that there was a correlation between height, weight, and strength); *Griggs*, 401 U.S. at 425–26 (striking down a high school diploma requirement for blue-collar jobs at a power plant because this requirement was not “significantly related to successful job performance”).

officers did not use illegal drugs.²³⁸ The First Circuit thus remanded the case to the district court to consider the department's argument that its drug testing program was "consistent with business necessity."²³⁹

B. IF COURTS DO NOT WISH TO ABOLISH THE PRACTICAL SIGNIFICANCE REQUIREMENT, THEY SHOULD ADOPT A MORE CONCRETE TEST FOR MEASURING PRACTICAL SIGNIFICANCE

A less drastic approach would be to keep the practical significance requirement as a way for employers to rebut the *presumption* that a statistically significant disparity is a "substantial disparity" for the purposes of Title VII. If the courts keep the practical significance test, however, they should develop a more concrete test for measuring practical significance. This would avoid many of the problems with the current formulation of practical significance as an ad hoc "I know it when I see it" test.

To that end, this Note lists several factors that courts *could* consider in determining whether a disparity is substantial enough to be significant in practical terms.²⁴⁰

Courts could consider the percentage difference between the effects of an employment practice on minority and non-minority employees. In *Jones*, for example, where an employment practice disqualified 1.30% of black employees but only 0.27% of white employees, there was a 1.03% difference in how the employment practice affected black employees.²⁴¹

Alternatively, courts could evaluate the same disparity by dividing the percentage of affected non-minority employees by the percentage of affected minority employees. In *Jones*, for example, the challenged employment practice disqualified almost five times as many black employees as white employees, as $1.30\% \div 0.27\% = 4.81$.²⁴² This, incidentally, was how the plaintiffs in *Jones* characterized the statistical disparity.

Another option would be to look at the absolute percentage of individuals affected by the challenged employment practice.

238. *Jones v. City of Boston*, 752 F.3d 38, 54 (1st Cir. 2014).

239. *Id.* at 54–55.

240. This list is illustrative only. Brighter minds can debate the relative merits of different tests for practical significance. This Note only suggests that a multi-factor test for practical significance would be an improvement over a circular definition of practical significance.

241. *Jones*, 752 F.3d at 44.

242. *Id.*

In *Jones*, for example, the challenged employment practice disqualified 1.30% of the department's black officers and cadets.²⁴³

Yet another option would be to consider the absolute *number* of individuals affected by the challenged employment practice. The higher this number, the more practically significant the disparity, even if the actual percentage difference of individuals affected is relatively small. For example, if "a series of clinical trials produced statistically significant results suggesting that the survival rates associated with the two treatments were 67.2% and 67.3%," although a difference of 0.1% might seem relatively unimportant, "if 1 million people suffer from the type of heart disease in question, using the better treatment will save, on average, 1000 lives a year."²⁴⁴ In *Jones*, the department's drug testing policy disqualified fifty-five African-American police officers over an eight-year period.²⁴⁵

A slightly different approach would be to focus on the qualitative nature of a disparate impact rather than its qualitative size. After all, in the medical field, "a small percentage increase in [the] mild side effects" of a drug would probably be considered less significant than "the same percentage increase in fatalities."²⁴⁶ In the employment context, courts could similarly evaluate the qualitative effect of the challenged employment practice. Has it just resulted in some relatively "minor" adverse employment action? Or has it had more serious effects, like requiring the outright firing of several minority employees?

This is a "second-best" solution. This Note primarily argues that courts should simply abolish the practical significance requirement and rely on other rules to address the policy concerns behind the practical significance requirement.²⁴⁷ But if a court decides not to abolish the practical significance requirement, it should consider adopting a set of factors for measuring practical significance rather than simply telling district court judges to use their "common sense" to determine when a certain disparity is significant enough to matter.²⁴⁸ That test for practical significance is unhelpful.²⁴⁹

243. *Id.*

244. COPE, *supra* note 56, at 524.

245. *Jones*, 752 F.3d at 44.

246. *Id.* at 52.

247. *See supra* Part III.A.

248. *See supra* Part II.F.

249. *See supra* Part II.F.

In short, courts should either abolish the practical significance requirement or come up with a more concrete way for measuring practical significance.

CONCLUSION

Statistical and practical significance are two distinct concepts. Statistical significance measures the likelihood that a statistical anomaly is attributable to something more than chance. Practical significance focuses on the size of the observed disparity. Ignoring the distinction between these two concepts has obscured the development of disparate impact theory and has made it more difficult for judges to apply (and employers to comply with) Title VII.

The circuit courts are split on whether disparate impact claims require proof of both statistical and practical significance. The Supreme Court has also given very little guidance in this area. However, the plain meaning of the term “disparate impact” includes *any* discrepancy between two sets of results and thus suggests that any statistically significant disparity should satisfy the plaintiffs’ burden of proof on their disparate impact claims. The legislative history also indicates that the term “disparate impact” was meant to refer to all statistically significant disparities, without any additional requirement of practical significance. Further, there is no sound, objective standard for measuring practical significance, and the concerns behind the practical significance can be addressed more consistently and objectively under existing Title VII rules—specifically, the statistical significance requirement, the causation requirement, and the business necessity defense. This Note thus urges courts to require only proof of statistical significance in disparate impact claims, putting to rest the fuzzy concept of practical significance that has hitherto plagued our Title VII jurisprudence. Courts that insist on keeping the practical significance requirement should consider adopting a more concrete test for measuring practical significance.