Faculty Scholarship at Penn Carey Law

2015

Can We Learn Anything About Pleading Changes from Existing Data?

Jonah B. Gelbach
University of Pennsylvania Carey Law School

CAN WE LEARN ANYTHING ABOUT PLEADING CHANGES FROM EXISTING
DATA?

Jonah B. Gelbach[*]
(202) 427-6093
jgelbach@law.upenn.edu

This version: July 30, 2015

Abstract
        In light of the gateway role that the pleading standard can play in
our civil litigation system, measuring the empirical effects of pleading
policy changes embodied in the Supreme Court's controversial *Twombly*
and *Iqbal* cases is important. In my earlier paper, *Locking the Doors to
Discovery*, I argued that in doing so, special care is required in formulating
the object of empirical study. Taking party behavior seriously, as *Locking
the Doors* does, leads to empirical results suggesting that *Twombly* and
Iqbal have had substantial effects among cases that face Rule 12(b)(6)
motions post-*Iqbal*. This paper responds to potentially important critiques
of my empirical implementation made by the FJC's Joe Cecil and Professor
David Engstrom. An additional contribution of the present paper is to
elucidate some important challenges for empirical work in civil procedure.
First, researchers should carefully consider which covariates belong in
statistical models, while also taking care in assessing the empirical
importance of controlling for covariates. Second, data collection protocols
should be designed with behavioral assumptions in mind. But third,
researchers should not let the perfect be the enemy of the good: even data
protocols that are less than perfectly designed may be broadly useful.

There has been much disagreement concerning the effects of the Supreme Court's decisions in *Bell Atl. Corp. v. Twombly*[1] and *Ashcroft v. Iqbal*.[2] In *Twombly*, the Court retired *Conley v. Gibson*'s "no set of facts" standard for adjudicating a Rule 12(b)(6) motion to dismiss for failure to state a claim, at least as to parallel conduct antitrust actions.[3] *Iqbal* then extended *Twombly's* plausibility pleading standard trans-substantively, to "all civil actions." Many observers—academics, other professional researchers, and practitioners—have debated the empirical effects of *Twombly* and *Iqbal*. Most of the studies involved in this debate are based on assessment of how (some measure of the) Rule 12(b)(6) grant rate changed between the pre-*Twombly* and post-*Iqbal* time periods. Debates over these studies have been nearly as controverted as the *Twombly* and *Iqbal* cases themselves, with some arguing that the evidence shows *Twombly* and *Iqbal* have substantially changed Rule 12(b)(6) adjudication practice and others arguing the opposite.[4]

In my own paper on the topic, *Locking the Doors to Discovery* (hereinafter "*Locking the Doors*"),[5] I argued that perceived changes in the pleading standard can be expected to cause parties to change their behavior—whether plaintiffs file suit, and whether defendants challenge filed actions with Rule 12(b)(6) motions, and whether parties to a dispute are able to settle. If party behavior is endogenous, then pre/post comparisons in grant rates might involve an apples-to-oranges problem when these comparisons are used as a way to measure the *ceteris paribus* changes in the probability that a judge would grant a Rule 12(b)(6) motion in a given set of cases—which I term "judicial behavior effects".

Indeed, evidence in a study released by the Federal Judicial Center ("FJC") indicates that while Rule 12(b)(6) motions aren't filed all that often, in those district courts the FJC studied, the share of filed actions in which Rule 12(b)(6) motions were filed rose by more than 50%.[6] Such a change is consistent with what I referred to as defendant selection effects: other things equal, rational defendants should be more willing to bear the costs of litigating a Rule 12(b)(6) motion if they expect to be more likely to win on that motion. I considered two other types of selection, as well. Plaintiff selection effects occur in those disputes such that plaintiffs would

---

[1] Bell Atl. Corp. v. Twombly, 550 U.S. 544 (2007).

[2] Ashcroft v. Iqbal, 129 S. Ct. 1937 (2009).

[3] Conley v. Gibson, 355 U.S. 41, 45-46 (1957) ("In appraising the sufficiency of the complaint we follow, of course, the accepted rule that 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.").

[4] Citations for these propositions may be found in (Gelbach, 2012).

[5] (Gelbach, 2012).

[6] (Cecil, et al., March 2011) (hereinafter, "FJC initial report").

file suit under the *Conley* pleading standard, but not under the plausibility pleading standard. And settlement selection effects occur in those disputes that would be settled before the answer/Rule 12(b)(6) stage under one pleading standard but that would be litigated through that stage under the other pleading standard.

In *Locking the Doors*, I showed that party selection effects "unidentify" Rule 12(b)(6) motion grant rate comparisons when these are used. That is, I show that even the possibility that parties might change their behavior in response to changes in the pleading standard implies that grant rate comparisons do not tell us anything discernible about how *Twombly* and *Iqbal* have changed judicial behavior in any fixed set of cases.

Moreover, I showed in *Locking the Doors* that judicial behavior effects are too limited a measure of *Twombly* and *Iqbal*'s effects in any event. This is true because changes in party behavior that arise from perceived changes in the pleading standard can affect parties' welfare directly, even holding constant judicial behavior. To account for such effects, I constructed the category of "negatively affected" cases among those in which a Rule 12(b)(6) motion would be filed under the plausibility pleading standard. This category comprises the set of disputes in which, as a but-for result of *Twombly* and *Iqbal*, either (i) the plaintiff fails to get to discovery, or (ii) a dispute that would be settled before the Rule 12(b)(6) stage instead winds up with a Rule 12(b)(6) motion being both litigated and granted.

As I discussed in *Locking the Doors*, it is not possible to estimate this negatively affected share of cases without very strong assumptions— assumptions strong enough to pin down the frequency of various types of selection.[7] However, I also showed how one can use observable data to calculate a lower bound on this negatively affected share. As I shall discuss in Part I, doing so requires only data on the numbers (or frequency) of Rule 12(b)(6) motions filed and granted in the pre-*Twombly* and post-*Iqbal* periods—data that are available from the two FJC reports. The formula for my negatively affected share can be conveniently decomposed into the sum of two components. The first component is the change in the Rule 12(b)(6)

---

[7] The underlying problem is that observing the numbers of disputes exhibiting selection of various types—defendant selection, plaintiff selection, or settlement selection—would require observing how a dispute would be resolved under each of two mutually exclusive pleading standards. This impossibility is sometimes known as the fundamental problem of evaluation. This problem can be solved only via sufficiently powerful assumptions; for a discussion of the role of behavioral assumptions in empirical work in civil procedure, *see* (Gelbach, 2014) (hereinafter "*Dark Arts*"). While assumptions should be as weak as possible, in order to minimize the extent to which the assumptions lead one to incorrect results, still there may be no way to both (i) avoid making substantively restrictive assumptions and (ii) learn anything of policy relevance.

grant rate—i.e., the same measure that most other studies have used to measure the effects of *Twombly* and *Iqbal*. The second component measures how substantial the change in the number of filed Rule 12(b)(6) motions has been; the greater the relative post-*Iqbal* increase in the number of such motions filed, the greater will be this component.

The resulting lower bounds indicate that *Twombly* and *Iqbal* negatively affected a substantial share of cases in which Rule 12(b)(6) motions were filed post-*Iqbal*: plaintiffs were negatively affected in at least 18.1% of civil rights cases, 15.4% of employment discrimination cases, and 21.5% of contract, tort, and various other cases. Further, all three of these estimates are statistically significantly different from zero. These findings occur even though the grant rate itself does not change discernibly for either the employment discrimination or contract, tort, and "other" case categories. Importantly, for all three categories I studied, a substantial part of the estimated negatively affected share would be missed if one did not account for selection effects via the second component of my lower bound formula.

The approach I took in *Locking the Doors*, and by extension the results just discussed, have been criticized in two thoughtful articles. The first article, to which I shall refer as "*Twiqbal Puzzle*," was written by Professor David Engstrom;[8] the other, to which I shall refer as "*Waves*," was written by Joe Cecil of the FJC,[9] who was the lead author on both FJC reports. Between them, Cecil and Engstrom lodge three types of criticisms. Because I have elsewhere addressed the first critique, which concerns the appropriateness of the substantive behavioral framework (including my choice of unit of analysis), I shall not do so here.[10]

The purpose of the present paper is to respond in detail to the other two critiques. Filing this response is worthwhile partly because the question of whether *Twombly* and *Iqbal* had substantial effects is itself important. But it is also worthwhile for the broader reason that Cecil's and Engstrom's critiques raise issues with relevance to the design and implementation of future empirical work concerning civil litigation. The second critique concerns whether and how one should use multivariate models in an effort to control for non-*Twombly*/*Iqbal* changes that might have contributed to measured changes in Rule 12(b)(6) motion practice and adjudication. I respond to this critique in detail in Part III, *infra*. The third critique concerns whether it is appropriate to use the FJC data as I do—combining information from separate studies of the frequency with which Rule 12(b)(6) motions are filed, and of the frequency with which such motions are granted among cases in which they are filed. I respond to this critique in

---

[8] (Cecil, 2012) [hereinafter "*Waves*"].
[9] (Engstrom, 2013).
[10] Interested readers should see *Dark Arts*.

detail in Part IV, *infra*.

The overall takeaway point from my response to these critiques is straightforward. I do not believe that either critique detracts from the empirical validity or relevance of the estimates I reported in *Locking the Doors*. That is not to say that the issues Cecil and Engstrom raise couldn't be important in principle. But the arguments they make are less problematic for my approach than is at first apparent, and the relevant empirical evidence strongly suggests that the qualitative conclusions in *Locking the Doors* are appropriate and robust to Cecil and Engstrom's critiques.

I. A QUICK SUMMARY OF THE EMPIRICAL APPROACH IN *LOCKING THE DOORS*

Here I briefly summarize the empirical approach and results in *Locking the Doors*.[11]

I define a dispute as involving a negatively affected plaintiff if (i) pre-*Twombly*, the plaintiff would either reach discovery as to all claims or achieve some sort of settlement as to all claims, while (ii) post-*Iqbal*, under the plausibility pleading standard, the plaintiff would lose on at least one claim as a result of a Rule 12(b)(6) motion. In *Locking the Doors*, I focused on saying something about the share of cases that are negatively affected in this sense, among those cases that have a Rule 12(b)(6) motion filed in the post-*Iqbal* period studies.

Define the number of cases in which Rule 12(b)(6) motions are filed pre-*Twombly* as M_{pre} and the number of such motions granted post-*Iqbal* as M_{post}. Define the number of Rule 12(b)(6) motions granted pre-*Twombly* as G_{pre} and the number of such motions granted post-*Iqbal* as G_{post}. In *Locking the Doors*, I demonstrate that the negatively affected share in question is never less than the following simple ratio:

$$(1) \qquad\qquad LB_{NAS} \equiv \frac{G_{post} - G_{pre}}{M_{post}},$$

so that LB_{NAS} is a lower bound on the negatively affected share.

Here is an intuitive explanation for this claim. By definition of but-for causation, if a real or perceived change in the pleading standard but-for causes a dispute to have a Rule 12(b)(6) motion granted post-*Iqbal*, then that dispute (i) must have a Rule 12(b)(6) motion granted post-*Iqbal* and (ii) must not have a Rule 12(b)(6) motion granted pre-*Twombly*. Thus, the numerator of equation (1), which is the change in the number of cases with Rule 12(b)(6) grants, must include all such but-for caused cases. Note that

---

[11] Readers interested in more detail should see my 2012 paper *Locking the Doors* or my 2014 paper *Dark Arts*.

the numerator does *not* include any cases that would have Rule 12(b)(6) motions granted under both pleading standards, since those cases are by definition present in equal numbers in G_{pre} and G_{post}. Thus, if the only types of cases were those but-for caused by the plausibility pleading standard to have Rule 12(b)(6) motions granted under *Twombly/Iqbal*, and those entirely unaffected, then the numerator of equation (1) would identify the number of cases with plaintiffs negatively affected in the way I defined above.

However, there is one other set of negatively affected cases to consider. This third set consists of those cases that would have Rule 12(b)(6) motions granted under the plausibility pleading standard but not under the *Conley* standard; such cases could occur due to defendant selection or settlement selection.[12] The possible presence of such cases means that the numerator of (1) does not equal the number of negatively affected cases among those facing Rule 12(b)(6) motions post-*Iqbal*. Note, though, that there must be at least zero cases in our third category, i.e., there cannot be a negative number. Consequently, the number of negatively affected cases can never be less than the numerator of (1), which means this numerator is a lower bound on my number of negatively affected cases.[13]

As noted above, the data I used in *Locking the Doors* come from the two FJC reports. As I shall discuss in detail in Part IV, *infra*, the FJC data on motion adjudication and motion filing come from different data-collection exercises, which somewhat complicates the use of formula (1). For this reason, it is useful to observe that this formula can be rewritten as follows:

(2) $$LB_{NAS} \equiv (g_{post} - g_{pre}) + g_{pre}m,$$

$$\text{where } m \equiv \frac{M_{post} - M_{pre}}{M_{post}}.$$

Here, g_{post} is the share of post-*Iqbal* cases in which a Rule 12(b)(6)

---

[12] To illustrate how such cases could arise due to defendant selection, consider a dispute in which, under *Conley*, a plaintiff would file suit, and the defendant would answer; this suit would not be represented in G_{pre}. If the defendant would file a Rule 12(b)(6) motion under *Twombly/Iqbal*, then the case would be represented in G_{post}. In the settlement selection version of this story, the same events occur under *Twombly/Iqbal* as in our plaintiff selection story, but the case settles pre-Rule 12(b)(6) adjudication under *Conley*.

[13] *Locking the Doors* provided a considerably more detailed exposition of this argument, because doing so has certain other methodological advantages. Given the simplicity of the argument as I have provided it here, it does not require technical elaboration via probability theory. However, it is worth noting that there is an extensive econometric literature on the identification of treatment-effect bounds; for early salvos, *see* (Manski, 1989) and (Manski, 1990).

motion is granted, among those in which such a motion is filed; g_{pre} is the corresponding share of pre-*Twombly* cases.[14] Formula (2) shows that the lower bound can be expressed as the sum of two components. The first component is the change in the Rule 12(b)(6) grant rate, which is the measure on which other studies have primarily focused. The second component, which helps account for selection-related effects, is the product of the pre-*Twombly* Rule 12(b)(6) grant rate and *m*, which is a measure of the increase in Rule 12(b)(6) motion filings: *m* equals the share of post-*Iqbal* cases facing a Rule 12(b)(6) motion that is accounted for by increases in the number of Rule 12(b)(6) motions filed.

Table 1: The Change in the Percentage of Movants Prevailing and the Lower bound on the Negatively Affected Share

	Percentage of Movants Prevailing				Second component in formula (2) (product of first and fourth columns)	Lower bound (sum of third and fifth columns)
	2006[a]	2010[a]	Change	m^{b}		
Civil rights	60.3	68.1	7.8	0.17	10.3	18.1
Employment discrimination	60.9	61.1	0.2	0.25	15.2	15.4
Contract, tort and other	55.2	56.3	1.1	0.37	20.4	21.5

[a] *Source*: Table 4 of *Locking the Doors*, at 2331.
[b] *Source*: Table 5 of Locking the Doors, at 2333.

Table 1 reports data sufficient to calculate my lower bound for three case categories using the FJC reports' data for 2006 (the pre-*Twombly* period) and 2010 (the post-*Iqbal* period).[15] The table's first column shows that, across all categories, the percentage of movants prevailing in the pre-*Twombly* period varies between roughly 55% and roughly 61%. The other factor here is the ratio *m*, which measures the relative importance of increased Rule 12(b)(6) motion filing; as the figures reported in the fourth

---

[14] Thus, $g_{post}=G_{post}/M_{post}$, and $g_{pre}=G_{pre}/M_{pre}$.
[15] Note that I here exclude cases involving financial instruments or ADA discrimination claims; *see* discussion in (Gelbach, 2012).

column of the table show, this increase was substantial, especially for employment discrimination and contract, tort, and other cases.

The results in the final column of Table 1 indicate that plaintiffs must have been negatively affected in a substantial share of cases that faced Rule 12(b)(6) motions post-*Iqbal*: the lower bounds are 18.1% for civil rights cases, 15.4% for employment discrimination cases, and 21.5% for contract, tort and other cases. These estimates are statistically significantly different from zero.[16] In light of the sizable second-component estimates reported in the table's fifth column, the results also indicate that party selection effects played a substantial role in this negative-effects story.

Before I turn to the methodological issues that are at the heart of this paper, I briefly address a number of erroneous characterizations Cecil has lodged concerning my treatment of party selection effects in *Locking the Doors*. First, Cecil points out that between the filing study's pre-*Twombly* and post-*Iqbal* observation periods, there was only a small drop in the number of "total other" cases filed, alongside an increase in the number of filings of civil rights and employment discrimination cases. From this he concludes that in *Locking the Doors* I

> "find[] little or no plaintiff selection effect, which is quite a surprise since such a plaintiff selection effect is a fundamental component of [*Locking the Doors*'] model of pretrial litigation and the foundation of Priest/Klein model of litigation from which [the] model is derived."[17]

But my model is not derived from the Priest/Klein model.[18] And I do not actually "find," or even claim to find, anything as to the presence of the plaintiff selection effect. The interplay of all three types of party selection effects is such that, without making substantive assumptions about the distribution of various types of party beliefs, it is impossible to say whether the number of cases filed will go up, go down, or stay the same following a perceived change in the pleading standard.[19]

Elsewhere in this part of his critique, Cecil explains that the "evidence for [a small plaintiff selection] effect seems very tenuous," citing specific observed numbers of filings.[20] But among the central methodological problems raised in *Locking the Doors* is precisely the fact that gross effects

---

[16] For details related to statistical significance*, see* Appendix B to *Locking the Doors*, available at http://www.yalelawjournal.org/images/documents/gelbach_appendix_b.pdf.

[17] (Cecil, 2012, p. 44) (footnote omitted).

[18] *See Dark Arts.*

[19] *See, e.g., Locking the Doors,* at 2311.

[20] (Cecil, 2012, pp. 42-43) (n. 159).

cannot be isolated from existing data. And that is *precisely because* of the "ambiguous predictions,"[21] about empirically observable objects that arise once we take seriously the possibility that parties will respond to perceived changes in the litigation incentives.

Finally, Cecil makes the startling claim that in *Locking the Doors*, I "assume[] that the courts will respond to motions to dismiss for failure to state a claim in cases filed after *Iqbal* by granting such motions at approximately the same rate a[s] in the past."[22] But I do no such thing. I simply point out that when parties respond to perceived changes in the pleading standard, there is no way to determine how courts, in particular, respond. This is just another instance in which we cannot use data that reflect *net* changes to measure multiple *gross* effects—here, it is how courts respond, other things equal, whereas previously it was how plaintiffs respond, other things equal.

I turn now to some methodological concerns that the approach in *Locking the Doors* does need to confront.

II. CONFOUNDING FACTORS

One assumption necessary to justify the approach in *Locking the Doors* is that the composition of disputes that actually occur in the post-*Iqbal* period is not importantly different from the composition of disputes that would have occurred in this period if *Twombly* and *Iqbal* had never happened. Roughly speaking, this is equivalent to assuming that *Twombly* and *Iqbal* were the only causes of the differences in the numbers of cases and Rule 12(b)(6) motions filed, and in adjudication of those Rule 12(b)(6) motions that were filed.

If, for example, the Great Recession caused a big uptick in disputes related to alleged employment discrimination, then even in the absence of *Twombly* and *Iqbal* we would expect to observe more Rule 12(b)(6) motions filed and granted.[23] Moreover, as I also discuss in *Locking the Doors*, *Twombly* and *Iqbal* might embolden potential defendants (e.g., employers) to engage in more aggressive behavior (e.g., layoffs that have a disparate impact on minority or female workers).[24] If such confounding factors are present, my approach could misattribute at least some of the change in the number of post-*Iqbal* Rule 12(b)(6) motions granted or filed. Whether such a problem leads to upward or downward bias in my reported lower bounds would necessarily depend on how frequently Rule 12(b)(6)

---

[21] (Cecil, 2012, p. 44) (n. 163).
[22] (Cecil, 2012, p. 44).
[23] *See* Part VI.B.1 of (Gelbach, 2012, pp. 2336-37) for more on this point.
[24] *See id.*, at 2337.

motions are filed and granted in cases that appear in the data due to confounding factors.

The FJC reports' authors, Cecil on his own, and Engstrom all have suggested an additional source of problematic compositional change: that changes in the geographic distribution of cases have occurred, that these changes were caused by something other than *Twombly/Iqbal*, and that they are important in understanding observed changes in who wins at the Rule 12(b)(6) stage. Cecil argues that for this reason, the only way to reliably measure *Twombly* and *Iqbal*'s impact is to use estimates from multivariate models that include dummy variables indicating the judicial district in which each case was filed:

> As noted in both of our reports, the corrections for factors unrelated to *Twombly* and *Iqbal* often account for the statistically significant differences that appear in the simple comparison between the pre-*Twombly* and post-*Iqbal* periods. To ignore the findings of these multi[variate] models and rely on the raw frequencies confounds changes that can be attributed to *Twombly* and *Iqbal* with numerous other changes that are unrelated to the effect of those decisions on the substantive standards of pleadings and the extent to which cases may progress beyond the pleading stage to discovery.[25]

Cecil is not the only one to take this view. Engstrom writes that

> variables designed to control for variation in outcomes by judicial district and case type may … control for a … general concern about unobserved case heterogeneity—that is, the possibility that simple shifts across the pre- and post-*Twiqbal* periods in the distribution of case types, litigants, or judges are behind observed differences in outcomes.[26]

These are familiar and good arguments for using covariates. It goes without saying that when confounding factors exist and can be accounted for using measurable variables without introducing any further problems, covariates should be used. On the other hand, some covariates are inappropriate to include. For example, a general rule of applied statistics is that variables that are themselves partly determined by the outcome variable are inappropriate to use as independent variables.

Consider an apparently unrelated example from micro-econometrics—the problem of demand estimation. Suppose one wants to estimate how county-level demand for a good changes with the presence of county ordinances that regulate its sale. To do so, one might regress county-level

---

[25] (Cecil, 2012, p. 39) (footnote omitted).
[26] (Engstrom, 2013, pp. 1217-18).

observed quantity on a dummy indicating the presence of the county ordinance. Should one include county-level price? On the one hand, price is related to quantity if one accepts the theory of supply and demand, which makes it seem like price belongs in the model as a regressor. On the other hand, the theory of supply and demand implies that price and quantity are endogenously determined together; thus, price cannot be included if the researcher uses ordinary estimation methods (such as ordinary least squares). One approach here is to exclude price, in which case the estimated coefficient on income should be interpreted as a *reduced form* effect—the association of county ordinances with quantity including not only any direct relationship between the ordinances and quantity, but also any indirect relationship that operates through variations in price.[27] The alternative approach is to find a valid instrumental variable for price, in which case one is able to isolate the direct effect of ordinances on quantity.[28] Either way, simply including price in ordinary (non-instrumental variables) estimation is likely to be a problematic approach.

Further, sometimes even when a set of covariates has a strong relationship with the outcome variable, the estimated effect of other variables will not be affected by whether the researcher includes or excludes the first set of variables. I show below that this is exactly the case in the context of my lower bound estimates: the qualitative and numerical conclusions I reported above in *Locking the Doors* (and repeated in Part II above) are entirely robust to using the FJC authors' multivariate estimates.

III. WHY GEOGRAPHICAL CONTROLS MIGHT NOT BE APPROPRIATE IN STUDYING THE EFFECTS OF *TWOMBLY* AND *IQBAL*

For practical purposes any important differences between using the FJC's multivariate estimates and the use I make of the FJC's raw data would have to be connected to systematic changes in the geographical pattern of motion filings, between 2006 and 2010, that are not caused by *Twombly* and *Iqbal*. To control for such changes, the FJC authors included as covariates in their models a set of 20 judicial district dummy variables.[29]

What good reason is there to include these dummies? In other words, why should anyone worry about an exogenous shift in the geographical pattern of motion filings? Engstrom suggests one possible answer:

---

[27] *See* (Greene, 2008) for a discussion of reduced form coefficients in linear regression models.

[28] *See id.*

[29] There are a few other variables in these models, but for reasons I explain in section 2, *infra*, they are unimportant to the present discussion.

> an idiosyncratic corporate event in the post-*Twiqbal* period—perhaps a large company moves its corporate headquarters to another district, producing a substantial downsizing of its white-collar workforce in the district—could yield a large number of job discrimination filings that are high-value compared to the pre-*Twiqbal* run of cases and so are also lower probability cases relative to pre-*Twiqbal* cases under standard assumptions that the litigant filing calculus turns, at least in part, on a case's expected value. Under this scenario, a regression analysis that does not include covariate controls for judicial district would wrongly suggest a larger *Twiqbal* effect than is warranted.[30]

Engstrom does not point to any such corporate events, and it is unclear how prevalent they were in the relevant time period. It is also unclear why such an event would lead to especially high-value employment discrimination cases.[31] It is also unclear why the litigant's filing calculus should lead to a predictable increase in case quality among actually filed cases, because a factor that makes a case obviously high-value likely is observable to both sides.[32]

More generally, the question lurking behind this discussion is simple: why do we think it might be true that "some of the districts with the highest grant rates were also the districts that showed the greatest increase

---

[30] (Engstrom, 2013, p. 1218) (n. 51).

[31] If anything, moving a whole corporate headquarters would seem to insulate rather than expose an employer, since both the defendant and any plaintiffs can expect the defendant to have an easy time establishing a legitimate business necessity basis under Title VII. *See* 42 U.S.C. § 2000e-2(k)(1)(A)(i). That could lead to *low* expected damage values for the plaintiffs in the event such cases make it to judgment, even if the salaries of would-be plaintiffs are high. Thus it is possible that such cases would be low-value, rather than high-value, by comparison to the pre-*Twombly/Iqbal* run of cases.

[32] For a detailed elaboration of this point in the summary judgment context, see (Gelbach, 2014, p. "Rethinking Summary Judgment Empirics"). Engstrom offers a second example of how nettlesome geographic changes in the case distribution might have occurred: "Covariate controls would similarly be indicated if some districts were to implement new case management practices post-*Twiqbal* that mute [*TwIqbal*'s] effects as to all or certain case types." (Engstrom, 2013, p. 1218) (n. 51) (citing FED. JUDICIAL CTR., PILOT PROJECT REGARDING INITIAL DISCOVERY PROTOCOLS FOR EMPLOYMENT CASES ALLEGING ADVERSE ACTION (2011), available at http://www.fjc.gov/public/pdf.nsf/lookup/DiscEmpl.pdf/$file/DiscEmpl.pdf, as "describing new pretrial procedure for job discrimination cases to be piloted by particular district judges"). I agree that implementation of meaningfully different discovery protocols by judges in some districts but not others during the post-*Twombly/Iqbal* data period could be a good reason to control for judicial district. But the pilot project report Engstrom cites is dated November 2011—after the latest possible activity in any case coded for the FJC's reports.

in the number of orders"?[33] To validate their use of judicial district dummies as regressors in their multivariate models, Cecil and his co-authors need this relationship to be exogenous—not caused by *Twombly* and *Iqbal*. But if we accept that parties will change their behavior in response to pleading standard changes, how can we also assume that these behavioral responses are unrelated to characteristics of local districts? In fact, "controlling" for the change in the geographical pattern of motion filings could yield misleading results. It is possible that it is no accident that "some of the districts with the highest grant rates were also the districts that showed the greatest increase in the number of orders."[34] *Contra* Cecil, then, it is at least possible that "variations in motion practice in individual districts"[35] are not "unrelated"[36] to the *Twombly* and *Iqbal* decisions, but in fact are causal effects of them.

Further, there is the risk in this literature that both producers and consumers of the civil procedure research will project talismanic powers onto multivariate regression analysis. At most, such analysis is only as good as the regressors, and the ones that are feasible to include are far from comprehensive. For example, aside from variables that allow estimated changes in the grant probability to vary by case type categories and district-level dummies, which I discuss in detail *infra*, the only additional regressors used in the FJC reports are dummy variables allowing the change in the grant probability to vary with the presence of an amended complaint. It seems hard to believe that there are no other important determinants of grant probabilities besides this short list.[37]

Of course, in the best of all possible worlds one would have access to variables recording every conceivably important exogenous aspect of cases. The wish list includes case quality, the parties' beliefs concerning the probability a judge would grant a Rule 12(b)(6) motion, and so on. One would of course include these variables as covariates in multivariate analysis if one could, obviating every imaginable source of bias in estimating the effects of *Twombly* and *Iqbal* on grant rates. In the real world, though, available variables may not be exogenous, and it might be worse to include them than to ignore them. Unfortunately, none of the *Twombly*/*Iqbal* studies that uses multivariate models even addresses this

---

[33] (Cecil, et al., March 2011, p. 13).

[34] *Id.*

[35] (Cecil, 2012, p. 39).

[36] *Id.*

[37] Even Alex Reinert's longer list of covariates surely misses plenty of case detail; he includes regressors related to whether an amended complaint is present, the judicial district, the nominating president, whether the claimant is an individual, corporation, government, or other organization, and whether the movant fits each of these categories. *See* (Reinert, 2015 (forthcoming)).

issue.

All of that said, it is worth asking: *were* there important geographical shifts in the distribution of cases? And, would using the FJC's multivariate model-generated estimates make any difference? These are the questions to which I turn next.

IV. DOES IT EVEN MATTER WHETHER ONE "CONTROLS" FOR GEOGRAPHICAL SHIFTS VIA MULTIVARIATE MODELS?

In taking into account the multivariate models in this section, I shall focus only on those related to the share of movants prevailing. That is, I shall continue to use the initial report's raw data, rather than its multivariate-model estimates, to measure changes in the prevalence of motion *filing*.[38] The estimates in question appear in Table A-2 of the updated report,[39] which provides coefficients and standard errors from estimation of a binary logit model concerning whether the Rule 12(b)(6) movant prevailed. The estimated model includes the following variables as predictors: a Year 2010 dummy; case type dummies and their interactions with the Year 2010 dummy; an amended complaint dummy and its interaction with the Year 2010 dummy; judicial district dummies; and a constant.

Including case type dummies and their interactions with the post-*Iqbal* dummy is just a parameterized way of measuring grant rate differences separately by case category. In *Locking the Doors* and above, I separately considered employment discrimination cases and other civil rights cases, so using the raw data cannot cause any problem related to case type dummies.[40]

As a threshold matter, Table 1, *supra*, shows that for both employment discrimination cases and the combined category of cases involving contract, tort, or "other" causes of action, the change in the grant rate contributes essentially nothing to the lower bounds reported in *Locking the Doors*.[41] Consequently, one could replace these figures with 0.0 and no

---

[38] I do so in part to save space; in any case, a quick comparison of Tables 1 and 2 of the initial report suggests that using the FJC multivariate model for motion filing would slightly inflate, rather than reduce, my lower bound estimates.

[39] (Cecil, et al., March 2011, p. 8).

[40] I did combine most other case types into my contracts, torts, and "other" case category. Separately calculated lower bounds for each of these case categories are 18.4% for contracts, 23.4% for torts, and 24.7% for "other" cases, by comparison to a figure of 21.5% reported in *Locking the Doors*, at 2334, for the three case types considered together.

[41] *See* Table 1 in Part III.E, *infra* (showing that the change in the grant rate used in *Locking the Doors* was 0.2 percentage points for civil rights cases and 1.1 points for the contracts, tort, and other case category, by comparison to lower bound amounts of 15.4 and

conclusion in *Locking the Doors* would change for those case categories. So what is actually at issue is only the change in the grant rate for civil rights cases. Accordingly, I will concentrate attention on those cases only.

First consider the question of whether Cecil is right that shifts in the distribution of cases across judicial districts lead to important differences in my results. To see how Cecil *could* be right, consider the simple example in Table 2. The table's first column provides the predicted percentage of the time—based on the FJC updated report model—that a movant would prevail on a Rule 12(b)(6) motion in a 2006 civil rights case with no amended complaint filed in either of two districts—the Middle District of Florida, where the percentage is 43.9%, and the Southern District of New York, where it is much higher, at 73.3%.[42]

In Period 1 of Table 2's hypothetical example, half of all cases are filed in each of these two districts, which implies that the movant would prevail in 58.6% of motions filed in the two districts considered together (see the table's third row). Now imagine that between Period 1 and Period 2, there is a massive shift in the population of cases—either fewer are filed in the Middle District of Florida, more are filed in the Southern District of New York, or both. As a result, more than three-fourths (78.5%, to be exact) of cases in the Period 2 case population are filed in the Southern District of New York, where the movants prevail much more frequently. Consequently, the population-level percentage of movants prevailing in motions across the two districts considered together rises to 67%. In this example, then, the probability that movants prevail increases by 8.4 percentage points even though the probability that a movant prevails is unchanged in each district following *Twombly/Iqbal*; the increase occurs simply because of a change in the composition of the case population.

Did the sort of shift in the case population illustrated in Table 2 occur between the pre-*Twombly* and post-*Iqbal* periods in the updated report's outcomes study? Consider Table 3, which conducts a real-world version of the hypothetical analysis in Table 2. Its first row reports the actual percentage of civil rights cases in which the movant prevails, as reported in the updated report. This percentage was 58.6 in 2006 and 67.0 in 2010, so that the share of cases in which movants prevailed increased by 8.4 percentage points in this period. (I confess to rigging the hypo in Table 2 to match these actual figures.) The table's next two rows present predicted percentages based on the updated report's multivariate model. The table's first column provides the predicted percentage of movants prevailing in

---

21.5 points, respectively).

[42] These are the actual estimated probabilities for these districts, based on my calculations using the logit coefficients provided in Table A-2 of the FJC updated report, *supra* note 39, at 8.

civil rights cases filed in 2006, based on my best approximation to the 2006 geographic distribution of these cases across judicial districts.[43] When I use the 2006 geographic distribution of all cases in the FJC's outcomes study, the model predicts that movants would prevail in 58.7% of those civil rights cases that were filed in 2006 and had no amended complaint. In the table's second column, I calculate the corresponding percentage using the *2010 geographic distribution* of cases, but holding all else the same, so that each district's *2006 grant percentage* is used.[44]

---

[43] To calculate this value, I use the logit coefficients reported in Table A-2 of the FJC updated report, supra note 39, at 8, to predict each district's probability that the movant would prevail in a civil rights case with no amended complaint adjudicated in 2006 (I report these probabilities in percentage form in Table 7 in Appendix A, *infra*). I then calculate the weighted mean of these predicted probabilities using the number of cases in each district in 2006, according to the initial report's Table B-1, at 35 (repeated in percentage form in Table 8 of Appendix B, *infra*, which provides the total number of orders the FJC authors coded in *any* case type, including both civil rights cases and those in other categories).

It would be better to base the geographic distribution on only the set of civil rights cases with no amended complaint that were included in estimation of the model used to estimate the model whose coefficients are provided in the updated report's Table A-2. For this reason I requested such data from Joe Cecil, who was generous in meeting several earlier data requests. Cecil declined to accommodate this data request, though, explaining that

> [w]e are presently seeking authorization to make the research data more broadly available through a public archive, and we will wait until this issue is resolved before fulfilling any additional individual requests.…If we are unable to obtain authorization to make the data available through a public archive, we will then seek permission to respond individually to your request and consult with you about how we can best meet your needs.

E-mail from Joe Cecil to author, time-stamped Tue, 18 Feb 2014 09:39:15. While I would of course prefer to use the actual data in question, the data from Table B-1 of the initial report that I use here are the best possible approximation given the information the FJC authors have released publicly. Using these data amounts to imposing the assumption that the pattern of changes in the cross-district distribution of all cases civil rights cases with no amended complaint that were coded and included in the updated report's Table A-2 estimation changed in the same ways as this pattern changed in the subset of these cases reported as involving civil rights.

[44] *See* note 43, *infra*, for details on the geographic distribution in question.

Table 2: Example of How a Geographic Shift Case Population Could Cause an Increase in the Percentage of Movants Prevailing Even When this Percentage is Unchanged in Each District

	Model-Based 2006 Percentage of Time Movant Prevails[a]	Hypothetical Percentage of Case Population		
		Period 1	Period 2	Difference
Middle District of Florida	43.9	50	21.5	
Southern District of New York	73.3	50	78.5	
Percentage of movants prevailing among all cases in both districts		58.6	67.0	8.4

[a] *Source*: Author's calculations using coefficient estimates provided in Table A-2 of FJC updated report, at 8.

  With the 2010 geographic distribution of cases and the 2006 model-based prediction for movant-prevailing frequency, I find that movants would prevail in 58.8% of civil rights cases having no amended complaint—virtually identical to the 58.7% figure obtained using the actual 2006 geographic distribution. And the corresponding figures for cases with amended complaints suggest that in those cases, the predicted percentage of movants prevailing in 2006 is 58.4% with *both* geographic distributions. Contrary to claims in the updated report and Cecil's *Waves*, then, the change in the geographic distribution of cases explains virtually *none* of the actually observed increase in the percentage of movants prevailing for civil rights cases.

  Given the foregoing discussion, how can Cecil be right when he states that "corrections for factors unrelated to *Twombly* and *Iqbal* often account for the statistically significant differences that appear in the simple

comparison between the pre-*Twombly* and post-*Iqbal* periods"?[45] The answer is simple: with respect to the statistics that are actually at issue, Cecil *isn't* right.

Table 3: The Actual Geographic Shift in the Case Population Would Not Have Changed the Observed 2006 Grant Rate

	2006 Geographic Distribution	2010 Geographic Distribution	Difference
Actual data			
Raw share	58.6[a]	67.0[a]	8.4
Model-based Estimates			
No amended complaint	58.7[b]	58.8[b]	0.1
With amended complaint	58.4[b]	58.4[b]	0.0

[a] *Source*: Table A-1 of FJC updated report, at 8.
[b] *Source*: Based on author's calculations using coefficient estimates provided in Table A-2 of FJC updated report, at 8, and reported counts in Table B-1 of FJC initial report, (Cecil, et al., March 2011, p. 35) (*see* Table 7 of Appendix A, *infra*, and Table 8 of Appendix B, *infra*, for data used in these calculations).

In the first row of Table 4, I repeat the raw percentages of civil rights cases in which the updated report's authors coded movants as prevailing (once again, these figures are 58.6% for 2006 and 67.0% for 2010). In the table's second row, I report estimated percentages for the pre-*Twombly* and post-*Iqbal* periods in the second row of Table 4, for cases with no amended complaint filed in one of the updated report's baseline districts.[46] The predicted percentage of movants prevailing in these districts was 69.0% in 2006 and 77.4% in 2010.[47]

---

[45] (Cecil, 2012, p. 39).

[46] That is, the figures in the second row of Table 4 are the estimated percentage of cases in which movants would prevail under the assumption that the case has no amended complaint and that it occurred in the District of Maryland, the Eastern District of Michigan, or the District of Rhode Island, which is the baseline category considered in the two FJC reports. *See* Appendix A, *infra*, for further details on the computation of these marginal effects.

[47] These estimates are each 10.4 percentage points greater than the corresponding raw movant-prevails rates. The reason for this difference is that cases in the reference-category districts—those filed in the District of Rhode Island, the District of Maryland, or the

Taking the difference of these predicted percentages yields the critical result in Table 4. Calculations based on the updated report's *multivariate model* indicate that even after imposing "corrections for factors unrelated to *Twombly* and *Iqbal*,"[48] the percentage of movants prevailing in civil rights cases with no amended complaint filed in one of the updated report's baseline districts increased by the *identical, 8.4 percentage-point* margin as did the raw percentage.[49]

Table 4: Raw and Predicted Probabilities and Marginal Effects for Whether the Movant Prevails in Civil Rights Cases

	2006	2010	Difference
Raw	58.6	67.0	8.4
Adjusted	69.0	77.4	8.4
Difference	10.4	10.4	0

Source: Raw probabilities are taken from Table A-1 of the FJC updated report, at 7. Adjusted probabilities are for the FJC's baseline category (cases with no amended complaint that were filed in the District of Maryland, the Eastern District of Michigan, or the District of Rhode Island) and are based on author's calculations using the logit functional form together with coefficient estimates reported in Table A-2 of the FJC updated report, at 8.

Nor is the estimated marginal effect based on the updated report's multivariate model any less statistically precise. Table A-1 of the FJC authors' updated report provides a *p*-value of 0.092 for a test of the null hypothesis that civil rights case movants prevailed at the same rate in 2006 and 2010.[50] My own calculations using the coefficient and variance matrix

---

Eastern District of Maryland—evidently had higher grant rates than average. *See* Table 7 in the Appendix, *infra*, for corresponding estimates for the other 20 districts.

[48] (Cecil, 2012, p. 39).

[49] I have not cherry-picked by using the FJC authors' omitted-districts category; in fact, the opposite is true. Table 7 of Appendix A, *infra*, shows that among civil rights cases with no amended complaint, this category has the fourth-*lowest* increase in the percentage of cases in which movants prevail; the observed range is 7.6 to 10.8 percentage points. The same table shows that the district-level estimated increases in movants' prevailing percentages are 2-3 percentage points lower for cases with an amended complaint, with the district-specific percentage-point change ranging from 5.7 to 7.8. If I had access to the distribution of cases across districts and amended-complaint status for each year studied in the updated report (*see supra* note 43), I could calculate a model-adjusted nationwide average increase in the percentage of prevailing movants. It seems clear, though, that such an average would look very similar to the raw data's overall increase of 8.4 points.

[50] FJC updated report, supra note 39, at 7.

estimates from Table A-2's multivariate model yields a *p*-value of 0.080.[51] If anything, then, there is a small increase in precision based on using the multivariate model. Thus when Cecil claims that "the corrections for factors unrelated to *Twombly* and *Iqbal* often account for the statistically significant differences that appear in the simple comparison between the pre-*Twombly* and post-*Iqbal* periods,"[52] he is mistaken as to the one estimate I use where there is any material issue at stake.

Perhaps it is worth noting that the updated report's Table A-2 indicates that the estimated coefficient on the interaction of the Civil Rights and Year-2010 variables was statistically insignificant (its reported *p*-value is 0.272). However, in a nonlinear model such as logit, marginal effects usually depend on the estimated values of multiple coefficients. Consequently, any given marginal effect's variance depends not only on the variance of the Civil Rights-Year 2010 interaction, but also on the variances of other coefficients used to estimate the marginal effect, as well as all relevant covariance terms. Evidently the interaction coefficient in question co-varies negatively with other coefficients in the model. Because of this very possibility, it is not enough to look only at the *p*-value for the interaction coefficient, as the updated report's authors and Cecil appear to have done; the results discussed in the previous paragraph show that their approach yields incorrect statistical inferences.

In sum, I have shown that the change in the geographic distribution of cases appears to do virtually no work in explaining why there was an increase in the rate at which movants prevailed in civil rights cases, as this rate is measured in the table of the updated report data that I use. In addition, using estimates based on the updated report's multivariate models yields changes in the percentage of movants prevailing in civil rights cases that are virtually identical to those I used in *Locking the Doors* and above. These findings flatly contradict Cecil's insistence that my results are somehow confounded.[53]

---

[51] Appropriately estimating the variance of a logit model's marginal effect involves the delta method. Computing delta method variance estimates requires the full estimated covariance matrix, rather than only variance estimates for individual coefficients. The FJC's updated report provides the estimated coefficient vector $\hat{\beta}$ in its Table A-2. However, as is conventional, the table reports only the estimated standard errors for each coefficient estimate, rather than the full estimated variance matrix $\hat{V}$ (the estimated standard errors are the square-roots of the main-diagonal elements of this matrix). I requested and received the full estimated covariance matrix $\hat{V}$ for Table A-2 of the updated report; I am grateful to Joe Cecil and Margie Williams for providing me with this information.

[52] (Cecil, 2012, p. 39).

[53] (Cecil, 2012, p. 39).

V. Issues Related to the Data Collection

It goes without saying that the data used in an empirical study should reasonably represent the variables of interest. For example, a study directed at measuring the effects of *Twombly* and *Iqbal* that cannot distinguish adjudication of Rule 12(b)(6) motions and Rule 56 summary judgment motions would not be using appropriate data.

The most common approach in the literature has been to use the results of a search of the electronic data bases hosted by Westlaw and Lexis. A typical approach is to search for cases whose text includes strings indicating that *Conley*, *Twombly*, or *Iqbal* was cited, and/or that phrases such as "no set of facts" of variations on "plausible" appear.[54] All studies I shall mention have a pre-*Twombly* data period (which is sometimes referred to as the *Conley* period). In addition to this period, some studies have only a post-*Twombly*/pre-*Iqbal* data period,[55] others have cases from only a post-*Iqbal* period,[56] and a third group have cases from all three periods.[57] Such studies have been criticized because those cases that are included in electronic case data bases may be systematically skewed in important ways.[58]

A second approach is to base data collection on federal district court dockets. Three such studies have collected data relating to events occurring either pre-*Twombly* or post-*Iqbal*,[59] while a fourth uses data relating to these periods as well as the period between *Twombly* and *Iqbal*.[60] The advantage of this approach is that, in principle, it covers the universe of relevant cases, though the shortcut approach of "using CM/ECF codes entered by court clerks at the time that motions and orders are docketed" apparently caused

---

[54] *See, for example*, (Moore, 2012, p. 610).

[55] *See* (Hannon, 2008); (Seiner, 2010); (Seiner, 2009); and (Hubbard, 2013).

[56] *See* (Quintanilla, 2011); and (Dodson, 2012).

[57] *See* (Moore, 2012, p. 610); (Moore, 2010); (Brescia, 2011-2012); and (Brescia & Ohanian, 2013-2014).

[58] *See* (Engstrom, 2013, pp. 1214-1215); (Cecil, 2012, pp. text at notes 99-105); and (Cecil, et al., March 2011, p. 37). *But see* (Moore, 2012, p. 608) (arguing "that district court orders ruling on 12(b)(6) motions in Westlaw *are* fairly representative of the universe of all such district court orders") (emphasis in original).

[59] *See* (Cecil, et al., March 2011) (explaining that the CM/ECF code-based procedure the FJC authors used was "intended to be equivalent to identifying motions and orders through docket sheet entries and then reviewing documents linked to the docket entries."); Cecil, et al, FJC updated report, supra note 39 (following up on a subset of the cases included in the initial report); Reinert, *Measuring Iqbal*, supra note 37.

[60] *See* Morgan L.W. Hazelton, Procedural Postures: The Influence of Legal Change on Strategic Litigants and Judges (March 28, 2014) (unpublished manuscript prepared for 2014 Midwest Political Science Association Conference in Chicago). Hazelton studies changes in plaintiffs' pleading behavior following *Twombly* and *Iqbal*.

the FJC authors to miss some motions in some districts studied.[61]

Finally, in two studies, William Hubbard has used data on case terminations from the Administrative Office of the U.S. Courts (AOUSC).[62] One of these studies includes cases filed between April 6, 2006 and May 21, 2006, for its pre-*Twombly* period, and cases filed between the same calendar dates a year later for its post-*Twombly* period.[63] The idea here is that (i) *Twombly* was handed down on May 21, 2007, (ii) Rule 12(b)(6) motions are unlikely to be filed, briefed, and adjudicated in fewer than 45 days, and (iii) May 21 is 45 days after April 6. Since all cases in Hubbard's post-*Twombly* period will have been filed before *Twombly* was handed down, Hubbard reasons that this sample design will avoid any plaintiff selection effects, while at the same time allowing him to compare cases that face the pre-*Twombly* pleading standard (the included cases filed in calendar year 2006) to cases that face the post-*Twombly* standard (the included cases filed in calendar year 2007).[64] Hubbard's other study uses a similar approach, using cases filed in 2008 and 2009, to create a set of cases that were all filed in the 45 days before *Iqbal*; within these cases, he then compares those whose dispositive Rule 12(b)(6) motions would have been adjudicated before *Iqbal* to those for which the adjudication would have happened after *Iqbal*.

One quirk of Hubbard's design is that he considers cases to have been dismissed in response to a Rule 12(b)(6) motion only if the AOUSC codes them as terminating within 225 days. This 225-day cutoff will surely cause him to miss a large fraction of cases in which Rule 12(b)(6) motions are granted with leave for the plaintiff to amend her complaint, after which a defendant files another Rule 12(b)(6) motion that ultimately terminates the action. Evidence in the initial report suggests that grants with leave to amend became more common following *Twombly* and *Iqbal*, so this could be a consequential problem.[65] Unfortunately, there is no way to address this

---

[61] *See* (Cecil, 2012, pp. 3-4) (relating finding that "some districts included in [the FJC authors' initial report's filing] study employ idiosyncratic coding practices when entering the CM/ECF data"), and *id.*, at 4 (stating that the FJC authors "are presently locating these missing motions and orders using docket sheet entries and other sources of information, and will reanalyze the data to determine what effect this has on our original findings.").

[62] *See* (Hubbard, 2013); and Hubbard, *A Theory of Pleading*, Litigation, and Settlement, http://ssrn.com/abstract=2360723 (January 23, 2014).

[63] (Hubbard, 2013, p. 55).

[64] *But see* Fed. R. of Civ. P. 41(a) (allowing plaintiffs to voluntarily dismiss an action without court approval when opposing parties have not answered or moved for summary judgment, or when all parties that have appeared stipulate to the dismissal).

[65] *See* (Cecil, et al., March 2011, pp. 7, text at n. 12). Moreover, many cases may not have had a *first* Rule 12(b)(6) motion adjudicated in less than 225 days. The source Hubbard cites as support for his 225-day cutoff indicates that in cases that terminated in fiscal year 2006 in eight district courts, the mean time from Rule 12(b)(6) motion filing to ruling was 130 days. *See* Table 9 of Institute for the Advancement of the American Legal

problem using Hubbard's data, because the AOUSC assigns the same termination code to cases that are terminated on Rule 12(b)(6) motions as to those that are terminated on Rule 56 motions. The longer the time elapsed following case initiation, the greater the share of such cases that will have been terminated at summary judgment rather than for failure to state a claim,[66] and it is for this reason that Hubbard must use a relatively short cutoff period.

A second problem with Hubbard's data is that the AOUSC codes only for termination of an entire action. This means that his dismissal variable is limited neither to claim-specific dismissals, nor to dismissals that eliminate fewer than all plaintiffs. Only when *all* claims of *all* plaintiffs are eliminated, with judgment then entered and the action terminated, would Hubbard's data allow him to observe that a Rule 12(b)(6) motion has been decided.

In *Locking the Doors*, I used data from the FJC initial report's filing study to measure the number of Rule 12(b)(6) motions filed, where such measures are needed. Where I used information related to outcomes of Rule 12(b)(6) motions, I drew that information from the FJC updated report's frequency tabulations concerning numbers of cases in which the defendant ultimately prevails. Concerning the definition of "prevails", the updated report states that

> We identified cases in which the movant prevailed as those in which the court granted the last motion to dismiss in whole or in part and no opportunity to amend the complaint remained. This included all cases in which the motion was granted with leave to amend, but no amended complaint was submitted during the time allowed. We identified cases in which the respondent prevailed as those in which the last motion to dismiss was denied, or in which the respondent submitted an amended complaint and the movant chose not to respond with an additional motion

---

System, CIVIL CASE PROCESSING IN THE FEDERAL DISTRICT COURTS 48 (2009), http://www.uscourts.gov/uscourts/RulesAndPolicies/rules/Duke%20Materials/Library/IAA LS,%20Civil%20Case%20Processing%20in%20the%20Federal%20District%20Courts.pdf . The median time was considerably lower, at 97 days, *id.*, which indicates substantial right-skewness, so it is possible that a substantial share of cases take considerably longer than 130 days. And of course, there is a lag between case filing and the date when defendants file Rule 12(b)(6) motions.

[66] Cecil reports that 28% of cases included in the FJC updated report's post-*Iqbal* period had been initiated before May 19, 2009. (Cecil, 2012, p. 43) (n. 160). All such cases necessarily had at least one order resolving a Rule 12(b)(6) motion on or after January 1, 2010, which is 226 days after May 19, 2009. To the extent that the statistics are similar for the relevant dates surrounding *Twombly*, then, Hubbard's data construction approach may exclude a nontrivial number of cases that should be included.

to dismiss.[67]

Thus the outcomes data used in *Locking the Doors* correspond relatively well to the colloquial idea that the "grant" of a Rule 12(b)(6) motion precludes further litigation of those claims for which the motion was granted.[68]

For the balance of this section, I shall turn to a detailed discussion of the data in the two FJC studies. These studies are of particular note because the initial report provides the only available data on Rule 12(b)(6) motion filing, and the updated report provides comprehensive evidence on outcomes of Rule 12(b)(6) motions. There are some tricky features—one might go so far as to call them bugs—of the FJC data, and making sense both of my results and of certain criticisms of them is most easily done with a clear understanding of these characteristics.

A. *The FJC authors' data collection methods*

The initial FJC report involved what might be considered two distinct studies: a "filing study" and an "outcomes study." The updated FJC report concerned only outcomes. I discuss each of these studies in turn in sections 1-3 below. For reference, Table 5 summarizes the time periods during which cases included in each study were filed (middle column) or studied (final column).

---

[67] FJC updated report, supra note 39, at 3.

[68] Rule 41(b) provides that "[u]nless the dismissal order states otherwise, … [a dismissal under Rule 12(b)(6)] operates as an adjudication on the merits," so a grant without leave to amend carries prejudice. *Cf. Semtek v. Lockheed Martin*, 531 U.S. 497 (2001) (providing that in cases for which subject matter jurisdiction is founded only on diversity, Rule 41(b)'s effect bars refiling the same claim in the same U.S. district court but bars refiling in another court only to the extent that the courts of the state where the U.S. district court sits would accord the dismissal preclusive effect).

I note that there is one substantive failure of correspondence between the colloquial idea and the FJC's coding approach. I refer interested readers to the discussion of "Type Z" disputes in *Dark Arts*, *supra* note 7.

Table 5: Study Periods for the Filing and Outcomes Studies in the Initial and Updated Report

FJC Study	Cases Filed During:	Case Activity Study Period:
A. Initial report		
1. Filing	Pre-*Twombly* cases:	Pre-*Twombly* cases:
	October 2005 –	October 2005 –
	June 2006	September 2006
	Post-*Iqbal* cases:	Post-*Iqbal* cases:
	October 2009 –	October 2009 –
	June 2010	September 2010
2. Outcomes	Pre-*Twombly* cases:	Pre-*Twombly* cases:
	Dates before	January –
	January 2006	June 2006
	Post-*Iqbal* cases:	Post-*Iqbal* cases:
	Dates before	January –
	January 2010	June 2010
B. Updated report		
3. Outcomes	Pre-*Twombly* cases:	Pre-*Twombly* cases:
	Dates before	January 2006 –
	January 2006	September 1, 2011
	Post-*Iqbal* cases:	Post-*Iqbal* cases:
	Dates before	January 2010 –
	January 2010	September 1, 2011

1. <u>The initial report's filings study</u>

In what I call the "filing study," the FJC authors "used the courts' CM/ECF codes indicating the filing of motions to dismiss and related orders to identify electronic documents with relevant motions … that were in PDF format and were linked to the civil case docket sheets."[69] This search was conducted across all civil cases that were filed in any of 23 district courts[70] between October 2005 and June 2006 (the "pre-*Twombly* period"), and between October 2009 and June 2010 (the "post-*Iqbal* period"). These periods are indicated by the two boxes with solid outlines in the top part of

---

[69] (Cecil, et al., March 2011, p. 5). The authors explain that this procedure was "intended to be equivalent to identifying motions and orders through docket sheet entries and then reviewing documents linked to the docket entries." *Id.*

[70] These 23 district courts "account for 51% of all federal civil cases filed during" 2009. *Id.*

the timeline in Figure 1. The FJC authors coded these cases as having a Rule 12(b)(6) motion filed if such a motion was filed within 90 days of case filing. For the pre-*Twombly* data period, then, any Rule 12(b)(6) motion filing activity is observed for the period running from October 1, 2005, through September 30, 2006 (the latest date when a Rule 12(b)(6) motion could be coded for a case filed on June 30, 2005). Similarly, the post-*Iqbal* data period covers Rule 12(b)(6) motion filing activity for the period running from October 1, 2009, through September 30, 2010.[71] In Figure 1, the boxes with dashed outlines, adjacent to those discussed just above, represent the FJC authors' coding of motion-filing activity in the period after June 30 of 2006 and 2010.

---

[71] Cecil describes the FJC authors' choice to use a 90-day window as being "a consequence of trying to obtain current data on cases filed following *Iqbal*." (Cecil, 2012, p. 9). The initial report was issued in March, 2011, and in light of the high demand for information on Rule 12(b)(6) practice following *Twombly* and *Iqbal*, the FJC authors had to stop coding motion filing activity at some point.

To my knowledge, no one, including any of the FJC authors, has followed up this study to determine whether a longer follow-up period would have appreciably changed the conclusions that can be drawn from the data subject to the 90-day cutoff. Perhaps some sense of the importance of the 90-day cutoff can be gleaned from a quick consideration of the Rules. Defendants wishing to file a Rule 12(b)(6) motion must do so before filing an answer. Fed. R. Civ. P. 12(b). Under the Rules prevailing at the time (as of this writing, pending amendments are slated to take effect on December 1, 2015), defendants waiving service of process pursuant to Rule 4(d) generally had 60 days to file an answer, following the date the plaintiff sends a request for a waiver. Fed. R. Civ. P. 12(a)(1)(A)(ii). Thus defendants intending to both waive service and file a Rule 12(b)(6) motion had an incentive to file that motion quickly, in order to avoid being compelled to file an answer before the Rule 12(b)(6) motion can be heard. For a defendant not waiving service, Rule 4(m) provided that a plaintiff generally had 120 days to serve process, and such defendants had 21 days to answer following the date of service. Fed. R. Civ. P. 12(a)(1)(A)(i). Thus as many as 141 days might have passed before a defendant not waiving service would have had to answer; for a non-waiving defendant's answer to be due within 90 days, then, the plaintiff would have had to have served process within 69 days.

Note that defendants failing to file a Rule 12(b)(6) motion before their answer was due could still file a Rule 12(c) motion for judgment on the pleadings once the pleadings had closed, as courts have held the standard for adjudicating a Rule 12(c) motion to be the same as that for Rule 12(b)(6). Note, though, that discovery might already have begun by then; *see* Rule 16 and Rule 26 for the relationship between the date of the first pretrial scheduling conference required by Rule 16 and initial disclosures required by Rule 26(a) following the Rule 26(f) scheduling conference. To the extent that avoiding discovery is an important objective for defendants, then, they had incentives to file Rule 12(b)(6) motions quickly.

Figure 1: Timeline of Events Related to *Twombly*, *Iqbal*, and FJC Data Collection
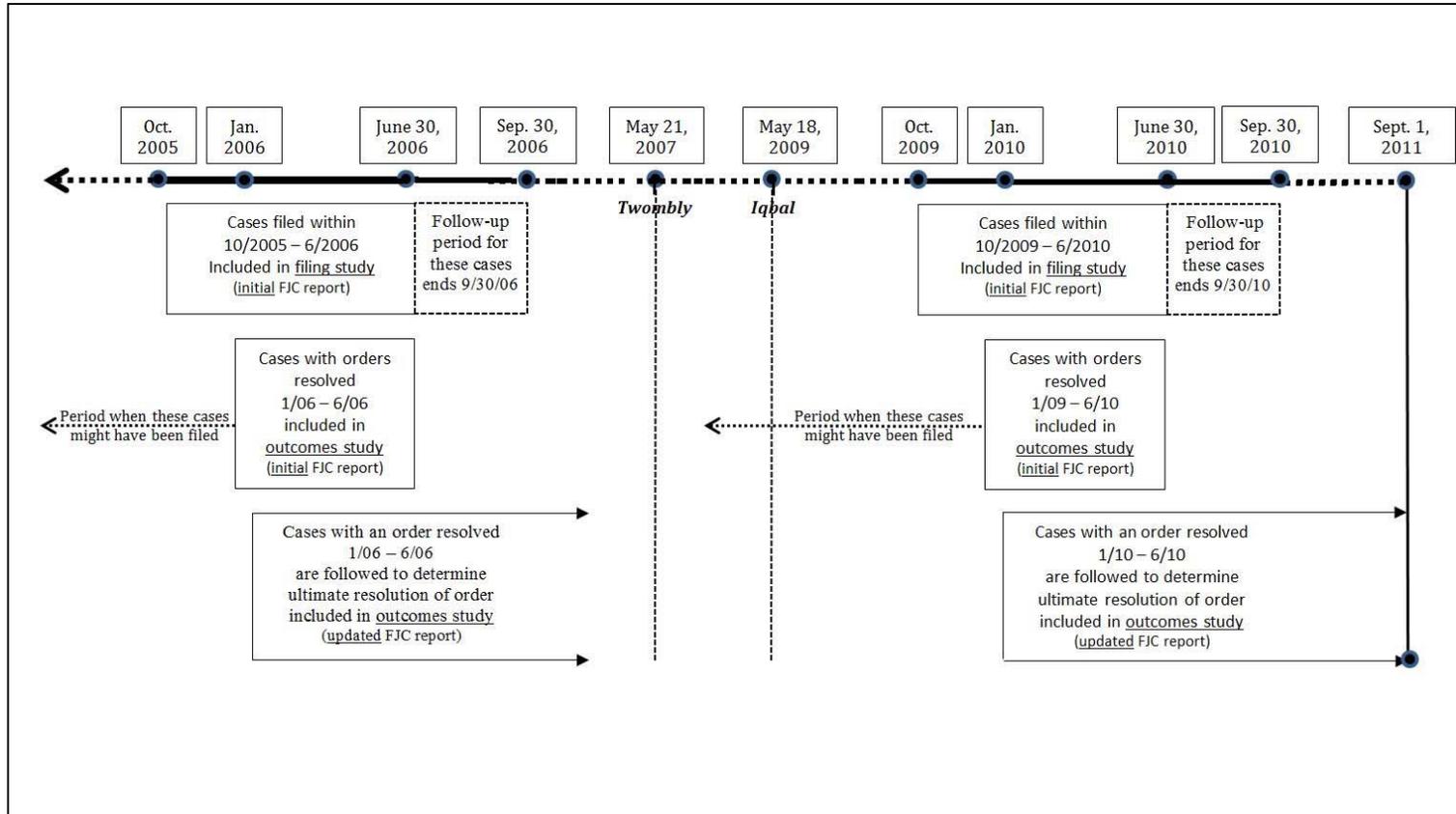
2. <u>The initial report outcomes study</u>

The second study in the initial report concerned outcomes of Rule 12(b)(6) motions. To understand the data collected for this study, it is easiest to quote the initial report at length:

> To assess the changes in the outcomes of motions to dismiss for failure to state a claim, we identified orders responding to motions decided in January through June of 2006 and 2010. … We indicated whether a motion was denied, was granted as to all relief requested by the motion, or was granted as to some but not all of the relief requested by the motion. These last two categories were often combined in the analyses and we simply noted that the motion was granted. In those instances in which the court granted at least some of the relief requested by the motion, we also coded whether the plaintiff was allowed to amend the complaint, and whether the motion eliminated only some claims or all claims of one or more plaintiffs.[72]

The key point to understand regarding the data in the initial report's outcomes study is that its unit of analysis is *orders resolved* (i) between January and June 2006 and (ii) between January and June of 2010. The outcomes study's data collection differs from the initial report's filing study in two ways. First, the calendar periods covered differ. The filing study includes information collected from case activity occurring between October 1, 2005 and September 30, 2006 for the pre-*Twombly* period, and between October 1, 2009 and September 30, 2010 for the post-*Iqbal* period. By comparison, the initial report's outcomes study includes information collected from case activity occurring between January 1 and June 30, 2006 for the pre-*Twombly* period, and between January 1 and June 30, 2010 for the post-*Iqbal* period.

Second, the structure of data collection differs. The filing study has a cohort structure: for both the pre-*Twombly* and post-*Iqbal* periods, outcomes for cases in the filing study are coded based on activity that occurs during a fixed and equal period of time (90 days) following case initiation. Data constructed in this way are sometimes called *flow samples*, since units are selected as they flow into a particular situation.[73] By contrast, the outcomes study is based on what is sometimes called a *stock sample* of cases, because that study selects cases that experience a given event during a period of time, regardless of when those cases were

---

[72] (Cecil, et al., March 2011, p. 5). Note that the initial report excludes prisoner cases and cases with *pro se* parties as well as information related to Rule 12(b)(6) motions that responded to counterclaims and affirmative defenses. *Id.*

[73] *See, e.g.,* (Lancaster & Chesher, 1981).

initiated.[74] The important distinction for present purposes is that a potentially substantial amount of time may have elapsed between a case's initiation and orders resolving any Rule 12(b)(6) motion.

My Figure 1 timeline indicates all this by (i) using boxes with solid outlines to represent the time period for which the outcomes study collects information on cases with orders resolving Rule 12(b)(6) motions, and (ii) using dotted arrows pointing leftward to represent the fact that the cases selected for inclusion in the outcomes study must have been filed at some earlier date. The FJC reports do not provide detailed information either concerning the dates on which the stock-sampled cases included in the outcomes study were filed, or concerning the filing dates of the Rule 12(b)(6) motions resolved by the orders the FJC authors coded. Cecil does report in *Waves* that 28% of cases included in the outcomes study's post-*Iqbal* period were filed before May, 2009.[75] For this reason, the leftward-pointing arrow in my Figure 1 timeline extends to the left of the date when *Iqbal* was filed for the initial report's post-*Iqbal* adjudication cases.

3.   The updated report's follow-up outcomes study

One finding in the initial report was that between the pre-*Twombly* and post-*Iqbal* periods studied, there was an increase in the frequency with which judges granting Rule 12(b)(6) motions also allowed plaintiffs leave to file an amended complaint.[76] In the FJC updated report, the FJC authors collected additional data for such cases in order to "determine the extent to which the respondents submitted amended complaints, and report the outcome of any subsequent motions to dismiss."[77] The updated report includes information on "any subsequent amended complaints, motions to dismiss, and orders resolving such motions,"[78] and it states that "[o]ne or more amended complaints were submitted in 347"[79] of 543 cases[80] that had grants with leave to amend.

This supplemental collection of information for cases included in the initial report's outcomes study thus extended data collection past the June 30, 2006 and June 30, 2010 dates. The updated report is silent as to the length of this extended collection period, but Cecil has informed me that

---

[74] *Id.*

[75] (Cecil, 2012, p. 43) (n. 160) (also reporting that "[o]ver 70% of the cases were filed before October, 2009.").

[76] (Cecil, et al., March 2011).

[77] (Cecil, et al., November 2011, p. 1).

[78] (Cecil, et al., November 2011, p. 3).

[79] *Id*.

[80] *Id*.

"the follow-up period is through September 1, 2011."[81] The arrows and text in the bottom row of my Figure 1 timeline indicate that the collection window for the updated report's extension of the outcomes study runs forward from January 2006 or January 2010.

B. *Combining data from the FJC's filing and outcomes studies*

Cecil relates that his "greatest concern" about the empirical work in *Locking the Doors* is the way I use data from the FJC's filing and outcomes studies together.[82] The underlying basis for this concern seems to be connected to the fact that the FJC's outcomes data set "includes many cases filed before the decision in *Iqbal* was handed down and will not accurately reflect the courts' response to the increased likelihood that defendants will file a motion to dismiss."[83] The reason for this supposedly distorted reflection is that "plaintiffs will be more selective in filing and pursuing cases, and such selectivity would likely remove the cases with weaker claims from the mix of cases considered by the courts."[84] Such additional selectivity "would tend to drive down the rate at defendants file motions to dismiss, or the rate at which judges grant such motions, or both." [85]

In *Locking the Doors*, I addressed issues related to the mis-match of data collection methods in the FJC authors' filing and outcomes studies. As I wrote there:

> some cases with MTDs adjudicated in the *Iqbal* period might have been filed before *Iqbal*, or even before *Twombly*, if the cases have had enough amended complaints. Consequently, the cross section of orders that the FJC analyzed might not fully represent the steady state that will ultimately develop over time. These are standard concerns when one compares cross sections of dynamic processes that are sampled on either side of a policy change.[86]

Cecil rejects this characterization of the problem in *Waves*, but the basis he provides for rejecting it is actually just a re-statement of the characterization itself.[87] The subtle issues here are worth discussing for any

---

[81] E-mail from Joe Cecil to author, time-stamped Wed, 7 Dec 2011 08:18:43 (on file with author).

[82] (Cecil, 2012, p. 42).

[83] (Cecil, 2012, p. 44).

[84] (Cecil, 2012, p. 44). Here Cecil pegs his argument to the presence of a *substantial* plaintiff selection effect, whose presence he rejects (or, perhaps more precisely, mischaracterizes me as rejecting) in his next paragraph. *Id.*

[85] (Cecil, 2012, p. 44).

[86] *Locking the Doors*, at 2338.

[87] (Cecil, 2012, p. 44) (n. 162) ("But the issue is not simply waiting [for] a steady state

who are unfamiliar with the identification of causal effects in dynamically evolving situations. After providing that discussion, I shall present some alternative lower bound estimates that involve only the data from the outcomes study. Since these estimates come from only one of the FJC data sets, any who reject the propriety of combining the data from the filing and outcomes studies should be prepared to accept these alternative estimates. The alternative estimates again indicate the presence of a substantial share of negatively affected plaintiffs among those involved in cases facing Rule 12(b)(6) motions adjudicated post-*Iqbal*.

1. A steady state/transition lens for understanding the implications of combining data from the two FJC data sets

Consider the stylized depiction of cases' lifecycles presented in Figure 2. In the figure's simplified world, Rule 12(b)(6) motions must be filed in the same year a case is filed, and all such motions are adjudicated the following year. Also for simplicity, assume a Rule 12(b)(6) motion will eliminate all claims if it is granted.
 In the figure, bars represent cases filed, and other key aspects are:

- Cases that will have a motion to dismiss filed are represented by shaded bars with superimposed triangles at the time of initiation, while those that will not have such a motion filed are represented by the hollow bars. Thus, the numerator of the share of cases, filed at a given time, that have a Rule 12(b)(6) motion filed equals the number of cases initiated at that time that are represented by the bars with triangles. The denominator equals the number of all cases filed at the time, which includes all those with or without a triangle.

- Among cases that have a Rule 12(b)(6) motion filed, the ones in which it is granted are represented by the bars bearing a superimposed dark "X"; those whose Rule 12(b)(6) motion is denied bear a superimposed dark circle. The numerator of the share in which the defendant prevails thus equals the number of such cases in the bar with the X. The denominator equals the number of such cases with *either* an X or a dark circle.

---

to evolve. The issue is that a sizeable portion of the cases on which Professor Gelbach estimates the judicial response to a motion to dismiss were filed before the change he is seeking to model. Under the terms of his model, motions with different characteristics are likely to be filed after *Iqbal*, possibly resulting in a different distribution of outcomes. The result is a model that relies on implausible assumptions to overcome internal inconsistencies, raising issues that are far removed from 'standard concerns.'")

The top set of four bars in Figure 2 represents cases filed in 2005, so that the ones with Rule 12(b)(6) motions filed have them adjudicated in 2006. Adjudication of these cases' Rule 12(b)(6) motions in 2006 is the subject of the outcomes studies described in the FJC's initial and updated reports, as represented by the solid oval surrounding the X and the filled-in circle. The next-lower set of bars represents cases filed in 2006. Whether these cases have Rule 12(b)(6) motions filed is the subject of the initial report's filing study, which I represent via the dashed oval around these cases.

Figure 2: Timeline of Events Related to *Twombly*, *Iqbal*, and FJC Data Collection
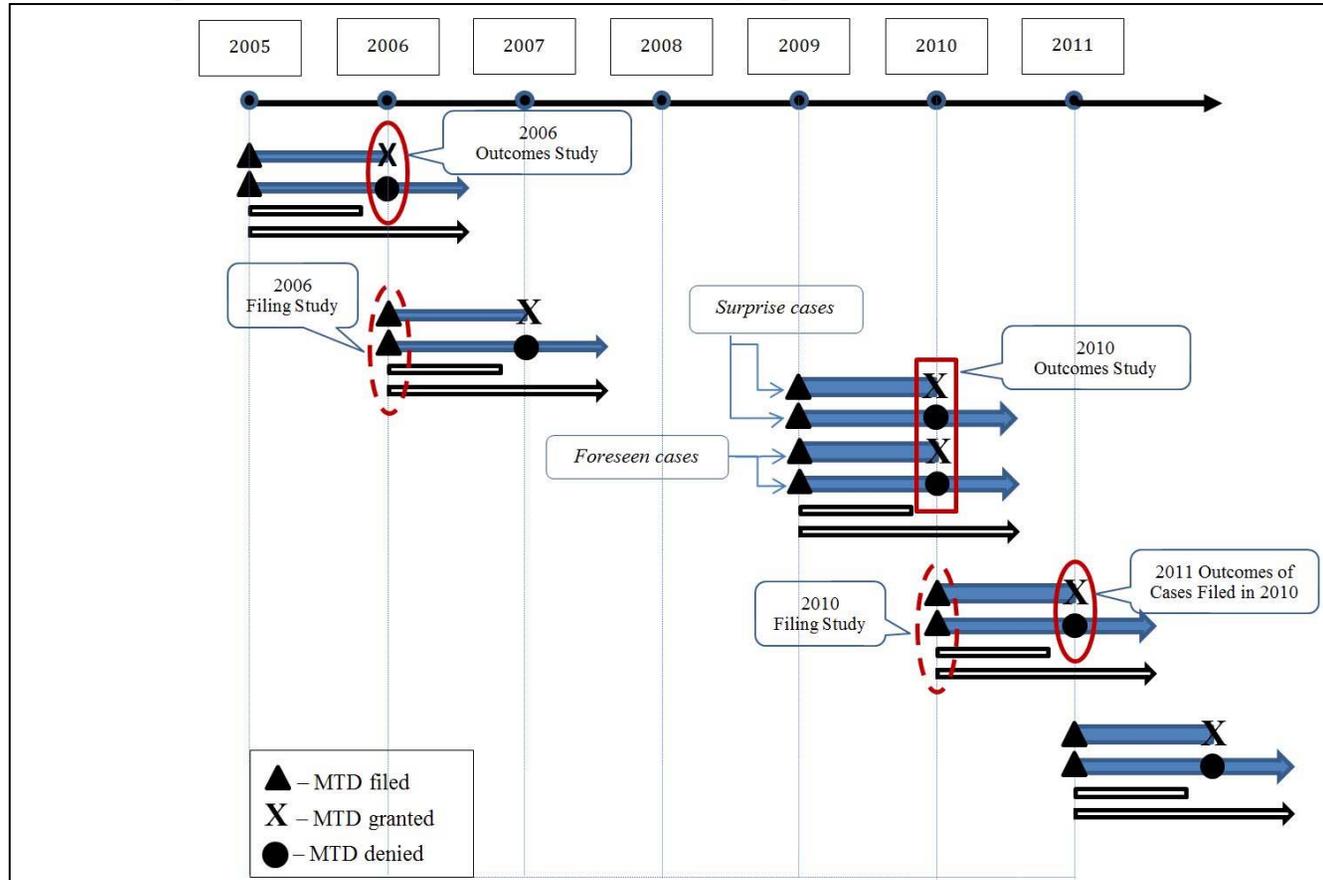
In steady state, each bar in the 2005 cohort of filed cases has the same width as the corresponding bar in the 2006 cohort, indicating that these two cohorts have the same numbers of cases, the same numbers with Rule 12(b)(6) motions filed, and the same numbers with Rule 12(b)(6) motions granted. Given that the litigation system is in steady state, one would obtain the same Rule 12(b)(6) motion filing rate if one studied (i) 2005 data on motion filing among cases filed in 2005, or (ii) 2006 Rule 12(b)(6) filing data using cases filed in 2006. Consequently, if the litigation system was in steady state as to pleading behavior in 2005 and 2006, it would be appropriate to use the pre-*Twombly* parts of the FJC's filing and outcomes studies together.

Now consider cases coded for the FJC reports' post-*Iqbal* period. As represented in Figure 2, data for the post-*Iqbal* part of the filing study come from the cohort of cases filed in 2010, whereas post-*Iqbal* outcomes data come from the cohort of cases filed a year earlier, in 2009. Since *Iqbal* was decided on May 22, 2009, the outcomes study data may include some "Surprise cases," whose parties did not anticipate the new pleading standard, alongside "Foreseen cases," whose parties did. (Note that the filing study's post-*Iqbal* data do not have this bug, since they are flow-sampled beginning after *Iqbal*.)

In Figure 2, the top two bars in the 2009 cohort of filed cases represent Surprise cases that face Rule 12(b)(6) motions, and the next two bars represent Foreseen cases that face such motions. Both types of cases are included in the 2010 part of the outcomes study, which concerns cases represented within the lone solid rectangle in Figure 2. The legitimate basis for concern about the FJC data is that the Rule 12(b)(6) outcomes of the cases inside the 2010 rectangle, considered together, might not reflect the outcomes of the Foreseen cases taken alone. The figure also shows that this problem would end quickly if the litigation system returned to steady state in 2011: even if the set of cases with Rule 12(b)(6) motions adjudicated in 2010 includes some Surprised parties, the set of cases with Rule 12(b)(6) motions adjudicated in 2011 would not. Clearly, then, a better approach to data collection would have been for the FJC authors to follow (i) those cases filed in 2010 in which (ii) Rule 12(b)(6) motions were observed in the filing study to determine (iii) the outcomes of those motions in the cohort of cases files in 2010.[88]

---

[88] Cecil has graciously conceded this point; *see* (Cecil, 2012, p. 45) ("We both agree that a better measure would be 'a cohort-based measure that followed a fixed set of cases from their filing, to the filing of initial Rule 12(b)(6) [motions to dismiss], and then over the period necessary to determine who ultimately prevails on these motions.'") (quoting from *Locking the Doors*, at 2338). He has also explained that the FJC authors did not take this approach because of "[o]ur need to file a prompt report with the Advisory Committee

An important implication of this analysis is that any problems created by the FJC's study design may well be isolated in the 2010 outcomes study data. The fundamental problem is not, then, related to combining cases studied as part of the collection of two different data sets. Rather, it is that one part of one of those data sets—the post-*Iqbal* period in the outcomes study—may not accurately represent the longer run post-*Iqbal* experience.

A second important implication is that if Surprise cases do create such an inconsistency problem, that problem is not limited to the combining of the FJC's filing and outcomes studies, as Cecil would have it. Rather, if the post-*Iqbal* outcomes data are inappropriate for my purposes, then they are inappropriate for other purposes, too. Indeed, such a problem would plague *any* use of the FJC's post-*Iqbal* outcomes data, including not only the FJC authors' original use of that data, but also Cecil's use of it to challenge both my work and that or other authors. Further, since most other studies in this literature all use the same stock-sampling approach to gathering post-*Iqbal* data, any such problem is endemic to this literature. The problem has nothing to do with combining data from multiple data sets, as such.[89]

2. Alternative estimates that use only the outcomes studies' data

Here I provide alternative empirical measures that use only data from the outcomes studies in the initial and final report. Thus these estimates do not combine data from the FJC filing and outcomes studies.

In the first column of Table 6, I once again provide the estimated lower bounds reported in *Locking the Doors*. In the second column, I provide a set of alternative estimates in which M_{post} in the denominator of equation (1), above, is replaced with G_{post}. That is, these estimates are based on the formula

$$(2) \qquad \qquad LB_{NAS,G} \equiv \frac{G_{post} - G_{pre}}{G_{post}},$$

where the numerator remains the change in the number of cases in which the movant prevails (as to one or more claims), and the denominator is the number of post-*Iqbal* cases in which the movant prevails (again as to one or more claims). These estimates do not rely on any data from the FJC's filings study, so they should be unobjectionable to anyone concerned by combining data from the FJC's two data sets.[90]

_____

on Civil Rules." *Id.*, at 43.

[89] A final interesting wrinkle is that Hubbard's approach is built on making *sure* to include Surprise cases. Only by doing so does he ensure that plaintiff selection cases do not disappear from his "post pleading standard reform" set of cases.

[90] On the other hand, as discussed supra, the estimand here differs from the one in *Locking the Doors*; *see* section III.F.3.b of *Dark Arts*, 281-84.

Table 6: Lower Bound Estimates Using Data from Only the FJC Authors' Outcomes study

[One-sided *p*-values in brackets][91]

	Locking the Doors estimates (among cases with Rule 12(b)(6) motion *filed* post-*Iqbal*)[a]	Using only data from outcomes study	
		Estimates among cases whose movants *prevail* on Rule 12(b)(6) motions post-*Iqbal*[a]	Alternative formula (1) estimates[b,c]
Civil rights	18.1 [0.004]	29.3 [0.005]	19.6 [0.008]
Employment discrimination	15.4 [0.033]	18.8 [0.141]	11.5 [0.146]
Contract, tort, and "other"	21.5 [0.000]	21.4 [0.000]	12.0 [0.001]

[a] *See* notes to Table 1 for source and other details.

[b] These estimates are calculated using the number of cases with Rule 12(b)(6) motion adjudicated—whether granted or denied—as a proxy for the appropriate number of cases with a post-*Iqbal* Rule 12(b)(6) motion filed.

[c] *See supra* note 91 for calculation of *p*-values.

---

[91] To calculate *p*-values for the third column of Table 6, first observe that if there are G_{2010} and G_{2006} cases in which the plaintiff loses, and D_{2010} and D_{2006} cases in which the plaintiff wins, at the Rule 12(b)(6) stage, then total number of cases with a Rule 12(b)(6) motion adjudicated is $N = G_{2010} + G_{2006} + D_{2010} + D_{2006}$. Dividing both numerator and denominator of formula (3) by N, we can write the lower bound as

$$LB = \frac{\hat{g}_{2010} - \hat{g}_{2006}}{1 - \hat{g}_{2006} - \hat{d}_{2006}},$$

where $\hat{g}_t = G_t/N$ is the share of all Rule 12(b)(6)-adjudicated cases that involve a movant's prevailing Rule 12(b)(6) grant in year t, and $\hat{d}_{2006} = D_{2006}/N$ is the share of all such cases that involve a respondent's prevailing at the Rule 12(b)(6) stage. Under the null hypothesis that no cases were negatively affected, both (i) $\hat{g}_{2010}$ and $\hat{g}_{2006}$ and (ii) $\hat{d}_{2010}$ and $\hat{d}_{2006}$ should be the same up to random error. It can be shown that under the null hypothesis that both (i) and (ii) hold, a consistent estimate of the variance of the lower bound is given by $\hat{V} \equiv 4 \times (G_{2010} + G_{2006})/N$, and that the lower bound is asymptotically normal with mean zero under the null hypothesis. Dividing each estimated lower bound by the square-root of $\hat{V}$ yields a t-statistic whose asymptotic null distribution is standard normal; the reported *p*-values equal the probability that a random variable with a standard normal distribution would take on a value greater than the t-statistic's realized value.

A second alternative that also does not rely on any data from the FJC's filing study is to implement formula (2) under the assumption that the change in the ratio *m*—which measures the extent to which Rule 12(b)(6) motion filings rose—can be appropriately estimated using the numbers of Rule 12(b)(6) adjudication orders coded for the FJC authors' outcomes study. One potential problem with this approach is that changes between 2006 and 2010 in the numbers of cases with relevant *orders* between January 1 and June 30 may be a poor proxy for the change in the number of cases that had such motions *filed*.[92]

Bearing this potential problem in mind, I report the resulting estimates in the final column of Table 6. For civil rights cases, the approach yields a lower bound of 19.6%, indicating that a plaintiff was negatively affected in at least a fifth of civil rights cases in which a Rule 12(b)(6) motion was adjudicated post-*Iqbal*. This estimate is essentially equivalent to the corresponding estimate provided in *Locking the Doors* (18.4%), and it is highly statistically significant. For employment discrimination cases and cases in the contract, tort, and "other" category, the outcomes data-only approach yields lower bound estimates that are lower than those in *Locking the Doors* by roughly 4 and 10 percentage points, respectively.[93] However, both lower bounds estimates still indicate that plaintiffs were negatively affected in at least one in nine of the cases under consideration—sizable effects.

In sum, both approaches that use only the outcomes study's data yield the same qualitative conclusion. *Twombly* and *Iqbal* appear to have negatively affected sizable shares of both those plaintiffs facing Rule 12(b)(6) motions post-*Iqbal* and the subset of them involved in cases in which movants prevailed on at least one claim via a Rule 12(b)(6) motion post-*Iqbal*. These findings should allay concerns related to the combining of the two FJC data sets in my preferred approach.

C. Do "Surprise" cases lead to upward bias?

I now take up the issue of whether, as Engstrom argues, the presence of "Surprise" cases should be expected to bias my estimates upward. Engstrom writes that

---

[92] For example, the increase in filing of Rule 12(b)(6) motions that is apparent from the initial report's filing study might cause bogged-down judges to take longer to handle any given case's activity. The result would be an increase in the number of orders filed in the first half of 2010, by comparison to the same part of 2006, that is smaller than the increase in motion filing.

[93] The estimated lower bound for employment discrimination cases is not statistically significant, while the contract, tort and "other" estimate is highly significant.

at least some portion of the orders on which [the *Locking the Doors*] estimates rely are directed at plaintiffs who may have been caught off guard by—and thus filed cases into the teeth of—*Twiqbal*'s elevated pleading standard." As a result, the FJC estimates on which Gelbach relies likely overstate the post-*Twiqbal* change in the 12(b)(6) grant rate, which will in turn inflate Gelbach's own selection-adjusted estimates (and also the alternate "Engstrom" calculations just presented).[94]

As sensible as Engstrom's intuition seems at first pass, it is mistaken. His claim is that when judges adjudicating Rule 12(b)(6) motions actually apply the *Twombly/Iqbal* standard, the probability of a Rule 12(b)(6) grant must be greater among Surprise cases than among Foreseen cases. Thus Engstrom is making a comparison *across* sets of cases, *within* a pleading standard. But the only *a priori* information we have about Surprise cases is that they are more likely to have Rule 12(b)(6) motions granted when the *Twombly/Iqbal* standard governs than when the *Conley* standard governs. That is, the information we have concerns a comparison *within* a set of cases, *across* pleading standards—not the same as Engstrom's comparison.

In fact, it is straightforward to construct a hypothetical example in which all of the following are true about a collection of disputes:

- The plaintiffs are caught off guard by *Twombly/Iqbal* and file suit, even though the parties would have settled their disputes had they known that the *Twombly/Iqbal* pleading standard would apply. Thus, these plaintiffs file "into the teeth of" the changed pleading standard, so that these disputes are Surprise cases.

- The defendants in these Surprise cases file Rule 12(b)(6) motions under *Twombly/Iqbal*, so that these Surprise cases would be represented in the FJC's outcomes study data set.

- The defendants in these cases prevail on Rule 12(b)(6) motions 40% of the time when *Twombly/Iqbal* governs.

Since the real-world rate at which movants prevail is in the 55-70% range for the case types I consider, movants in this hypo would prevail at a *lower* rate in Surprise cases than in Foreseen cases.[95] Thus Engstrom's premise—that there are Surprise cases—says nothing about how these cases

---

[94] (Engstrom, 2013, p. 1229).
[95] This is a result of the fact that when an overall group consists of two subgroups, the overall rate at which movants lies between the two subgroups' group-specific rates.

affect the observed rate at which movants prevail. Indeed, this example shows that the FJC data could just as easily understate the post-*Twiqbal* change in the 12(b)(6) grant rate as overstate it. That, in turn, would *deflate* my lower bound estimates—just the opposite of Engstrom's contention.

My point here is not that we know that Engstrom's criticism is wrong— just that we have no particular reason to believe that it is right. I shall state clearly that I self-consciously adopt the behavioral assumption that the rates at which movants prevail in Surprise and Foreseen cases is similar;[96] under this assumption, the results in *Locking the Doors* are valid. Of course, others are free to reject this assumption, but if they want to say anything other than "Who knows?", the burden is on them to offer *alternative* behavioral assumptions that are sufficient to yield informative conclusions. As I have stated elsewhere, the alternative to one set of assumptions is not *no assumptions*, but rather *some other set of assumptions.*[97]

CONCLUSION

The Supreme Court's *Twombly* and *Iqbal* cases have been highly controversial. In light of the gateway role that the pleading standard can play in our civil litigation system, measuring the empirical effects of pleading policy changes is important. One of the central messages of my earlier work, *Locking the Doors*, was that care is required in formulating the object of empirical study. Taking party behavior seriously, as *Locking the Doors* does, leads to empirical results suggesting that *Twombly* and *Iqbal* have had a substantial effect on plaintiffs (and thus, also, on defendants).

In their critiques of my empirical implementation, Joe Cecil and Professor David Engstrom have raised important questions about these results. Their questions are important enough to warrant the detailed answers I have provided in the present paper, which, I believe, show that the results provided and conclusions drawn in *Locking the Doors* continue to stand.

An additional contribution of the present paper is that, in confronting Cecil's and Engstrom's critiques, it elucidates some important points about empirical work in civil procedure. In particular, the discussion here suggests that researchers should carefully consider which covariates belong in statistical models, while also taking care in assessing the empirical importance of controlling for covariates. Further, data collection protocols should be designed with behavioral assumptions in mind;

---

[96] That is, I take my own advice from *Dark Arts*, where I argue for the importance of being clear about what assumptions one adopts when doing empirical work.

[97] *See Dark Arts*, at 248.

moreover, even less than perfectly designed data protocols may be broadly useful. Whether the fruits of particular data collection efforts are useful in evaluating policy changes—whether alone or in tandem with other such efforts—requires careful consideration. Such consideration must especially include attention to the ways in which changes in litigant behavior might be reflected in the data that are collected.

VI.REFERENCES

Brescia, R. H., 2011-2012. The Iqbal Effect: The Impact of New Pleading Standards in Employment and Housing Discrimination Litigation. *Kentucky Law Journal,* Volume 100, pp. 235-88.

Brescia, R. H. & Ohanian, E. J., 2013-2014. The Politics of Procedure: An Empirical Analysis of Motion Practice in Civil Rights Litigation Under the New Plausibility Standard. *Akron Law Review,* Volume 47, pp. 329-373.

Cecil, J. S., 2012. *Of Waves and Water: A Response to Comments on the FJC Study Motions to Dismiss for Failure to State a Claim after Iqbal,* s.l.: s.n.

Cecil, J. S., Cort, G. W., Wiliams, M. S. & Bataillon, J. J., November 2011. *Update on Resolution of Rule 12(b)(6) Motions Granted with Leave to Amend: Report to the Judicial Conference Advisory Committee on Civil Rules,* s.l.: Federal Judicial Center.

Cecil, J. S., Cort, G. W., Williams, M. S. & Bataillon, J. J., March 2011. *Motions to Dismiss for Failure to State a Claim After Iqbal: Report to the Judicial Conference Advisory Committee on Civil Rules,* s.l.: Federal Judicial Center.

Dodson, S., 2012. A New Look: Dismissal Rates of Federal Civil Claims. *Judicature,* Volume 96, pp. 127-135.

Engstrom, D. F., 2013. The Twiqbal Puzzle And Empirical Study Of Civil Procedure. *Stanford Law Review ,* Volume 65, pp. 1203-1248.

Gelbach, J. B., 2012. Locking the Doors to Discovery? Assessing the Effects of Twombly and Iqbal on Access to Discovery. *Yale Law Journal,* pp. 2270-2348.

Gelbach, J. B., 2014. Can The Dark Arts Of The Dismal Science Shed Light On The Empirical Reality Of Civil Procedure?. *Stanford Journal of Complex Litigation,* Volume 2, pp. 223-292.

Gelbach, J. B., 2014. Rethinking Summary Judgment Empirics: The Life Of The Parties. *University of Pennsylvani Law Review,* pp. 1663-89.

Greene, W. H., 2008. *Econometric Analysis.* s.l.:s.n.

Hannon, K. W., 2008. Note, Much Ado About Twombly? A Study on the Impact of Bell Atlantic Corp. v. Twombly on 12(b)(6) Motions. *NOTRE DAME LAW REVIEW ,* Volume 83, pp. 1811-46.

Hubbard, W. H. J., 2013. Testing for Change in Procedural Standards, with Application to Bell Atlantic v. Twombly. *Journal of Legal Studies,* Volume 42, pp. 35-68.

Lancaster, T. & Chesher, A., 1981. Stock and Flow Sampling. *Economics Letters,* Volume 8, pp. 63-65.

Manski, C. P., 1989. Anatomy of the Selection Problem. *Journal of Human Resources ,* Volume 24, p. 343.

Manski, C. P., 1990. Nonparametric Bounds on Treatment Effects. *American Economic Review,* p. 319.

Moore, P. H., 2010. The Tao of Pleading: Do Twombly and Iqbal Matter Empirically?. *American University Law Review,* Volume 59, pp. 554-633.

Moore, P. H., 2012. An Updated Quantitative Study of Iqbal's Impact on 12(b)(6) Motions. *University of Richmond Law Review,* January, 46(2), pp. 603-57.

Quintanilla, V. D., 2011. Beyond Common Sense: A Social Psychological Study of Iqbal's Effect on Claims of Race Discrimination. *Michigan Journal of Race & Law ,* Volume 17, pp. 1-61.

Reinert, A., 2015 (forthcoming). Measuring the Impact of Plausibility Pleading. *Virginia Law Review .*

Seiner, J. A., 2009. The Trouble with Twombly: A Proposed Pleading Standard for Employment Discrimination Cases. *University of Illinois Law Review,* pp. 1011-60.

Seiner, J. A., 2010. Pleading Disability. *Boston College Law Review ,* Volume 51 , pp. 95-149.

VII. APPENDIX

A. *Calculation of Estimated District-Specific Changes in the Percentage of Movants Prevailing in Civil Rights Cases Using the Updated Report's Binary Logit Model Estimates*

The estimated coefficients from binary logit models do not generally have a simple interpretation, because the dependent variable is a complicated function of all of them. To measure the impact of changing one or more regressors from one set of values to another, it is necessary to specify the value of all the regressors at once. As a first pass, I shall use the approach that the FJC authors took to reporting marginal effects in other nonlinear models they estimated in the initial report,[98] using a "baseline

---

[98] *See* Table A-3 of the FJC initial report, which reports marginal effects from a multinomial logit model whose outcome variable takes on one of three values that indicate

consist[ing] of … cases decided in 2006 with no amended complaint in the District of Maryland, the Eastern District of Michigan, or the District of Rhode Island."[99] Given the structure of the binary logit model, the probability of a 2006 movant's prevailing in a Civil Rights case filed in one of the baseline districts without an amended complaint is given by

$$\hat{P}_{CivilRights,2006} \quad = \quad \frac{\exp\left(\hat{\beta}_{constant} + \hat{\beta}_{CivilRights}\right)}{1 + \exp\left(\hat{\beta}_{constant} + \hat{\beta}_{CivilRights}\right)}.$$

This expression equals 0.690 since Table A-2 tells us that the coefficient on the constant is 0.9178 and the coefficient on the CivilRights dummy is -0.1177. The estimated probability of a 2010 movant's prevailing in a Civil Rights case filed in one of the baseline districts without an amended complaint is given by

$$\hat{P}_{CivilRights,2010} \quad = \quad \frac{\exp\left(\hat{\beta}_{constant} + \hat{\beta}_{CivilRights} + \hat{\beta}_{2010} + \hat{\beta}_{CivilRights\times 2010}\right)}{1 + \exp\left(\hat{\beta}_{constant} + \hat{\beta}_{CivilRights} + \hat{\beta}_{2010} + \hat{\beta}_{CivilRights\times 2010}\right)}$$

It is important in principle to recognize that $\hat{\beta}_{2010}$ must be included, since when we switch the year from 2006 to 2010, we turn on not only the Civil_Rights×2010 interaction dummy, but also the 2010 dummy for the omitted category, i.e., torts. Table A-2 tells us that $\hat{\beta}_{2010}$=0.0021 and $\hat{\beta}_{CivilRights\times 2010}$=0.4308. Plugging these values into the definition above yields a probability estimate of 0.774.

In the bottom panel of Table 7, I report the estimated probabilities (expressed in percentage terms) for each judicial district. The percentages in the table are reported separately according to whether an amended complaint was filed in the case (as coded by the FJC authors). The table's top panel reports the actual percentage of movants prevailing overall, according to the raw data reported in Table A-1 of the updated report.

The final columns of Table 7 report the actual overall raw change (top panel) and estimated changes by district (bottom panel) between 2006 and 2010. The overall raw change of 8.4 percentage points is identical to the baseline change (for the omitted districts with no amended complaint), which is represented in bold font. Seventeen of the 21 district-specific estimates for cases with no

---

whether a motion was denied in full, granted with leave to amend as to at least one claim, or granted without leave to amend as to at least one claim. (Cecil, et al., March 2011, p. 30).

[99] (Cecil, et al., November 2011, p. 8). This is the same baseline category used for all models in the two FJC reports.

amended complaint exceed this amount, while none of the district-specific estimates with no amended complaints do. The district-specific estimated changes range from a low of 5.7 percentage points (case with an amended complaint filed in the Southern District of New York) to a high of 10.8 percentage points (case with no amended complaint filed in the Middle District of Florida, the District of South Carolina, or the Northern District of Texas).

B. Computations Related to Geographical Distribution of Cases

In the first two columns of Table 8, I report the number of cases for which the initial report's authors coded orders resolving motions to dismiss for each district represented in the outcomes study. I emphasize that these are the numbers of coded cases for *all* cases coded for the initial report's outcomes study, rather than the actual numbers of civil rights cases included for each district (and amended complaint cell) in the logit model estimated in Table A-2 of the updated report. The latter set of numbers is what would be required to investigate the effects of the exact change in filings across judicial districts. As noted in footnote 43, *supra*, Cecil refused my request for these data; the numbers in Table 8 are the closest available approximation.

**Table 7: Predicted Percentage of Prevailing Movants for Civil
Rights Cases, Without and With an Amended Complaint**

	2006		2010		Change	
Overall (raw data)	58.6		67.0		8.4	
	Amended Complaint?		Amended Complaint?		Amended Complaint?	
District	No	Yes	No	Yes	No	Yes
NYS	73.3	73.0	80.9	78.7	7.6	5.7
MA	71.3	71.0	79.3	77.0	8.0	6.0
GAN	69.2	68.8	77.6	75.2	8.4	6.3
OMITTED (RI, MD, MIE)	**69.0**	68.7	**77.4**	75.0	**8.4**	6.3
OHS	68.9	68.5	77.3	74.9	8.5	6.4
DC	67.5	67.1	76.2	73.6	8.7	6.5
NYE	66.8	66.4	75.6	73.0	8.8	6.6
MN	66.2	65.8	75.1	72.5	8.9	6.7
NJ	65.7	65.3	74.7	72.1	9.0	6.7
INS	62.4	62.1	71.9	69.1	9.5	7.1
TXS	60.6	60.2	70.4	67.5	9.7	7.2
CAE	59.5	59.1	69.3	66.4	9.9	7.3
CAN	57.8	57.4	67.9	64.9	10.1	7.5
PAE	55.0	54.6	65.3	62.2	10.3	7.6
KS	54.4	54.0	64.8	61.7	10.4	7.7
CO	52.6	52.2	63.1	59.9	10.5	7.7
ARE	50.9	50.5	61.5	58.3	10.6	7.8
ILN	50.9	50.5	61.5	58.3	10.6	7.8
FLM	43.9	43.5	54.6	51.3	10.8	7.8
SC	45.0	44.7	55.8	52.5	10.8	7.8
TXN	44.8	44.4	55.6	52.2	10.8	7.8

Source: Author's calculations based on coefficients reported in Table A-2 of
FJC updated report, at 8.

Table 8: Number and Corresponding Share of Cases Included in Outcomes study, by District

District	Number of Cases in District		Share of Cases	
	2006[a]	2010[a]	2006	2010
FLM	84	124	0.120	0.101
TXN	14	30	0.020	0.025
SC	9	18	0.013	0.015
ILN	44	86	0.063	0.070
ARE	14	13	0.020	0.011
CO	23	19	0.033	0.016
KS	26	29	0.037	0.024
PAE	58	31	0.083	0.025
CAN	100	238	0.143	0.195
CAE	33	204	0.047	0.167
TXS	16	29	0.023	0.024
INS	24	28	0.034	0.023
NJ	45	71	0.064	0.058
MN	16	31	0.023	0.025
NYE	35	47	0.050	0.038
DC	9	17	0.013	0.014
OHS	27	55	0.039	0.045
OMITTED (RI, MD, MIE)	46	78	0.066	0.064
GAN	47	13	0.067	0.011
MA	14	23	0.020	0.019
NYS	16	38	0.023	0.031
Total	700	1222	1	1

[a] *Source*: Table B-1 of FJC initial report, (Cecil, et al., March 2011, p. 35).