

THE YALE LAW JOURNAL

KEVIN TOBIA

Disparate Statistics

ABSTRACT. Statistical evidence is crucial throughout disparate impact's three-stage analysis: during (1) the plaintiff's prima facie demonstration of a policy's disparate impact; (2) the defendant's job-related business necessity defense of the discriminatory policy; and (3) the plaintiff's demonstration of an alternative policy without the same discriminatory impact. The circuit courts are split on a vital question about the "practical significance" of statistics at Stage 1: Are "small" impacts legally insignificant? For example, is an employment policy that causes a one percent disparate impact an appropriate policy for redress through disparate impact litigation? This circuit split calls for a comprehensive analysis of practical significance testing across disparate impact's stages. Importantly, courts and commentators use "practical significance" ambiguously between two aspects of practical significance: the *magnitude* of an effect and *confidence* in statistical evidence. For example, at Stage 1 courts might ask whether statistical evidence supports a disparate impact (a confidence inquiry) and whether such an impact is large enough to be legally relevant (a magnitude inquiry). Disparate impact's texts, purposes, and controlling interpretations are consistent with confidence inquiries at all three stages, but not magnitude inquiries. Specifically, magnitude inquiries are inappropriate at Stages 1 and 3—there is no discriminatory impact or reduction too small or subtle for the purposes of the disparate impact analysis. Magnitude inquiries are appropriate at Stage 2, when an employer defends a discriminatory policy on the basis of its job-related business necessity.

AUTHOR. Yale Law School, J.D. expected; Yale Philosophy, Ph.D. expected; Rutgers University, B.A. 2012. I thank the *Yale Law Journal* staff, especially Notes Editors Greg Cui, Joe Falvey, and Urja Mittal. This argument's examples involve impacts on communities of which I am not a member. Such advocacy is "a touchy sort of subject," in the words of SJA Germanotta: "Can you stand up for people [when] you are not necessarily fully part of that community in a way that [members] can understand?" Most special thanks to Owen Fiss and the 2016 Community of Equals seminar participants who taught me a tremendous amount, including how to approach this question.



NOTE CONTENTS

INTRODUCTION	2384
I. FOUNDATIONS: DISPARATE IMPACT AND STATISTICAL CONCEPTS	2388
A. Disparate Impact: A Brief Overview	2388
B. Motivations and Purposes	2390
C. Statistical Concepts	2392
1. Statistical Significance	2392
2. Practical Significance	2394
II. DISPARATE STATISTICS	2397
A. The Statistical Standard of Prima Facie Disparate Impact	2398
B. The Statistical Standard of Job-Related Business Necessity	2407
C. The Statistical Standard of Showing a Suitable Alternative	2411
III. RECOMMENDATIONS AND IMPLICATIONS	2412
CONCLUSION	2419

INTRODUCTION

Statistical evidence is crucial in each stage of disparate impact's three-stage analysis: (1) the plaintiff's prima facie demonstration of a policy's disparate impact; (2) the defendant's job-related business necessity defense of the discriminatory policy; and (3) the plaintiff's demonstration of an alternative policy without the same discriminatory impact. There is a circuit split on the role of "practical significance" inquiries at the prima facie stage,¹ raising a fundamental question about disparate impact theory: Are such "small" — effects, about whose existence we are confident — legally insignificant? For example, is an employment policy that causes a one percent disparate impact an appropriate object of disparate impact litigation?

This question calls for a broader analysis of "practical significance" at each of disparate impact's three stages. Importantly, courts use "practical significance" in multiple ways. The present argument's primary focus is practical significance referring to the *magnitude* of an effect supported by statistical evidence. I call courts' evaluation of the size of an effect a "magnitude inquiry." Another sense of practical significance involves the strength of the inference from an empirical-statistical finding to the real world. I refer to a court's evaluation of this aspect of practical significance as a "confidence inquiry." This is an important distinction, and courts and commentators often use "practical significance" in ways that are ambiguous between these two aspects.² The second aspect — practical significance as the strength of the inference supported by statistical evidence — is obviously relevant to disparate impact analysis, in the same way that assessing the strength of the inference supported by evidence is al-

-
1. *E.g., compare* Jones v. City of Bos., 752 F.3d 38, 53 (1st Cir. 2014) (finding a prima facie disparate impact where there was a 1% difference in selection rates), *with* Frazier v. Garrison I.S.D., 980 F.2d 1514, 1524 (5th Cir. 1993) (holding that a 4.5% difference in selection rates was trivial); *see also* sources cited *infra* note 6. *Compare* 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 (2015) (arguing for practical significance testing at the prima facie stage), *and* Katie Eissenstat, Note, *Lies, Damned Lies, and Statistics: The Case To Require "Practical Significance" To Establish a Prima Facie Case of Disparate Impact Discrimination*, 68 OKLA. L. REV. 641 (2016) (same), *with* Elliott Ko, Note, *Big Enough To Matter: Whether Statistical Significance or Practical Significance Should Be the Test for Title VII Disparate Impact Claims*, 101 MINN. L. REV. 869, 881-87 (2016) (arguing for no practical significance testing at the prima facie stage).
 2. For example, Eissenstat, Stenger, and Ko focus on "practical significance" rather than distinguishing between magnitude and confidence inquiries. *See* sources cited *supra* note 1. While this argument agrees with Ko's conclusion regarding magnitude inquires at the prima facie stage — such inquiries are inappropriate — the analysis here employs different reasoning. These commentators also focus on only the first stage of disparate impact analysis.

ways relevant. A debate remains regarding “magnitude inquiries,” evaluations of whether some effect is sufficiently large, at each stage of analysis.

I argue that such magnitude inquiries are inappropriately used to evaluate whether a “large enough” prima facie disparate impact exists or whether an alternative policy with less discriminatory impact promises a “large enough” decrease in discriminatory impact, at the first and third stages of disparate impact litigation. However, magnitude inquiries are more appropriate when an employer defends a discriminatory policy on the basis of its job-related business necessity, at the second stage of disparate impact litigation. Thus, this argument’s primary contribution is an analysis of “magnitude inquiries,” one aspect of practical significance, across all three stages of disparate impact.

The Note proceeds in three parts. Part I describes disparate impact theory, highlighting the logic of the shifting burden of proof,³ and relevant statistical concepts. Part II analyzes statistics’ role at three stages of disparate impact analysis: the plaintiff’s establishment of prima facie disparate impact, the defendant’s rebuttal of establishing a test’s job-relatedness and business necessity, and the plaintiff’s proposal of a less discriminatory alternative policy. I argue that disparate impact law supports the rejection of magnitude inquiries for a plaintiff’s prima facie case of disparate impact and proposal of a less discriminatory alternative, but it supports a more robust magnitude inquiry during an employer’s establishment of a disparity-causing test’s job-relatedness and business necessity. Part III provides recommendations for improving the use of statistics in disparate impact analysis.

This Note contributes a defense of the First Circuit’s decision, which has previously been subjected to critical commentary.⁴ Importantly, it highlights the distinction between two aspects of “practical significance” sometimes obscured in disparate impact discussions: magnitude and confidence. The Note also contributes a comprehensive analysis of practical significance, providing recommendations for the use of statistics at all three stages of disparate impact litigation. In doing so, it calls for courts to reflect broadly about whether their use of statistics at each stage is consistent with their uses at the two other stages, their underlying theory of statistics and evidence, and their disparate impact theory.

3. The burden shifts from the plaintiff’s prima facie case of disparate impact to the defendant’s rebuttal (demonstrating job-relatedness and business necessity), then back to the plaintiff’s demonstration of an alternative measure that causes a lesser disparity.

4. See, e.g., Eissenstat, *supra* note 1; Stenger, *supra* note 1. *But see* Ko, *supra* note 1.

Given the amount⁵ and importance⁶ of disparate impact litigation, addressing key questions that can determine the outcome of these actions, such as courts' use of magnitude inquiries, can be of great consequence. Indeed, these issues have provoked controversy. Today, the role of "practical significance" in the prima facie stage of disparate impact analysis is at the heart of a circuit split. The First, Third, and Tenth Circuits oppose practical significance inquiries; the Second, Fourth, Fifth, Sixth, Ninth, and Eleventh Circuits endorse them; and the D.C., Seventh, and Eighth Circuits have no clear precedent.⁷

-
5. There have been hundreds of disparate impact cases. See Michael Selmi, *Was the Disparate Impact Theory a Mistake?*, 53 UCLA L. REV. 701 (2006).
 6. See, e.g., Robert Belton, *The Dismantling of the Griggs Disparate Impact Theory and the Future of Title VII: The Need for a Third Reconstruction*, 8 YALE L. & POL'Y REV. 223, 225-26 (1990).
 7. See Ko, *supra* note 1, at 881-87 (cataloging the circuit split in these terms). In support of the circuit split claim, Ko helpfully cites to (1) various cases opposing practical significance testing from the First Circuit: *Jones v. City of Bos.*, 752 F.3d 38, 53 (1st Cir. 2014); Third Circuit: *Meditz v. City of Newark*, 658 F.3d 364, 372 (3d Cir. 2011); *Stagi v. Nat'l R.R. Passenger Corp.*, 391 F. App'x 133, 139 (3d Cir. 2010); Tenth Circuit: *Apsley v. Boeing Co.*, 691 F.3d 1184, 1199 (10th Cir. 2012) (an Age Discrimination in Employment Act (ADEA) case); (2) various cases endorsing practical significance testing from the Second Circuit: *Burgis v. N.Y.C. Dep't of Sanitation*, 798 F.3d 63, 69 (2d Cir. 2015); *Chin v. Port Auth. of N.Y. & N.J.*, 685 F.3d 135, 153 (2d Cir. 2012); *Waisome v. Port Auth. of N.Y. & N.J.*, 948 F.2d 1370, 1376 (2d Cir. 1991); Fourth Circuit: *Brown v. Nucor Corp.*, 785 F.3d 895, 908 (4th Cir. 2015); Fifth Circuit: *Fisher v. Procter & Gamble Mfg. Co.*, 613 F.2d 527, 545 (5th Cir. 1980); *Ensley Branch of NAACP v. Seibels*, 616 F.2d 812, 818 n.15 (5th Cir. 1980); *Moore v. Sw. Bell Tel. Co.*, 593 F.2d 607, 608 (5th Cir. 1979); Sixth Circuit: *Isabel v. City of Memphis*, 404 F.3d 404 (6th Cir. 2005); Ninth Circuit: *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); and Eleventh Circuit: *Ensley Branch of NAACP v. Seibels*, 31 F.3d 1548, 1555 (11th Cir. 1994); and (3) various cases indicating no clear precedent on practical significance testing from the D.C. Circuit: *Delgado v. Ashcroft*, No. Civ. A. 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); Seventh Circuit: *Bew v. City of Chicago*, 252 F.3d 891, 894 (7th Cir. 2001); *Adams v. Ameritech Servs., Inc.*, 231 F.3d 414, 426-27 (7th Cir. 2000); *EEOC v. Sears, Roebuck & Co.*, 628 F. Supp. 1264, 1286-88 (N.D. Ill. 1986), *aff'd*, 839 F.2d 302 (7th Cir. 1988); *Coates v. Johnson & Johnson*, 756 F.2d 524, 536-40 (7th Cir. 1985); and Eighth Circuit: *Hameed v. Int'l Ass'n of Bridge, Structural & Ornamental Iron Workers, Local Union No. 396*, 637 F.2d 506, 514 (8th Cir. 1980).

There is some room for debate about whether all of these courts endorse, oppose, or remain neutral (respectively) on the question of prima facie practical significance. For instance, some of the Circuits that are counted as endorsing practical significance might instead be read as considering "practical significance" in the sense of whether the statistical evidence is good evidence of a disparity, not in the sense of whether the real-world disparity is of a certain magnitude. In *Waisome* and *Nucor*, the Second and Fourth Circuits, respectively, endorse case-by-case approaches in which statistical significance should often be interpreted with consideration of surrounding circumstances; this is far from a clear endorsement of

Before turning to the analysis, it is worth noting that these legal questions arise against a particular scientific and cultural backdrop: the danger of relying on mere statistical significance in interpreting empirical studies is the subject of scientific and increasingly popular concern, and looking to “practical significance” is a popular remedy.⁸ Calls to move science beyond simple statistical significance testing are not exclusive to the current moment,⁹ nor are calls to move toward some form of practical significance testing.¹⁰ Unreflective reliance on scientific trends might suggest that practical significance inquiries of all forms – including magnitude inquiries – are necessary parts of sound methodology, including throughout disparate impact analysis.

This Note cautions otherwise.¹¹

practical significance testing in the sense of requiring a particular magnitude of a prima facie disparate impact. There is certainly good evidence for the existence of *some* circuit split on the question of statistical significance in the sense of a disparity’s magnitude – compare, for example, the First and Fifth Circuits: *Jones* against *Moore*. But the magnitude of the circuit split on this question may not be as large as is sometimes suggested.

Of course, the more cautious interpretation of the circuit split’s breadth does not imply the unimportance of analyzing practical significance usage; in fact, it implies the opposite. The breadth of the split is difficult to assess precisely because courts use “practical significance” in ambiguous and divergent ways. To assess statistical evidence’s “practical significance” – in the sense of weighing how the evidence bears on the inference of a real-world disparate impact – is clearly a useful and legitimate inquiry. To assess statistical evidence’s “practical significance” – in the sense of weighing whether the real-world disparity is *large enough* – is more controversial.

8. See, e.g., Eric Loken & Andrew Gelman, *Measurement Error and the Replication Crisis*, 355 SCI. 584 (2017); Regina Nuzzo, *Statistical Errors*, 506 NATURE 150 (2014); Amy Gallo, *A Refresher on Statistical Significance*, HARV. BUS. REV. (Feb. 16, 2016), <http://hbr.org/2016/02/a-refresher-on-statistical-significance> [<http://perma.cc/Q96D-3PC3>]; see also Christie Aschwanden, *Science Isn’t Broken: It’s Just a Hell of a Lot Harder than We Give It Credit for*, FIVETHIRTYEIGHT (Aug. 19, 2015), <http://fivethirtyeight.com/features/science-isnt-broken> [<http://perma.cc/R6W6-ZUNA>] (providing a popular science commentary on statistical significance and “p-hacking”, the process of influencing or manipulating statistical significance testing results by making choices about data such as which groups to include in the analysis or what factors to control).
9. See, e.g., James K. Skipper, Jr., Anthony L. Guenther & Gilbert Nass, *The Sacredness of .05: A Note Concerning the Uses of Statistical Levels of Significance in Social Science*, 2 AM. SOCIOLOGIST 16 (1967).
10. See, e.g., Roger E. Kirk, *Practical Significance: A Concept Whose Time Has Come*, 56 EDUC. & PSYCHOL. MEASUREMENT 746 (1996).
11. As will become clear, I do not conclude that courts should *never* look to effect size when assessing whether there is a prima facie disparate impact or suitable alternative policy. For instance, in assessing a suitable alternative policy with less discriminatory impact, courts *should* look to a comparison of effect sizes, asking whether the suitable alternative policy’s disparate impact is of a magnitude lesser than that of the challenged policy. See *infra* Part II.

I. FOUNDATIONS: DISPARATE IMPACT AND STATISTICAL CONCEPTS

This Part provides an overview of disparate impact litigation and its three-stage burden-shifting framework: the plaintiff's prima facie demonstration of a disparate impact, the defendant's job-related business necessity defense, and the plaintiff's demonstration of a suitable alternative policy with less discriminatory impact. Then, I describe disparate impact theory's fundamental aims, the purpose of each stage, and the two key statistical concepts: statistical significance and practical significance. The discussion of practical significance outlines the fundamentally different aspects of practical significance testing that courts use: "magnitude inquiries" evaluate whether an effect is sufficiently large to be legally relevant, while "confidence inquiries" evaluate whether statistical evidence sufficiently supports a claim. For instance, in evaluating whether a prima facie showing of disparate impact has been made, a court might examine whether the impact is sufficiently large (for instance, is a one percent disparity legally relevant?) or whether the evidence supports the claim that the policy caused a disparity.

A. Disparate Impact: A Brief Overview

Title VII of the Civil Rights Act of 1964 prohibits workplace discrimination on the basis of protected characteristics: race, color, religion, sex, and national origin.¹² In early opinions, courts read the Act to protect individuals against intentional discrimination.¹³ In 1971, the Supreme Court articulated a broader understanding of Title VII in *Griggs v. Duke Power Co.*, the landmark decision that introduced disparate impact theory.¹⁴ *Griggs* held that Title VII prohibits "not only overt discrimination but also practices that are fair in form, but discriminatory in operation."¹⁵ This theory of disparate impact allows a plaintiff to recover when an employer implements a test or policy that adversely affects a protected group. Unlike disparate treatment, disparate impact does not require employer animus or particular intentions.¹⁶ The "touchstone" of disparate im-

12. Civil Rights Act of 1964, Pub. L. No. 88-352, 78 Stat. 241 (codified at 42 U.S.C. § 2000e (2012)).

13. See, e.g., *Int'l Bhd. of Teamsters v. United States*, 431 U.S. 324 (1977).

14. 401 U.S. 424 (1971).

15. *Id.* at 431.

16. See *Faulkner v. Super Valu Stores, Inc.*, 3 F.3d 1419, 1428 (10th Cir. 1993) (distinguishing disparate impact from intentional discrimination by noting that disparate impact "does not require proof of discriminatory motive or intent"). For a more recent statement comparing disparate impact and disparate treatment, see *Ricci v. DeStefano*, 557 U.S. 557 (2009).

pact theory, according to the *Griggs* Court, is business necessity.¹⁷ In order to justify a practice that has a discriminatory impact, an employer must show that the disparity-causing practice is a business necessity.

Post-*Griggs* decisions refined disparate impact theory. Notably, in 1975, the Court in *Albemarle Paper Co. v. Moody* outlined a three-part burden-shifting framework for disparate impact litigation.¹⁸ The Supreme Court stepped back from this approach in *Wards Cove Packing Co. v. Atonio*,¹⁹ limiting *Griggs* by modifying the standard of business necessity to require merely a “legitimate business justification” for a discriminatory practice.²⁰ But two years later, the Civil Rights Act of 1991 superseded *Wards Cove*, restoring the disparate impact framework preceding *Wards Cove*.²¹

The Civil Rights Act of 1991 codified disparate impact theory developed in case law, including the three-part burden-shifting framework from *Albemarle Paper Co.*²² Under this framework, the plaintiff (the employee) must first make a prima facie demonstration that a policy or practice has a disparate impact on the plaintiff’s protected class.²³ Next, the defendant (the employer) must demonstrate that its policy or practice is “job related” and “consistent with business necessity.”²⁴ If the defendant meets this burden, the plaintiff has the burden of demonstrating that there is a suitable alternative employment practice with less discriminatory impact.²⁵ The plaintiff can recover if the employer fails to meet its burden at the second stage or if the plaintiff meets his or her burden at the third stage.

17. *Griggs*, 401 U.S. at 431-32 (“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 . . . Congress has placed on the employer the burden of showing that any given requirement must have a manifest relationship to the employment in question.”).

18. 422 U.S. 405, 425 (1975).

19. 490 U.S. 642 (1989).

20. *Id.* at 660.

21. Civil Rights Act of 1991, Pub. L. No. 102-166, § 105, 105 Stat. 1071, 1074 (codified in scattered sections of 42 U.S.C.) (rejecting the *Wards Cove* standard and adopting the *Griggs* and *Albemarle* standard).

22. *Id.*; see also *Ricci v. DeStefano*, 557 U.S. 557, 624 (2009) (explaining that the Civil Rights Act of 1991 formally codified Title VII disparate impact theory as it stood prior to *Wards Cove*).

23. 42 U.S.C. § 2000e-2(k)(1)(A)(i) (2012).

24. *Id.*

25. *Id.* § 2000e-2(k)(1)(A)(ii).

This language in the Civil Rights Act of 1991 indicates an intention to codify the principles of *Griggs* and its legacy,²⁶ including the three-part burden-shifting test articulated in *Albemarle Paper*.²⁷ It also echoes the language of job-relatedness and business necessity: after the prima facie demonstration of adverse impact, a defendant must “demonstrate that the challenged practice is job related for the position in question and consistent with business necessity.”²⁸

B. Motivations and Purposes

Griggs articulates a simple but powerful antisubordination principle:²⁹ employment practices must be revised such that protected classifications become irrelevant.³⁰ Any unnecessary and discriminatory employment practice—

-
26. Civil Rights Act of 1991, Pub. L. No. 102-166, 105 Stat. 1071 (codified in scattered sections of 42 U.S.C.).
27. *Albemarle Paper Co. v. Moody*, 422 U.S. 405 (1975); *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971); see also *Frazier v. Garrison I.S.D.*, 980 F.2d 1514, 1525 n.34 (5th Cir. 1993) (marking on the law’s return to a pre-*Wards Cove* standard after the Civil Rights Act’s passage); Joseph A. Seiner, *Disentangling Disparate Impact and Disparate Treatment: Adapting the Canadian Approach*, 25 YALE L. & POL’Y REV. 95, 103 n.55 (2006) (“Congress provided that the statute should be read ‘in accordance with the law as it existed on June 4, 1989, with respect to the concept of ‘alternative employment practice’ for claims of disparate impact. *Wards Cove* was decided on June 5, 1989.” (citation omitted)).
28. 42 U.S.C. § 2000e-2(k)(1)(A)(i); see also *Albemarle Paper*, 422 U.S. at 425; *Griggs*, 401 U.S. at 431.
29. There are various important refinements of antisubordination, equal status, and anticaste theory. See, e.g., Owen M. Fiss, *Groups and the Equal Protection Clause*, 5 PHIL. & PUB. AFF. 107 (1976); see also CATHERINE A. MACKINNON, TOWARD A FEMINIST THEORY OF THE STATE (1989); Ruth Colker, *Anti-Subordination Above All: Sex, Race, and Equal Protection*, 61 N.Y.U. L. REV. 1003 (1986); Reva Siegel, *Why Equal Protection No Longer Protects: The Evolving Forms of Status-Enforcing State Action*, 49 STAN. L. REV. 1111 (1997); Cass R. Sunstein, *The Anticaste Principle*, 92 MICH. L. REV. 2410 (1994); cf. JOSEPH FISHKIN, BOTTLENECKS: A NEW THEORY OF EQUAL OPPORTUNITY 120-21 (2014) (“[P]art of the distinctive appeal of equal opportunity is that it enables people to pursue goals in life that are to a greater degree their own, rather than being dictated by the limited opportunities that were available to them. Unequal opportunities, most obviously when they take the form of social structures like a caste system, a class system, or a gender role system, limit the kinds of lives people can lead [E]qual opportunity . . . gives each of us more of a chance to . . . become . . . ‘part author of his life.’” (citing JOSEPH RAZ, THE MORALITY OF FREEDOM 370 (1986)). But see Jennifer L. Levi, *Misapplying Equality Theories: Dress Codes at Work*, 19 YALE J. L. & FEMINISM 353 (2008) (arguing that there are limits to antisubordination theory).
30. “Congress has not commanded that the less qualified be preferred over the better qualified simply because of minority origins. Far from disparaging job qualifications as such, Congress has made such qualifications the controlling factor, so that race, religion, nationality, and sex become irrelevant. What Congress has commanded is that any tests used must meas-

intentional or unintentional, of large or small magnitude—must be removed. The primary purpose of disparate impact is antisubordination. Discriminatory employment practices should be removed so that protected classifications become irrelevant. As the *Griggs* Court put it, the fundamental aim of Title VII is “the removal of artificial, arbitrary, and unnecessary barriers to employment when the barriers operate invidiously to discriminate on the basis of racial or other impermissible classification.”³¹

Of course, the goal of antisubordination has an unavoidable limit. It does not entirely “preclude[] the use of [employment] testing or measuring procedures.”³² In the absence of a less discriminatory alternative, policies that have a disparate impact may be permitted if “they are demonstrably a reasonable measure of job performance.”³³ Therefore, when the goal of antisubordination and a legitimate business interest clash, disparate impact is tolerated—to an extent—for the sake of business interests that are sufficiently substantial and in the absence of an alternative policy of less discriminatory impact.

The overarching antisubordination aim and the business necessity limit inform the structure of the three-part burden-shifting framework.³⁴ First, the plaintiff must demonstrate a prima facie case of disparate impact: “that a [defendant] uses a particular employment practice that causes a disparate impact on the basis of race, color, religion, sex, or national origin.”³⁵ Then, the plaintiff must identify a discriminatory employment practice, one that functions to make a protected status like race relevant. The employer can also demonstrate that the practice does not cause the disparate impact: “If the [defendant] demonstrates that a specific employment practice does not cause the disparate impact, the [defendant] shall not be required to demonstrate that such practice is required by business necessity.”³⁶ In rebuttal, the defendant must demonstrate that “the challenged practice is job related for the position in question and consistent with business necessity.”³⁷ Note that the practice must not only be related to the job, but must also be a reasonable measure of job performance, one that justifies a departure from disparate impact’s primary aim to

ure the person for the job and not the person in the abstract.” *Griggs*, 401 U.S. at 436 (emphasis added).

31. *Id.* at 431.

32. *Id.*

33. *Id.*

34. 42 U.S.C. § 2000e-2(k) (2012).

35. *Id.* § 2000e-2(k)(1)(A)(i).

36. *Id.* § 2000e-2(k)(1)(B)(ii).

37. *Id.* § 2000e-2(k)(1)(A)(i).

make factors like race and religion irrelevant. Finally, even if the discriminatory practice is job related and consistent with business necessity, the plaintiff may succeed by presenting an alternative employment practice³⁸ that also serves the employer's legitimate interests "without a similarly undesirable [discriminatory] effect" but that the respondent refuses to adopt.³⁹ The fundamental purpose of this three-part framework is to eliminate unnecessary and discriminatory employment barriers. Some discriminatory barriers might be business necessities—barriers that have been permitted despite the motivation to make factors like race irrelevant. Yet if there is an alternative policy that serves the same purpose without equal discriminatory impact, the employer must adopt that policy instead.

The fundamental aim of antistatutory discrimination might be achieved in court or out of court. Although it is easy to focus primarily on disparate impact litigation, successful lawsuits are only one way through which disparate impact law might dismantle unnecessary and discriminatory barriers to employment. Another, less costly way that disparate impact law serves its function is by creating incentives for employers to remove problematic and unlawful barriers to employment before litigation commences.

C. Statistical Concepts

Several statistical concepts are relevant to disparate impact analysis. Here, I detail the most important concepts for the purposes of this Note: statistical significance and practical significance.

1. Statistical Significance

Statistical significance is a concept that is frequently applied to empirical results. One of the most common forms in which statistical significance is expressed is through a *p*-value (e.g., " $p < .05$ "). A *p*-value is the probability of obtaining results that are at least as extreme as if the null hypothesis were true. Smaller *p*-values provide evidence that is less consistent with the null hypothesis.

In the context of Title VII employment discrimination litigation, a null hypothesis might assume equal selection rates by an employer among different racial applicant groups. For instance, suppose the evidence shows that a policy differentially rejects blacks and that this difference is statistically significant

38. *Id.* § 2000e-2(k)(1)(A)(ii).

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

with a p -value of five percent. This means that, assuming equal selection rates for each group, there is a five percent chance of arriving at a difference in selection rates of equal or greater magnitude.

In disparate impact analysis (and elsewhere), p -values should be interpreted cautiously; statistical significance testing should not be relied upon in isolation.⁴⁰ In a recent volume, the American Statistical Association summarized some of the key principles and flaws in how p -values have been used in empirical analysis:⁴¹

1. p -values can indicate how incompatible the data are with the model being tested.
2. p -values do not tell you the probability the model is true or the probability the data are random.
3. No decision—scientific, business, legal or otherwise—should be based solely on p -values passing a cutoff value (i.e., a “bright line,” such as $p < .01$ or $.05$).
4. The proper understanding of statistical tests requires full reporting and transparency (i.e., report all statistical analyses and p -values; do not cherry-pick results to be reported).
5. A p -value does not indicate the size or importance of an effect that is obtained, no matter how small the p -value is (and large p -values do not tell you that an effect does not exist, only that it is not supported by the data).
6. The p -value does not tell you how good your model or hypothesis is (i.e. a high p -value may support the null hypothesis, yet many other models might also be supported by the data).⁴²

-
40. See, e.g., Rick Jacobs, Kevin Murphy & Jay Silva, *Unintended Consequences of EEO Enforcement Policies: Being Big Is Worse than Being Bad*, 28 J. BUS. & PSYCHOL. 467, 468 (2013) (explaining that in “statistical power analysis, with large samples even very small differences in outcomes will be statistically significant”) (citation omitted); Kevin R. Murphy & Rick R. Jacobs, *Using Effect Size Measures To Reform the Determination of Adverse Impact in Equal Employment Litigation*, 18 PSYCHOL. PUB. POL’Y & L. 477, 496 (2012) (recommending that “effect size measure be combined with tests of statistical significance, either through the joint reporting of effect sizes and p values or through minimum-effect tests when evaluating adverse impact”).
41. Ronald L. Wasserstein & Nicole A. Lazar, *The ASA’s Statement on p -Values: Context, Process, and Purpose*, 70 AM. STATISTICIAN 129 (2016).
42. Frederick L. Oswald, Eric M. Dunleavy & Amy Shaw, *Measuring Practical Significance in Adverse Impact Analysis*, in ADVERSE IMPACT ANALYSIS: UNDERSTANDING DATA, STATISTICS, AND RISK (Scott B. Morris & Eric M. Dunleavy eds., 2017).

These lessons highlight the dangers of relying solely on p -values or interpreting them inappropriately.⁴³ For instance, “ $p = .05$ ” does not mean that the null hypothesis has only a five percent chance of being true, nor does it mean that the observed data would occur only five percent of the time under the null hypothesis.⁴⁴ A p -value is simply the probability of the observed result or a more extreme result occurring, given that the null hypothesis is true. It is important to remember that a p -value is calculated on the assumption that the null hypothesis is true. Therefore, the p -value is *not* the probability that the null hypothesis is false.

Consider what p -values can tell us in disparate impact analysis. Suppose our null hypothesis is that there is no racial effect of a business’s hiring policy. That is, the null hypothesis is that any difference in hiring rates between two racial groups is simply due to chance. If the real-world data indicate a statistically significant difference in the employer’s hiring rates between black and white groups with a p -value of less than five percent, we have learned that, assuming no racial effect, we would find a difference in white and black hiring rates at least this extreme less than five percent of the time. The data do not tell us that there is less than a five percent chance that the racial disparity is due to chance.

2. Practical Significance

Practical significance refers to the real-world import of a statistical finding. In disparate impact cases, the term is used in two notably different ways. One is to refer to a “magnitude inquiry,” an analysis of the magnitude of a result supported by statistical evidence—for instance, the size of the effect indicated by a statistically significant finding. The other is a “confidence inquiry,” an analysis of the strength of the inference drawn between statistical evidence and the conclusion one draws from it about the real world.

43. See Kingsley R. Browne, *The Strangely Persistent “Transposition Fallacy”: Why “Statistically Significant” Evidence of Discrimination May Not Be Significant*, 14 LAB. LAW. 437 (1998); D.H. Kaye, *Is Proof of Statistical Significance Relevant?*, 61 WASH. L. REV. 1333 (1986); Ramona L. Paetzold, *Problems with Statistical Significance in Employment Discrimination Litigation*, 26 NEW ENG. L. REV. 395 (1991).

44. See Steven Goodman, *A Dirty Dozen: Twelve P-Value Misconceptions*, 45 SEMINARS HEMATOLOGY 135, 136-37 (2008).

A magnitude inquiry is an assessment of the size of an effect.⁴⁵ For instance, a statistically significant effect can be small in size. Suppose there is evidence that an employer had a hiring pool of ten thousand applicants. A five percent racial disparate impact might be statistically significant given the large sample size, but nevertheless deemed to have a small effect size, since some may think that a five percent difference is “small” in size.⁴⁶ Of course, whether an effect size is “large” or “small” is fundamentally a conventional or normative judgment and not derived purely from statistical analysis.

In contrast, a confidence inquiry is an assessment of the strength of the evidence, which asks how strong the inference is between the evidence and the claim it supports about the world. For instance, we might evaluate the statistical evidence of an observed disparity by asking whether it really supports the existence of a real-world disparity caused by the hiring policy in question. Imagine that statistical evidence suggests a three percent disparity in the hiring rates of black and white applicants. Courts might ask whether this result is practically significant in the sense of whether this evinces any real-world disparity. This aspect of practical significance is important, but it is also a standard inquiry: we can, should, and do regularly ask whether any piece of evidence is practically significant in this second sense.

Even an effect with a size that is considered “medium” or “large” in the first sense might be deemed as having little practical significance in the second sense, especially when the evidence is based on a small sample size. For instance, suppose an employer has a hiring pool of ten applicants, half from one group and half from another, and a hiring test excludes all but three.⁴⁷ Even if the difference in hiring rates suggested by this evidence is of large magnitude, we might doubt the real-world inference of a disparate impact supported by these results.

This distinction – practical significance as a measure of a disparity’s magnitude vs. practical significance as a measure of confidence in the strength of evi-

45. For the seminal conventional standards of effect size, see Jacob Cohen, *A Power Primer*, 112 PSYCHOL. BULL. 155 (1992), which describes “small,” “medium,” and “large” effects for various statistical tests.

46. In the context of disparate impact, the EEOC’s four-fifths rule is essentially an effect size rule: a selection rate for any protected group that is less than four-fifths of the rate for the group with the highest rate is generally regarded as evidence of adverse impact. 29 C.F.R. § 1607.4(D) (2015). Based on the guideline, if the size of the rate difference is big enough, this supports prima facie adverse impact. See *infra* Section II.A.

47. Even in the best case, this produces a ratio less than that advised by the four-fifths rule. Assuming three are selected, in the best case, the test admits one member from one class and two from the other. The ratio (.5) of acceptances is less than .8. For more on the four-fifths rule, see *infra* Section II.A.

dence—is crucial. This Note focuses on magnitude inquiries. This is not to say, however, that evaluation of the inference between statistical evidence and the real world is irrelevant. To the contrary, such evaluations should remain fundamental at each stage.

Consider another example. Suppose a company has reviewed applications from one hundred candidates, forty-five of whom are white and fifty-five of whom are black. The application requires a hair follicle drug test, which more white applicants pass. By conventional standards (significance determined by $p < .05$), the effect of race is on the border of statistical significance. Depending on the assumptions, different statistical tests lead to different results.⁴⁸ This demonstrates an important but overlooked feature of statistical significance testing: Despite its allure of objectivity, its results vary based on its assumptions.

Regardless of statistical significance, the effect's practical significance remains. First, consider the "magnitude" aspect of practical significance: what do these statistical analyses imply about the magnitude of the disparity? A conventional measure of effect size suggests that this is a "small" or "weak" effect size.⁴⁹ But we can still ask about the "confidence" aspect of practical significance: how strongly do these statistical facts (including our analysis of effect size) support the existence of any real-world disparity? In other words, how strong is the evidence of a disparity?

Although there is an important distinction between two aspects of practical significance—magnitude and confidence—authorities sometimes emphasize only one aspect. Consider the Federal Judicial Center's definition, which understands practical significance only in terms of magnitude: practical significance means that "the magnitude of the effect being studied is not de minimis—it is sufficiently important substantively for the court to be concerned."⁵⁰ Some courts have adopted a similar understanding of practical significance. In *Frazier v. Garrison I.S.D.*, the Fifth Circuit held that a 4.5% difference in selection rates did not have sufficient practical significance when 95% of applicants were selected.⁵¹ The *Frazier* Court justified its decision by citing a case in which it had previously held

48. Fisher's Exact Test indicates this is "statistically significant," $p = .049$. A χ^2 test indicates that this falls above the standard ($p < .05$) cutoff: $p = .079$.

49. For example, Cramér's $V = .176$, indicating a small effect size.

50. FED. JUDICIAL CTR., REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 292 (3d ed. 2011), [http://www.fjc.gov/public/pdf.nsf/lookup/SciMan3Do1.pdf/\\$file/SciMan3Do1.pdf](http://www.fjc.gov/public/pdf.nsf/lookup/SciMan3Do1.pdf/$file/SciMan3Do1.pdf) [<http://perma.cc/5WJG-64PA>].

51. 980 F.2d 1514, 1526 (5th Cir. 1993).

DISPARATE STATISTICS

that employment examinations having a 7.1 percentage point differential between black and white test takers do *not*, as a matter of law, state a prima facie case of disparate impact. Therefore [in this case in which the difference is 4.5 percentage points], there is no significant statistical discrepancy between minority and non-minority pass rates.⁵²

Thus, the court applied a practical significance requirement in the sense of a magnitude inquiry. This was not an inquiry into how strongly the evidence supported the possibility of a real-world disparity. The *Frazier* Court was essentially performing a logical deduction: since a 7.1% difference was not big enough to constitute prima facie disparate impact, a 4.5% difference was also insufficiently large.

II. DISPARATE STATISTICS

Statistics play a crucial role at each of the three stages of disparate impact litigation: the plaintiff's prima facie case of disparate impact, the defendant's rebuttal relating to job-relatedness and business necessity, and the plaintiff's demonstration of a suitable alternative practice. This Part outlines the role of statistics at each stage and presents arguments for the appropriate use of statistics and "practical significance" inquiries at each stage.

Section II.A argues that many courts inappropriately conduct magnitude inquiries at the prima facie stage of disparate impact analysis. Scrutinizing a disparity's "practical significance" through a magnitude inquiry at the prima facie stage is to ask whether the disparity is big enough to warrant the court's attention. This question is antithetical to the statutory text, purpose, and precedents of disparate impact law.

Section II.B argues that a robust magnitude inquiry is more appropriate at the second stage of disparate impact analysis. Although such a requirement is incongruous at the prima facie stage, it is apt when assessing the merit of an employer's rebuttal that some disparity-causing policy is a job-related business necessity—employers must demonstrate that a disparity-causing test has large enough relevance to justify permitting discriminatory impact on the basis of certain legitimate business interests. A magnitude inquiry at this rebuttal stage is more consistent with disparate impact law.

Comparatively fewer cases proceed to the third stage of disparate impact analysis: the plaintiff's proposal of a less discriminatory alternative policy. Sec-

52. *Id.* at 1524 (citing *Moore v. Sw. Bell Tel. Co.*, 593 F.2d 607, 608 (5th Cir. 1979) (per curiam)).

tion II.C argues that the logic underlying the elimination of magnitude inquiries during the prima facie stage applies to the third stage as well. Just as the aim of the first stage is to identify a policy that causes any disparity, the aim of the third stage is to identify an alternative policy that provides any decrease in discriminatory effect. Plaintiffs should not be required to show that their proposal reduces discrimination by a particular magnitude. As long as the proposal satisfies the employer's legitimate interest without a similarly undesirable effect on potential or current employees, the plaintiffs should be found to have met their burden.

A. The Statistical Standard of Prima Facie Disparate Impact

The plaintiff typically provides evidence of statistically significant disparities to help support the prima facie demonstration of a disparate impact.⁵³ Courts adopt a variety of approaches in assessing these disparities. One common approach is to adopt thresholds based on standard deviations.⁵⁴ Some courts hold that disparities not rising to a certain level of statistical significance are insufficient proof of disparate impact.⁵⁵

53. See, e.g., *Int'l Bhd. of Teamsters v. United States*, 431 U.S. 324, 339 (1977) (noting that statistical analysis is probative in demonstrating prima facie disparate impact). The Court suggests that statistically significant tests are probative since they can uncover covert discrimination: "Statistics showing racial or ethnic imbalance are probative in a case such as this one only because such imbalance is often a telltale sign of purposeful discrimination; absent explanation, it is ordinarily to be expected that nondiscriminatory hiring practices will in time result in a work force more or less representative of the racial and ethnic composition of the population in the community from which employees are hired." *Id.* at 339 n.20. The Court also quotes *United States v. Ironworkers Local 86*, stating, "In many cases the only available avenue of proof is the use of racial statistics to uncover clandestine and covert discrimination by the employer or union involved." *Id.* (quoting 443 F.2d 544, 551 (9th Cir. 1971)). This is a blurring of disparate impact and disparate treatment theory. But the reasoning extends: statistics can uncover intentional or covert causes of discrimination and also causes of discrimination that are unintentional.

54. See, e.g., *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 308 n.14 (1977) (stating that "two or three standard deviations" indicates a disparate impact of gross significance, making "suspect" the hypothesis that hiring was conducted "without regard to race" (citing *Castaneda v. Partida*, 430 U.S. 482, 497 n.17 (1977))).

55. See, e.g., *Mems v. City of St. Paul, Dep't of Fire & Safety Servs.*, 224 F.3d 735, 741 (8th Cir. 2000) (race); *Bennett v. Total Minatome Corp.*, 138 F.3d 1053, 1062 (5th Cir. 1998) (age); *Palmer v. Shultz*, 815 F.2d 84, 95 (D.C. Cir. 1987) (gender).

Other courts adopt the EEOC's "four-fifths" or "eighty percent" rule as a standard for measuring prima facie disparate impact.⁵⁶ The four-fifths rule compares the ratio of selection rates between the rate of selection for the protected class and the greatest rate of selection for any group and asks whether this ratio is less than four-fifths. The Supreme Court famously branded the four-fifths rule as one that "has not provided more than a rule of thumb."⁵⁷ Moreover, the EEOC guidance itself acknowledges that "[s]maller differences in selection rate may nevertheless constitute adverse impact, where they are significant in both statistical and practical terms or where a user's actions have discouraged applicants disproportionately on grounds of race, sex, or ethnic group."⁵⁸ In other words, the four-fifths rule yields at most a first cut of easily decided cases of prima facie disparate impact: a prima facie case is demonstrated by group selection rates with a ratio below four-fifths, but smaller differences (i.e., larger ratios) require further scrutiny.

These guidelines—statistical significance (and other measures like standard deviation analysis) and the four-fifths rule—can be combined with each other. For instance, a court might adopt an analysis that looks first to the four-fifths rule and then to statistical significance for data failing the four-fifths rule. The four-fifths rule is essentially a guideline that takes practical significance into account, allowing prima facie impact to be established when the effect size (disparity) is large enough. The guideline might also be supplemented by an interpretation that holds practical significance is not established where the disparity is insufficiently large. This is the magnitude inquiry debate at the heart of the circuit split.

The Supreme Court has consistently stated that the essence of demonstrating a prima facie disparate impact is showing statistically significant evidence of a disparity.⁵⁹ This view was recently reaffirmed in *Ricci v. DeStefano*: "[A]

56. 29 C.F.R. § 1607.4(D) (2016) ("A selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact, while a greater than four-fifths rate will generally not be regarded by Federal enforcement agencies as evidence of adverse impact."). For a history of the four-fifths test, see DAN BIDDLE, *ADVERSE IMPACT AND TEST VALIDATION: A PRACTITIONER'S GUIDE TO VALID AND DEFENSIBLE EMPLOYMENT TESTING* (2d ed. 2006).

57. *Watson v. Fort Worth Bank & Tr.*, 487 U.S. 977, 995 n.3 (1988); see also *Ricci v. DeStefano*, 557 U.S. 557, 587 (2009) (noting that the four-fifths rule "is 'a rule of thumb for the courts'" (quoting *Watson*, 487 U.S. at 995 n.3)).

58. 29 C.F.R. § 1607.4(D).

59. For Supreme Court decisions emphasizing the fundamentality of a "significantly" different statistical disparity, see *Connecticut v. Teal*, 457 U.S. 440, 446 (1982); *N.Y.C. Transit Auth. v.*

prima facie case of disparate-impact liability [is] essentially, a threshold showing of a significant statistical disparity . . . and nothing more”⁶⁰ The Court characterized the prima facie demonstration as one not requiring a disparity of any particular magnitude. This reaffirms the core commitment of disparate impact theory: “Title VII tolerates no racial discrimination, subtle or otherwise.”⁶¹

There are various considerations weighing against magnitude inquiries at the prima facie stage of disparate impact analysis, such as definitions of “disparate impact”⁶² and the legislative history of the relevant statutes.⁶³

Statutory text and Supreme Court precedent demonstrate that practical significance is irrelevant at the prima facie stage. According to Title VII, a complaining party must demonstrate “that a respondent uses a particular employment practice that causes a disparate impact on the basis of race, color, religion, sex, or national origin”⁶⁴ The text indicates that a plaintiff must show a disparate impact, not a substantial, notable, large, or even significant disparate impact. A confidence inquiry is relevant in determining whether the evidence presented supports causation, but there is no basis in the text for a magnitude inquiry, which asks whether the evidence supports a disparate impact that is big enough to be worth proceeding.

Supreme Court precedent supports the same interpretation. *Griggs* interprets Title VII as aimed at the removal of unnecessary barriers that have a discriminatory impact in employment:

What is required by Congress is the removal of artificial, arbitrary, and unnecessary barriers to employment when the barriers operate invidiously to discriminate on the basis of racial or other impermissible classification. Congress has now provided that tests or criteria for employment or promotion may not provide equality of oppor-

Beazer, 440 U.S. 568, 586 (1979); *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 425 (1975); and *Griggs v. Duke Power Co.*, 401 U.S. 424, 426 (1971).

60. 557 U.S. at 587 (citing *Teal*, 457 U.S. at 446).

61. *McDonnell Douglas Corp. v. Green*, 411 U.S. 792, 801 (1973).

62. For discussions of the relevance of dictionary definitions to textualist arguments, see *Jones v. City of Boston*, 752 F.3d 38, 50 (1st Cir. 2014), which discusses definitions; Ko *supra* note 1, at 888–90, which argues that the dictionary definition of “disparate impact” also supports declining practical significance inquiries at the prima facie stage. *But see* Eissenstat, *supra* note 1, at 670 (contesting the First Circuit’s analysis of definitions in *Jones*).

63. For legislative history considerations, see Ko, *supra* note 1, at 890–92, which argues that the legislative history of the Civil Rights Acts of 1964, 1990, and 1991 supports declining practical significance inquiries at the prima facie stage.

64. 42 U.S.C. § 2000e-2(k)(1)(A) (2012).

tunity merely in the sense of the fabled offer of milk to the stork and the fox. On the contrary, Congress has now required that the posture and condition of the job seeker be taken into account. It has—to resort again to the fable—provided that the vessel in which the milk is proffered be one all seekers can use. The Act proscribes not only overt discrimination, but also practices that are fair in form, but discriminatory in operation. 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.⁶⁵

This well-known passage is worth careful attention. As *Griggs* interprets Title VII, Congress is not concerned only about the barriers that cause the largest disparities; rather, if an unnecessary barrier causes any disparity, the barrier must be removed. *Albemarle Paper* reinforces this early understanding.⁶⁶

Considering practical significance at the prima facie stage is equally inconsistent with recent Supreme Court opinions on disparate impact. Recall *Ricci*'s straightforward avowal: “[A] prima facie case of disparate-impact liability [is] essentially, a threshold showing of a significant statistical disparity . . . and nothing more”⁶⁷ Inquiry into a disparity's size is unambiguously something more. It is also worth noting the unanimity in understanding given the ideological diversity represented among the authors and signers of just these two opinions. Chief Justice Burger wrote the opinion in *Griggs* on behalf of a unanimous Court; four decades later, Justice Kennedy wrote the *Ricci* opinion on behalf of the Court's conservatives. Requiring a demonstration of this sufficient magnitude aspect of practical significance entails a subjective verdict on the importance of some (“small”) disparity. This is at odds with the textual basis, aims, and precedent (from *Griggs* to *Ricci*⁶⁸) of prima facie disparate impact demonstration.

This argument raises two important questions: (1) how does the argument square with the four-fifths rule, a commonly accepted mode of inquiring into practical significance; and (2) if magnitude inquiries are so clearly inappropri-

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

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

67. *Ricci v. DeStefano*, 557 U.S. 557, 587 (2009) (citing *Connecticut v. Teal*, 457 U.S. 440, 446 (1982)).

68. *Id.*; *Griggs*, 401 U.S. 424.

ate at the prima facie stage, why is there a controversial circuit split on the issue?⁶⁹

Although rejecting a prima facie case on the basis of practical significance is inappropriate, many courts look to practical significance as a shorthand to demonstrate a prima facie disparate impact through the four-fifths rule.⁷⁰ The four-fifths rule has an air of objectivity: if the hiring rate for the impacted group is lower than this sharp cut-off—eighty percent of the rate for the favored group—then there is a prima facie disparate impact. But this rule has different effects depending on selection rates. For instance, if a favored group is hired at a rate of twenty percent, then any impacted-group hiring rate less than sixteen percent would establish prima facie disparate impact. But if a favored group is hired at a ninety-five percent rate, then any impacted-group hiring rate less than seventy-six percent would establish the prima facie case. In other words, based on the hiring base rate, the four-fifths rule's guidance fluctuates between a group-group difference of zero to twenty percent.

Crucially, the EEOC's characterization of the four-fifths rule advises that any rate less than four-fifths of the higher selection rate establishes the prima facie case without showing further practical significance, but smaller differences may “nevertheless constitute adverse impact” if those differences are statistically and practically significant.⁷¹ In other words, the four-fifths rule advises granting the demonstration of prima facie disparate impact under certain conditions, but it never advises denying it on such a basis. Smaller differences should be considered in further detail to determine whether they evince prima facie disparate impact.

The four-fifths rule is essentially a practical significance guideline that functions as a ceiling, not a floor. If the effect size is large enough, there is a prima facie disparate impact. A number of other courts have suggested that something beyond mere statistical significance should be required in demon-

69. See discussion *supra* note 7.

70. 29 C.F.R. § 1607.4(D) (2016) (“A selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact, while a greater than four-fifths rate will generally not be regarded by Federal enforcement agencies as evidence of adverse impact. Smaller differences in selection rate may nevertheless constitute adverse impact, where they are significant in both statistical and practical terms or where a user's actions have discouraged applicants disproportionately on grounds of race, sex, or ethnic group.”). See generally 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, 182-85 (2008) (describing the use of the four-fifths rule among several circuits).

71. 29 C.F.R. § 1607.4(D).

strating the prima facie case of disparate impact.⁷² This requirement is a demonstration of a certain form of “practical significance”: the statistically significant result must evince a substantial disparity.

Now consider the second question. If practical significance inquiries are so clearly inappropriate at the prima facie stage, why is there a circuit split? Recall that the First Circuit rejected a practical significance requirement in *Jones v. City of Boston*, but the Fifth Circuit held that a disparate job selection rate was too small to establish a prima facie case in *Frazier v. Garrison I.S.D.*⁷³ Part of the answer, I suspect, is that some courts prefer a thoughtful, contextual analysis of the evidence that supports the prima facie disparate impact. A contextualized inquiry—for instance, examining sample size, statistical significance, and effect size—is appropriate in a confidence inquiry. It is inappropriate, however, for courts to smuggle a magnitude inquiry floor into a confidence inquiry. At the prima facie stage, courts should ask whether the evidence supports a finding of disparate impact, not what amount of disparate impact merits attention.

Magnitude inquiries are a necessarily subjective practice. *Frazier* held that a 4.5% difference was trivial, when ninety-five percent of applicants were selected.⁷⁴ The justification for such reasoning is unclear: Would a 4.5% difference be more relevant if only eighty percent of applicants were selected? What if only thirty percent of applicants were selected?

Confidence inquiries are appropriately contextual. There are many factors to consider in a confidence inquiry. When evaluating how strongly the evidence supports the existence of a disparate impact, courts might look to the statistical evidence’s sample size, the size of the respective group categories, and even the effect size.

But the subjective contextualism of a magnitude inquiry is more dangerous. Determining what magnitude of disparate impact is sufficient to demonstrate a prima facie disparate impact allows—and invites—judgment about the importance of some disparate impact on a protected class. This allows lines to be drawn differently in different contexts. For instance, some jurisdictions might consider a five percent hiring difference significant, while others might

72. See, e.g., *Bos. Chapter, NAACP, Inc. v. Beecher*, 504 F.2d 1017, 1019-20 (1st Cir. 1974).

73. See sources cited *supra* note 1 (describing critical commentary of the First Circuit’s decision in *Jones*). Compare *Jones v. City of Bos.*, 752 F.3d 38, 53 (1st Cir. 2014) (rejecting a practical significance requirement), with *Frazier v. Garrison I.S.D.*, 980 F.2d 1514, 1524 (5th Cir. 1993) (holding that a 4.5% difference in selection rates was trivial when 95% of applicants were selected and citing *Moore v. Southwestern Bell Telephone Co.*, 593 F.2d 607, 608 (5th Cir. 1979), for the proposition that the use of employment examinations with a 7.1% difference between black and white examinees does not constitute a prima facie case of disparate impact).

74. 980 F.2d at 1524.

consider the difference trivial. This injects subjectivity into the core of disparate impact analysis. Moreover, it contradicts the text of Title VII and Supreme Court precedent, which require plaintiffs to identify a disparate impact—and nothing more.

To be more precise, one reason that practical significance testing at the prima facie stage is ever invoked is that courts consider an impact's magnitude in the name of practical significance, when they really are invoking the confidence inquiry aspect of practical significance. This is a statistical fallacy. While confidence inquiries are an appropriate consideration at the stage of prima facie disparate impact, and effect size can serve as relevant evidence for a confidence inquiry, a magnitude inquiry is not in itself necessary to satisfy a confidence inquiry. It may be that courts commit the fallacy of requiring consideration of what is merely one source of possible evidence. The relevant, crucial question at the prima facie demonstration stage is this: is there good evidence that the policy caused some disparity? Evidence of a large disparity helps build confidence in the proof of some (perhaps even smaller) disparity. But evidence of a large disparity is not required. In some cases, we expect it to be absent—namely, when there is a small real-world disparity.

A similar confusion underlies appeals to the four-fifths rule. The theory of disparate impact does not privilege “large” disparities over “smaller, insubstantial” disparities. The appropriate justification for recommending acceptance of “big” disparities as clear evidence of prima facie disparate impact is not that they reflect big real-world disparities. Rather, such evidence typically inspires more confidence than evidence of smaller disparities that some real disparity exists. Smaller differences are no less important, but smaller differences generally provide less confidence that any difference exists (sample size and all else equal). The exception that proves this rule is a case like *Jones*,⁷⁵ where there is a small effect size but a very large sample size, supporting the court's confidence that the disparity is not the product of chance.

Thus, the four-fifths rule really ought to be a “rule of thumb.”⁷⁶ As the EEOC guidance recommends, smaller differences than advised by the rule should not be rejected as insufficient proof of prima facie disparate impact; instead, they should be scrutinized more closely.⁷⁷

75. 752 F.3d 38.

76. *Watson v. Fort Worth Bank & Tr.*, 487 U.S. 977, 995 n.3 (1988).

77. Note that this interpretation is consistent with deference to the EEOC, a concern of some commentators. See, e.g., McKinley, *supra* note 70, at 182-83; Eissenstat, *supra* note 1, at 669; Ko, *supra* note 1, at 894.

These considerations also indicate an important way in which the standard for prima facie disparate impact demonstration should be strong. It is possible that some statistical evidence for “large” differences over the four-fifth rule’s cutoff are actually unconvincing evidence. The most intuitive example is evidence involving a small sample. Imagine five people, two white and three black, apply for a job. The two white applicants and one black applicant are not excluded by the company’s policy. This involves an enormous disparity between white-applicant and black-applicant hiring rates. Yet, this does not give us confidence that the defendant’s policy caused a disparate impact. Accordingly, courts have recognized the limited value of small sample sizes in disparate impact cases.⁷⁸ This exemplifies the appropriateness of a confidence inquiry.

This is made all the more complicated by the multiple meanings of “practical significance.” Some courts use it to analyze the magnitude of a disparity,⁷⁹ which I argue is inappropriate at the prima facie stage. Yet other courts refer to practical significance when pointing to a worry about the confidence in a statistically significant difference.⁸⁰ Unlike the former, the latter is a legitimate inquiry at the prima facie stage of disparate impact.

This issue is not merely terminological. Judges writing in support of a “practical significance” requirement or inquiry should investigate which meaning of practical significance they intend to employ. For instance, when discussing whether a disparity is “substantial” (read in the magnitude sense),⁸¹ the First Circuit was concerned with whether the disparity was “due to chance” (closer to the sense of a confidence inquiry), not whether the disparity was of a

78. See, e.g., *Connecticut v. Teal*, 457 U.S. 440, 463 n.7 (1982); *Int’l Bhd. of Teamsters v. United States*, 431 U.S. 324, 340 n.20 (1977); *Mayor of Phila. v. Educ. Equal. League*, 415 U.S. 605, 620-21 (1974); *Dendy v. Wash. Hosp. Ctr.*, 431 F. Supp. 873, 876 (D.D.C. 1977); *Rogillio v. Diamond Shamrock Chem. Co.*, 446 F. Supp. 423, 427-28 (S.D. Tex. 1977).

79. See, e.g., *Jones*, 752 F.3d 38.

80. For instance, the court in *United States v. Virginia*, 454 F. Supp. 1077 (E.D. Va. 1978) noted that adding two people from the not-passing-the-test group to the passing-the-test group changed a finding of statistical significance. The court held the results were not practically significant. See also *Waisome v. Port Auth.*, 948 F.2d 1370, 1376 (2d Cir. 1991) (endorsing a case-by-case approach to fit the circumstances). In such cases, the court invokes practical significance where they should invoke a worry about certainty.

In *Contreras v. City of Los Angeles*, 656 F.2d 1267, 1273 (9th Cir. 1981), two practical significance tests were discussed. One dealt with the effect of adding three people to the plaintiff group in a favorable way that eliminated four-fifths rule of thumb conclusion. When the four-fifths rule of thumb conclusion can be changed by adding only three people, the sample is considered unreliably small and of no practical significance.

81. *Fire v. Providence Fire Dep’t*, 766 F.2d 650 (1st Cir. 1985).

certain magnitude.⁸² It is misleading to interpret these decisions as support for a practical significance requirement in the sense of an inquiry into the sufficiency of a disparity's size. Practical significance, in the sense of a disparity of requisite size, is distinct from confidence in a statistically significant result.

Commentators also commit this error:

On the one hand, statistical significance allows plaintiffs to demonstrate that a particular practice causes some disparity between classes (the "disparate" prong of the inquiry); on the other, practical significance determines if that disparity is large enough to have real-world implications (the "impact" prong of the inquiry). Practices that do in fact create a noticeable disparate impact would implicate both of these considerations.⁸³

A prima facie case of disparate impact does not depend on whether we care sufficiently about the size of the impact; evidence of adverse impact establishes the prima facie case, even if that adverse impact is small.

A tempting policy counterargument is that prohibiting magnitude inquiries at the prima facie stage would incentivize frivolous disparate impact litigation.⁸⁴ But this claim underestimates the strength of the statistical significance requirement. As *Jones* explained,⁸⁵ requirements to show statistical significance will frequently eliminate frivolous lawsuits, since small-sized impacts will require large sample sizes to demonstrate statistical significance.⁸⁶ Second, if the defendant shows job-related business necessity, the plaintiff will still have to prove an alternative practice with less impact. This will be relatively easier when the magnitude of the disparity is large, providing a balanced corrective.⁸⁷ In cases in which the prima facie impact is small, the plaintiff will still have a larger burden in the demonstration of an alternative practice since the alternative policy has less room to reduce the disparity than if the disparity were large.

82. *Id.* at 657-58 ("Where the use of employment tests results in differential pass rates for blacks and whites, even an apparently substantial differential, the discrepancy may be due to chance. Statistical significance and, in the case of so small a sample as the 1974 sample, we believe judicial significance, can be attributed to an observed discrepancy only where there is a low probability that the differential in pass rates would be expected to occur simply by chance." (internal citation omitted)).

83. See Stenger, *supra* note 1, at 436.

84. See *id.*

85. *Jones v. City of Bos.*, 752 F.3d 38 (1st Cir. 2014).

86. *Id.* at 53.

87. *Id.*

A further reason not to fear a rise in trivial claims is that complainants have little personal incentive to bring disparate impact claims. Disparate impact relief is limited to equitable relief and back pay. Compensatory and punitive damages are not available, as they are for disparate treatment claims.⁸⁸

A final, but important, response to this counterargument concerns the logic of burden shifting in disparate impact litigation. The previously articulated responses address worries about an increase in frivolous employee complaints by justifying the unlikelihood of an effect on the complainant. But disallowing consideration of practical significance at the prima facie stage might instead have an effect on employers. Knowing that any robustly proven disparity can shift the burden to the defendants can have an important effect outside of litigation, encouraging employers to reflect on whether their policies and procedures that have such an impact are actually job-related business necessities, or whether less discriminatory alternatives exist. One way of responding to potential litigation is to reinforce incentives for employers to eliminate the very practices and procedures that unnecessarily impact protected classes.⁸⁹ Thus, concern about the litigation effects of changing the statistical burden in fact provides an additional reason for condemning the use of magnitude inquiries at the prima facie stage.

B. The Statistical Standard of Job-Related Business Necessity

Statistics are also relevant to the second stage of disparate impact litigation, in which a defendant must prove the job-relatedness and business necessity of a policy that has been shown to have a prima facie disparate impact. Compared to the plaintiff's prima facie standard, the defendant's proof of job-related business necessity is typically described as a more stringent standard. Prima facie disparate impact requires a plaintiff to "only show" that a policy causes a

88. 42 U.S.C. § 1981a(a)(1) (2012) (providing for compensatory and punitive damages in cases of intentional discrimination); *id.* § 2000e-5(g)(1) (describing the remedies that are available for disparate impact claims); see also Martha Chamallas, *The Market Excuse*, 68 U. CHI. L. REV. 579, 600 (2001) (contrasting intentional discrimination and disparate impact claims and noting that the latter "are limited to equitable relief"); Meghan E. Changelo, *Reconciling Class Action Certification with the Civil Rights Act of 1991*, 36 COLUM. J.L. & SOC. PROBS. 133, 138 (2003) ("[P]laintiffs are not entitled to compensatory or punitive damages when they succeed in showing disparate impact discrimination.").

89. See, e.g., William M. Landes, *The Economics of Fair Employment Laws*, 76 J. POL. ECON. 507 (1968); Michael Selmi, *Testing for Equality: Merit, Efficiency, and the Affirmative Action Debate*, 42 UCLA L. REV. 1251 (1995).

“discriminatory pattern,” while job-related business necessity requires proof that the policy has “a manifest relationship to the employment in question.”⁹⁰

This distinction supports the aims of disparate impact. The *prima facie* stage only identifies a disparity-causing policy. The second stage offers the employer the opportunity to prove that the discriminatory policy falls within the subset of policies that Title VII is willing to tolerate on the basis of business necessity. As such, the second stage requires a more robust consideration of the policy’s significance to business interests; a mere relation is not necessarily sufficient to permit discrimination.

At the second stage, the defendant must show that the contested policy is “job related” and “consistent with business necessity.”⁹¹ The EEOC’s Uniform Guidelines indicate three measures of validation in assessing this demonstration of job-related business necessity: criterion-related, content, and construct validation. Criterion-related validation requires empirical data showing that the selection procedure “is predictive of or significantly correlated with important elements of job performance.”⁹² Content validation requires “data showing that the content of the selection procedure is representative of important aspects of performance on the job.”⁹³ Construct validation requires data showing that the selection “procedure measures the degree to which candidates have identifiable characteristics which have been determined to be important in successful performance in the job.”⁹⁴

Many courts require showing both statistical and practical significance in defending a discriminatory test.⁹⁵ To assess these showings, courts often look to the correlation coefficient, a numeral measure from -1 to 1 of the relation be-

90. See, e.g., *Dothard v. Rawlinson*, 433 U.S. 321, 329 (1977) (“[T]o establish a *prima facie* case of discrimination, a plaintiff need only show that the facially neutral standards in question select applicants for hire in a discriminatory pattern. Once it is shown that the employment standards are discriminatory in effect, the employer must meet ‘the burden of showing that any given requirement [has] . . . a manifest relationship to the employment in question.’” (quoting *Griggs v. Duke Power Co.*, 401 U.S. 424, 432 (1971))).

91. 42 U.S.C. § 2000e-2(k)(1)(A)(i).

92. 29 C.F.R. § 1607.5(B) (2016); see also *Hamer v. City of Atlanta*, 872 F.2d 1521 (11th Cir. 1989) (holding that a test had been properly validated based on a sufficient correlation between test scores and job performance); 29 C.F.R. § 1607.14 (discussing the minimum standards for validity studies).

93. 29 C.F.R. § 1607.5; see also *id.* § 1607.14 (discussing the minimum standards for validity studies).

94. 29 C.F.R. § 1607.5; see also *id.* § 1607.14 (discussing the minimum standards for validity studies).

95. See, e.g., *Hamer*, 872 F.2d at 1525-26; *Jones v. Pepsi-Cola Metro. Bottling Co.*, 871 F. Supp. 305, 313 n.25 (E.D. Mich. 1994).

tween two values, between the test and job performance.⁹⁶ But many do not look specifically at the practical significance of the (statistically significant) correlations presented as evidence of validation. Although defendants often have to show a moderate correlation between a policy (e.g., a test) and the outcome (e.g., job performance), this is often not interpreted through the lens of a magnitude inquiry, asking how big the actual relationship is.

The use of correlation coefficients ought to be accompanied by a practical significance analysis of the policy or procedure's job-relatedness and business necessity. Specifically, it ought to be accompanied by consideration of both the relevance of the evidence to a real-world job-related business necessity and the magnitude of this relation. A test that is merely correlated with job performance might not actually be related to the job or a "business necessity." For instance, achieving a certain score on a general standardized achievement test might be correlated with some aspect of job performance, even though that achievement is not actually a strong predictor of job success.

Taking practical significance into account means rejecting implausible claims of job-relatedness in which there is no strong relation between the policy and outcome. For instance, in *Dickerson v. U.S. Steel Corp.*, a statistically significant correlation between a policy and job performance of 0.3 was rejected since it was found to have little practical significance—indicating only nine percent of job success attributable to the disparity-causing policy.⁹⁷

In other cases, paltry consideration of practical significance permits evidence of job-relatedness that has little practical significance. Consider *United States v. City of Garland*.⁹⁸ There, the court determined that police and firefighter job examinations were job related on the basis of a significant correlation between those exams and performance on academy exams and state certification exams. Yet an important practical significance question was obscured: what magnitude of significance do these exams have to the job? The practical signifi-

96. See, e.g., *EEOC v. Atlas Paper Box Co.*, 868 F.2d 1487 (6th Cir. 1989); *Ensley Branch of the NAACP v. Seibels*, 13 Emp. Prac. Dec. (CCH) ¶ 11,504 (N.D. Ala. 1977).

97. See *Dickerson v. U.S. Steel Corp.*, 472 F. Supp. 1304 (E.D. Pa. 1978). Regarding the validity coefficients in the case, the court noted, "[A] low coefficient, even though statistically significant, may indicate a low practical utility." *Id.* at 1348. The court further stated: "[O]ne can readily see that even on the statistically significant correlations of .30 or so, only 9% of the success on the job is attributable to success on the [test] batteries. This is a very low level, which does not justify use of these batteries, where correlations are all below .30. In conclusion, based upon the guidelines and statistical analysis . . . the Court cannot find that these tests have any real practical utility. The guidelines do not permit a finding of job-relatedness where statistical but not practical significance is shown. On this final ground as well, therefore, the test batteries must be rejected." *Id.* at 1351.

98. No. Civ. A. 3:98-CV-0307-L, 2004 WL 741295 (N.D. Tex. Mar. 31, 2004).

cance of an exam's results for job-relatedness and consistency with business necessity cannot be inferred simply from a correlation between the exam and another exam. Moreover, there is little rigorous consideration of the magnitude of this effect; even if the evidence supports a good inference for the job-relation, does it support evidence of a sufficiently large effect consistent with business necessity?

This particular decision is even more problematic. In the same decision, the court determined that there was insufficient practical significance to establish a *prima facie* case of disparate impact,⁹⁹ and there would be sufficient proof of job-relatedness consistent with business necessity, without serious consideration of the practical significance of this evidence.¹⁰⁰ This provides an example of a bizarre practice: a relatively high practical significance requirement in the *prima facie* stage of disparate impact, but a paltry one in the job-relatedness business necessity stage.

99. *Id.* at *23 (“Dr. Stoikov analyzed the practical significance or the magnitude of the effect of rank-order hiring. She found that the percentage[s] of minority and white test-passers who were hired are not significantly different, and thus concluded that ranking could not have adversely affected the chances of being hired. Dr. Stoikov also analyzed the hiring consequences of the City’s use of [test scores]. She concluded that in the Police Department, there was no significant shortfall in expected Hispanic hires, and one additional black hire would have eliminated the significant shortfall in expected black hires. She further concluded that in the Fire Department from 1992 through 1998, there was no significant shortfall in expected Hispanic hires, and a significant shortfall of 1.2 black hires.”). The Note does not aim to challenge all the details of the court’s analysis or even its holding. The main purpose of the example is to show the rigor with which some courts analyze *prima facie* disparate impact. The plaintiff should be required to show nothing more than a disparity.

100. *Id.* at *23 n.74 (“Even if the United States has established a *prima facie* showing, which it has not, the City presented sufficient evidence to establish a business necessity and job relatedness for its practice of ‘rank order hiring.’ Specifically, Dr. Wollack testified that rank-order hiring based on [test] scores is psychometrically appropriate because there was an adequate job analysis, the examinations were reliable, and there was a useful spread of scores. Dr. Wollack further testified that rank-order hiring is appropriate because the relationship between test scores and job performance is assumed to be linear. The court incorporates by reference its discussion of Dr. Wollack’s testimony regarding the validity of the . . . examinations.”). This is *dicta* (since the court holds that the *prima facie* demonstration fails), but it reveals a serious lack of engagement with the practical significance of an employer’s rebuttal. While the plaintiff’s *prima facie* case is scrutinized for its practical significance, this hypothetical determination of job-related business necessity does not nearly as robustly address practical significance – in either the magnitude or confidence aspects.

C. The Statistical Standard of Showing a Suitable Alternative

At the third stage, courts use statistics to evaluate the plaintiff's demonstration of a nondiscriminatory alternative policy.¹⁰¹ This stage provides a final opportunity for the plaintiff to rebut the defendant's job-related business necessity defense by offering an alternative policy that could serve the business's legitimate interests without the same discriminatory impact. Comparatively few disparate impact cases proceed to this third stage, but the cases that do consider alternative proposals may look to statistics to evaluate the merits of the alternative proposal. Courts may conduct practical significance inquiries following the logic of the Stage 1 inquiry, asking whether the evidence indicates that the alternative proposal will have large practical significance, greatly or sufficiently reducing discriminatory impact. Such an inquiry is a magnitude inquiry, assessing the sufficiency of the effect size. As in the other stages, there is also room at the third stage for a confidence inquiry, asking whether the court is confident that the alternative proposal will reduce discrimination (at any rate).

The argument here follows directly from the logic of Section II.A. Given disparate impact's foundational texts, purposes, and interpretations, practical significance testing—in the sense of measuring the magnitude of some disparity reduction—should not be relevant to assessing an alternative policy.¹⁰² A suitable alternative policy should be accepted regardless of whether it decreases discriminatory impact by a small or large amount. What matters is that the policy can be expected to actually reduce discriminatory impact.¹⁰³ Disparate impact aims to remove all unnecessary discriminatory barriers, not just the largest ones. The third stage inquiry asks whether the employer “refuses to adopt an available alternative employment practice that has less disparate im-

^{101.} See 42 U.S.C. § 2000e-2(k)(1)(C) (2012).

^{102.} At least it is not relevant to the most obvious way in which practical significance testing might enter: evaluating the degree to which discrimination will be reduced. A more radical recommendation might be made: practical significance testing is relevant for the purpose of evaluating the degree to which the alternative policy realizes the legitimate business interests. I do not pursue a full treatment of this proposal here, but simply offer the suggestion. If practical significance is relevant in determining the degree of business necessity that justifies discrimination, might it also be relevant in determining the degree to which an alternative policy falls (slightly or severely) short of serving those interests?

^{103.} Cf. *Jones v. City of Bos.*, 118 F. Supp. 3d 425 (D. Mass. 2015) (holding the plaintiffs did not meet their burden of showing a suitable alternative practice without a similarly undesirable racial effect that served the city's legitimate interest), *rev'd in part*, 845 F.3d 28 (1st Cir. 2016) (holding the plaintiffs did demonstrate a suitable alternative practice).

pact and serves the employer's legitimate needs."¹⁰⁴ The employment practice need not have dramatically, substantially, or even significantly less disparate impact—just “less.”

Because comparatively fewer cases proceed to this stage of disparate impact analysis, this Section first considers a stylized example, an imagined case in which this issue of practical significance is clearly implicated at the third stage of disparate impact analysis. Imagine, as an example, that a disparate impact case has proceeded to the third stage. The contested policy results in a twenty-percentage-point racial difference between black and white test takers, but the alternative policy would still result in a fifteen-percentage-point differential. Given some courts' treatment of practical significance at the *prima facie* stage, one could imagine courts deciding that a five-percentage-point reduction does not support a suitable alternative policy because it is not “practically significant.”

The fundamental aim of disparate impact is the removal of unnecessary and discriminatory barriers so as to make factors like race irrelevant in employment practices. This means that assessment of an alternative practice's magnitude of disparity reduction is dubious. Does the fact that some alternative practice would be only modestly less discriminatory justify rejecting it? The broader disparate impact framework clearly recommends any policy that offers even modest discrimination reduction. While confidence inquiries are relevant—at this stage and all the others—magnitude inquiries are not appropriate in weighing an alternative policy.

III. RECOMMENDATIONS AND IMPLICATIONS

How should courts resolve these questions of disparate statistics? This Part proposes three solutions. The most theoretically justifiable solution is to strike magnitude inquiries from the first stage's *prima facie* demonstration and the third stage's evaluation of a suitable alternative, but to adopt a more robust analysis at the second stage when the defendant presents a job-relatedness and business-necessity rebuttal.

104. *Ricci v. DeStefano*, 557 U.S. 557, 578 (2009) (citing 42 U.S.C. § 2000e-2(k)(1)(A)(ii), (C) (2012)); see also *Jones*, 845 F.3d 28 (“Application of this prong in this case turns on the answers to three questions: First, does the record contain evidence that would allow a jury to find that there was an ‘alternative’ method of meeting the Department’s legitimate needs? Second, does the record also allow a jury to find that adopting that alternative method would have had less of a disparate impact? And finally, could a jury find that the Department ‘refuses to adopt’ that alternative method?”).

DISPARATE STATISTICS

Before turning to these recommendations, it is worth summarizing the framework that has been established and the broad conclusions. First, recall the distinctions between magnitude inquiries—those asking whether some disparity, business interest, or impact reduction is big enough—and confidence inquiries—those asking whether how strong an inference can be drawn from statistical evidence to the real world.

FIGURE 1.
SUMMARY OF PRACTICAL SIGNIFICANCE INQUIRIES AT EACH STAGE OF
DISPARATE IMPACT ANALYSIS

	Magnitude Inquiry	Confidence Inquiry
Stage 1	<p>Is the prima facie disparity sufficiently large?</p> <p><i>Ex. “[E]mployment examinations having a 7.1 percentage point differential between black and white test takers do not, as a matter of law, state a prima facie case of disparate impact.”</i></p>	<p>Is there sufficient evidence for the existence of a prima facie disparity?</p> <p><i>Ex. “To determine the practical significance of statistical results, a court must look at the theories and assumptions underlying the analysis”</i>¹⁰⁵</p>
Stage 2	<p>[Inappropriate]</p> <p>Is the policy’s job-relatedness and business necessity sufficiently large?</p> <p><i>Ex. “[O]nly 9% of the success on the job is attributable to success on the [test] batteries [T]he Court cannot find that these tests have any real practical utility. The guidelines do not permit a finding of job-relatedness where statistical but not practical significance is shown.”</i>¹⁰⁶</p> <p>[Appropriate]</p>	<p>[Appropriate]</p> <p>Is there sufficient evidence for the existence of a job-related business necessity?</p> <p><i>Ex. “[T]he employer must establish two elements of correlation: . . . the degree to which test scores relate to job performance . . . [and] the confidence that can be placed on the practical significance”</i>¹⁰⁷</p> <p>[Appropriate]</p>

¹⁰⁵. EEOC v. Sears, Roebuck & Co., 628 F. Supp. 1264, 1286 (N.D. Ill. 1986).

¹⁰⁶. Dickerson v. U.S. Steel Corp., 472 F. Supp. 1304, 1351 (E.D. Pa. 1978).

¹⁰⁷. Jones v. Pepsi-Cola Metro. Bottling Co., 871 F. Supp. 305, 313 n.25 (E.D. Mich. 1994) (citing Hamer v. City of Atlanta, 872 F.2d 1521, 1525-26 (11th Cir. 1989)).

	Magnitude Inquiry	Confidence Inquiry
	Is the suitable alternative's disparity mitigation sufficiently large?	Is there sufficient evidence for the existence of a suitable alternative with (some degree of) lesser disparate impact?
Stage 3	<i>Ex. The contested policy results in a twenty-percentage-point racial difference between black and white test takers, but the alternative policy would still result in a fifteen-percentage-point differential; as a matter of law, a five-percentage-point reduction does not support a suitable alternative policy.</i> ¹⁰⁸	<i>Ex. In evaluating statistical results, a court must look at the theories and assumptions underlying the analysis; this includes weighing confidence placed in statistical results supporting that a policy has reduced discriminatory impact.</i>
	[Inappropriate]	[Appropriate]

This figure highlights that courts that conduct magnitude inquiries at Stage 1 but not Stage 2 are not only conducting investigations at odds with the text and interpretation of disparate impact, as I have argued, but are also conducting unjustifiably asymmetric inquiries.

The primary recommendation is that courts should reject a magnitude inquiry for the demonstrations of a prima facie disparate impact and a suitable alternative stages of analysis, but insist on a more robust inquiry during the defendant's job-related business necessity rebuttal. This solution is the most faithful to the text, purpose, and precedent of disparate impact law. Requiring a showing of prima facie disparate impact of a particular size is out of line with the original interpretations of Title VII,¹⁰⁹ as well as with more recent Supreme Court precedents.¹¹⁰ Equally inappropriate is a thin or absent investigation into the practical significance of a defendant's defense that a disparity-causing impact is a job-related business necessity,¹¹¹ requiring proof of a manifest¹¹² rela-

¹⁰⁸. Because comparatively fewer courts have considered Stage 3 of analysis, I use stylized examples for the magnitude and confidence inquiries in this table.

¹⁰⁹. *E.g.*, *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971).

¹¹⁰. *E.g.*, *Ricci v. DeStefano*, 557 U.S. 557 (2009).

¹¹¹. *See* 42 U.S.C. § 2000e-2(k)(1)(A)(i) (2012).

tionship between the contested policy and a job-related function that is not simply justified by a business interest,¹¹³ but is a “business necessity.”¹¹⁴

Even if this primary recommendation is not adopted on the basis of Part II’s arguments, there are two more modest possibilities that deal with the asymmetry of conducting magnitude inquiries at some but not all stages of disparate impact analysis. The previous Part justifies a disparity in the use of statistics in one direction—magnitude inquiries during the second stage, but not during the first or third stages—given the aims of disparate impact theory. However, an asymmetry in the other direction is entirely lacking in justification.

One option to deal with the asymmetry is to “level down” the magnitude inquiry requirement across the stages of disparate impact analysis, for example by removing it from the prima facie disparate impact standard. If a court engages in no serious magnitude analysis at the business-necessity-rebuttal stage, it should not engage in a rigorous one at the prima facie stage (or third stage). The other option is to “level up” the practical significance requirement across stages of disparate impact litigation, most crucially by raising the level of magnitude inquiry at the business-necessity rebuttal. If a court engages in a robust magnitude analysis at the prima facie stage or evaluation of a less-discriminatory alternative stage, it ought to do the same at the job-related and business-necessity-rebuttal stage.

These two options represent each half of the primary recommendation. They are less justifiable than the primary recommendation, but they do correct the unjustified asymmetry and they may be more immediately attainable. And they would be preferable to current practices that require a magnitude demonstration for the plaintiff’s prima facie demonstration, but not for the defendant’s business-necessity demonstration.

If adopted, any of these recommendations will likely affect both disparate impact litigation and business practices out-of-court. First, the recommendations may result in different decisions in disparate impact cases. The *Jones*¹¹⁵ decision is recent, but already there are suggestions that removal of a prima facie stage magnitude inquiry will result in different outcomes for disparate impact cases. For instance, *Smith v. City of Boston* held that a police department’s lieutenant-selection process had a racially disparate impact and was not a job-

112. See *Griggs*, 401 U.S. at 430-31.

113. See *Wards Cove Packing Co. v. Atonio*, 490 U.S. 642 (1989).

114. 42 U.S.C. § 2000e-2(k)(1)(A)(i).

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

related business necessity that survived disparate impact analysis.¹¹⁶ Like many analyses of disparate impact, the discussion of the prima facie stage involved much consideration of statistics. The court frequently cited to *Jones* in requiring that the impact not be the result of chance,¹¹⁷ declining to require a strict statistical significance cutoff,¹¹⁸ and rejecting the necessity of the four-fifths rule, interpreted as a practical significance requirement.¹¹⁹ Removing such magnitude requirements from the prima facie stage will allow a broader range of legitimate disparate impact claims to go forward. Just because a policy has a “small” discriminatory impact does not mean the interests of those people are less deserving of protection.

Requiring a more robust analysis of the magnitude and confidence aspects of practical significance at the business necessity stage would also affect disparate impact litigation. Cases like *United States v. City of Garland*¹²⁰ would require a more rigorous analysis. Discriminatory policies justified on the basis of business necessity would be held to a more robust inquiry into the degree to which the business necessity affects the legitimate interest.

Similarly, prohibiting magnitude inquiries at the third stage would allow disparate impact to better realize its aims. The third stage targets any real reduction in discrimination. Together these recommendations remedy inappropriate uses of statistics in disparate impact litigation and provide opportunity for greater plaintiff success in legitimate employment discrimination suits.

Second, the recommendations would also have out-of-court effects. The recommendations might seem likely to increase disparate impact litigation. Regardless of whether that would actually occur,¹²¹ removing barriers to successful disparate impact claims should incentivize employers to reflect more thoughtfully on their employment practices. Disparate impact litigation is just one mechanism to achieve the aims of greater employment equality. Another mechanism—in many ways a preferable one—is for employers to simply remove disparity-causing policies and procedures that are not business necessities or replace necessary policies with less discriminatory ones. Whether or not disparate impact litigation increases, it is plausible that there is power in the mere threat of litigation; *any* real disparity—no matter how small—caused by

116. *Smith v. City of Bos.*, 144 F. Supp. 3d 177 (D. Mass. 2015).

117. *Id.* at 191.

118. *Id.*

119. *Id.* at 192.

120. No. Civ. A. 3:98-CV-0307-L, 2004 WL 741295 (N.D. Tex. Mar. 31, 2004).

121. *Cf. supra* notes 84-89 and accompanying text (arguing that the Note’s recommendations are unlikely to encourage frivolous disparate impact claims).

an employment practice could be the target of disparate impact litigation. It may be difficult to quantify the number of discriminatory employment practices that the employer might remove to comply with disparate impact law or to minimize the threat of litigation. But that effect is equally important.

These arguments are not partisan; they do not unfairly help plaintiffs bring disparate impact claims, whether meritorious or not. In fact, the recommendations might result in reduced litigation—for example, if employers respond to the threat of litigation by removing unnecessary and discriminatory employment barriers. Rather, these arguments seek to correct the use of statistics to achieve the proper understanding of both disparate impact's antistatutory aim and its business necessity limit. Whether the acceptance of such recommendations would help plaintiffs, defendants, employees, or employers depends on empirical questions about the parties' responsive behaviors. The removal of discriminatory employment practices, subject to the limit of business necessity, whether through litigation or other incentives, is at the heart of disparate impact law. Remedying disparate and inappropriate uses of statistics moves the law closer to this fundamental aim.

Consider again the very project of paying attention to such "small" disparities. It might seem unimportant to focus on ameliorating employment practices where the local discriminatory impact reduction is relatively small, such as reducing a discriminatory impact by just five percent. But such reasoning belies the theory of disparate impact: no discriminatory effect is too small to matter. But even one unconvinced by disparate impact theory and compelled by a singular focus on big effects must address a second practical consideration. Small discriminatory effects at multiple points—in an individual's life or across a group—result in large cumulative disadvantages.¹²² This Note's recommendations suggest broad changes in the way courts treat numerous small disparities. It may be the case that the current social landscape does not consist primarily of policies that cause infrequent and large disparities but rather, an enormous web of smaller-disparity-causing policies, the combination of which results in large disparities overall. If so, the total effect of attending more judiciously to "small" disparities may not be so small.

122. See MEASURING RACIAL DISCRIMINATION 69 (Rebecca Blank et al. eds., 2004) ("[E]ffects of discrimination may cumulate over time through the course of an individual's life . . ."); Anthony G. Greenwald et al., *Statistically Small Effects of the Implicit Association Test Can Have Societally Large Effects*, 108 J. PERSONALITY & SOC. PSYCHOL. 553 (2015); cf. Robert P. Abelson, *A Variance Explanation Paradox: When a Little Is a Lot*, 97 PSYCHOL. BULL. 129 (1985) (explaining how, although skill strongly predicts baseball batting success, the correlation between skill and getting a hit at any given at-bat is low).

CONCLUSION

Statistics, particularly “practical significance,” play a crucial role in disparate impact analysis. This Note distinguishes between two types of practical significance inquiries: magnitude inquiries—questions about the magnitude of a finding supported by statistical evidence—and confidence inquiries—questions about the strength of statistical evidence. Looking across the three stages of disparate impact analysis, I argue for the inappropriateness of magnitude inquiries at the first prima facie stage of demonstrating disparate impact and at the third stage of providing a less discriminatory alternative, but that such a robust inquiry should come at the second stage of a defendant’s job-related business necessity rebuttal.

This buttresses recent court decisions to not require demonstration of a particular magnitude of disparity at the prima facie stage. It also outlines a holistic conception of practical significance testing across every area of disparate impact analysis, a project bearing on the current circuit split and also the doctrine’s future challenges.

The consequences of these conclusions should not be underestimated. A universal rejection of magnitude inquiry at the prima facie stage of disparate impact would have a large effect. Cases like *Moore v. Southwestern Bell Telephone Co.*¹²³ and *Frazier v. Garrison*¹²⁴ would require different justifications. Requiring a more robust analysis at the defendant’s rebuttal stage would be equally impactful, requiring more thoroughgoing analysis in cases like *United States v. City of Garland*.¹²⁵

These changes are justified. Requiring that a prima facie disparate impact be of a certain magnitude invites inappropriate subjective weighing, asking judges to assess whether a disparity is big enough. Failing to inquire robustly about the practical significance of a defendant’s rebuttal is equally problematic—resulting in the justification of policies that have a discriminatory impact on the basis of slack correlations. So too is requiring that an alternative proposal be sufficiently less discriminatory, rather than simply less discriminatory.

All of these practices are at odds with the motivation and aims of disparate impact: a prima facie disparate impact must simply demonstrate a disparity caused by the contested policy on a protected class; a job-related business necessity defense is meant to show the weighty significance of the contested policy, which must bear a manifest relationship to the employment, justifying the

123. 593 F.2d 607 (5th Cir. 1979).

124. 980 F.2d 1514 (5th Cir. 1993).

125. No. Civ. A. 3:98-CV-0307-L, 2004 WL 741295 (N.D. Tex. Mar. 31, 2004).

permission of a discriminatory policy; and an alternative proposal is meant to provide a policy that serves the legitimate business interests, with a large or even small degree of lesser discriminatory impact.

The Note recommends analyzing and correcting these uses of practical significance testing across the three stages of disparate impact analysis. The recommendations advance disparate impact's fundamental aim: removing artificial and arbitrary barriers that operate to discriminate on the basis of a protected classification.