

2016

Material Facts in the Debate Over *Twombly* and *Iqbal*

Jonah B. Gelbach

University of Pennsylvania Law School, jgelbach@law.upenn.edu

Follow this and additional works at: http://scholarship.law.upenn.edu/faculty_scholarship



Part of the [Civil Procedure Commons](#), [Courts Commons](#), [Judges Commons](#), [Law and Economics Commons](#), [Litigation Commons](#), and the [Public Economics Commons](#)

Recommended Citation

Gelbach, Jonah B., "Material Facts in the Debate Over *Twombly* and *Iqbal*" (2016). *Faculty Scholarship*. Paper 1539.
http://scholarship.law.upenn.edu/faculty_scholarship/1539

This Article is brought to you for free and open access by Penn Law: Legal Scholarship Repository. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of Penn Law: Legal Scholarship Repository. For more information, please contact PennlawIR@law.upenn.edu.



ARTICLE

**Material Facts in the Debate over
Twombly and *Iqbal***

Jonah B. Gelbach*

Abstract. This Article presents empirical evidence concerning the adjudication of defendant-filed summary judgment motions. Using nearly 2000 randomly selected employment discrimination and contract cases, the Article tries to assess the performance of *Twombly* and *Iqbal*, which raised the federal pleading standard, in filtering cases according to merit. This Article begins by explaining how such data might be helpful in making such an assessment, taking into account the possibility that parties' behavior might have changed following *Twombly* and *Iqbal*.

This Article ultimately concludes that even using this large collection of data—the most comprehensive data assembled to date on the subject—it is not possible to determine whether “*Twiqbal*’s” supporters or critics are more accurate in their assessments of the

* Associate Professor, University of Pennsylvania Law School. This Article is based on data collected with the assistance of the Searle Civil Justice Institute (SCJI) and the Law and Economics Center (LEC) at George Mason University. An early version of the results was posted as a policy report, of which I was the primary author, as SEARLE CIVIL JUSTICE INST., GEORGE MASON UNIV. SCH. OF LAW, MEASURING THE EFFECTS OF A HEIGHTENED PLEADING STANDARD UNDER *TWOMBLY* AND *IQBAL* (2013), <http://www.masonlec.org/programs/46>. I am grateful to the SCJI at the George Mason University LEC for its generous financial support of this project as well as for providing coders. I am also grateful for the many helpful comments I received from the judges, practitioners, and academics on the SCJI Board of Overseers, participants in the SCJI research workshop at George Mason University School of Law, and two anonymous referees provided by the LEC. I also thank Rita Choi for help managing early data collection and James Cooper for his management expertise and intellectual support for my work on this project, as well as current and former GMU School of Law students Dale Baker, Josh Branson, Corey Carpenter, Wei Fan, Patrick Vincent, and Chris Woolley, and current and former University of Pennsylvania Law School students Meredith Gage, Gary Ho, Jonathan Kaplan, James Klima, and Jonelle Saunders for all their hard work coding cases. In addition, this project could not have been completed without the data (and helpful discussions concerning it) provided by Thomson Reuters, nor the generous financial support in acquiring those data that was provided by the Oscar M. Ruebhausen Fund at the Yale Law School. I am also very grateful to Yale Law School's William N. Eskridge and Dean Robert C. Post for their enthusiastic support in acquiring these data. Finally, I had many valuable conversations with legal scholars too numerous to name; at the risk of offending those left out, I offer special thanks to Steve Burbank, Henry Butler, Joe Cecil, David Engstrom, Eric Helland, William Hubbard, Daniel Klerman, Jon Klick, Bruce Kobayashi, Alan Morrison, and Joshua Wright.

efficacy of the new plausibility pleading regime with respect to its ability to filter cases according to merit at the pre-discovery stage. This null result points to the very real possibility that plausibility pleading's case-quality effects—a quintessential empirical question—simply cannot be answered using data.

This Article's basic message, then, is that empirical evidence is unlikely to settle the debate over the case-quality effects of the new pleading regime ushered in by *Twombly* and *Iqbal*.

Table of Contents

Introduction.....	372
I. <i>Twombly</i> , <i>Iqbal</i> , and What We Can Learn from Summary Judgment Results	377
A. The Evolution of Pleading Standards and How They Relate to Discovery.....	377
B. Defining “Merit” Postdiscovery	381
C. Understanding the Summary Judgment Link when Only Judges’ Behavior Changes Following <i>Twombly</i> and <i>Iqbal</i>	382
D. Party Selection Effects of <i>Twombly</i> and <i>Iqbal</i>	389
1. Party selection effects that involve filtering out at the Rule 12(b)(6) stage.....	389
2. A party selection effect that involves selection into summary judgment.....	391
II. Data	393
III. Employment Discrimination Cases.....	398
A. Basic Characteristics of the Employment Discrimination Sample.....	398
B. Summary Judgment Adjudication Results for Employment Discrimination Cases	400
1. The percentage of cases in which plaintiffs win as to all challenged claims or issues	401
2. The percentage of cases in which plaintiffs win on at least one challenged claim.....	403
3. Measuring the quality-filtering effect using changes in the number of employment discrimination cases facing summary judgment motions.....	404
IV. Contract Cases.....	407
A. Basic Characteristics of the Contracts Sample.....	408
B. Summary Judgment Adjudication Results for Contract Cases	410
1. The percentage of cases in which plaintiffs win as to all aspects raised by defense summary judgment motions	410
2. The percentage of cases in which plaintiffs win on at least one claim challenged via defense summary judgment motion.....	411
3. Challenges in measuring the quality-filtering effect using changes in the number of contract cases facing summary judgment motions.....	413

Material Facts in the Debate over Twombly and Iqbal
68 STAN. L. REV. 369 (2016)

V. Further Issues.....	416
A. Why It Would Be Problematic to Focus Only on Cases in Which a Rule 12(b)(6) Motion Was Filed.....	416
B. Threats to the Validity of the Empirical Approach	418
1. The recession.....	418
2. <i>Ledbetter v. Goodyear Tire & Rubber Co.</i> and the Lilly Ledbetter Fair Pay Act	419
3. <i>Scott v. Harris</i>	419
4. Amendments to Rule 56.....	421
5. Changes in primary behavior	422
6. Other factors that might lead to invalid estimates.....	422
Conclusion.....	424

Introduction

Much controversy has surrounded the Supreme Court's opinions in *Bell Atlantic Corp. v. Twombly*¹ and *Ashcroft v. Iqbal*.² Critics of the plausibility standard introduced through these cases argue that it will reduce access to the federal courts for meritorious suits.³ In some disputes, the critics argue, the defendant controls the information that would be necessary to plead in sufficient factual detail to meet the plausibility standard. Under *Conley v. Gibson*'s now-retired "no set of facts" pleading standard,⁴ plaintiffs could allege wrongdoing generally and then rely on discovery to unearth the facts necessary to establish the elements of such causes of action. By requiring plaintiffs to allege such facts *before* Rule 12(b)(6) adjudication,⁵ critics argue, *Twombly*'s plausibility standard establishes a catch-22: pleading sufficiently to reach discovery requires access to information that is available only through discovery.⁶

On the other side of the "*Twombly*" debate, supporters of the plausibility standard argue that too many plaintiffs intentionally bring low-merit lawsuits for settlement value only. According to this view, defendants must agree to pay off plaintiffs in such cases because the burden of discovery is greater for defendants than for plaintiffs. Thus, plaintiffs in such suits have little disincentive to proceed through discovery, leaving defendants to choose between either settling beforehand or bearing the high discovery costs that precede the summary judgment phase where the defendants are very likely to win. Supporters of the switch to *Twombly*'s plausibility standard believe it will help eliminate low-merit cases, whose plaintiffs, they believe, will be unable to plead with sufficient factual detail before discovery costs mount.

1. 550 U.S. 544 (2007).

2. 556 U.S. 662 (2009).

3. For purposes of this Article, a case has merit if, following discovery, there would be sufficient facts that either (i) are in dispute, or (ii) point in the plaintiff's favor if not in dispute, such that the defendant would not be entitled to judgment as a matter of law. That is, a case has merit if, following discovery, its factual posture would either present an issue of triable fact or entitle the plaintiff to judgment as a matter of law. To put it another way, a case has merit if, following discovery, the plaintiff could demonstrate that she would be able to meet her burden of production at trial. Given this definition of merit, it is possible for even meritorious suits to fail the plausibility pleading standard. A judge deciding a Rule 12(b)(6) motion might believe that the complaint's allegations are implausible, so that she grant the motions, even though it happens to be true that discovery, were it to occur, would turn up evidence sufficient to meet the plaintiff's burden of production.

4. 355 U.S. 41, 45-46 (1957).

5. See FED. R. CIV. P. 12(b)(6).

6. For an interesting discussion of the challenges in designing a pleading standard that accounts both for such catch-22-style asymmetries and for the possibility that pretrial litigation costs are asymmetric in the opposite direction, see Paul Stancil, *Balancing the Pleading Equation*, 61 BAYLOR L. REV. 90 (2009).

Both the critics and the supporters hold theoretically coherent views: logically, both views could be correct. Further, it is possible that there are sizable numbers of *both* meritorious cases likely to face a catch-22 problem under the plausibility standard and strike suits likely to be filtered out by the plausibility standard. Therefore, the question as to which effect predominates ultimately is an empirical one.

While there has been no shortage of empirical work on *Twombly*, it has overwhelmingly focused on the question whether judges have indeed applied a higher standard when adjudicating Rule 12(b)(6) motions.⁷ But almost no work besides the present Article has even attempted to measure the extent to which

7. See, e.g., JOE S. CECIL ET AL., FED. JUDICIAL CTR., MOTIONS TO DISMISS FOR FAILURE TO STATE A CLAIM AFTER *IQBAL*: REPORT TO THE JUDICIAL CONFERENCE ADVISORY COMMITTEE ON CIVIL RULES (2011), [http://www.fjc.gov/public/pdf.nsf/lookup/motioniqbal.pdf/\\$file/motioniqbal.pdf](http://www.fjc.gov/public/pdf.nsf/lookup/motioniqbal.pdf/$file/motioniqbal.pdf) [hereinafter FJC FIRST STUDY]; JOE S. CECIL ET AL., FED. JUDICIAL CTR., UPDATE ON RESOLUTION OF RULE 12(B)(6) MOTIONS GRANTED WITH LEAVE TO AMEND: REPORT TO THE JUDICIAL CONFERENCE ADVISORY COMMITTEE ON CIVIL RULES (2011), [http://www.fjc.gov/public/pdf.nsf/lookup/motioniqbal2.pdf/\\$file/motioniqbal2.pdf](http://www.fjc.gov/public/pdf.nsf/lookup/motioniqbal2.pdf/$file/motioniqbal2.pdf) [hereinafter FJC SECOND STUDY]; Raymond H. Brescia, *The Iqbal Effect: The Impact of New Pleading Standards in Employment and Housing Discrimination Litigation*, 100 KY. L.J. 235 (2011-2012); Scott Dodson, *A New Look: Dismissal Rates of Federal Civil Claims*, 96 JUDICATURE 127 (2012); Patricia W. Hatamyar, *The Tao of Pleading: Do Twombly and Iqbal Matter Empirically?*, 59 AM. U. L. REV. 553 (2010); Lonny Hoffman, *Rulemaking in the Age of Twombly and Iqbal*, 46 U.C. DAVIS L. REV. 1483 (2013); William H.J. Hubbard, *Testing for Change in Procedural Standards, with Application to Bell Atlantic v. Twombly*, 42 J. LEGAL STUD. 35 (2013); Patricia Hatamyar Moore, *An Updated Quantitative Study of Iqbal's Impact on 12(b)(6) Motions*, 46 U. RICH. L. REV. 603 (2012); Victor D. Quintanilla, *Beyond Common Sense: A Social Psychological Study of Iqbal's Effect on Claims of Race Discrimination*, 17 MICH. J. RACE & L. 1 (2011); Alexander A. Reinert, *Measuring the Impact of Plausibility Pleading*, 101 VA. L. REV. 2117 (2015); Joseph A. Seiner, *Pleading Disability*, 51 B.C. L. REV. 95 (2010) [hereinafter Seiner, *Pleading Disability*]; Joseph A. Seiner, *The Trouble with Twombly: A Proposed Pleading Standard for Employment Discrimination Cases*, 2009 U. ILL. L. REV. 1011; Kendall W. Hannon, Note, *Much Ado About Twombly?: A Study on the Impact of Bell Atlantic Corp. v. Twombly on 12(b)(6) Motions*, 83 NOTRE DAME L. REV. 1811 (2008); Joe S. Cecil, *Of Waves and Water: A Response to Comments on the FJC Study Motions to Dismiss for Failure to State a Claim After Iqbal* (Mar. 19, 2012) (unpublished manuscript), <http://ssrn.com/abstract=2026103>; see also Theodore Eisenberg & Kevin M. Clermont, *Plaintiphobia in the Supreme Court*, 100 CORNELL L. REV. 193, 195-96 (2014) (reporting trends in "the difference between . . . pretrial-adjudication judgment rates for the defendant and for the plaintiff" and associating these trends with the Supreme Court's "trilogy" of summary judgment cases as well as *Twombly* and *Iqbal*); Samuel Issacharoff & Geoffrey Miller, *An Information-Forcing Approach to the Motion to Dismiss*, 5 J. LEGAL ANALYSIS 437, 439, 452 (2013) (proposing "a use of fee and cost shifting in the initial process of factual claiming and denials that would create incentives for the parties to acquire or reveal information germane to the motion to dismiss by forcing them to absorb the costs of either producing the information or denying its existence" and engaging in "a mild empirical inquiry" to determine how much such a policy might matter (capitalization altered)); Roger Michalski & Abby K. Wood, *Twombly and Iqbal at the State Level* (USC Law Legal Studies Paper No. 14-30, 2015), http://papers.ssrn.com/abstract_id=2468864 (using data from states to conduct before-after comparisons of the sort other authors have done at the federal level).

Twiqbal has affected the merit of cases that proceed beyond the Rule 12(b)(6) stage.⁸

This Article takes up that task, using outcomes of defense summary judgment motions to attempt to measure *Twiqbal*'s effects on case quality. More specifically, the analysis defines a case as having no merit postdiscovery if a reasonable factfinder could not find for the plaintiff on the basis of the postdiscovery record.⁹ Defendants in such cases can be expected to file and win summary judgment motions. Assuming that the plausibility pleading standard is effective at snagging low-merit cases at the Rule 12(b)(6) stage, such no-merit cases will be eliminated from the set of cases that make it to summary judgment in the post-*Iqbal* period. Thus, (i) there should be fewer cases in the summary judgment population post-*Iqbal*, and (ii) those cases should be more likely to survive summary judgment. Consequently, the plaintiff's win rate against defense summary judgment motions should be greater following *Twiqbal* than before. On the other hand, if *Twiqbal*'s critics are correct, then the plaintiff's win rate will have either dropped (indicating that *Twiqbal* filters out higher-quality, rather than lower-quality, cases) or stayed the same (indicating that *Twiqbal* filters equal proportions of high- and low-quality cases).

In reality, things are more complicated. There are good reasons to believe that parties' case strategy and Rule 12(b)(6)-motion-filing behavior, as well as their inclination to settle, will change following a perceived change in the pleading standard.¹⁰ Accordingly, Part I.D also discusses the role such party selection effects play in the empirical analysis.

8. Alexander Reinert has written two articles that in whole or in part are directed toward this question. Reinert's *Measuring the Impact of Plausibility Pleading*, *supra* note 7, focuses primarily on the same question that most of the *Twiqbal* empirical literature has considered—how the outcomes of Rule 12(b)(6) motions themselves have changed. However, the article also includes a brief section, titled “Plausibility Pleading as Filter?,” based on Reinert's coding of post-Rule 12(b)(6) outcomes in the cases he studies. Reinert, *supra* note 7, at 2162 (italics omitted). I discuss this article further below in Part V.A. In the other article, Alexander A. Reinert, *The Costs of Heightened Pleading*, 86 IND. L.J. 119 (2011), Reinert attempts to measure the effects that a heightened pleading standard would have had on the ultimate dispensation of pre-*Twombly* cases that had Rule 12(b)(6) motions granted and then reversed on appeal. By design, this article uses only pre-*Twombly* cases, so it cannot provide information about cases that are actually litigated after the switch to *Twiqbal*'s plausibility pleading standard. Moreover, by coding cases that settle as meritorious, *id.* at 138, Reinert begs the very question raised by supporters of the plausibility pleading standard: whether settlements occur because of the threat of costly discovery rather than the presence of a genuine basis for suit.

9. See *supra* note 3. Part I.B further elaborates on this definition.

10. See Jonah B. Gelbach, *Can the Dark Arts of the Dismal Science Shed Light on the Empirical Reality of Civil Procedure?*, 2 STAN. J. COMPLEX LITIG. 223, 228 (2014) [hereinafter Gelbach, *Dark Arts*]; Hubbard, *supra* note 7, at 36; Jonah B. Gelbach, Note, *Locking the Doors to Discovery?: Assessing the Effects of Twombly and Iqbal on Access to Discovery*, 121 YALE L.J. 2270, 2277 (2012) [hereinafter Gelbach, *Locking the Doors to Discovery?*]. For discussions of changes in party behavior induced by changing pleading standards under the Private Securities Litigation Reform Act, see Stephen J. Choi, *Do the Merits Matter*
footnote continued on next page

A critical question that arises in this discussion is whether there are any disputes that “select into summary judgment,” in the sense that they would develop into lawsuits that ultimately face summary judgment motions under *Twiqbal*, but not under the pre-*Twiqbal* pleading standard. While counterintuitive, the possibility that some such lawsuits exist is real.¹¹ This possibility is important because the empirical approach adopted in this Article to measure the quality-filtering effect of *Twiqbal* requires the assumption that there are no such cases—an underlying premise subsequently referred to as the “no-selection-into-summary-judgment” assumption. This assumption is strong in that it amounts to a restriction on the scope of party selection effects. Without the assumption, differences in the plaintiff’s win rate at summary judgment involve not only the quality-filtering effect discussed above, but also a component related to the quality of cases that are selected into summary judgment as a result of *Twiqbal*.

If there were such a mixture of quality filtering among some cases and selection among others, it would be impossible to use empirical evidence to isolate *Twiqbal*’s quality-filtering effects. For example, a finding of no change in the plaintiff’s win rate at summary judgment would be consistent with *either* no quality-filtering effect and no selection effect *or* a large quality-filtering effect that is exactly counteracted by a large selection effect operating in the opposite direction. In such a situation, neither possibility can be ruled out. Consequently, it would be impossible to assess commentators’ claims about *Twiqbal*’s quality-filtering effects in particular. Thus, the no-selection-into-summary-judgment assumption is necessary to assess commentators’ claims about *Twiqbal*’s effects on the mix of cases that make it past the Rule 12(b)(6) stage.¹²

Finally, as to the no-selection-into-summary-judgment assumption, I note that it is partially testable. If this assumption is correct, then *Twiqbal* would have eliminated some cases from the set facing summary judgment motions while not causing any new cases to face summary judgment motions. Thus, *Twiqbal* would have to reduce the share of filed cases that face summary judgment motions, all else being equal. If the no-selection-into-summary-

Less After the Private Securities Litigation Reform Act?, 23 J.L. ECON. & ORG. 598 (2007); and Stephen J. Choi et al., *The Screening Effect of the Private Securities Litigation Reform Act*, 6 J. EMPIRICAL LEGAL STUD. 35 (2009).

11. Because the details are a bit involved, they are deferred to Part I.D.2. In short, though, such a lawsuit need only satisfy two counterfactual conditions. First, if it were adjudicated under the old pleading standard, the lawsuit would be settled relatively early—before the answer/Rule 12(b)(6) stage. Second, if this lawsuit were adjudicated under the *Twiqbal* pleading standard, the parties would fail to settle after all, with the defendant then filing a summary judgment motion after discovery.
12. On the unavoidable role of making behavioral assumptions in empirical research about human behavior, including the effects of changes in legal rules, see Gelbach, *Dark Arts*, *supra* note 10, at 248-49.

judgment assumption is correct, then there should have been a drop in the share of cases facing summary judgment motions following *Twiqbal*.

The empirical work in this Article is based on a random sample of cases in which defendants filed Rule 56 summary judgment motions. This sample was created using a unique database of docket reports.¹³ To construct the dataset, all civil cases filed in the federal district courts in the periods of October 1, 2005 to June 30, 2006 (the pre-*Twombly* period) and October 1, 2009 to June 30, 2010 (the post-*Iqbal* period) were combed to find those which had at least one motion for summary judgment filed.¹⁴ After restricting attention to cases with a PACER code indicating that the nature of the suit was employment discrimination or contracts, cases were then sampled randomly from these two periods.¹⁵

This Article focuses its analysis on employment discrimination and contract cases for a simple reason: employment discrimination cases may hinge on the underlying motivation for an employer's adverse action, and these motivations may be difficult for an employee to ascertain without discovery. Thus, employment discrimination cases are relatively likely to involve the catch-22 information asymmetry problem embodied by what I call the "aggressive critics' view."¹⁶ By contrast, it seems reasonable to think that plaintiffs should be able to plead the bases for a breach of contract claim without the benefit of discovery. Thus, the aggressive critics' view of *Twiqbal* seems unlikely to hold in the contract litigation setting.

Somewhat unsatisfyingly, the results of the analysis in both contexts ultimately illustrates that empirics cannot conclusively resolve the case-quality aspects of the *Twiqbal* debate. The results for employment discrimination cases are statistically insignificant and generate wide confidence intervals. This fact suggests that data for employment discrimination cases do not allow us to distinguish between very different views about *Twiqbal*. For contract cases, there is relatively weak evidence in favor of an increase in the quality of cases that make it to summary judgment due to *Twiqbal*. However, for this category of cases, the data indicate that there is reason to doubt the assumption that there is no selection into summary judgment.¹⁷ In context, this means that the *Twiqbal*-induced change in the plaintiff win rate likely measures a combination of both (i) the case-quality effects of interest, and (ii) case quality among cases that select into summary judgment as a result of *Twiqbal*. This dual causality creates uncertainty as to whether the results for contract cases truly address the central debate over *Twiqbal*'s effects on case quality.

13. Part II discusses the source of these data in more detail.

14. These are the filing periods used in the reports by FJC researchers. See FJC FIRST STUDY, *supra* note 7, at 5.

15. Current and/or recently graduated law students then downloaded relevant case documents for the sampled cases and coded them.

16. See *infra* Part I.C.

17. The basis for such doubt is a bit involved and is explained in detail in Part IV.B.3.

These results underscore a disappointing but unavoidable fact: there are some empirical questions that cannot be clearly answered using feasible data. This Article's basic message, then, is that data are unlikely to settle the debate over the case-quality effects of the new pleading regime ushered in by *Twombly* and *Iqbal*.¹⁸

Part I of this Article discusses the methodological ideas that underpin the Article's analysis. Part II describes the data in detail. Parts III and IV present empirical results for employment discrimination and contract cases, respectively. Part V discusses some further methodological considerations and challenges.

I. *Twombly*, *Iqbal*, and What We Can Learn from Summary Judgment Results

A. The Evolution of Pleading Standards and How They Relate to Discovery

Before *Twombly*, *Conley v. Gibson* set the standard for Rule 12(b)(6) motions to dismiss in federal courts. According to *Conley*: "[A] complaint should not be dismissed for failure to state a claim unless it appears beyond doubt that the plaintiff can prove no set of facts in support of his claim which would entitle him to relief."¹⁹ This standard is both objective and relatively easy to satisfy.

In the half century between *Conley* and *Twombly*, a number of lower courts imposed effective pleading standards more demanding than *Conley*'s no-set-of-facts standard.²⁰ In response, the Supreme Court more than once reversed courts of appeals, affirming the *Conley* standard in no uncertain terms. One

18. The direction of *Twombly* and *Iqbal*'s quality-filtering effects is hardly alone on the list of important empirical questions that feasible data and methods are unlikely to resolve. See, for example, the debate over the existence of a deterrent effect of the death penalty, as investigated in John J. Donohue & Justin Wolfers, *Uses and Abuses of Empirical Evidence in the Death Penalty Debate*, 58 STAN. L. REV. 791, 794 (2005):

Our estimates suggest not just "reasonable doubt" about whether there is any deterrent effect of the death penalty, but profound uncertainty. We are confident that the effects are not large, but we remain unsure even of whether they are positive or negative. The difficulty is not just one of statistical significance: whether one measures positive or negative effects of the death penalty is extremely sensitive to very small changes in econometric specifications. Moreover, we are pessimistic that existing data can resolve this uncertainty.

19. *Conley v. Gibson*, 355 U.S. 41, 45-46 (1957). To be precise, *Conley* thereby set the pleading standard under Federal Rule of Civil Procedure 8(a)(2), which then has implications for Rule 12(b)(6) motions. Similarly, the *Twiqbal* cases change the Rule 8(a)(2) pleading standard, thereby affecting Rule 12(b)(6) motion practice.

20. For discussions of such cases, see Christopher M. Fairman, *The Myth of Notice Pleading*, 45 ARIZ. L. REV. 987 (2003); Richard L. Marcus, *The Puzzling Persistence of Pleading Practice*, 76 TEX. L. REV. 1749 (1998); and Richard L. Marcus, *The Revival of Fact Pleading Under the Federal Rules of Civil Procedure*, 86 COLUM. L. REV. 433 (1986).

such reversal occurred in 1993 in a constitutional civil rights case,²¹ while another occurred as recently as 2002 in a Title VII employment discrimination case.²²

In 2007's now-famous *Twombly* decision—regarding a putative class action involving allegations of conspiracy founded only on allegations of parallel conduct²³—the Court switched direction. Although some of the arguments in *Twombly* could be read as applying specifically to antitrust's substantive prohibition on drawing inferences of conspiracy from evidence indicating only the presence of parallel conduct, the Court was categorical in rejecting *Conley's* no-set-of-facts standard, holding that it had “earned its retirement.”²⁴ Two years later, the *Iqbal* Court eliminated any residual doubt concerning the reach of *Twombly's* new standard, straightforwardly holding that *Twombly's* plausibility standard governed pleading in “all civil actions.”²⁵ *Twombly* and *Iqbal* have been cited an enormous number of times: as of February 2016, Westlaw reported that *Twombly* had been cited in federal cases over 125,000 times; for *Iqbal*, the corresponding figure was over 100,000 times.

Twombly and *Iqbal* have also touched off a firestorm of debate among judges, practitioners, and academics.²⁶ Critics believe that these cases have

-
21. *Leatherman v. Tarrant Cty. Narcotics Intelligence & Coordination Unit*, 507 U.S. 163, 168 (1993) (“[I]t is impossible to square the ‘heightened pleading standard’ applied by the Fifth Circuit in this case with the liberal system of ‘notice pleading’ set up by the Federal Rules [and elaborated by the Court in *Conley*].”).
 22. *Swierkiewicz v. Sorema N.A.*, 534 U.S. 506, 512 (2002) (“[I]mposing the Court of Appeals’ heightened pleading standard in employment discrimination cases conflicts with Federal Rule of Civil Procedure 8(a)(2), which provides that a complaint must include only ‘a short and plain statement of the claim showing that the pleader is entitled to relief.’ Such a statement must simply ‘give the defendant fair notice of what the plaintiff’s claim is and the grounds upon which it rests.’” (first quoting FED. R. CIV. P. 8(a)(2); then quoting *Conley*, 355 U.S. at 47)).
 23. *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 550-51 (2007).
 24. *Id.* at 563.
 25. *Ashcroft v. Iqbal*, 556 U.S. 662, 684 (2009) (quoting FED. R. CIV. P. 1).
 26. See, e.g., *Barriers to Justice and Accountability: How the Supreme Court’s Recent Rulings Will Affect Corporate Behavior: Hearing Before the S. Comm. on the Judiciary*, 112th Cong. 2 (2011) (statement of Andrew Pincus, Partner, Mayer Brown LLP), <http://www.judiciary.senate.gov/download/testimony-of-pincuspdf> (“Two years ago, many asserted that the Court’s ruling in *Ashcroft v. Iqbal* . . . would dramatically restrict plaintiffs’ access to court and that Congressional action was needed to overturn that decision. That speculation has been proven wrong . . .” (citing FJC FIRST STUDY, *supra* note 7, at vii)); Mark Herrmann & James M. Beck, Debate, *Plausible Denial: Should Congress Overrule Twombly and Iqbal?*, 158 U. PA. L. REV. PENNUMBRA 141, 142 (2009), <http://www.pennlawreview.com/online/158-U-Pa-L-Rev-PENnumbra-141.pdf>; Colleen McMahon, *The Law of Unintended Consequences: Shockwaves in the Lower Courts After Bell Atlantic Corp. v. Twombly*, 41 SUFFOLK U. L. REV. 851, 852-53 (2008) (“[N]o one quite understands what the case holds. . . . We district court judges suddenly and unexpectedly find ourselves puzzled over something we thought we knew how to do with our eyes closed: dispose of a motion to dismiss a case for failure to
- footnote continued on next page*

destabilized the pleading system, and thus federal litigation generally.²⁷ Especially relevant to this Article, critics have argued that the plausibility pleading standard will reduce access to justice for plaintiffs with meritorious claims, especially in disputes whose alleged wrongdoers control access to the information necessary to meet the heightened pleading standard.²⁸ By contrast, supporters of *Twombly* and *Iqbal* point to their hoped-for role in reducing the burdens presented by meritless lawsuits.²⁹

The Court's opinions in the two cases themselves raise policy concerns related to the discovery burdens that defendants face. In *Twombly*, Justice Souter suggested that caution related to prediscovery dismissal must be counterbalanced against the expense of discovery in antitrust cases. He cited a law review student note that focused on the special discovery burdens antitrust

state a claim.”); Arthur R. Miller, *From Conley to Twombly to Iqbal: A Double Play on the Federal Rules of Civil Procedure*, 60 DUKE L.J. 1, 9-10 (2010).

27. See, e.g., Robert G. Bone, *Plausibility Pleading Revisited and Revised: A Comment on Ashcroft v. Iqbal*, 85 NOTRE DAME L. REV. 849, 852 (2010) (“*Iqbal* applies a thick screening model that aims to screen weak as well as meritless suits, whereas *Twombly* applies a thin screening model that aims to screen only truly meritless suits. The thick screening model is highly problematic on policy grounds[. . .]”); Kevin M. Clermont & Stephen C. Yeazell, *Inventing Tests, Destabilizing Systems*, 95 IOWA L. REV. 821, 840 & n.70 (2010) (writing that a defense attorney “commits legal malpractice if he or she fails to move to dismiss with liberal citations to *Twombly* and *Iqbal*” and quoting “experienced litigator” Tom Goldstein as “predict[ing] that *Iqbal* will be ‘the basis for an attempt to dismiss more than 50 percent of all the complaints filed in federal court” (quoting Melinda Hanson, *Civil Cases: Iqbal, Ricci, Wyeth, NAMUDNO Tagged as Landmark Opinions of the Term*, 78 U.S. L. WEEK 3025, 3026 (2009)); Michael C. Dorf, *Iqbal and Bad Apples*, 14 LEWIS & CLARK L. REV. 217, 218 (2010) (“*Iqbal* called into question an entire basis for civil liability for unconstitutional discrimination.”).
28. See, e.g., JOSHUA CIVIN & DEBO P. ADEGBILE, AM. CONSTITUTION SOC’Y, RESTORING ACCESS TO JUSTICE: THE IMPACT OF IQBAL AND TWOMBLY ON FEDERAL CIVIL RIGHTS LITIGATION 2 (2010), http://www.acslaw.org/sites/default/files/Civin_Adegbile_Iqbal_Twombly.pdf (expressing fear that *Twombly* and *Iqbal* might “create an undesirable safe harbor that effectively places some defendants beyond the reach of civil rights laws”); Arthur R. Miller, *Are the Federal Courthouse Doors Closing?: What’s Happened to the Federal Rules of Civil Procedure?*, 43 TEX. TECH L. REV. 587, 596 (2011) (“In many modern litigation contexts the critical information is in the possession of the defendant and unavailable to the plaintiff. . . . Discovery was designed to let each side have access to that type of information so that the litigation playing field would be level to promote more informed settlements and trials.”); Reinert, *supra* note 8, at 159 (“[C]ases in which state of mind plays a large role or in which there are large information asymmetries, such as civil rights, constitutional, and employment discrimination cases, are most likely to be vulnerable to accusations of thin pleading.”).
29. See, e.g., Richard A. Epstein, *Of Pleading and Discovery: Reflections on Twombly and Iqbal with Special Reference to Antitrust*, 2011 U. ILL. L. REV. 187, 196 (“[A] bad set of legal rules also leads to bad settlements. As a general matter, these settlements reflect the probable outcomes of cases that go to final judgment. Any errors in the overall procedural rules, therefore, are likely to be embedded in the settlements.”); Herrmann & Beck, *supra* note 26, at 146-47.

defendants face,³⁰ the Manual for Complex Litigation,³¹ and a Judicial Conference Committee document emphasizing discovery's high share of litigation costs in cases where it is used.³² Moreover, in responding to Justice Stevens's suggestion in dissent that judicial management of cases could address these issues, Justice Souter declared defeat in the use of case management "given the common lament that the success of judicial supervision in checking discovery abuse has been on the modest side."³³

Justice Souter's concern regarding the role of discovery expense is well summarized by the following passage:

[T]he threat of discovery expense will push cost-conscious defendants to settle even anemic cases before reaching [summary judgment]. Probably, then, it is only by taking care to require allegations that reach the level *suggesting* conspiracy that we can hope to avoid the potentially enormous expense of discovery in cases with no "reasonably founded hope that the [discovery] process will reveal relevant evidence" to support [an antitrust claim under section 1 of the Sherman Act].³⁴

Ashcroft v. Iqbal involved a defendant facing a different kind of discovery burden. The plaintiff alleged that after the attacks of September 11, 2001, he was arrested and held in administrative detention where he had suffered "brutal mistreatment and discrimination,"³⁵ including having been:

30. *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 558 (2007) (citing William H. Wagener, Note, *Modeling the Effect of One-Way Fee Shifting on Discovery Abuse in Private Antitrust Litigation*, 78 N.Y.U. L. REV. 1887, 1898-99 (2003)).

31. *Id.* at 558-59 (citing MANUAL FOR COMPLEX LITIGATION (FOURTH) § 30 (2004)).

32. *Id.* at 559 (citing Memorandum from Paul V. Niemeyer, Chair, Advisory Comm. on Civil Rules, to Anthony J. Scirica, Chair, Comm. on Rules of Practice & Procedure of the Judicial Conference of the U.S. (May 11, 1999), 192 F.R.D. 354, 357 (2000)).

33. *Id.* For this proposition, Justice Souter cited a law review article by Judge Frank Easterbrook, *see id.*, whose argument may be worth quoting at length:

The timing is all wrong. The plaintiff files a sketchy complaint (the Rules of Civil Procedure discourage fulsome documents), and discovery is launched. A judicial officer does not know the details of the case the parties will present and in theory *cannot* know the details. Discovery is used to find the details. The judicial officer always knows less than the parties, and the parties themselves may not know very well where they are going or what they expect to find. A magistrate supervising discovery does not—cannot—know the expected productivity of a given request, because the nature of the requester's claim and the contents of the files (or head) of the adverse party are unknown. Judicial officers cannot measure the costs and benefits to the requester and so cannot isolate impositional requests. Requesters have no reason to disclose their own estimates because they gain from imposing costs on rivals (and may lose from an improvement in accuracy). The portions of the Rules of Civil Procedure calling on judges to trim back excessive demands, therefore, have been, and are doomed to be, hollow.

Frank H. Easterbrook, *Discovery as Abuse*, 69 B.U. L. REV. 635, 638 (1989).

34. *Twombly*, 550 U.S. at 559 (third alteration in original) (emphasis added) (quoting *Dura Pharmaceuticals, Inc. v. Broudo*, 544 U.S. 336, 347 (2005)).

35. First Amended Complaint & Jury Demand at 1, *Elmaghraby v. Ashcroft*, No. 1:04-cv-01809 (E.D.N.Y. Sept. 27, 2005), *aff'd in part, rev'd in part sub nom. Iqbal v. Hasty*, 490 F.3d 143 (2d Cir. 2007), *rev'd sub nom. Ashcroft v. Iqbal*, 556 U.S. 662 (2009), 2004 WL 3756442.

deliberately and cruelly subjected to numerous instances of excessive force and verbal abuse, unlawful strip and body cavity-searches, denial of medical treatment, the denial of adequate nutrition, extended detention in solitary confinement, the denial of adequate exercise, and deliberate interference with . . . rights to counsel and to exercise of . . . sincere religious beliefs.³⁶

In addition to low-ranking defendants such as correctional officers, Iqbal sued then-Attorney General John Ashcroft and then-FBI Director Robert Mueller, alleging that the government had treated him unlawfully “as a matter of policy, solely on account of [his] religion, race, and/or national origin and for no legitimate penological interest.”³⁷

Echoing Justice Souter’s skepticism of managerial judging as a solution to discovery costs, Justice Kennedy wrote that the “rejection of the careful-case-management approach is especially important in suits where Government-official defendants are entitled to . . . [be free of] the concerns of litigation, including ‘avoidance of disruptive discovery.’”³⁸ Justice Kennedy also linked the Court’s earlier holding in *Twombly* explicitly to discovery, writing that although the Federal Rules of Civil Procedure’s pleading standard in Rule 8 “marks a notable and generous departure from the hypertechnical . . . regime of a prior era, . . . it does not unlock the doors of discovery for a plaintiff armed with nothing more than conclusions,” rather than hard factual information provided in the lawsuit’s complaint.³⁹

In sum, the majority opinions in both *Twombly* and *Iqbal* focus on the discovery burdens defendants can be expected to face in a system that allows merits determinations only after discovery. Each opinion suggests a belief that district courts will be able to usefully forecast, on the basis of the plaintiffs’ complaints, the set of cases in which discovery will yield evidence of liability, signaling the useful link between summary judgment results and the case-quality views held by supporters and opponents of *Twombly* and *Iqbal*.

B. Defining “Merit” Postdiscovery

Under Rule 56, summary judgment should be granted only when “there is no genuine dispute as to any material fact and the movant is entitled to judgment as a matter of law.”⁴⁰ Thus, summary judgment against a plaintiff’s claim is appropriate only when the plaintiff has been able to come up with no

36. *Id.* at 3.

37. *Iqbal*, 556 U.S. at 669 (alteration in original) (quoting First Amended Complaint & Jury Demand, *supra* note 35, at 17-18).

38. *Id.* at 685 (quoting *Siegert v. Gilley*, 500 U.S. 226, 236 (1991) (Kennedy, J., concurring in the judgment)).

39. *Id.* at 678-79.

40. FED. R. CIV. P. 56.

evidence that would allow a reasonable jury to find in her favor.⁴¹ One might say that where pleading standards involve judgments about the possibility (*Conley*) or plausibility (*Twiqbal*) that sufficient facts will materialize to entitle the plaintiff to relief, the summary judgment standard involves the *actuality* of entitlement to relief given those facts that have been discovered.

This Article defines a claim as having no merit postdiscovery—or just “no merit”—if summary judgment for the defendant is appropriate. Conversely, this Article assumes a claim “has merit postdiscovery” if, after discovery, the plaintiff would be able to survive a summary judgment motion. These definitions are not intended to, and do not, implicate a claim’s *pre*discovery merit. Importantly, there is no reason why a case that turns out to be meritless postdiscovery might not have merit *pre*discovery. To take an extreme example, suppose Paula has actually been injured due to someone’s negligence. After a reasonable investigation, she sues DeAndre, who appears to be responsible, only to find out through discovery that Dave was actually responsible.⁴² By my definition, Paula’s claim against DeAndre has no merit postdiscovery, even though many, if not all, would agree that it was meritorious prediscovery.

Of course, such a case is not the sort that *Twombly* and *Iqbal* purport to target; by their own terms, *Twombly* and *Iqbal* would not lead to Rule 12(b)(6) dismissal of Paula’s case since her complaint meets the plausibility standard. Nevertheless, my definition of merit postdiscovery is a good one for understanding *Twombly* and *Iqbal*’s effects on case quality. Even though it is true that *Twombly* and *Iqbal* are directed at early termination of cases based on a judge’s *pre*discovery assessment, the object of that assessment is whether, *after* discovery, there is likely to be any evidence of entitlement to relief. This object is precisely what the definition of postdiscovery merit is meant to encapsulate.

C. Understanding the Summary Judgment Link when Only Judges’ Behavior Changes Following *Twombly* and *Iqbal*

This Subpart begins to explain the empirical link between pleading standards and summary judgment, as well as how the arguments for and against higher pleading standards relate to this link. Because the possibility that parties might change their litigation behavior following a change in the pleading standard makes this analysis a bit complicated, it is helpful to start by

41. Note that “no evidence that would allow a reasonable jury to find in her favor” is not the same thing as no evidence at all. It is not sufficient for a plaintiff to come up with a mere “scintilla” of evidence, because courts impose a burden of production on parties that requires some minimal level of evidence that is more than simply a shred. *See Anderson v. Liberty Lobby, Inc.*, 477 U.S. 242, 252 (1986) (“The mere existence of a scintilla of evidence in support of the plaintiff’s position will be insufficient; there must be evidence on which the jury could reasonably find for the plaintiff.”).

42. *See, e.g., Zielinski v. Phila. Piers, Inc.*, 139 F. Supp. 408, 410-11 (E.D. Pa. 1956).

imagining that party behavior is unaffected by *Twiqbal*. The issue of changes in party behavior is subsequently taken up and addressed in Subpart D below.

In a suit with no merit postdiscovery, there will by definition be no genuine dispute concerning any material fact after discovery, and the defendant will be entitled to judgment as a matter of law. Therefore, defendants should always win their summary judgment motions in meritless suits. In his opinion for the *Twombly* Court, Justice Souter wrote that the plausibility standard “calls for enough fact to raise a reasonable expectation that discovery will reveal evidence of [entitlement to relief].”⁴³ Thus, one useful way to understand the plausibility standard is that it (i) asks trial judges to forecast what will result from discovery, and (ii) expects them to be able to forecast that result with some level of success. Translated into empirical terms, one may read Justice Souter’s opinion to mean that the plausibility standard “calls for enough fact to raise a reasonable expectation *that the case will have merit postdiscovery.*” This reading of Justice Souter’s opinion in *Twombly* allows us usefully to connect the plausibility pleading standard to summary judgment adjudication.

Accordingly, under the “supporters’ view,” judges are presumed to be good at forming expectations regarding the expected fruits of discovery. Consider a judge adjudicating a Rule 12(b)(6) motion to dismiss. If the supporters’ view is right, then the judge will be at least partly able to forecast whether the case will have no merit postdiscovery. Thus, under the *Twiqbal* pleading standard, judges will be more likely to dismiss cases that would turn out to have no merit postdiscovery. Elsewhere I have called such changes in what judges do “judicial behavior effects.”⁴⁴

However, the supporters’ view is not the only possible way of looking at the link between pleading and postdiscovery merit. As discussed above, *Twiqbal* critics worry that judges will erroneously dismiss at least some high-merit cases under the plausibility pleading standard because plaintiffs will be unable to allege sufficient information at the pleading stage. That is, these critics fear that some cases that would actually turn out to have merit postdiscovery will be dismissed at the Rule 12(b)(6) stage because judges wrongly forecast that discovery will indicate that such cases lack merit.

Among these *Twiqbal* critics, one might imagine two polar strains of critical views. Under what may be termed an “aggressive critics’ view,” judges will do such an inaccurate job of forecasting discovery results that they will be systematically more likely to dismiss high-merit cases than to dismiss low-

43. *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 556 (2007). The Court actually said “calls for enough fact to raise a reasonable expectation that discovery will reveal evidence of illegal agreement,” in reference to *Twombly*’s section 1 antitrust context. *Id.*

44. *See, e.g., Gelbach, Locking the Doors to Discovery?*, *supra* note 10, at 2276 (italics omitted).

merit cases.⁴⁵ Under an alternative “moderate critics’ view,” judges might be neither particularly good nor particularly bad at forecasting discovery results.⁴⁶ Accordingly, the moderate critics’ view holds that Rule 12(b)(6) dismissals will be essentially random with respect to postdiscovery case quality: the probability of dismissing high- and zero-merit cases will track their relative incidence in the population of cases facing Rule 12(b)(6) motions.⁴⁷

45. Such a view might be held by one who believes that at least some judges are unwarrantedly skeptical of plaintiffs in certain classes of cases, such as those involving allegations of employment discrimination. *See, e.g.,* Elizabeth M. Schneider, *The Changing Shape of Federal Civil Pretrial Practice: The Disparate Impact on Civil Rights and Employment Discrimination Cases*, 158 U. PA. L. REV. 517, 530 (2010) (endorsing observation by Stephen Burbank that *Iqbal* “obviously licenses highly subjective judgments” and that such a license constitutes “a blank check for federal judges to get rid of cases they disfavor” (quoting Adam Liptak, *Case About 9/11 Could Lead to a Broad Shift on Civil Lawsuits*, N.Y. TIMES, July 21, 2009, at A10)).

46. For an expression that is consistent with the moderate critics’ view, see Reinert, *supra* note 8, at 161, stating that “there is no reason to think that courts will be skillful in determining which potentially dismissible case is most likely to be shown to be meritorious.”

47. Notice that I have assumed that the applicable pleading standard has no bearing on whether a given plaintiff would win at summary judgment, once a case actually gets there. This assumption can be justified from two more primitive assumptions. The first primitive assumption is that the pleading standard does not affect the record that will exist once the parties get through discovery and to summary judgment, if indeed a case survives to discovery. The second primitive assumption is that, for a given record, judges will adjudicate summary judgment motions the same way post-*Iqbal* as pre-*Twombly*. If both these assumptions hold, then a given case that makes it to summary judgment under both pleading standards will have the same record, which will be adjudicated the same way, under both pleading standards.

One might challenge the first assumption on the ground that *Twombly* and *Iqbal* give plaintiffs an incentive to conduct a more comprehensive investigation before filing suit in order to draft a complaint that is more likely to be found plausible in the event that a Rule 12(b)(6) motion is filed. Would more intensive prefiling investigations lead to a more plaintiff-friendly record post-*Iqbal* than pre-*Twombly*? Not necessarily, because it is possible—indeed, arguably likely—that evidence relevant to the summary judgment record that a plaintiff could uncover in a pre-filing investigation under the post-*Iqbal* pleading standard would also turn up *during discovery* under the pre-*Twombly* pleading standard.

Since the summary judgment standard was unaffected by *Twombly*, the second assumption is really quite undemanding. Still, one might imagine arguments to the contrary. For example, numerous scholars argue that *Twombly* and *Iqbal* are part of a broad campaign of retrenchment against plaintiffs. *See, e.g.,* Stephen B. Burbank & Sean Farhang, *Litigation Reform: An Institutional Approach*, 162 U. PA. L. REV. 1543, 1614 (2014); Eisenberg & Clermont, *supra* note 7, at 193; Schneider, *supra* note 45, at 520. If judges respond to such a campaign by “tightening up” on plaintiffs across the board, then the second primitive assumption would fail to hold. It is worth noting that if such “zeitgeist” effects do occur, then it is highly unlikely that there will exist *any* empirical strategy that allows us to learn about the quality-filtering effects of *Twombly*. Some behavioral assumptions are inevitably necessary to allow one to estimate the effects of legal rules. *See* Gelbach, *Dark Arts*, *supra* note 10, at 245.

Table 1 summarizes the analysis so far. The first row of the Table illustrates the *Twiqbal* supporters' view that cases that would turn out to have no merit postdiscovery will be eliminated by switching to the plausibility pleading standard. As a result, (i) there will be fewer cases facing summary judgment motions, and (ii) among those cases that do face such motions, an *increased* share will have merit postdiscovery. Consequently, under the supporters' view, the plaintiff's win rate on summary judgment motions raised by the defendant should be greater among cases filed under the *Twiqbal* pleading standard.

Table 1
Predictions Flowing from Supporters' and Critics' Views
when There Is No Selection into Summary Judgment

View	Impact on:		
	(1) Number of Cases Facing Summary Judgment Motions	(2) Share of Cases that Have Merit Postdiscovery Among Those Facing Defense Summary Judgment Motions	(3) Plaintiff's Win Rate Against Summary Judgment Motions
Supporters	Drops	Increases	Increases
Aggressive Critics	Drops	Drops	Drops
Moderate Critics	Drops	Stays Unchanged	Stays Unchanged

Table 1's second row indicates the aggressive critics' view—that switching to a plausibility pleading standard will result in the elimination of meritorious cases. Under this view, (i) there will be fewer cases facing summary judgment motions, and (ii) among those cases that do face such motions, a *reduced* share will have merit postdiscovery. Consequently, under the aggressive critics' view, the plaintiffs' win rate against defense summary judgment motions should be lower among cases filed under the *Twiqbal* pleading standard. The third row indicates the moderate critics' view—that cases will be randomly filtered out as a result of *Twombly* and *Iqbal*, leading to no change in the quality of cases that actually face defense summary judgment motions. As a result, (i) there will be fewer cases facing summary judgment motions, and (ii) among those cases that do face such motions, there will be *no change* in the share that would have merit postdiscovery. Consequently, under the moderate critics' view, the plaintiff's win rate against defense summary judgment motions should be unaffected by switching to the *Twiqbal* pleading standard.

Of course, the three views detailed here—supporters', aggressive critics', and moderate critics'—are ideal types. The reality is presumably somewhere in

the middle. For example, judges' use of *Twiqbal* to do quality filtering might lead to some erroneous dismissals of meritorious cases, which is in line with both types of critics' views, even as they dismiss proportionally more cases that have no merit, which is in line with the supporters' view. One way of thinking about the empirical approach used in this Article, then, is that it has the potential to identify not so much which of the three views in Table 1 is entirely correct, but which is *more* correct.

Table 1 allows us to connect an observable fact and a key question of interest in understanding *Twiqbal*'s effects. The observable fact is the direction in which the plaintiff's win rate against defense summary judgment motions changes. The question of interest is whether *Twiqbal* has led district court judges to systematically filter out cases that would have no merit postdiscovery, as the supporters' view predicts, or whether one of the critics' views holds instead.

In fact, under our maintained assumption of no change in party behavior, empirical evidence would enable even further illumination of the issue. As we have discussed, *Twiqbal*'s more demanding pleading standard forces judges to filter out some cases at the Rule 12(b)(6) stage. But *Twiqbal* certainly should not cause judges to deny any Rule 12(b)(6) motions they would otherwise grant. Accordingly, while *Twiqbal* should cause some cases to be filtered out of the set of cases that would face summary judgment motions if allowed to get to discovery, *Twiqbal* should not cause any cases to select into summary judgment absent party selection effects: there should not be any cases that would face summary judgment under the *Twiqbal* pleading standard but not under the pre-*Twiqbal* standard. This means that any difference in the pre-*Twombly*/post-*Iqbal* plaintiff's win rate against defense summary judgment motions would have to be driven by case-quality differences between filtered-out cases, on the one hand, and those that face summary judgment motions under both pleading standards, on the other.

Consider a simple, concrete example. Suppose that 900 cases would face summary judgment motions if litigated at a time when the pre-*Twombly* pleading standard applied. Now suppose that only 600 of these cases would face summary judgment motions if litigated at a time when the *Twiqbal* standard applied. This means that there are 300 cases that would be filtered out of the system before facing summary judgment motions as a result of *Twiqbal*. Now suppose that the plaintiff's win rate against defense summary judgment motions in these cases is 35% if the cases are litigated at a time when the pre-*Twombly* pleading standard applied, and 40% if litigated under *Twiqbal*. The impact of *Twiqbal* on the overall plaintiff's win rate at summary judgment would thus be 5 percentage points (40% minus 35%). However, by construction, there are 600 cases whose summary judgment outcome remains unaffected by *Twiqbal*. Thus, the impact of 5 percentage points is driven by the 300 cases that are filtered out of summary judgment due to *Twiqbal*. These 300 cases constitute one-third of the original 900 cases. Intuition suggests that if filtering

out one-third of cases cause a 5-point increase in the plaintiff's win rate, then the plaintiff's win rate among the filtered-out cases must be less than the plaintiff's win rate among other cases by 15 percentage points (3 times 5 percentage points).

Thus, we should surmise that the plaintiff's win rate among the 300 filtered-out cases is 15 percentage points less than plaintiff's win rate in the cases whose summary judgment status is unaffected by *Twiqbal* (in this example the supporters' view is correct). The same reasoning suggests that if the filtered-out cases constituted one-fourth of the total, then we should multiply the observed difference in the plaintiff's win rate by four; were they half of the total number of cases, then we should multiply by two; and so on. This intuition is exactly right. It can be proved mathematically that when there is no selection into summary judgment, the difference between the plaintiff's win rate among filtered-out cases and those that would have summary judgment motions under both pleading standards equals the ratio of (i) the change in the plaintiff's win rate against defense summary judgment motions, to (ii) the share of cases that are filtered out from facing summary judgment as a result of *Twiqbal*.⁴⁸

48. Let W_{both} be the plaintiff's win rate among cases that would face summary judgment motions under both pleading standards, and let W_{fo} be the corresponding plaintiff's win rate among cases that would be filtered out of facing summary judgment motions due to *Twombly* and *Iqbal*. Define the quality-filtering effect $Q = W_{both} - W_{fo}$; this is the effect we wish to identify. I shall now prove that Q may be determined from observable data under the assumption of no selection into summary judgment.

Let W_{pre-T} and W_{post-I} be the observed plaintiff's win rate against defense summary judgment motions under the pre-*Twombly* and post-*Iqbal* pleading standards, respectively. Let W_{both} be the plaintiff's win rate against defense summary judgment motions among cases that would face summary judgment motions under both pleading standards, and let W_{si} be the corresponding plaintiff's win rate among any cases that select into summary judgment. Lastly, denote by π the share of pre-*Twombly* cases with summary judgment motions that would be filtered out by *Twiqbal*, and denote by λ the share of post-*Iqbal* cases with summary judgment motions that select into summary judgment (note that $\lambda = 0$ under the assumption of no party selection effects, which I have maintained in the discussion in this section, but it is useful to allow it to be nonzero for expositional purposes in this footnote). Then we may write the observed plaintiff's win rate as:

$$\begin{aligned} W_{pre-T} &= (1 - \pi)W_{both} + \pi W_{fo} \\ W_{post-I} &= (1 - \lambda)W_{both} + \lambda W_{si} \end{aligned} \tag{1}$$

It follows that the change in the observed plaintiff's win rate is:

$$\begin{aligned} W_{post-I} - W_{pre-T} &= (\pi - \lambda)W_{both} + \lambda W_{si} - \pi W_{fo} \\ &= (\pi - \lambda)[W_{both} - W_{fo}] + \lambda\{W_{si} - W_{fo}\} \end{aligned} \tag{2}$$

where the second line follows by adding and subtracting λW_{fo} to the right hand side of the first line. If we then adopt the assumption of no selection into summary judgment, so that λ is zero, the term in curly braces disappears, and we obtain:

$$W_{post-I} - W_{pre-T} = \pi[W_{both} - W_{fo}] \tag{3}$$

Now observe that if we divide both sides of equation (3) by π and rearrange slightly, we get:

footnote continued on next page

To recap the discussion in this Part, this Article has adopted two assumptions and derived three key analytical results:

- Assumption 1: There is no change in party behavior as a result of *Twiqbal*.
- Assumption 2: How cases are adjudicated at summary judgment is not affected by the pleading standard that governs through the Rule 12(b)(6) stage.⁴⁹

$$Q = W_{both} - W_{fo} = \frac{W_{post-I} - W_{pre-T}}{\pi} \quad (4)$$

In other words, the quality-filtering effect, $W_{both} - W_{fo}$, equals the ratio of the observed change in the plaintiff's win rate against defense summary judgment motions divided by the filtered-out share π . To make this equation useful requires us to have a way to determine π using observable data.

Let N_{pre-T} and N_{post-I} be the number of cases that would face summary judgment motions under the pre-*Twombly* and post-*Iqbal* pleading standards, respectively. Let N_{both} , N_{fo} , and N_{si} respectively be the number of cases that would face summary judgment motions under both pleading standards, the number that would be filtered out due to *Twiqbal*, and the number that would select into summary judgment due to *Twiqbal*. Then it is always true that:

$$\begin{aligned} N_{post-I} &= N_{both} + N_{si} \\ N_{pre-T} &= N_{both} + N_{fo} \end{aligned} \quad (5)$$

Subtracting each side of the bottom equation in (5) from the corresponding side of the top equation then yields:

$$N_{post-I} - N_{pre-T} = N_{si} - N_{fo} \quad (6)$$

so that the observed change in the number of cases facing summary judgment motions equals the difference in the number of cases that select into summary judgment and the number that are filtered out. Under the assumption of no selection into summary judgment, we have $N_{si} = 0$; adopting this assumption and dividing both sides of (6) by N_{pre-T} then yields:

$$\frac{N_{post-I} - N_{pre-T}}{N_{pre-T}} = -\frac{N_{fo}}{N_{pre-T}} \quad (7)$$

and since the ratio on the right hand side of (7) is the filtered-out share, we have:

$$\pi = -\frac{N_{post-I} - N_{pre-T}}{N_{pre-T}} \quad (8)$$

In other words, π may be calculated from observable data: it equals negative-1 times the change in the number of cases that face summary judgment motions, expressed as a ratio of the number of cases that would face summary judgment motions under the pre-*Twombly* pleading standard. Combining equations (4) and (8) thus yields:

$$Q = W_{both} - W_{fo} = \frac{W_{post-I} - W_{pre-T}}{-\left(\frac{N_{post-I} - N_{pre-T}}{N_{pre-T}}\right)} \quad (9)$$

All terms on the right hand side of equation (9) are observable quantities. This proves that under the assumption of no selection into summary judgment, the quality-filtering effect of *Twiqbal*, $W_{both} - W_{fo}$, may be found using observable data.

49. For a discussion of this assumption, see note 47 above.

- Result 1: The number of cases that face summary judgment motions will be lower under *Twiqbal* than under the pre-*Twiqbal* pleading standard.
- Result 2: If the supporters are correct, the plaintiff's win rate against defense summary judgment motions will be greater under *Twiqbal* than under the pre-*Twiqbal* pleading standard; if the aggressive critics are correct, this rate will be lower under *Twiqbal*; and if the moderate critics are correct, the plaintiff's win rate will be unaffected by the switch to *Twiqbal*.
- Result 3: The difference between the plaintiff's win rate among cases that are filtered out of facing summary judgment motions and those that would face these motions regardless of the pleading standard equals (i) the observed difference in the plaintiff's win rate across pleading standards, divided by (ii) the percentage drop in the number of cases facing summary judgment motions.

Subpart D below turns to the important question of how party selection effects affect the analysis described herein.

D. Party Selection Effects of *Twombly* and *Iqbal*

The discussion above in Subpart C assumed that *Twombly* and *Iqbal*'s filtering effect operates only via changes in judicial behavior among cases that face a Rule 12(b)(6) motion. But *Twombly* and *Iqbal* can be expected to have broader effects, via changes in party behavior that can be thought of as "party selection effects."⁵⁰ Three of these types of party selection effects involve filtering out cases at the pre-Rule 12(b)(6) stage; these are discussed in Part IV.D.1. The fourth type involves the selection of cases into summary judgment; this type is discussed in Part IV.D.2.

1. Party selection effects that involve filtering out at the Rule 12(b)(6) stage

To illustrate party selection effects, consider a hypothetical dispute in which Smith believes that Jones, Inc. has discriminated against her. Assume for the moment that under the pre-*Twombly* pleading standard, Jones will forego filing a Rule 12(b)(6) motion, and the case would go through discovery and then face a summary judgment motion. However, under the post-*Iqbal* pleading standard, Jones might become sufficiently optimistic about prevailing on Rule 12(b)(6) that Jones would file such a motion if Smith filed a complaint, and knowing this, and being pessimistic about her chances of surviving a Rule 12(b)(6) motion under *Twiqbal*, Smith might choose not to file suit in the

50. See, e.g., Gelbach, *Locking the Doors to Discovery?*, *supra* note 10, at 2280.

first place. This is an example of what I have elsewhere referred to as “plaintiff selection.”⁵¹ In this example, a dispute that would make it to summary judgment under the pre-*Twombly* pleading standard would not get there post-*Iqbal*.

Consider a variation on the previous example’s theme. The pre-*Twombly* situation is the same as in the previous example. But under the post-*Iqbal* standard, Smith is more optimistic about her Rule 12(b)(6) chances, and so she files suit even after *Twiqbal*. If Jones files a Rule 12(b)(6) motion and prevails on it, then once again we have a dispute that would make it to summary judgment under the pre-*Twombly* pleading standard but that would not get there post-*Iqbal*. This is an example of what I have elsewhere referred to as “defendant selection.”⁵²

For a second variation on the original example, suppose both parties expect that under the post-*Iqbal* standard the judge would be very likely to grant a Rule 12(b)(6) motion. Then under that pleading standard, the parties might be able to find a mutually agreeable settlement amount before the Rule 12(b)(6) stage. In this example, too, then, we have a dispute that would make it to summary judgment pre-*Twombly* but that would not get there post-*Iqbal*; this is an example of what I have elsewhere referred to as “settlement selection.”⁵³

These three examples involve three different types of party selection, but in each case, *Twiqbal*’s selection effects cause a reduction in the number of cases that make it to summary judgment. Thus, these types of party selection operate in tandem with the judicial behavior effects discussed above in Part I.C,⁵⁴ to filter cases out of the set that face summary judgment motions.

Accordingly, the analysis above applies to these types of party selection as well as to judicial behavior effects. However, it is important to observe that when there are party selection effects, *Twiqbal*’s quality-filtering effect must be interpreted more carefully. This effect can no longer be viewed as indicating only whether judges are good at forecasting postdiscovery case merit. With party selection effects of the types discussed above, *Twiqbal*’s quality-filtering effects operate through both judicial filtering and party filtering. Because there is no evident way to determine which filtered-out cases are filtered out due to

51. *Id.* at 2276 (italics omitted).

52. *Id.* (italics omitted). Such a dispute is one in which the plaintiff files suit under either pleading standard, with the defendant allowing the case to go to discovery under the pre-*Twiqbal* standard but filing a Rule 12(b)(6) motion under the post-*Twiqbal* standard. *See id.*

53. *See id.* at 2277. For a recent empirical study concerning settlement rates before *Twombly* and after *Iqbal*, see Victor Abel Pereyra & Benjamin Sunshine, *Access-to-Justice v. Efficiency: An Empirical Study of Settlement Rates After Twombly & Iqbal*, 2015 U. ILL. L. REV. 357.

54. *See supra* note 44 and accompanying text.

changes in judicial behavior at Rule 12(b)(6) and which are filtered out due to party selection, all we can hope to measure is the overall net change resulting from the combined effects of both variables.

Table 1 remains germane, but the channels through which *Twiqbal* operates are broader: they encompass changes in both judicial and party behavior.⁵⁵

2. A party selection effect that involves selection into summary judgment

There remains one type of party selection effect to discuss. Consider again our hypothetical dispute between Smith and Jones. But now assume that under the notice pleading standard, Smith and Jones would settle the dispute before Smith files suit, in part because Jones would not expect to be able to win a pre-discovery dismissal of a lawsuit, should Smith file suit. If the same dispute occurred when the post-*Twiqbal* pleading standard applied, though, Jones might believe it could win a Rule 12(b)(6) motion, making the company unwilling to settle for an amount Smith would accept. Now suppose that Jones, having been overly optimistic, lost its post-*Twiqbal* Rule 12(b)(6) motion. If the parties did not then settle, then under the post-*Twiqbal* pleading standard, *Smith v. Jones, Inc.* would reach the discovery phase and possibly also the summary judgment phase.⁵⁶

This example shows that changes in the pleading standard can eliminate settlements, causing cases to select into summary judgment. An immediate consequence of this fact is that it is possible for the number of cases that face summary judgment motions to rise as a result of *Twombly* and *Iqbal*. Whether that happens will depend on the relative magnitudes of filtering out by judges

55. Insofar as *Twiqbal* shifted the pre-Rule 12(b)(6) balance of power in favor of defendants, cases that are filtered out of summary judgment due to party selection can be expected to have done so on terms that are both beneficial to defendants and detrimental to plaintiffs. That places party-filtered cases and judge-filtered cases on a similar footing—both involve *Twombly* and *Iqbal* working to cause cases to end before summary judgment in ways that benefit defendants at plaintiffs' expense.

56. It is possible that the parties would settle following the denial of Jones's Rule 12(b)(6) motion; after all, under the notice pleading standard, they would settle before Smith even filed suit. But it is also possible that the parties would not settle. One inducement to settle before either party files suit is that it allows the parties to avoid all costs not yet sunk into the filing and defending of the suit. Once the plaintiff files suit, she has sunk some costs, reducing the scope for a settlement to leave both parties better off. Similarly, by litigating its Rule 12(b)(6) motion, Jones might have made some investigations that also reduce its cost of post-Rule 12(b)(6) litigation; sinking such costs further reduces the scope for mutually beneficial settlement. Thus, some cases that would (i) settle without the filing of a complaint under the notice pleading standard but (ii) face a Rule 12(b)(6) motion under the plausibility pleading standard might not settle following denial of a Rule 12(b)(6) motion.

and parties, on the one hand, and selection into summary judgment due to failed settlements, on the other.

One thing that can be said is that if a change from the pre-*Twiqbal* pleading standard to the post-*Twiqbal* pleading standard does cause an increase in the number of cases facing summary judgment motions, then there must have been selection into summary judgment.⁵⁷ Beyond that, the analysis of what can be gleaned from the data when there is selection into summary judgment is relatively complicated. Differences in the plaintiff's win rate at summary judgment across cases filed under different pleading standard regimes then involve both (i) the quality-filtering effect discussed above, and (ii) the difference in case quality between cases that are filtered out and cases that select into summary judgment.⁵⁸ In this situation, empirical evidence will not be able

57. Note that the converse is not true: even if the change in the pleading standard causes a decrease in the number of cases facing summary judgment motions, it is still possible that there is some selection into summary judgment; there might just be more filtering out than selection into summary judgment.

58. This fact may be seen using the second line of equation (2) from note 48 above, which shows that the difference in the post-*Iqbal* and pre-*Twombly* plaintiff's win rates may be written as:

$$W_{post-I} - W_{pre-T} = (\pi - \lambda)[W_{both} - W_{fo}] + \lambda\{W_{si} - W_{fo}\}$$

The parameter λ is the share, among cases that faces summary judgment motions under *Twiqbal*, of cases that would not have had summary judgment motions filed under the pre-*Twiqbal* pleading standard; in other words, λ is the share of *Twiqbal*-standard cases that were selected into summary judgment. When λ is zero—when there is no selection into summary judgment—this equation is necessarily proportional to the quality-filtering effect $[W_{both} - W_{fo}]$ discussed above. But when there is selection into summary judgment, λ is not zero, and the term $\{W_{si} - W_{fo}\}$ comes into play as well. This term is the difference in the plaintiff's win rate—the case-quality difference—between cases that select into summary judgment and those that are filtered out.

An example will help illustrate how important selection into summary judgment can be in determining what conclusions may be drawn from the observed difference in the plaintiff's win rate. Suppose that the number of cases with summary judgment motions filed under the two pleading standards would be exactly the same. This equality necessarily means that the number of filtered-out cases and the number of cases that are selected into summary judgment are the same; since there are the same total number of cases with summary judgment motions filed under the two pleading standards, this result in turn means that π and λ must be equal. Consequently, what I have called the quality-filtering effect, $[W_{both} - W_{fo}]$, drops out of the equation above. What is left is the term $\lambda\{W_{si} - W_{fo}\}$. Since λ is a proportion, it is necessarily positive. Thus, when there is no change in the number of cases with summary judgment motions filed, so that $\pi = \lambda$, the observed difference in the plaintiff's win rate has the same sign as $\{W_{si} - W_{fo}\}$. In other words, the information in the observable data tells us whether cases that are selected into summary judgment or cases that are filtered out of summary judgment have greater quality, as measured by the plaintiff's win rate against defense summary judgment motions. In this example we would be unable to learn anything at all about quality among cases that would face summary judgment motions under both pleading standards. That would prevent us from determining whether either the filtering-out or selecting-in channels of *Twiqbal* improve quality. When $\pi = \lambda$ —when there is no difference in the number of summary judgment

footnote continued on next page

to resolve the debates over *Twombly*'s case-quality effects that have occurred in recent years.

II. Data

The data used here originally became available as a result of a grant funded by the Oscar M. Ruebhausen Fund at the Yale Law School.⁵⁹ That grant funded a contract between Yale and Thomson Reuters, owner of Westlaw, to provide direct access to the universe of federal district court docket reports for civil cases filed beginning on January 1, 2005. These are the docket-sheet data one can search on Westlaw via its “DCT” database.⁶⁰

The sample used for this study was constructed by first searching the docket reports in all employment discrimination or contract cases for text suggesting that a docket entry involved the filing of a motion for summary judgment.⁶¹ Next, docket-sheet entries were collected from each selected case, and the cases were sorted by the number generated using a computerized pseudorandom number generator. Cases were then served in this random order to coders who logged on to a secure web-based coding site. The coder assigned to particular cases then read through those cases' docket sheets, looking for entries that appeared to be motions for summary judgment and orders resolving them. These coders then downloaded the relevant documents—motions and resolving orders—from a legal research site, read the documents, and entered details concerning the motions and their resolutions.

Employment discrimination cases were chosen both because there are a large number of them and because many critics of *Twombly* and *Iqbal* have focused on employment discrimination suits as among those most prone to the information asymmetry that might create a need-discovery-to-get-to-discovery catch-22.⁶² Contract cases, on the other hand, were chosen because

motions faced by cases under the two pleading standards—the data simply cannot tell us anything about such questions.

59. Yale Law School's William N. Eskridge and I jointly submitted this grant.

60. Pursuant to the contract, Thomson Reuters delivered docket report data in raw form. Various database and scripting tools were used to select cases with at least one docket entry whose text indicated that the docketed event was a motion for summary judgment under Rule 56. These cases were uploaded into a database connected to a web-based coding tool. A number of current J.D. students and recently graduated students working with the project then read and coded judicial opinions and orders related to cases' summary judgment motions (as well as, where needed, the underlying motions or other case documents).

61. Roughly speaking, docket entries were searched to determine whether they began with the phrase “motion for summary judgment” or certain variations thereon.

62. Cases involving employment discrimination are those with PACER nature-of-suit code 442. All cases were dropped for which coding indicated the presence of any claims, among those challenged by a defense summary judgment motion, that were potentially related to the Americans with Disability Act (ADA). See *Americans with Disabilities Act*

footnote continued on next page

plaintiffs in these cases would seem to be less likely to require information that can only be acquired in discovery to satisfy *Twombly* and *Iqbal*'s plausibility standard.⁶³

This study includes cases that were initially filed in either of two time periods: (i) October 1, 2005 to June 30, 2006 (the "Pre-*Twombly* observation period"), and (ii) October 1, 2009 to June 30, 2010 (the "Post-*Iqbal* observation period"). These time periods were chosen because they were the ones used by the Federal Judicial Center (FJC) to evaluate changes in Rule 12(b)(6) filing rates in its initial study of the effects of *Twombly* and *Iqbal* on Rule 12(b)(6) practice.⁶⁴ These time periods are appropriate because the pre-*Twombly* observation period ends well in advance of *Twombly* (the Supreme Court released its opinion on May 21, 2007), and the post-*Iqbal* observation period begins several months after *Iqbal* (the Supreme Court released its opinion on May 18, 2009).⁶⁵

of 1990, Pub. L. No. 101-336, 104 Stat. 327 (codified as amended at 42 U.S.C. §§ 12101-213 (2014)). This exclusion of cases is warranted because the ADA Amendments Act of 2008 expanded the set of people protected by the ADA. See ADA Amendments Act of 2008, Pub. L. No. 110-325, § 2(b)(1), 122 Stat. 3553, 3554 (2008) (codified as amended at 42 U.S.C. § 12101(b)); see also Seiner, *Pleading Disability*, *supra* note 7, at 108 (noting that the amendments in question "explicitly seek to 'reinstat[e] a broad scope of protection'" (alteration in original) (quoting ADA Amendments Act of 2008, § 2(b)(1))). This change could be expected to induce more ADA-related cases and claims all else equal, so that defense summary judgment adjudication might differ across pre-*Twombly* and post-*Iqbal* cases involving the ADA for reasons unrelated to the conceptual issues discussed in Part I. There are reserved PACER codes for ADA cases (code 445 is for ADA employment claims, and 446 is for other ADA claims). However, some cases with ADA claims are coded under PACER's nature-of-suit code 442, the omnibus employment discrimination code. These cases can be identified using the brief textual description of challenged claims that coders were asked to create. Claims were coded as ADA-related when this description contained any of the following strings: "isabil," "isable," or "ADA." It is possible that not all ADA-related cases were flagged this way by coders, but as many have been excluded as could have been.

63. Cases involving contracts are those with PACER nature-of-suit codes between 110 and 190.

64. See FJC FIRST STUDY, *supra* note 7, at 25.

65. It is possible that with many delays or repeated pleading amendments, a case in my pre-*Twombly* sample could have been at risk of facing a Rule 12(b)(6) motion *after Twombly* was decided. To assess this possibility, the earliest date on which each case had a summary judgment motion filed was coded. Among the cases used in the final analysis pre-*Twombly* sample, 87% of employment discrimination and 84% of contract cases had a summary judgment motion filed before *Twombly*, so none of these cases could have faced a Rule 12(b)(6) motion after *Twombly* (since Rule 12(b)(6) motions must be filed earlier than summary judgment motions). Note also that any Rule 12(b)(6) motion converted to a summary judgment motion must be decided under Rule 56. Results calculated without the remaining 13% and 16% of cases were very similar to those reported below. Finally, since discovery takes time, it is perhaps reasonable to assume that a case would not face a summary judgment motion fewer than ninety days following the filing of a Rule 12(b)(6) motion. Among cases in the final analysis pre-*Twombly* sample, 96% of employment discrimination and 94% of contract cases had their first summary judgment motion filed within ninety days following May 21, 2007.

footnote continued on next page

Further, the FJC characterizes the period between January and June of 2006 as one “of stable motion practice,” and its report suggests that the period between January and June of 2010 is appropriate because it occurs after “each of the circuits had had a chance to publish at least one appellate court opinion interpreting *Ashcroft v. Iqbal* and offering guidance to the district courts.”⁶⁶ Thus the observation periods used here are reasonable. Given the data that were available at the time case coding was undertaken, the sample used included only cases with summary judgment motions adjudicated within 731 days of case filing.⁶⁷ Finally, like the FJC reports, this study uses only cases in which plaintiffs were represented by counsel.⁶⁸

Thus there seems little basis for concern about whether the cases in my pre-*Twombly* sample were actually affected by *Twombly*.

66. See FJC FIRST STUDY, *supra* note 7, at 36.

67. The latest date on which a case could have been filed and included in the analysis was June 30, 2010. At the time the coders began work, the database had up-to-date information on cases through June 30, 2012. Therefore, docket-report information was available for up to 731 days (2012 having been a leap year). For cases filed on dates earlier than June 30, 2010, there were more days of information, but such information was disregarded in order to allow the same period of observation for all cases considered. Consequently, the sample includes only cases with summary judgment motions adjudicated within 731 days of case filing. Without the 731-day limit, noncomparability problems could occur. To illustrate, consider two cases—one filed on June 30, 2006, and one filed on June 30, 2010. The docket reports’ text is up-to-date through June 30, 2012 for both cases. Therefore, there are 2192 days of docket information for the earlier case (four years having 365 days, plus two leap years having 366 days), by comparison to the 731 days of docket information for the later-filed case. Suppose that in general there are two types of cases, simple and complex, and suppose that simple cases always have motions for summary judgment filed and adjudicated within 731 days of case filing, while complex cases have these motions filed and adjudicated between 731 and 2192 days of case filing. Then an unrestricted search of earlier cases’ docket reports would yield a data set that included both simple and complex cases, whereas such a search of later cases’ docket reports would yield a data set including only simple cases. If claims challenged by defense summary judgment motions have different average merit levels in simple and complex cases, then ignoring the different lengths of data availability would bias the results. To avoid this potential problem and maintain comparability of the pre-*Twombly* and post-*Iqbal* data sets, only cases whose summary judgment motions are adjudicated within 731 days of case filing are considered.

68. Here I simply follow the approach in FJC FIRST STUDY, *supra* note 7, at 6 n.10. Shortly after handing down *Twombly*, the Supreme Court in *Erickson v. Pardus* reaffirmed that “a *pro se* complaint, however inartfully pleaded, must be held to less stringent standards than formal pleadings drafted by lawyers.” 551 U.S. 89, 94 (2007) (per curiam) (quoting *Estelle v. Gamble*, 429 U.S. 97, 106 (1976)). *Estelle v. Gamble*, in turn, cites directly to *Conley’s* no-set-of-facts language. See 429 U.S. at 106 (quoting *Conley v. Gibson*, 355 U.S. 41, 45-46 (1957)). For this reason, some empirical studies concerning *Twombly* and *Iqbal* exclude *pro se* cases, see, e.g., FJC FIRST STUDY, *supra* note 7, at 6 n.10, though the approach Cecil and his coauthors took has been somewhat controversial, see Moore, *supra* note 7, at 607. There were many fewer *pro se* than counseled cases in my sample, and excluding them does not have important effects on my results.

There is one case-coverage difference between the FJC reports⁶⁹ and the present study. The FJC reports contain data from only twenty-three district courts, which according to the FJC, account for roughly half the civil cases filed in the U.S. district courts in 2009.⁷⁰ By comparison, the sample employed by the study presented herein draws from cases in a considerably larger number of districts: seventy-five of the seventy-eight districts that had adopted the electronic case filing (ECF) system before October 1, 2005.⁷¹

69. See FJC FIRST STUDY, *supra* note 7; FJC SECOND STUDY, *supra* note 7.

70. FJC FIRST STUDY, *supra* note 7, at 5.

71. The Article considers only cases filed after their district courts had adopted the ECF system because docket report information tends to be much sparser among cases from district courts that had not yet adopted the ECF system. The date on which a district court adopted ECF was determined by visiting the *Individual Court Sites* webpage at the PACER website, see *Individual Court Sites*, PACER, <http://www.pacer.gov/psc0/cgi-bin/links.pl> (last visited Feb. 2, 2016), and clicking on the information icon (which looks like this: ) next to each district court's name. The resulting webpage lists the date on which the court began using the ECF system in the field "ECF Go Live Date." The courts that had not yet gone live as of October 1, 2005, were the District of Nevada, District of Montana, District of North Dakota, District of Hawaii, District of Alaska, Eastern District of Oklahoma, Western District of Texas, Southern District of California, District of Vermont, Southern District of Florida, District of New Mexico, District of Virgin Islands, Central District of California, Western District of Wisconsin, District of Northern Mariana Islands, and Western District of Tennessee. I use only the seventy-eight districts that had fully implemented the CM/ECF system in order to avoid any problems that might arise if the ECF system were adopted between my pre-*Twombly* and post-*Iqbal* periods; such problems could occur to the extent that coders are able to obtain case documents more for cases filed after adoption than before, in late-adopting districts.

For a study of summary judgment activity in fiscal year 2006 by Federal Judicial Center researchers that uses the same universe of courts, see Memorandum from Joe Cecil & George Cort, Fed. Judicial Ctr., to Michael Baylson, Judge, U.S. Dist. Court for the E. Dist. of Pa., Estimate of Summary Judgment Activity Fiscal Year 2006 (June 15, 2007), <https://bulk.resource.org/courts.gov/fjc/sujufy06.pdf>. I will not discuss the voluminous literature on summary judgment practice more broadly; for contributions concerning summary judgment practice, see, for example, Stephen B. Burbank, *Vanishing Trials and Summary Judgment in Federal Civil Cases: Drifting Toward Bethlehem or Gomorrah?*, 1 J. EMPIRICAL LEGAL STUD. 591 (2004); Joe S. Cecil et al., *A Quarter-Century of Summary Judgment Practice in Six Federal District Courts*, 4 J. EMPIRICAL LEGAL STUD. 861 (2007); Jonah B. Gelbach, *Rethinking Summary Judgment Empirics: The Life of the Parties*, 162 U. PA. L. REV. 1663 (2014); Theodore Eisenberg & Charlotte Lanvers, *Summary Judgment Rates over Time, Across Case Categories, and Across Districts: An Empirical Study of Three Large Federal Districts* (Cornell Law Sch. Research Paper No. 08-022, 2008), <http://ssrn.com/abstract=1138373>; Memorandum from Joe Cecil & George Cort, *supra*.

This Article's sample contains no cases from the district courts for the Northern District of Illinois, the Northern Mariana Islands, or the Southern District of West Virginia. It became evident after coding had been completed that docket entries involving summary judgment motions follow a different textual structure in these districts from the structure typically used elsewhere. Consequently, the text search used to find summary judgment motion entries did not detect cases in these districts.

Finally, the study had to determine how best to measure the plaintiff's win rate. Plaintiffs can state multiple claims in a lawsuit, and defendants can challenge none, all, or some subset of the claims in both Rule 12(b)(6) motions to dismiss and Rule 56 motions for summary judgment. Further, parties can move for,⁷² or be granted,⁷³ summary judgment as to only certain aspects of a claim. This discussion raises the question of how to measure the plaintiff's win rate in defense summary judgment motions. Should it be the fraction of cases in which the plaintiff defeats a defense summary judgment motion on all challenged claims? Or should it be the fraction of cases in which the plaintiff defeats a defense summary judgment motion on at least one challenged claim?⁷⁴

A plaintiff, the saying goes, is the master of her complaint, so the number of claims lodged in an action is subject to party choice. Along the same lines, defendants are masters of their summary judgment motions: they choose for themselves which claims to attack. Thus, there is no convincing reason to prefer using the plaintiff-wins-on-all-claims or the plaintiff-wins-on-any-claim measure of the plaintiff's win rate. Accordingly, the empirical work presented herein reports the change in the plaintiff's win rate for both the fraction of cases in which the plaintiff defeats a defense summary judgment motion on all challenged claims and the fraction of cases in which the plaintiff defeats a defense summary judgment motion on at least one challenged claim. This approach allows readers to choose which approach they favor, if either; in

72. FED. R. CIV. P. 56(a) ("A party may move for summary judgment, identifying each claim or defense—or *the part of each claim or defense*—on which summary judgment is sought." (emphasis added)).

73. *Id.* 56(g) ("If the court does not grant all the relief requested by the motion, it may enter an order stating *any material fact*—including an item of damages or other relief—that is not genuinely in dispute and treating the fact as established in the case." (emphasis added)).

74. A third alternative would be to use the fraction of all challenged claims in which the plaintiff defeats a defense summary judgment motion. Experience shows that it can sometimes be difficult to determine exactly how many claims are challenged via a defense summary judgment motion. An important reason for this difficulty is that defense summary judgment motions sometimes state that they challenge "all claims," or some other phrasing with similar meaning, when some claims have already been withdrawn by the plaintiff or otherwise dismissed. To determine exactly how many claims are challenged for each case thus could require reading multiple case documents (e.g., complaints). There was not enough time to ask coders to engage in such extraordinarily detailed work. Also, the judicial orders or opinions resolving motions for summary judgment could not always be retrieved, but the disposition of the summary judgment motion in question could sometimes still be coded because it was docketed. In sum, a percentage-of-all-claims-challenged measure might well be beset by unavoidable measurement error. Moreover, sometimes particular issues, rather than claims as such, are challenged, and it is not always clear how to map general fact issues into claims. For these reasons, I restrict my attention to the two measures discussed in the text—whether the plaintiff prevails at summary judgment as to either all claims or at least one claim that the defendant challenges.

practice, though, my substantive conclusions are the same using either approach.

III. Employment Discrimination Cases

This Part of the Article reports the study's empirical findings for employment discrimination cases. Part III.A discusses the details of the employment discrimination sample, while Part III.B discusses the empirical results for cases in this sample. These results indicate that regardless of whether a plaintiff's win is defined as winning with respect to all or simply at least one challenged claim or issue, and regardless of whether one adjusts for certain potential confounding variables, the plaintiff's win rate did not change in a statistically significant way following *Twombly*.

A. Basic Characteristics of the Employment Discrimination Sample

Coders processed a total of 2511 employment discrimination cases. Of these, 185 were dropped because they were from districts that had not adopted the ECF system as of October 1, 2005.⁷⁵ An additional 335 were dropped because at least one pro se plaintiff was involved in at least one motion for summary judgment.⁷⁶ A further 138 were dropped because they appeared to involve ADA-related claims. The ADA Amendments Act of 2008 expanded the set of people covered by the ADA, which introduces a risk of misattributing effects of this change in ADA law to changes in pleading standards. This study therefore dropped from its sample those cases with claims that could be identified as ADA related.⁷⁷ This left a sample of 1853 employment discrimination cases with summary judgment motions in which all plaintiffs had counsel and none of their claims appeared to be ADA related. In 32 of these, there was no defendant's summary judgment motion filed.⁷⁸ This left a total sample of 1821 non-ADA employment discrimination cases in which there were no pro se plaintiffs and at least one defense summary judgment motion.⁷⁹

The first column of Table 2 reports the number of such cases in the sample by year of filing (Panel A) and according to whether the filing year was in the pre-*Twombly* or post-*Iqbal* period (Panel B). All told, the sample contained 1189 employment discrimination cases coded in the pre-*Twombly* period that had defense summary judgment motions and no pro se plaintiffs; there were 632

75. See *supra* note 71.

76. See *supra* note 68.

77. For a discussion of the method for identifying cases with ADA-related claims, see note 62 above.

78. Summary judgment motions filed by defendants in their capacity as cross-claimants or counterclaimants were not treated as defense summary judgment motions.

79. Only 74 of these 1821 cases also had a plaintiff's summary judgment motion filed.

such cases in the post-*Iqbal* period. The second column of Table 2 reports the number of these cases that had at least one summary judgment motion adjudicated within 731 days of case filing. It is not an accident that there are a disproportionate number of cases included in the 2005 filing period. Law school student coders were ready to work before all years of data were loaded into the database. Thus, while database code necessary to load the other years' cases was completed, the students worked on coding cases that were filed in 2005. To avoid any unrepresentativeness in the results, this study uses weights that adjust for the difference in sample sizes. These weights appear in the final column of Table 2; each case from 2005 effectively gets just under twenty-seven percent as much weight as cases filed in other years.⁸⁰

Within each year, just below sixty percent of cases had motions that had been adjudicated before the 731-day cutoff. All told, the pre-*Twombly* period had 700 employment discrimination cases with no pro se plaintiff and an adjudicated defense summary judgment motion, while the post-*Iqbal* period had 368 such cases. These are the cases on which the study's primary analysis is performed below.

80. The weight for cases filed in 2005 was constructed so that it equals:

$$\frac{N_{2006}}{2 \times N_{2005}} = \frac{243}{2 \times 457} = 0.266$$

where N_{2006} and N_{2005} are the numbers of cases in the second column of Table 2 for 2006 and 2005, respectively; the resulting value ensures that cases filed in 2005 receive one-third of the weight accorded to cases in my pre-*Twombly* period, reflecting the relative number of days from 2005 in that period. Note that this is the opposite of the approach one would take in a stratified sampling design, wherein the weight a stratum receives is an increasing function of the fraction of sampling fractions that were used, since sample means within these strata have lower variance. In the usual case, one is trying to obtain lower-variance estimates in a sample for which the population mean is assumed not to vary across strata. Here, though, my concern is with the representativeness of the sample, and I am unwilling to assume that the 2005 and 2006 population means of outcome variables necessarily are the same.

Table 2
Number of Employment Discrimination Cases with Summary Judgment Motions Filed in Observed Period and with Adjudications Within 731 Days

	Defense Summary Judgment Motion		
	Filed in Observed Period	Adjudicated Within 731 Days	Sampling Weight
A. By Year Case Filed			
2005 (Oct. 1—Dec. 31)	769	457	0.266
2006 (Jan. 1—June 30)	420	243	1
2009 (Oct. 1—Dec. 31)	206	119	1
2010 (Jan. 1—June 30)	426	249	1
B. By Pre-<i>Twombly</i>/Post-<i>Iqbal</i>			
Total Pre- <i>Twombly</i>	1189	700	
Total Post- <i>Iqbal</i>	632	368	

B. Summary Judgment Adjudication Results for Employment Discrimination Cases

This Subpart reports the core results for employment discrimination cases through both an unadjusted and an adjusted estimate of the change in the plaintiff’s win rate following *Twombly* and *Iqbal*. The unadjusted estimates are based on the simple win rate for plaintiffs among observations in the analysis sample. The adjusted estimates are based on binary logit models.⁸¹ In these models, the outcome variable serves as a dummy variable equal either to one for a case wherein the plaintiff wins on a defense summary judgment motion (however a win is measured) or equal to zero if the plaintiff does not. The variable of primary interest is a dummy variable indicating whether the case was filed in the post-*Iqbal* period. The adjusted estimate of the change in the plaintiff’s win rate following *Twombly* and *Iqbal* is calculated using the following three steps. First, the binary logit model is estimated. Second, for each observation the estimated coefficients from that model are used to calculate the estimated change in the plaintiff’s win probability due to *Twombly* and *Iqbal*. Such an estimated change is commonly known as a “marginal effect” of a change in circumstances—here, the change in the pleading standard. Finally, these observation-specific estimated marginal effects are averaged over all cases in the relevant sample, yielding a single adjusted effect.⁸² Besides the *Twombly* and *Iqbal*

81. A binary logit model is a commonly used model for measuring the relationship between a binary outcome variable of interest—such as whether a motion was denied or not—and a set of predictor variables, such as the post-*Iqbal* dummy and the district court dummies.

82. Estimating a logit model yields estimated coefficients relating to each of the included predictor variables (the post-*Iqbal* dummy variable, plus the district court dummies). These coefficient estimates can then be used to estimate the probability that the

footnote continued on next page

dummy variable, the other predictor variables—whose inclusion in the model is the source of the adjustment—are the following:

- A dummy variable indicating whether the judge was appointed by a Republican President.⁸³
- A dummy variable indicating whether any defendant in the case was a business organization.⁸⁴
- The employment-to-population ratio for the year when the case was filed, in the state where the judicial district is located.
- The employment-to-population ratio for each of the two years preceding the year in which the case was filed, in the state where the judicial district is located.
- The population for the year when the case was filed, for the state where the judicial district is located.

The appointing-President variable is meant to capture the effects of any ideological variation in judging, to the extent that such a variable does so.⁸⁵ The business organization variable is included to capture any effects of differences in party type. The employment-to-population variables are designed to control for business cycle variation that could have important effects on litigation behavior independent of *Twombly* and *Iqbal*.⁸⁶ Finally, the population variable is included to control for the possibility that cases in bigger states might be systematically different from those in smaller districts.

1. The percentage of cases in which plaintiffs win as to all challenged claims or issues

This Subpart reports results for the percentage of cases in which plaintiffs win as to all challenged claims or issues. As the discussion to come will show, the results indicate that this measure of the plaintiff's win rate was lower

plaintiff in each case would win against a defense summary judgment motion in each of the pre-*Twombly* and post-*Iqbal* periods. The difference in these estimated probabilities for a given case is the estimated value of the marginal effect of switching pleading standards. The average marginal effect is then the average of this estimate over all cases included in the analysis. That average value is what appears in the tables below.

83. My data on the judge assigned to the case are current as of the time that Thomson Reuters pulled the cases' docket information from PACER. Since cases might have been reassigned from one judge to another between the time that summary judgment motions were adjudicated and the time when the docket information was retrieved, it is possible that the judge data are not entirely accurate.

84. I coded cases as having a defendant business organization if a defendant party's name ended with the string "Inc" or "Corp" or included any of the following strings: "Inc.," "Incorporated," "Corp.," "L.L.P.," "LLP," "Limited Liability Partnership," "L.L.C.," "LLC," or "Limited Liability Company."

85. See Gelbach, *supra* note 71, at 1673-75, for more discussion of why judicial ideology need not lead to predictable differences across cases in litigation behavior or outcomes.

86. For more on this issue, see Part V.B.1 below.

following *Twiqbal*. However, this drop in the plaintiff’s win rate is estimated imprecisely, so that the results are not statistically significantly different from a zero effect.

The first column of Table 3 reports percentages of employment discrimination cases, among those with summary judgment motions adjudicated within 731 days of case filing, in which the plaintiff won on all challenged claims. For the pre-*Twombly* sample, 18.3% of motions were denied as to all claims raised in the motions. For the post-*Iqbal* period, this figure was slightly lower, at 17.1%. Table 3’s second column reports estimated standard errors for these percentages, which are 1.8% and 2.0% respectively.

The first column of Table 3’s third row reports information for the difference in the all-claims-denied percentage. This difference is -1.2%, nominally suggesting that case quality fell. However, the second column shows that the difference in win rates has an estimated standard error of 2.7%—great enough so that the difference is far from being statistically significantly different from zero.⁸⁷ Another way to put this is via a 90% confidence interval for the change in the plaintiff’s win rate against all challenged claims. This confidence interval, reported in the bottom row of Table 3, includes all values between a drop of 5.5% and an increase of 3.2%, which makes it clear that the results are not precise enough to allow us to conclude that *Twiqbal* caused either a drop or an increase in this measure of the plaintiff’s win rate.

Table 3
Percentage of Cases in Which Plaintiffs Won on
All Challenged Claims in Employment Discrimination Cases
(among those with motions resolved in 731 or fewer days)

	Unadjusted		Adjusted*	
	Percentage	Estimated Standard Error	Percentage	Estimated Standard Error
Pre- <i>Twombly</i>	18.3	1.8		
Post- <i>Iqbal</i>	17.1	2.0		
Difference	-1.2	2.7	-4.4	6.1
90% CI	[-5.5, 3.2]		[-14.4, 5.7]	

* A number of cases were dropped by the logit estimate’s statistical routine because of either missing data on the party of the President who appointed the judge or because

87. The *p*-value based on a standard *t*-test is 0.659, whereas conventional levels of significance in the social sciences would require a value less than 0.10. Note that the reported *p*-value is for a test of the null hypothesis of zero difference in the plaintiff’s win rate, against the two-sided alternative of a change not equal to zero. The two-sided alternative is appropriate here since the supporters’ view predicts a positive difference, while the aggressive critics’ view predicts a negative difference.

these district courts have no variation in the plaintiff's win rate, preventing the inclusion of these observations in logit estimation. The adjusted estimates in this table are thus based on 968 observations, by comparison to 1068 observations in the unadjusted estimation. Unadjusted estimates calculated using only the 968 observations included in the logit estimation yielded a difference in the plaintiff's win rate of -0.8%, with an estimated standard error of 2.8%.

The third and fourth columns show that adjusting for party of the appointing President, business status of defendants, state population, and trends in the state employment-to-population ratio leads to a substantially larger estimate of the change in the plaintiff's win rate—a drop of 4.4%. However, the adjusted difference is still quite far from being statistically significant: its estimated standard error is 6.1%, yielding a *p*-value well above conventional significance levels.⁸⁸ The greater estimated standard error leads to an even wider 90% confidence interval, which includes all values between a drop of 14.4% and an increase of 5.7%. Thus the adjusted results also are not precise enough to allow us to conclude with typical levels of confidence that *Twiqbal* caused either a drop or an increase in this measure of the plaintiff's win rate.

2. The percentage of cases in which plaintiffs win on at least one challenged claim

This Subpart reports results for the percentage of cases in which plaintiffs win as to at least one challenged claim or issue. As the discussion to come indicates, the results indicate that this measure of the plaintiff's win rate was a bit higher following *Twiqbal*. However, this increase in the plaintiff's win rate is estimated imprecisely, so that the results are not statistically significantly different from a zero effect. The first column of Table 4 reports the percentages of claims on which plaintiffs won on at least one challenged claim in employment discrimination cases. For the pre-*Twombly* sample, plaintiffs won on at least one claim in 37.0% of cases. For the post-*Iqbal* period, this percentage was slightly greater, at 38.0%. The second column reports estimated standard errors for these percentages, which are 2.2% and 2.5% respectively. The unadjusted difference of 1.0% has an estimated standard error of 3.4% and is thus far from statistically significant. Table 4's third row also reports information for the adjusted difference in the plaintiff's win rate. The difference of 2.7% is likewise statistically insignificant given its estimated standard error of 7.8%. Once again, another way of framing the statistical insignificance of the estimates is that 90% confidence intervals are wide enough to include zero for both the unadjusted and adjusted results.

⁸⁸ The *p*-value, testing a null of zero effect against the two-sided alternative of nonzero effect, is 0.474.

Table 4
Percentage of Cases in which Plaintiffs Won on
at Least One Challenged Claim in Employment Discrimination Cases
(among those with motions resolved in 731 or fewer days)

	Unadjusted		Adjusted*	
	Percentage	Estimated Standard Error	Percentage	Estimated Standard Error
Pre- <i>Twombly</i>	37.0	2.2		
Post- <i>Iqbal</i>	38.0	2.5		
Difference	1.0	3.4	2.7	7.8
90% CI	[-4.5, 6.6]		[-10.2, 15.6]	

* A number of cases were dropped by the logit estimate's statistical routine because of either missing data on the party of the President who appointed the judge or because these district courts have no variation in the plaintiff's win rate, preventing the inclusion of these observations in logit estimation. The adjusted estimates in this table are thus based on 968 observations, by comparison to 1068 observations in the unadjusted estimation. Unadjusted estimates calculated using only the 968 observations included in the logit estimation yielded a difference in the plaintiff's win rate of 0.8%, with an estimated standard error of 3.5%.

3. Measuring the quality-filtering effect using changes in the number of employment discrimination cases facing summary judgment motions

As discussed in Parts I.C and I.D, the percentage change in the number of cases facing summary judgment motions plays an important role in understanding the quality-filtering effect of *Twiqbal*. To measure this change for employment discrimination cases filed in each judicial district in each study period, this study compiled the number of cases that appeared to have a summary judgment motion filed via the search defined above in Part II. This is an imperfect method because some of the summary judgment motions would have been filed by plaintiffs rather than defendants. However, as discussed above, the overwhelming majority of summary judgment motions filed in employment discrimination cases are filed by defendants, so this method is reasonable for employment discrimination cases.⁸⁹ Of the cases included in the

89. Recall that cases in the pre-*Twombly* period have been ongoing for four years longer than cases in the post-*Iqbal* period. While it is possible that the drop in the number of employment discrimination cases facing summary judgment motions is due partly to different case durations, that seems unlikely. If there were substantial censoring in the filing of summary judgment motions in the post-*Iqbal* period by comparison to the pre-*Twombly* period, I would expect the share of cases with motions adjudicated before the 731-day cutoff to be substantially greater in the post-*Iqbal* period than in the pre-*Twombly* period. But as I wrote above in Part III.A, within each year, just below sixty
footnote continued on next page

above analysis, docket sheets indicated that a summary judgment motion was filed in 2695 cases during the pre-*Twombly* study period. On the flip side, docket sheets indicated that 2218 cases filed in the post-*Iqbal* study period had such filings. Thus, the number of summary judgment motions filed in employment discrimination cases fell by 477 during the period. This is a drop of roughly 18%, using the pre-*Twombly* study period as the denominator ($477 \div 2695 = 0.177$).⁹⁰ If we adopt the assumption that there was no selection into summary judgment, then as discussed above in Part I.C, the quality-filtering effect equals the ratio of the change in the plaintiff's win rate to the drop of 0.18 in the share of cases with summary judgment motions.⁹¹ Dividing by 0.18 is the same as multiplying by 5.6.⁹² For concreteness, this study employed a multiplication factor of 5.

Finally, recall that under the assumption of no selection into summary judgment, the quality-filtering effect should be interpreted as the difference between the plaintiff's win rate among cases that would face summary judgment motions under both pleading standards and the plaintiff's win rate

percent of cases had motions that had been adjudicated before the 731-day cutoff. I conclude that it is reasonable to compare the number of cases with summary judgment motions filed in the pre-2012 period, even though this allows four more years of observation time for cases filed in the pre-*Twombly* period.

90. One might worry that the recession affected litigation behavior, especially in employment discrimination cases. See John J. Donohue III & Peter Siegelman, *Law and Macroeconomics: Employment Discrimination Litigation over the Business Cycle*, 66 S. CAL. L. REV. 709, 710 (1993). To account for this possibility, I first calculated the number of cases with a summary judgment motion filed in the judicial districts of each state of the union, for each year of the study. I then used ordinary least squares to estimate a model relating the natural logarithm of the district-level number of cases to a dummy variable for the post-*Iqbal* period and the natural logarithm of the state's population, as well as another model that also includes the log of the employment-to-population ratio in the year the case was filed, as well as two annual lags of this (logged) variable. The results of the first specification imply a value of the percentage drop in the number of cases with summary judgment motions filed of roughly 15%, which is highly statistically significantly different from zero. The second specification's results imply a larger percentage drop—roughly 30%. However, the coefficients on the employment-to-population variables are both individually and jointly insignificant in the second model, suggesting the economy is not significantly associated with changes in summary judgment incidence, and the coefficient on the post-*Iqbal* dummy variable is statistically insignificantly different from zero. Presumably the loss of precision in the post-*Iqbal* dummy's coefficient, coupled with an increase in this estimated coefficient's magnitude, is due to the addition of the employment-to-population variables, which vary greatly across the pre-*Twombly*/post-*Iqbal* period and thus chew up a substantial amount of the information in the post-*Iqbal* dummy. Thus the simpler approaches to estimating the percentage drop appear more credible than the second multivariate model. I therefore use the simple percentage drop of eighteen points, discussed in the main text above.

91. An 18% drop is a drop of 0.18.

92. The multiplicative inverse of 0.18 is $\frac{1}{0.18} = 5.6$.

among cases that are filtered out; when the quality-filtering effect is positive, *Twiqbal* operates to filter out lower-quality cases, and vice versa.

Table 5 reports estimated ranges for the quality-filtering effect that are consistent with various approaches to estimating the effect of *Twombly* and *Iqbal* on the plaintiff's win rate at summary judgment in employment discrimination cases. Table 5 reports ranges for both approaches to defining the plaintiff's win rate discussed in Table 3 (defining a plaintiff win as requiring the plaintiff to win on all challenged claims) and Table 4 (defining a plaintiff win as requiring the plaintiff to win only one challenged claim), in each case presenting ranges based both on unadjusted and adjusted-for-covariates estimation approaches. For each of the four approaches, the study presented herein reports both a lower bound and an upper bound on the quality-filtering effect. Each estimation method's lower bound equals the multiplication factor of five, discussed above, times the lower bound of the 90% confidence interval for each method. The upper bound is calculated the same way except using the upper bound of the 90% confidence interval.

Leaving the numerical particulars aside, the central message of Table 5 is that, even under the assumption of no selection into summary judgment, the data reported above are consistent with a wide range of quality-filtering effect values. The range in the first row and first two columns of the table were calculated using the approach in which a plaintiff must win on all challenged claims for the case to be considered high merit, with the unadjusted method of estimating the change in the plaintiff's win rate. This is the approach with the narrowest of the four bounding ranges. Following *Twiqbal*, plaintiffs won in 17.1% of summary judgment motions. If the lower bound impact calculation in the first row and column of Table 5—a *Twiqbal*-caused drop of 28% in the plaintiff's win rate at summary judgment—were correct, then plaintiffs would have won in roughly 45% of filtered-out cases.⁹³ Thus under the lower bound impact calculation, *Twiqbal*'s quality-filtering effect would have caused the filtering out of cases with a high-merit share that is 28% *higher* than the high-merit share among cases that would face summary judgment motions under both pleading standards.

93. Given a particular estimate of Q , the quality-filtering effect, we can back out the values of W_{both} and W_{fo} under the assumption of no selection into summary judgment. Under that assumption, $W_{both} = W_{post}$, so $W_{fo} = W_{post} - Q$. When a plaintiff must win on all challenged claims for the case to be considered high merit and we use the unadjusted method of estimating the change in the plaintiff's win rate, we obtain an estimate of $W_{post} = 17.1\%$. If $Q = -28$, then $W_{fo} = 17.1 - (-28) = 45.1$. This calculation implies that plaintiffs would have won at summary judgment on all claims in 45.1% of cases that were filtered out of summary judgment due to *Twiqbal*—by comparison to 17.1% in those cases whose summary judgment status is unaffected by *Twiqbal*.

Table 5
Range of Quality-Filtering Effect Estimates for Employment Discrimination Cases, Assuming No Selection into Summary Judgment*

Measure of Plaintiff's Win Rate	Impact of <i>Twiqbal</i> on Plaintiff's Win-Rate Estimation Method			
	Unadjusted		Adjusted	
	Lower Bound	Upper Bound	Lower Bound	Upper Bound
Plaintiff wins on all challenged claims	-28	16	-72	29
Plaintiff wins on at least one challenged claim	-23	33	-51	78

* Bounds calculated by multiplying 90% confidence interval bounds in Table 3 and Table 4 by a multiplication factor of 5.

Table 5 shows that the corresponding upper bound when a plaintiff must win on all challenged claims for the case to be considered high merit, again using the unadjusted estimation method, is a quality-filtering effect of 16%. If this figure is accurate, and if there is no selection into summary judgment, plaintiffs would have won at summary judgment on all claims in only 1% of the cases that were filtered out of summary judgment due to *Twiqbal*—by comparison to 17.1% in those cases whose summary judgment status is unaffected by *Twiqbal*.⁹⁴

Thus, even if we use the narrowest confidence interval in Table 5, we find that the data are consistent with a range of plaintiff's win rates of between 1% and 45% for the cases filtered out of summary judgment adjudication by *Twiqbal*. Given the observed plaintiff's win rate of 17.1% for cases not filtered out, this range is potentially consistent with any of the case-quality views held by *Twiqbal*'s supporters or its critics, whether aggressive or moderate. And that means that, even under the assumption that there is no selection into summary judgment, the data on employment discrimination cases simply cannot settle the dispute over *Twiqbal*'s quality-filtering effect: this empirical evidence is consistent with the position of partisans on all sides of the *Twiqbal* quality debate.

IV. Contract Cases

This Part of the Article reports the study's empirical findings for contract cases. Subpart A discusses the details of the contract cases sample, while Subpart B discusses the empirical results for cases in this sample. On the whole, these results are similar to those for employment discrimination cases in that they indicate the plaintiff's win rate did not change in a statistically significant

94. If $Q = 12.8$, then $W_{fo} = 17.1 - 12.8 = 4.3$.

way following *Twiqbal*. As discussed below in Part IV.B.2, there is some weak statistical evidence of a statistically significant effect: with a plaintiff's win defined to include any case in which the plaintiff wins at summary judgment as to at least one challenged claim, the adjusted estimates yield an estimated increase in the plaintiff's win rate following *Twiqbal* that is statistically significant at the 90% level. For contextual reasons discussed in Part IV.B.3, though, this finding does not much alter the basic conclusion that *Twiqbal* did not have a clear and discernible effect on the plaintiff's win rate at summary judgment in contract cases.

A. Basic Characteristics of the Contracts Sample

Coders processed a total of 2478 contract cases. Of these, 236 were dropped because they originated in districts that had not adopted the ECF system before October 1, 2005.⁹⁵ An additional 53 were dropped because at least one pro se plaintiff was involved in at least one motion for summary judgment,⁹⁶ and 8 more were dropped because coders indicated that at least one claim involved the ADA.⁹⁷ That leaves 2181 contract cases with summary judgment motions in which all plaintiffs had legal representation. In 750 of these, there was no defendant's summary judgment motion filed,⁹⁸ so there were 1431 contract cases in which there were no pro se plaintiffs involved in any motion for summary judgment and at least one defense summary judgment motion was filed.⁹⁹ Table 6 reports the breakdown of these 1431 cases across subtypes of contract suits: 48.2% (689 cases) involved insurance, 7.1% (103 cases) involved other enumerated types of cases, and the remaining 44.7% (639 cases) involved the "Other Contract" PACER category.

95. *See supra* note 71.

96. Concerning the rationale for excluding cases with pro se plaintiffs, see note 68 above.

97. *See supra* note 62.

98. As with employment discrimination cases, summary judgment motions filed by defendants in their capacity as either cross-claimants or counterclaimants were not treated as defense summary judgment motions.

99. Of the 2181 contract cases with summary judgment motions in which all plaintiffs were counseled, 85 had summary judgment motions filed only by a party classified as neither a plaintiff nor a defendant, and another 665 had summary judgment motions filed by plaintiffs but not defendants. Together, these two sets of cases make up the 750 contract cases with summary judgment motions in which all plaintiffs were counseled, but in which no defense summary judgment motion was filed. The share of all contract cases with summary judgment motions in which all plaintiffs were counseled and in which a plaintiff filed a summary judgment motion thus was 36% ($100\% \times \frac{750}{2181-85}$), which is much greater than the rate for filing by plaintiffs in employment discrimination cases. *See supra* note 79.

Table 6
Served Contract Cases, by Individual Nature of Suit Code

PACER Code	Nature of Suit	Number of Cases Served
110	Insurance	689
120	Marine	27
130	Miller Act	4
140	Negotiable Instrument	12
150	Recovery of Overpayment and Enforcement of Judgment	10
151	Medicare Act	15
152	Recovery of Defaulted Student Loans (Excluding Veterans)	1
160	Stockholders' Suits	5
190	Other Contract	639
195	Contract Product Liability	22
196	Franchise	7
	All Contract Cases	1431

The first column of Table 7 reports the distribution of contract cases with a defense summary judgment motion and no pro se plaintiff by year (Panel A) and by pre-*Twombly*/post-*Iqbal* period (Panel B).¹⁰⁰ There are somewhat more coded cases in the pre-*Twombly* period (758 cases) than in the post-*Iqbal* period (673 cases). This discrepancy arose because the coders worked on one year of cases at a time, and all coders had to stop working before they managed to code as many 2010 cases as 2006 cases.¹⁰¹

100. Unlike the employment discrimination sample, the contracts sample does not have a disproportionate number of cases included in the 2005 filing period because all relevant years of contract cases had been loaded into the database when the coders began coding these cases.

101. There were no qualitative changes in the results when the samples were weighted so that the 2005 and 2009, and 2006 and 2010, samples had the same effective number of cases.

Table 7
Number of Contract Cases with Summary Judgment Motions
Observed Filed, and with Adjudications Within 731 Days

	Defense Summary Filed in Observed Period	Judgment Motion Adjudicated Within 731 Days
A. By year case filed		
2005 (Oct. 1—Dec. 31)	238	124
2006 (Jan. 1—Jun. 30)	520	276
2009 (Oct. 1—Dec. 31)	254	144
2010 (Jan. 1—June 30)	419	237
B. By pre-<i>Twombly</i>/post-<i>Iqbal</i>		
Total, pre- <i>Twombly</i>	758	400
Total, post- <i>Iqbal</i>	673	381

The second column of Table 7 reports the number of the contract cases with any defense summary judgment motion filed that had at least one such motion adjudicated within 731 days of case filing. The discrepancy in the number of cases across the pre-*Twombly*/post-*Iqbal* periods was largely eliminated because the share of cases that were adjudicated within 731 days was greater in the post-*Iqbal* period (57%) than in the pre-*Twombly* period (53%). Overall, the pre-*Twombly* period had 400 contract cases with an adjudicated defense summary judgment motion, while the post-*Iqbal* period had 381.

B. Summary Judgment Adjudication Results for Contract Cases

1. The percentage of cases in which plaintiffs win as to all aspects raised by defense summary judgment motions

Below, I will report results for the percentage of contract cases in which plaintiffs win as to all challenged claims or issues. As the discussion to come illustrates, the results indicate that this measure of the plaintiff's win rate was higher following *Twiqbal*. However, this increase in the plaintiff's win rate is estimated imprecisely, so that the results are not statistically significantly different from a zero effect.

The first column of Table 8 reports the percentages of contract cases in which plaintiffs won on all claims challenged via a defense summary judgment motion. For the pre-*Twombly* sample, plaintiffs won on all challenged claims in 36.3% of cases. For the post-*Iqbal* period, this figure was slightly greater, at 37.3%. The table's second column reports estimated standard errors for these percentages, which are 2.4% and 2.5% respectively.

Table 8
Percentage of Cases in Which Plaintiffs Win on *All Challenged Claims* in Contract Cases (among those with motions resolved in 731 or fewer days)

	Unadjusted		Adjusted*	
	Percentage	Estimated Standard Error	Percentage	Estimated Standard Error
Pre- <i>Twombly</i>	36.3	2.4		
Post- <i>Iqbal</i>	37.3	2.5		
Difference	1.0	3.5	6.2	7.0
90% CI	[-4.7, 6.7]		[-5.4, 17.7]	

* A number of cases were dropped by the logit estimate's statistical routine because of either missing data on the party of the President who appointed the judge or because these district courts have no variation in the plaintiff's win rate, preventing the inclusion of these observations in logit estimation. The adjusted estimates in this table are thus based on 707 observations, by comparison to 781 observations in the unadjusted estimation. Unadjusted estimates calculated using only the 746 observations included in the logit estimation yielded a difference in the plaintiff's win rate of 0.1%, with an estimated standard error of 3.6%.

Table 8's next row reports information for the difference in the percentage of cases in which plaintiffs won on all challenged claims. The unadjusted increase of 1.0% is in line with the supporters' view of *Twombly* and *Iqbal*, but this increase is very small. Moreover, its estimated standard error, 3.5%, is large enough that 90% confidence intervals include both positive and negative numbers. Once again, this difference is not statistically different from zero using either a two-sided¹⁰² or one-sided test of significance.¹⁰³ The adjusted difference, which is based on the same logit model described above in Part III, is much larger, at 6.2%. However, this estimate is quite imprecise, with an estimated standard error of 7.0%, so that it, too, is statistically insignificant.¹⁰⁴

2. The percentage of cases in which plaintiffs win on at least one claim challenged via defense summary judgment motion

The first column of Table 9 reports the percentages of contract cases in which the plaintiff won at summary judgment on at least one claim challenged

102. The *p*-value for a two-sided significance test is 0.768.

103. A one-sided alternative hypothesis is arguably appropriate for contract cases, since the aggressive critics' concern that only defendants will have the information necessary to plead seems considerably less likely to hold in contract actions. Since the *p*-value for a one-sided significance test is 0.384, far above conventional levels for a finding of statistical significance, the type of alternative adopted is practically irrelevant.

104. This conclusion holds regardless of whether one uses a two- or one-sided alternative hypothesis.

by a defense summary judgment motion. For the pre-*Twombly* sample, the plaintiff won on at least one challenged claim in 52.0% of cases. For the post-*Iqbal* period, this percentage was 56.4%. The second column reports estimated standard errors for these percentages, which are 2.5% in both periods. Table 9's third row reports information for the difference in the denial percentage.

The difference of 4.4% provides some support for the supporters' view of *Twombly* and *Iqbal*, since it suggests that average merit increased. However, the difference has an estimated standard error of 3.6%, indicating that this difference, like those above, is not statistically significantly different from zero.¹⁰⁵

Table 9
Percentage of Cases in Which Plaintiff Won on *at Least One Claim* in Contract Cases (among those with motions resolved in 731 or fewer days)

	Unadjusted		Adjusted*	
	Percentage	Estimated Standard Error	Percentage	Estimated Standard Error
Pre- <i>Twombly</i>	52.0	2.5		
Post- <i>Iqbal</i>	56.4	2.5		
Difference	4.4	3.6	13.6	7.0
90% CI	[-1.4, 10.3]		[2.1, 25.2]	

*A number of cases were dropped by the logit estimate's statistical routine because of either missing data on the party of the President who appointed the judge or because these district courts have no variation in the plaintiff's win rate, preventing the inclusion of these observations in logit estimation. The adjusted estimates in this table are thus based on 707 observations, in comparison to 781 observations in the unadjusted estimation. Unadjusted estimates calculated using only the 707 observations included in the logit estimation yielded a difference in the plaintiff's win rate of 3.5%, with an estimated standard error of 3.8%.

The final contracts results appear in the third and fourth columns of Table 9. The adjusted difference, in the third column, indicates that after adjustments are made for measures related to the judge, for the business organization status of defendants, and for state population and business cycle variables, *Twombly* and *Iqbal* are associated with a 13.6% increase in the plaintiff's win rate. This estimated difference in the plaintiff's win rate is statistically significantly different from zero, using either a two-sided (p -value = 0.052) or one-sided (p -value = 0.026) alternative hypothesis.¹⁰⁶

105. Even the one-sided p -value for this estimate is 0.108, above the cutoff level of 0.10 necessary for a finding of statistical significance.

106. See note 103 above for an explanation of why a one-sided test is arguably appropriate for contract cases.

Moreover, the point estimate is quite sizable: 13.6% amounts to nearly a 25% increase over the pre-*Twombly* plaintiff's win rate of 52.0% (see the first row of the first column in Table 9). Thus, using the conventional approach to significance testing, the results for contract cases do provide some evidence that the quality of cases adjudicated at summary judgment was greater following *Twiqbal* than before.

3. Challenges in measuring the quality-filtering effect using changes in the number of contract cases facing summary judgment motions

As I discussed above in Part I.C conceptually, and in Part III.B as applied to employment discrimination cases, the change in the number of cases facing a defendant's summary judgment motion plays an important role in understanding the quality-filtering effect of *Twiqbal*. Among contract cases with at least one summary judgment motion, there is no defendant-filed motion in more than a third.¹⁰⁷ Moreover, an initial review of the coded data suggests that this fraction might have been greater in the post-*Iqbal* period than in the pre-*Twombly* period. Thus, it seems considerably more problematic to use pre-*Twombly* and post-*Iqbal* counts of the number of cases with summary judgment motions filed, as this study did for employment discrimination cases in Part III.B, to measure the change in the number of cases with defendant-filed summary judgment motions.

For the sake of discussion, the analysis here ignores that concern and assumes that the number of cases with a defendant-filed summary judgment motion is proportional to the number of cases with any summary judgment motion. Under this assumption, the percentage drop in the number of cases with any summary judgment motion (regardless of which party filed it) will equal the percentage drop in the number of cases with a defendant-filed summary judgment motion (our drop of interest).

Using the search defined above in Part II to count the number of contract cases filed in the districts included in this study that appeared to have at least one summary judgment motion filed, there were 3127 such cases in the pre-*Twombly* period and 3005 in the post-*Iqbal* period. These figures imply a 3.9% drop, whose inverse—which is the multiplication factor discussed above in Part III.B.3—is roughly 25.¹⁰⁸ Thus under the assumption that there was no selection into summary judgment, the difference in the plaintiff's win rate among filtered-out cases and those that would face defense summary judgment motions under both pleading standards would equal 25 times whatever “difference” measure we choose from Table 8 or Table 9.

107. See *supra* note 99. By comparison, there are precious few employment discrimination cases in which only plaintiffs file summary judgment motions.

108. That is, $\left| \frac{3005-3127}{3127} \right| = 0.039$, and $\frac{1}{0.039} = 25.6$.

Consider the first column of Table 8,¹⁰⁹ whose estimated difference of 1% has an associated 90% confidence interval that includes all difference values between a drop of 4.7% and an increase of 6.7%. Multiplying each of these figures by 25 would suggest a confidence interval that includes all values between a drop of more than 100 points and an increase of more than 100 points.¹¹⁰ But the maximum logically possible difference in the plaintiff's win rate between any two sets of cases is 100 points in either direction. This means that even when we impose the assumption of no selection into summary judgment, the results involved in the first column of Table 8 are entirely uninformative about *Twiqbal's* quality-filtering effect for contract cases.

The same argument applies to the second column of Table 8 as well, and a variation on it applies to the first column of Table 9. Thus, if we were to adopt any of these measures of the observed change in the plaintiff's win rate against defense summary judgment motions, then even under the assumption of no selection into summary judgment we should conclude that the data for contract cases are consistent with such a wide range of results that they are uninformative about *Twiqbal's* quality-filtering effects.

Now consider the remaining results in the second column of Table 9. It is impossible for *Twiqbal* to have increased the plaintiff's win rate against defense summary judgment motions by the reported 13.6% if both (i) there was no selection into summary judgment and (ii) only 3.9% of cases were filtered out of facing defense summary judgment motions: 13.6 points times the multiplication factor of 25 exceeds 100 points, which is the maximum logically possible quality-filtering effect. In fact, it can be shown that a filtered-out share of 3.9% would be consistent with the absence of selection into summary judgment only if the actual effect of *Twiqbal* on the plaintiff's win rate against defense summary judgment motions is roughly 2%—the very bottom of the 90% confidence interval reported in the second column of Table 9.¹¹¹

109. In this Table, a plaintiff is considered to have won against a defense summary judgment motion only if she wins on all challenged claims; the first column uses the unadjusted approach to measuring the change in the plaintiff's win rate resulting from *Twiqbal*.

110. That is, 25×4.7 is more than 100, as is 25×6.7 .

111. Under the assumption of no selection into summary judgment, the plaintiff's win rate against defense summary judgment motions before *Twombly* can be written as a weighted average of the win rate among filtered out cases and the win rate among cases that would face summary judgment motions under both pleading standards. With 3.9% of cases having been filtered out by *Twiqbal*, as discussed above, we have: $W_{pre-T} = 0.039 \times W_{fo} + 0.961 \times W_{both}$. In the most extreme situation, plaintiffs would win in none of the filtered out cases, so that $W_{fo}^{min} = 0$. That would imply $W_{both}^{max} = W_{pre-T} \div 0.961$.

footnote continued on next page

Thus, unless the adjusted estimates in the second column of Table 9 are greatly overstated, there must have been some selection into summary judgment. If that is so, then the quality-filtering effect cannot be isolated using the data discussed here.¹¹² As a result, even if there really was a quality improvement, it likely is due at least in part, and potentially in total, to the replacement of relatively low-quality filtered-out cases by cases that have selected into summary judgment due to *Twiqbal*. It is difficult to assess the welfare implications of such a result. First, while the filtered-out cases would be lower quality than those selected into summary judgment, there is no way to know whether filtered-out cases were lower quality than those that face summary judgment motions under both pleading standards. Second, the result would imply that a relatively substantial number of high-quality cases that would settle pre-discovery but for *Twiqbal* instead are going through discovery, and then summary judgment adjudication. At the very least, such a result seems inconsistent with *Twiqbal*'s meta-policy rationale related to reducing litigation costs.

In sum, the results for contract cases do not paint any clearer a picture than those for employment discrimination cases. On the one hand, there is some evidence that *Twiqbal* caused an increase in the quality of contract cases that

From the first row and column of Table 9, we know that plaintiffs win 52.0% of the time pre-*Twombly*, so $W_{pre-T} = 0.52$. Thus, W_{both} is no greater than $0.52 \div 0.961$, or 0.54. Thus the maximum quality-filtering effect is $W_{both}^{max} - W_{pre-T}^{min} = 0.54 - 0 = 0.54$. Dividing this by the multiplication factor of 25 yields a result of roughly 2% for the maximum possible change in the observed plaintiff's win rate that is consistent with both (i) no selection into summary judgment and (ii) a filtered out share of 3.9%. This same argument applies even if the filtered out share of cases is, say, 8%. In that case, the maximum possible change in the observed plaintiff's win rate that is consistent with no selection into summary judgment is below 5%—still far below the point estimate in the second column of Table 9.

Perhaps one should not make too much of the statistical significance of the result in the second column of Table 9 in the first place. There are four different models estimated for contract cases (two in each of Table 8 and Table 9). It is well known in statistics that when a researcher conducts what are known as multiple inferences—here, the construction of four confidence intervals—the standard confidence interval bounds are too narrow. The norm in much social science scholarship, and also in empirical legal studies, is simply to ignore this fact. I shall largely follow that norm here, as it is complicated to explain how to construct confidence intervals that properly account for multiple inferences. However, it takes little gumption to conjecture that the one 90% confidence interval in these tables that doesn't include zero—the one statistically significant result—would not likely be deemed statistically significant once multiple inferences are taken into account.

112. The example considered in note 58 above shows that when there is no change in the number of cases with summary judgment motions filed, the observed difference in the plaintiff's win rate equals $\lambda\{W_{si} - W_{fo}\}$. However, there is no way to separately identify the part of this difference that is due to λ , which measures the extent of selection into summary judgment, and the part that is due to $\{W_{si} - W_{fo}\}$, which is the difference in case quality between cases that select into and that are filtered out of summary judgment.

survive through to the summary judgment stage. On the other hand, this evidence is relatively weak in context.¹¹³ Further, it appears likely that there was selection into summary judgment in contract cases, which would defeat our ability to draw meaningful inferences limited to *Twiqbal's* quality-filtering effect.

V. Further Issues

This Part discusses some additional methodological issues. First, this Part discusses why it would be problematic to focus only on cases in which a Rule 12(b)(6) motion was filed, as one other recent paper does in its assessment of *Twiqbal's* performance as a filter for summary judgment. Second, this Part discusses a number of factors that might render the estimates presented above unreliable as measures of *Twiqbal's* overall causal effect on the change in the plaintiff's win rate against defense summary judgment motions: the recession; the Supreme Court's resolution of the *Ledbetter* case and its subsequent override by Congress; *Scott v. Harris*; amendments to Rule 56 that took effect in 2007, 2009, and 2010; changes in primary behavior; and the possibility of other confounding factors.

A. Why It Would Be Problematic to Focus Only on Cases in Which a Rule 12(b)(6) Motion Was Filed

As mentioned in the Introduction, there is one other paper that attempts to draw inferences concerning *Twiqbal's* effects on case quality. Alexander Reinert read and coded information related to the resolution of Rule 12(b)(6) motions in more than 4000 cases.¹¹⁴ Most of his paper is directed at comparing the rate of Rule 12(b)(6) dismissals before *Twombly* and after *Iqbal*.¹¹⁵ But Reinert also includes a section titled, "Plausibility Pleading as Filter?"¹¹⁶ There, he reports the "ultimate outcome" of cases after adjudication of an initial Rule 12(b)(6) motion.¹¹⁷ One of the categories of ultimate outcomes he considers is "Dismissed ([Summary Judgment])."¹¹⁸

113. See *supra* note 111.

114. Reinert, *supra* note 7, at 2121.

115. *Id.* For critical discussions of such an approach to measuring the impact of changes in pleading standards, see Gelbach, *Dark Arts*, *supra* note 10, at 246-48; and Gelbach, *Locking the Doors to Discovery?*, *supra* note 10, at 2311-16.

116. Reinert, *supra* note 6, at 2162 (italics omitted).

117. *Id.* at 2164 tbl.12 (bolding omitted) (capitalization altered). It is potentially meaningful to follow up even those cases in which the initial Rule 12(b)(6) motion is granted, because in some cases plaintiffs receive leave to amend their complaints even after a Rule 12(b)(6) dismissal.

118. *Id.*

Reinert's analysis in that section differs from the one presented here in three key ways. First, he does not account for the fact that cases in the pre-*Twombly* period are more likely to have had time to reach a definitive result than cases filed in the post-*Iqbal* period, simply due to the fact that the pre-*Twombly* cases were filed earlier.¹¹⁹ For this reason, his summary judgment dismissal comparisons might be skewed due to the longer follow-up period available for cases filed in his pre-*Twombly* period.¹²⁰ Second, Reinert does not report the share of cases that are challenged via summary judgment, which is enough to render his results and mine noncomparable.

Third, and possibly most importantly, Reinert's study design focuses only on cases in which a Rule 12(b)(6) motion is filed. It might seem that this approach usefully focuses attention where it is warranted. But that view is mistaken. Consider a defendant-selection case, which the defendant would answer if the *Conley* standard applied but would challenge with a Rule 12(b)(6) motion if instead the *Twiqbal* standard applied.¹²¹ If the study presented here considered only cases in which a Rule 12(b)(6) motion is filed and denied, then the post-*Twiqbal* sample would *include* those cases in which the defendant loses the Rule 12(b)(6) motion and then files for summary judgment. But the pre-*Twiqbal* sample would *exclude* such cases. Defendant selection cases thus would be represented in the post-*Twiqbal* sample, but not the pre-*Twiqbal* sample. If case quality is correlated in important ways with changes in parties' litigation choices, as it might be, then this sample-definition rule would induce noncomparability across the two periods.

There are other selection-related types of cases that also make Reinert's approach to comparing ultimate outcomes problematic. Here is another example: Pre-*Twiqbal*, the parties agree that the defendant would be unlikely to win a Rule 12(b)(6) motion, and none would be filed. The parties disagree enough on the case's merits, however, such that rather than settle, the defendant would answer, allowing the case to go to discovery, and then move for summary judgment. Under the *Twiqbal* pleading standard, the parties in this same case would expect the defendant to file a Rule 12(b)(6) motion and to have a substantial probability of winning on it. Rather than spend the resources necessary to litigate this motion, the parties would settle even before the answer/Rule 12(b)(6) stage. Such a case would yield the opposite result to

119. As discussed above in note 67, my main focus is on only those summary judgment motions that are adjudicated within 731 days of case filing in order to avoid differential follow-up periods.

120. Indeed, the share of cases in Reinert's "Unresolved" category for cases filed in his post-*Iqbal* period dwarfs the corresponding share in his pre-*Twombly* period. Reinert, *supra* note 7, at 2164 tbl.12.

121. See Gelbach, *Dark Arts*, *supra* note 10, at 228; Gelbach, *Locking the Doors to Discovery?*, *supra* note 10, at 2307.

the defendant selection example discussed just before it: the second case would be represented in the *pre-Twiqbal* sample, but not the *post-Twiqbal* sample.

These examples show that it is problematic to measure *Twiqbal*'s case-quality effects using only cases in which Rule 12(b)(6) motions are filed. Such an approach unavoidably introduces the possibility of noncomparable samples in the *pre-Twiqbal* and *post-Twiqbal* periods.

B. Threats to the Validity of the Empirical Approach

1. The recession

Earlier empirical work has suggested that the characteristics of employment discrimination cases differ over the business cycle.¹²² It seems reasonable to think that contract disputes, or simply undisputed breaches, might also be more likely during economic downturns, with some parties being less able to perform than they had anticipated at the time they formed the contracts. As it happens, the worst recession in six decades gripped the United States in the period between the *pre-Twiqbal* and *post-Twiqbal* study periods. Thus, it is reasonable to worry that changes in the economy might have had important effects on the incidence of disputes, the filing of civil suits, or the nature of litigation.¹²³

It is for this reason that both the contemporaneous employment-to-population variable and one- and two-year lags were included in this study to account for the fact that lawsuits need not be filed immediately. It is certainly possible that this approach fails to capture important aspects of the business cycle, but the employment-to-population ratio seems as good a measure as any. Unlike the unemployment rate, the employment-to-population ratio does not move in the direction of "good economic news" when more workers become so discouraged that they drop out of the labor force altogether.¹²⁴ Further, by including both the contemporaneous employment-to-population ratio and two annual lags, this study is able to account for not just the current level, but also ongoing trends in labor market conditions in the state where a district court is located.

122. See Donohue & Siegelman, *supra* note 90, at 710.

123. See, e.g., David Freeman Engstrom, *The Twiqbal Puzzle and Empirical Study of Civil Procedure*, 65 STAN. L. REV. 1203, 1212 (2013) (suggesting a need to take the business cycle seriously in discussing a very early version of the results presented in the present study).

124. The unemployment rate is vulnerable to such changes because its denominator includes only those looking for work. Thus when workers become discouraged, all else equal, the unemployment rate rises. In contrast, the denominator of the employment-to-population ratio is unaffected by such changes as long as unemployed people do not move out of state.

2. *Ledbetter v. Goodyear Tire & Rubber Co.* and the Lilly Ledbetter Fair Pay Act

The functioning of the Title VII statute of limitations was in a state of some flux between the two study periods considered here. In 2007, the Supreme Court held that the Title VII statute of limitations does not reset following an initial act of employment discrimination.¹²⁵ Congress overrode this holding in 2009, retaining the six-month statute of limitations but allowing it to reset with each new paycheck embodying a discriminatory act.¹²⁶

It is possible that these changes caused differences in the number and composition of employment discrimination cases filed in the pre-*Twiqbal* and post-*Twiqbal* periods of study, but I am unaware of other empirical *Twiqbal* studies that have considered this possibility. Of course, the general failure of the literature to consider this issue might simply have been an oversight. Unfortunately, aside from controlling for business cycle fluctuations, there does not appear to be a good strategy for overcoming any bias potentially caused by this change. However, the circuit split that the Supreme Court addressed in *Ledbetter* appears to have involved only the Eleventh Circuit's earlier resolution of *Ledbetter* itself, on the one hand, and the D.C. Circuit and the Second Circuit, on the other.¹²⁷ Both the Second Circuit and the D.C. Circuit cases involved in this split were decided in 2005, so it is possible that there was relatively little on-the-ground awareness of this issue before *Ledbetter*. Moreover, the quick override of the Supreme Court's decision in *Ledbetter*—President Obama signed the legislation in a high-profile ceremony less than two weeks after taking office¹²⁸—meant that the law on this issue was nationally uniform well before my post-*Twiqbal* study period.

3. *Scott v. Harris*

In *Scott v. Harris*, the Supreme Court reversed the lower court's denial of summary judgment to a police officer who rammed his car into the subject of a high-speed car chase.¹²⁹ In holding that the officer's use of force was reasonable,

125. *Ledbetter v. Goodyear Tire & Rubber Co.*, 550 U.S. 618, 625 (2007).

126. Lilly Ledbetter Fair Pay Act of 2009, Pub. L. No. 111-2, § 3, 123 Stat. 5, 5-6 (codified as amended at 42 U.S.C. § 2000e-5(e)(3)(A) (2014)).

127. See *Ledbetter*, 550 U.S. at 623 (explaining the grant of certiorari as occurring “[i]n light of disagreement among the Courts of Appeals as to the proper application of the limitations period in Title VII disparate-treatment pay cases” and contrasting *Ledbetter v. Goodyear Tire & Rubber Co.*, 421 F.3d 1169 (11th Cir. 2005), with *Forsyth v. Federation Employment & Guidance Service*, 409 F.3d 565 (2d Cir. 2005), and *Shea v. Rice*, 409 F.3d 448 (D.C. Cir. 2005), for this proposition).

128. See Sheryl Gay Stolberg, *Obama Signs Equal-Pay Legislation*, N.Y. TIMES (Jan. 29, 2009), <http://nyti.ms/1Jjh7jy>.

129. 550 U.S. 372, 376 (2007).

the Court relied on a videotape of the car chase.¹³⁰ The Court held that the videotape so thoroughly discredited the plaintiff's account of events that only an unreasonable jury would find for him.¹³¹ Even though courts handling summary judgment motions usually construe all facts in the light most favorable to the nonmovant, the Court held that when confronting the particular facts of *Scott v. Harris*, courts "should not adopt [the discredited] version of the facts for purposes of ruling on a motion for summary judgment."¹³²

Scott v. Harris was handed down three weeks before *Twombly*, so most of the summary judgment motions considered in my pre-*Twiqbal* period were filed before *Scott*. While pending motions could have been affected by *Scott*, it is clear that more cases in the post-*Twiqbal* period would be affected. Thus, if *Scott* created substantial change in district courts' summary judgment practice, it would also render the two periods noncomparable for purposes of summary judgment adjudication.

This seems unlikely, however, because the factual predicate for *Scott*'s applicability is unusual. Few cases can be expected to involve a documentary record that directly and, according to the Supreme Court, inarguably controverts testimonial evidence that would be the only evidence a nonmovant could provide.¹³³ Justice Scalia's opinion for the Court's 8-1 majority says as much, characterizing the videotape's role in *Scott* as "an added wrinkle."¹³⁴ In his concurrence, Justice Breyer refers to the Court's determination concerning the underlying Fourth Amendment issue at stake in the case as "highly fact-dependent."¹³⁵ While it is possible that *Scott* might signal the Supreme Court's approval of lower courts' use of added discretion at summary judgment, it seems at least as likely that the case's impact on the great mass of summary judgment motions is quite limited,¹³⁶ especially in the two substantive areas considered here.

130. *Id.* at 383-84.

131. *Id.* at 380.

132. *Id.*

133. *See id.* at 380-81 ("When opposing parties tell two different stories, one of which is blatantly contradicted by the record, so that no reasonable jury could believe it, a court should not adopt that version of the facts for purposes of ruling on a motion for summary judgment. . . . Respondent's version of events is so utterly discredited by the record that no reasonable jury could have believed him. The Court of Appeals . . . should have viewed the facts in the light depicted by the videotape.").

134. *Id.* at 378.

135. *Id.* at 387 (Breyer, J., concurring). It is worth noting that Justice Stevens issued a spirited dissent. *See id.* at 389-97 (Stevens, J., dissenting).

136. *See Tolan v. Cotton*, 134 S. Ct. 1861, 1863 (2014) (per curiam) (vacating the Fifth Circuit's affirmance of the grant of a defendant's summary judgment motion for having "failed to adhere to the axiom that in ruling on a motion for summary judgment, [t]he evidence of the nonmovant is to be believed, and all justifiable

footnote continued on next page

4. Amendments to Rule 56

Following 1986's famed trilogy of summary judgment cases,¹³⁷ Rule 56 was amended in 1987 and then left untouched by the Advisory Committee for twenty years. There were then three amendments in as many years, with changes taking effect on December 1 of 2007, 2009, and 2010. The 2007 amendment was part of an overall restyling project, which was intended to clarify the Rules without changing the actual meaning of any of the Rules.¹³⁸ The 2009 amendment "consolidated and substantially revised" the timing provisions for Rule 56, allowing parties to file summary judgment motions earlier than before.¹³⁹ The 2010 amendment revised various aspects of "the procedures for presenting and deciding summary-judgment motions . . . to make the procedures more consistent with those already used in many courts."¹⁴⁰ It also restored the word "shall," which had been replaced by "should" as part of the 2007 restyling project, and replaced "no genuine issue" with "no genuine dispute."¹⁴¹ However, the Advisory Committee's note for the 2010 amendment states that "[t]he standard for granting summary judgment remains unchanged" and that the "amendments will not affect continuing development of the decisional law construing and applying these phrases."¹⁴² By the Advisory Committee's own lights, then, nothing important about the incidence and ultimate results of summary judgment adjudication should have

inferences are to be drawn in his favor" (alteration in original) (quoting *Anderson v. Liberty Lobby, Inc.*, 477 U.S. 242, 255 (1986)).

137. *Celotex Corp. v. Catrett*, 477 U.S. 317 (1986); *Anderson*, 477 U.S. 242; *Matsushita Elec. Indus., Co. v. Zenith Radio Corp.*, 475 U.S. 574 (1986).

138. See, e.g., Edward H. Cooper, *Restyling the Civil Rules: Clarity Without Change*, 79 NOTRE DAME L. REV. 1761, 1761 (2004) (characterizing "the goal of the Style Project" as being "to translate present text into clear language that does not change the meaning").

139. FED. R. CIV. P. 56 advisory committee's note to 2009 amendment.

140. *Id.* 56 advisory committee's note to 2010 amendment.

141. *Id.* Before the 2010 Amendment to Rule 56, Rule 56(c) stated that "[t]he judgment sought *should* be rendered if . . . there is no genuine *issue* as to any material fact and that the movant is entitled to judgment as a matter of law." FED. R. CIV. P. 56(c) (West 2009) (emphasis added). After 2010, the relevant language, now found in Rule 56(a), is "[t]he court *shall* grant summary judgment if the movant shows that there is no genuine *dispute* as to any material fact and the movant is entitled to judgment as a matter of law." FED. R. CIV. P. 56(a) (emphasis added). According to the Advisory Committee's note to the 2010 amendment:

The word "shall" in Rule 56 acquired significance over many decades of use. Rule 56 was amended in 2007 to replace "shall" with "should" as part of the Style Project, acting under a convention that prohibited any use of "shall." Comments on proposals to amend Rule 56, as published in 2008, have shown that neither of the choices available under the Style Project conventions—"must" or "should"—is suitable in light of the case law on whether a district court has discretion to deny summary judgment when there appears to be no genuine dispute as to any material fact.

Id. 56 advisory committee's note to 2010 amendment.

142. FED. R. CIV. P. 56 advisory committee's note to 2010 amendment.

changed due to the amendments interposed between my pre-*Twiqbal* and post-*Twiqbal* periods.

5. Changes in primary behavior

Some fear that due to the catch-22 problem, *Twiqbal* will “create an undesirable safe harbor that effectively places some defendants beyond the reach of civil rights laws,”¹⁴³ with an increased amount of unlawful discrimination as one possible result. Alternatively, it is possible to construct examples in which employers are dissuaded from implementing employment policies that would be lawful, if adjudicated on the merits, because the employers fear plaintiffs will file suit alleging unlawful disparate impact. If employers expect the shift to *Twiqbal* to make it easier to get such lawsuits dismissed earlier and at lower cost, then employers might institute the policies in the first place.¹⁴⁴ These are examples of *Twiqbal*-induced changes in primary behavior that could lead to changes in the number and quality of lawsuits filed. Such changes could render the estimation approach used here invalid. Thus, this Article’s approach relies on the assumption that any such primary behavior changes must have been small enough so as not to have importantly changed the distribution of cases over the relevant period studied here.

6. Other factors that might lead to invalid estimates

The recession, the *Ledbetter* case and its override, *Scott v. Harris*, and amendments to Rule 56 are all identifiable reasons why one might worry about comparing plaintiffs’ win rates in cases filed in the pre-*Twiqbal* and post-*Twiqbal* periods. Even if one accepts my arguments for concluding that the present study is not importantly affected by these issues, it is possible that some other, heretofore unaddressed factors unrelated to changes in the pleading standard might have affected plaintiffs’ win rates in these two study periods. Such concerns are an unavoidable part of attempting to draw causal inferences from real-world data when the variables of interest are not under researchers’ control.¹⁴⁵

143. CIVIN & ADEGBILE, *supra* note 28, at 2.

144. This example is drawn from Jonah B. Gelbach, Selection in Motion: A Formal Model of Rule 12(b)(6) and the *Twombly-Iqbal* Shift in Pleading Policy 43 (Aug. 29, 2012) (unpublished manuscript) (on file with author).

145. One consequence is that there has been a movement among legal scholars in favor of “randomizing law.” See Michael Abramowicz et al., *Randomizing Law*, 159 U. PA. L. REV. 929, 933 (2011) (arguing that where it is feasible to do so, legal rules should be randomized in order to facilitate measurement of policies’ effects). For a article focusing on the patent law context, see Lisa Larrimore Ouellette, *Patent Experimentalism*, 101 VA. L. REV. 65 (2015). As Abramowicz and his coauthors acknowledge, though, randomization is not always feasible, and it does have drawbacks. Abramowicz et al., *supra*, at 948-74. Consequently, recommendations for randomizing wherever possible are actually controversial among empirically oriented

footnote continued on next page

One approach that is often advocated, and nearly as often used, is to draw an analogy to medical trials with the units of empirical interest conceptualized as an experimental “treatment group.” Causal inferences are then based on finding a comparison group that is supposed to be entirely similar to the treatment group aside from exposure to some treatment of interest.¹⁴⁶ In the present situation, the “treatment” is the post-*Twigbal* pleading standard. So the treatment group is the set of all cases exposed to this treatment—including both my employment discrimination and contract cases.

Since *Twombly* and *Iqbal* apply transsubstantively due to the transsubstantivity of the Federal Rules of Civil Procedure,¹⁴⁷ precious few comparison groups are conceivable. One example involves securities litigation. The Private Securities Litigation Reform Act of 1995 (PSLRA)¹⁴⁸ prescribes a pleading standard that applies only to certain securities cases, requiring that a complaint’s allegations raise *more* than a plausible inference as to relevant elements.¹⁴⁹ Since *Twombly* and *Iqbal* purport to interpret only Rule 8 of the Federal Rules of Civil Procedure, they should not have affected the pleading standard that applies to the relevant cases covered by the PSLRA. Thus, one possibility would be to use these securities cases as a comparison group.

economists, with some highly influential scholars emphasizing the problems with an exclusive focus on randomization as the basis for credible empirical work. *See, e.g.,* Angus Deaton, *Instruments, Randomization, and Learning About Development*, 48 J. ECON. LITERATURE 424, 426 (2010); James J. Heckman & Jeffrey A. Smith, *Assessing the Case for Social Experiments*, 9 J. ECON. PERSP. 85, 86 (1995). For more on issues related to randomization in the empirical law and economics context in particular, see Jonah B. Gelbach & Jonathan Klick, *Empirical Law and Economics*, in OXFORD HANDBOOK OF LAW & ECONOMICS (Francesco Parisi ed., forthcoming), <http://ssrn.com/abstract=2507324>.

146. Researchers often refer to the comparison group as a “control group.” This practice is unfortunate, because the whole problem in a nonexperimental study is that assignment to the treatment is *not* controlled by the researcher. Thus it is an abuse of analogy to use the term “control group.”
147. *See* *Ashcroft v. Iqbal*, 556 U.S. 662, 684 (2009) (“Our decision in *Twombly* expounded the pleading standard for ‘all civil actions,’ . . . [including] antitrust and discrimination suits alike.” (quoting FED. R. CIV. P. 1)).
148. Pub. L. No. 104-67, 109 Stat. 737 (1995) (codified as amended in scattered sections of 15 U.S.C.).
149. *See* 15 U.S.C. § 78u-4(b)(1) (2014) (requiring the plaintiff to plead with particularity in the complaint when alleging that the defendant untruthfully stated a material fact or omitted a material fact necessary to make a statement not misleading); *id.* § 78u-4(b)(2)(A) (stating that when recovery of money damages requires proof of the defendant’s state of mind, “the complaint shall . . . state with particularity facts giving rise to a strong inference that the defendant acted with the required state of mind”); *see also* *Tellabs, Inc. v. Makor Issues & Rights, Ltd.*, 551 U.S. 308, 314 (2007) (interpreting a “strong inference” in § 78u-4(b)(2) to mean that “an inference of scienter must be more than merely plausible or reasonable—it must be cogent and at least as compelling as any opposing inference of nonfraudulent intent”).

However, this approach seems unlikely to generate fruitful results. There are surely many, many differences between the archetypal employment discrimination (or contract) case, on the one hand, and even the typical securities fraud case, on the other. Many employment discrimination cases are relatively simple affairs, involving a single plaintiff, and such cases need not involve much procedural wrangling other than the adjudication of straightforward Rule 12(b)(6) and Rule 56 motions. By comparison, securities fraud litigation typically involves an effort to certify a class under Rule 23, and many of these cases become very complicated and drawn out. Thus, the effects of changes in procedural law might be quite difficult to identify. In sum, there is little reason to believe that securities fraud cases constitute a comparison group that is entirely similar to the treatment group aside from exposure to some treatment of interest, as required to draw any meaningful inferences from such an approach.

Conclusion

This Article investigates the effects of *Twombly* and *Iqbal* on the quality of cases that face summary judgment motions filed by defendants. The methodological discussion shows that, under the assumption that the *Twiqbal* pleading standard does not cause any selection into summary judgment, the views of both *Twiqbal* supporters and critics line up cleanly with the direction of change in the plaintiff's win rate against defense summary judgment motions. The analysis also reveals how, again under the assumption of no selection into summary judgment, one can use observable data to compare the quality of cases that are filtered out of summary judgment as a result of *Twiqbal* with the quality of cases that would face summary judgment motions under both pleading standards. The results suggest that even with the relatively strong assumption related to selection into summary judgment, and even with usable data on nearly 2000 cases, it is not possible to clearly determine the quality-filtering effects of *Twiqbal*. The data are consistent with the positions of all three views described above—supporters, aggressive critics, and moderate critics. This finding suggests the disappointing, but nevertheless real, prospect that it might not be possible to settle the controversy over *Twiqbal*'s quality-filtering effects using empirical evidence.