2018

What We Don't Know About Class Actions but Hope to Know Soon

Jonah B. Gelbach
University of Pennsylvania Law School

Deborah R. Hensler
Stanford Law School

Recommended Citation

WHAT WE DON'T KNOW
ABOUT CLASS ACTIONS
BUT HOPE TO KNOW SOON

Jonah B. Gelbach* & Deborah R. Hensler**

INTRODUCTION

Legislation that would alter class action practice in the federal courts has been pending in Congress. Nearly a decade's worth of U.S. Supreme Court cases have restricted the scope and ease of use of the class action device. Class action critics argue that class litigation is a "racket" that fails to compensate plaintiffs and instead enriches plaintiffs' lawyers at the expense of legitimate business practices.[1] On the other hand, defenders of class actions decry the legislative and judicial forces aligned against them, warning that trends in class action law will eviscerate the practical rights held by consumers and workers.[2] In short, there is considerable controversy over whether class actions are an economic menace or a boon to the little guys.

Notwithstanding the fierceness of the class action debate and the apparent confidence of opponents and proponents in their factual assertions, there is a lot we do not know about federal court class actions, and we have no data that can be used to reliably determine whether class actions are good, bad, or some of each. In fact, *we do not even know how many federal class actions there are*.[3] The absence of data on federal (and state) class actions, various explanations for the absence, and the import of this absence, were elaborated upon by one of us in another article not long ago.[4]

---

1. *See* Myriam Gilles & Gary Friedman, *After Class: Aggregate Litigation in the Wake of* AT&T Mobility v Concepcion, 79 U. CHI. L. REV. 623, 626 (2012); David Marcus, *The History of the Modern Class Action, Part II: Litigation and Legitimacy, 1981–1994*, 86 FORDHAM L. REV. 1785, 1828 (2018) (quoting other sources that describe class actions as "shakedowns" and "bludgeon[ings]").

2. *See generally* Elizabeth Chamblee Burch & Myriam Gilles, *Congress's Judicial Mistrust*, Prod. Safety & Liab. Rep. (BNA) (Apr. 3, 2017).

3. *See infra* Part I for a discussion of several unknown factors concerning federal class action lawsuits.

4. *See generally* Deborah R. Hensler, *Happy 50th Anniversary, Rule 23! Shouldn't We Know You Better After All This Time?*, 165 U. PA. L. REV. 1599 (2017).

We have two purposes in this brief Article. First, we wish to focus continuing attention on the need for more empirical information about the actual functioning of the federal class action system. Second, we wish to share our current efforts to use a one-of-a-kind collection of docket reports, originally harvested from Public Access to Court Electronic Records (PACER),[5] to fill the empirical gap. Presentation of empirical findings resulting from this effort awaits a future article. However, this Article includes suggestions as to how the federal judiciary and Administrative Office of the United States Courts ("AO") could improve data management and data reporting so as to make information about federal class actions more accessible to scholars and others interested in how the class action device operates in practice and what reforms, if any, would be advisable.

I. WHAT WE STILL DON'T KNOW, AND WHY IT MATTERS[6]

For a case to proceed as a class action in federal court, a class must be certified by judicial order pursuant to Rule 23 of the Federal Rules of Civil Procedure.[7] Are there many such cases? In what fraction of the 250,000 to 300,000 civil actions filed in U.S. district courts in a typical year[8] do plaintiffs even seek a certification order?[9] We do not know. Of those class complaints that are certified, we do not know how many are dismissed before reaching the merits. Of those that reach judgment on the merits, we do not know how many are disposed of on summary judgment and how many reach trial (although we do know that trial is generally very rare).[10] Nor do we know how often certified class actions are terminated due to settlement at any stage of litigation.

We also do not know the distribution of putative class actions across different areas of the law, whether state or federal, common law or statutory in origin. We do not know how much litigation effort is expended on these cases, either by the lawyers or the judiciary. We do not know how often class actions and multidistrict litigation (MDL) overlap.

---

5. PACER: PUBLIC ACCESS TO COURT ELECTRONIC RECORDS, https://www.pacer.gov [https://perma.cc/KK7T-KASL] (last visited Aug. 24, 2018).

6. This Part largely follows Hensler, *supra* note 4.

7. FED. R. CIV. P. 23(c).

8. For the twelve-month period ending September 30, 2017, 267,769 civil actions were filed in U.S. district courts. *Table C-2: U.S. District Courts—Civil Cases Commenced, by Basis of Jurisdiction and Nature of Suit, During the 12-Month Periods Ending September 30, 2016 and 2017*, U.S. CTS., at 1, http://www.uscourts.gov/sites/default/files/data_tables/ jb_c2_0930.2017.pdf [https://perma.cc/X3D8-YJGW] [hereinafter *Table C-2*]. In the year prior, 291,851 civil actions had been filed. *Id.*

9. Rule 23 allows any party to seek class certification, but defendant classes are very rare.

10. *See Table C-4: U.S. District Courts—Civil Cases Terminated, by Nature of Suit and Action Taken, During the 12-Month Period Ending September 30, 2017*, U.S. CTS., at 1, http://www.uscourts.gov/sites/default/files/data_tables/jb_c4_0930.2017.pdf [https://perma.cc/3N4M-X82W] (last visited Aug. 24, 2018) (reporting that only 0.9 percent of federal civil cases terminated during the twelve-month period ending September 30, 2017, reached trial); *see also* Nora Freeman Engstrom, *The Diminished Trial*, 86 FORDHAM L. REV. 2131, 2131–37 (2018) (discussing the general downward trends in federal civil trials).

In short, we do not know a lot of things that would seem critical to know for anyone arguing anything about the performance of class actions, or how to improve that performance, much less to support rational policy reform.

To be sure, there is some empirical research on class actions. The most well-researched area is securities fraud litigation, thanks in part to the Securities Class Action Clearinghouse (SCAC). SCAC collects docket information and case documents for securities class actions filed since 1995.[11] Anyone can search the data collection, and numerous research papers have been written using that data. Perhaps as a result, there has been a large amount of scholarship concerning securities class actions.[12]

What about outside the securities arena? There are a number of empirical studies of various aspects of the class action system, including attorney's fees in class action settlements[13] and the role of docket congestion in determining whether proposed settlements are approved.[14] Many other examples abound.[15] We do not mean to devalue these empirical studies. To the contrary, we laud their authors for producing them despite the general difficulty of acquiring data on class actions.

Still, it is remarkable how few basic facts about class actions we actually know. Following recent suggestions one of us made, we propose that to begin to make sense of the performance of the class action system, we need, at a minimum, to know:

> 1. The number of class complaints filed annually, by case type (e.g., securities, anti-trust, consumer fraud, product liability), party characteristics, venue and category of certification (i.e., (b)(1)(a), (b)(1)(b), (b)(2), or (b)(3)).
>
> 2. The mode of disposition of these complaints, whether dropped, dismissed, decided by summary judgment, tried to verdict or settled.
>
> 3. Time to disposition.

---

11. *About Us*, SEC. CLASS ACTION CLEARINGHOUSE, http://securities.stanford.edu/about-the-scac.html#about [https://perma.cc/8W4L-NE52] (last visited Aug. 24, 2018).

12. *See, e.g.*, Lynn Baker et al., *Is the Price Right? An Empirical Study of Fee-Setting in Securities Class Actions*, 115 COLUM. L. REV. 1371 (2015); Lynn Baker et al., *Setting Attorneys' Fees in Securities Class Actions: An Empirical Assessment*, 66 VAND. L. REV. 1677, 1691 (2013); Stephen J. Choi et al., *The Price of Pay to Play in Securities Class Actions*, 8 J. EMPIRICAL LEGAL STUD. 650 (2011); Michael Klausner, Jason Hegland & Matthew Goforth, *When Are Securities Class Actions Dismissed, When Do They Settle, and for How Much?—An Update*, 26 PLUS J., Apr. 2013.

13. *See, e.g.*, Theodore Eisenberg & Geoffrey P. Miller, *Attorney Fees and Expenses in Class Action Settlements: 1993–2008*, 7 J. EMPIRICAL LEGAL STUD. 248 (2010); Brian T. Fitzpatrick, *An Empirical Study of Class Action Settlements and Their Fee Awards*, 7 J. EMPIRICAL LEGAL STUD. 811 (2010).

14. *See, e.g.*, Eric Helland & Jonathan Klick, *The Effect of Judicial Expedience on Attorney Fees in Class Actions*, 36 J. LEGAL STUD. 171 (2007).

15. Hensler, *supra* note 4, at 1603 n.19 ("There is a voluminous literature on Rule 23 class actions. Although much of the literature is limited to doctrinal and normative analysis, a significant fraction of publications cite empirical data. Reviewing and synthesizing the empirical evidence presented in scholarly and public policy literature is beyond the scope of this Article.").

4. Whether or not these complaints were ever certified as class actions, either for all purposes or for settlement only.

5. Whether certification occurred prior to or following judicial decision on pre-trial dispositive motions.

6. Whether there were *Daubert* or other evidentiary hearings on certification.

7. Whether the complaint was resolved as part of a multidistrict litigation (MDL).

8. Whether there was appellate litigation related to the certification decision and the outcome of this litigation.

9. Whether the final outcome of the complaint at the district court level was appealed and the outcome of the appeal.[16]

In the long run, information sufficient to answer these questions should be gathered through the Case Management/Electronic Case Filing (CM/ECF) system run by the AO. The AO should make available such information, in an appropriate format, for academic and government researchers, and, perhaps, the general public (see below regarding the PACER revenue stream). A complementary plan would be for the Judicial Conference to ask either the AO or the Federal Judicial Center to regularly report on class action statistics.

II. OUR NASCENT EMPIRICAL PROJECT ON CLASS ACTIONS

John Maynard Keynes once wrote that "In the long run, we are all dead,"[17] and we should not wait that long to learn the basics of what is going on in class action practice. Part II sketches how we are using docket report information to begin to answer the questions above.[18]

In 2011, Thomson Reuters entered into an agreement with Yale Law School to provide docket report text from Westlaw's database of U.S. district court civil docket reports, which Westlaw refers to as the "DCT" database. This contract was funded through a grant generously provided by Yale's Oscar M. Ruebhausen Fund. The grant proposal was submitted jointly by one of us, Gelbach, and William N. Eskridge, Jr. Under the agreement, which lasted for three years, Thomson Reuters provided the universe of docket report activity for civil actions filed on or after January 1, 2005, with coverage continuing through the end of 2014. Thus, the resulting database includes ten years of docket activity for all civil actions filed in U.S. district courts between 2005 and 2014.[19]

The docket report text is exactly what Thomson Reuters receives from "scraping" the PACER website (except that certain privacy filtering is done

---

16. *Id.* at 1605–06 (footnotes omitted).

17. JOHN MAYNARD KEYNES, A TRACT ON MONETARY REFORM 80 (1923).

18. *See supra* Part I.

19. Information is not available for docket activity after the end of 2014 for cases still pending at that time.

to remove information such as social security numbers). Thus, with respect to docket activity, this database contains the functional equivalent of PACER itself.[20] There are more than 2.8 million unique cases represented in this database.[21] Altogether, Thomson Reuters provided the text of over 170 million docket entries. Excluding duplicates, there are a bit more than 85 million textually unique docket entries.

Some of the information is provided via case-level metadata, including the docket number, information about the district court where the suit was filed, and the PACER nature of suit. The database also provides the names of and other information about all parties and their attorneys, and the date, if any, on which the involvement of a party or an attorney terminated. Also included in the metadata is the name of the judge assigned to the case, but this information changes when a case is reassigned or transferred, such as via the MDL panel. Therefore, presiding judge information likely will have to be recovered via analysis of the docket report itself. The database does not include all the information that appears in a case's civil cover sheet.[22] As a result, we have to identify putative class actions from searches of entry-level text. This turns out to be an involved process.

Entry-level text has important structure, which makes it possible to use pattern-matching text searches for some tasks. For example, entries that represent the docketing of a first complaint begin with the word "COMPLAINT."[23] Below is a result from searching the relevant table in our database for docket report entries that begin with the word "COMPLAINT":

> COMPLAINT AGAINST MEDVENTIVE LLC FILING FEE: $ 350, RECEIPT NUMBER 743327, FILED BY ROBERT J. MONIZ. (ATTACHMENTS: #1 CIVIL COVER SHEET)(BOYCE, KATHY) (ENTERED: 08/14/2006).[24]

Although searching for complaints may be simple to do using pattern-matching text, identifying putative class actions is much more complicated due to the significant variation in the way relevant text appears in the docket entries of these cases. A simple search for "RULE 23," "FRCP 23," or "FED. R. CIV. P. 23" will be both underinclusive and overinclusive. On the

---

20. The database provided by Thomson Reuters does not include any docketed documents, such as complaints, answers, briefs, or judicial memoranda, filed with the clerk of court. Where necessary or useful, those documents can be obtained from PACER or electronic databases that have previously obtained them from PACER.

21. There are several hundred thousand additional cases in the database that appear to have been filed before 2005; it is unclear why these cases were included.

22. If the cover sheet reliably captured whether the complaint was filed in class form, this would be a significant loss. However, extensive analysis conducted by the Federal Judicial Center indicates that the information provided by the attorney in Section VII of the cover sheet (the class action check box) is frequently in error. *See* Hensler, *supra* note 4, at 1616–17.

23. Text in PACER typically is capitalized.

24. Moniz v. MedVentive LLC, No. 1:06-cv-11404 (D. Mass. Aug. 9, 2006), ECF No. 1. This is the first docket entry in *Moniz v. MedVentive LLC*, case number 1:06-cv-11404, filed in the District of Massachusetts on August 9, 2006. The case's nature of suit code is 360, which indicates it involves "TORTS: Other Personal Injury." Civil Cover Sheet, Moniz v. MedVentive LLC, No. 1:06-cv-11404 (D. Mass. Aug. 9, 2006), ECF No. 1-2.

overinclusive side, the phrase "RULE 23" sometimes appears in docket entries that involve something other than Rule 23, such as shareholder derivative suits brought under Rule 23.1.[25] On the other hand, the variations on "RULE 23" listed above also yield an underinclusive set of matches because the person using the CM/ECF system to enter text into the docket often has significant discretion in how to characterize the document. Class-relevant docket entries might therefore never mention Rule 23, instead being framed as "MOTION TO CERTIFY CLASS ACTION," "CLASS ACTION COMPLAINT," or "CLASS COMPLAINT."[26]

Because of the lack of textual structure in class-action-relevant docket entries, we had to experiment quite a bit before settling on a pattern-matching algorithm with which to identify putative class actions. We believe we now have a high-quality algorithm. This algorithm should allow us to answer arguably the most fundamental empirical question: how many putative class actions were filed in various courts, in various nature of suit categories, and in various years? The next question we will tackle will be what happens to these putative class actions. This question likely will be even more challenging to answer.

Although docket reports contain substantial information related to dispositive events, that information does not necessarily come in a neat form. For example, a motion for class certification might be granted as to some claims but not others, or the motion might be granted but for a narrowed class definition relative to the movants' request. Such results could be docketed in forms such as "MOTION GRANTED IN PART AND DENIED IN PART" or "MOTION GRANTED AS TO CLAIMS 1–3 AND DENIED AS TO ALL OTHER CLAIMS," and so on.[27] Similarly, a motion to dismiss or a motion for summary judgment might be granted as to some claims but not others, as to some movants but not others, as to some respondents and not others, or with or without prejudice.

Such textual variation means that we will again have to experiment with pattern-matching algorithms. As we did with the simple counts of putative class actions, we are employing human research assistants to code random samples of cases as a way to cross-check and, in some ways, inform our pattern-matching approach. If necessary, we may use machine learning algorithms to leverage the results of human coding.[28] The upshot to all this

---

25. *See, e.g.*, Kent v. Stoelting, No. 2:17-cv-00893-PP (E.D. Wis. Nov. 17, 2017), ECF No. 18 ("BRIEF in Opposition filed by Jesse Kent re 13 MOTION to Dismiss pursuant to Rule 23.1 MOTION TO DISMISS FOR FAILURE TO STATE A CLAIM.").

26. *See, e.g.*, Hardy v. Ducote, No. 1:02-cv-1520 (W.D. La. Sept. 18, 2002), ECF No. 33 ("MOTION . . . to certify class action"); Lewert v. Boiron, Inc., No. 2:11-cv-10803 (C.D. Cal. Dec. 29, 2011), ECF No. 1 ("CLASS ACTION COMPLAINT against Defendants Boiron Inc, Boiron USA Inc, Laboratories Boiron.").

27. *See, e.g.*, *Lewert*, ECF No. 217 ("ORDER GRANTING IN PART PLAINTIFF'S MOTION FOR CLASS CERTIFICATION . . . .").

28. This could work because the results of human coding can serve as the basis for what is known as classification, which allows us to use machine learning algorithms to predict how cases not coded by humans would be classified if they were so coded. Such algorithms perform very well in many applications involving natural language processing.

is that this project will be time-consuming and, very possibly, subject to error. The reason is that the data we are using were created with the practicalities of litigation, rather than research, in mind.

One final question for future consideration as part of this project is whether the Judicial Conference could make small changes to the CM/ECF system that would make measurement of class action results less difficult, obviating the need for such extensive efforts. For example, the CM/ECF system could be modified so that the clerk of the court flags cases in which class certification has been sought. It could be modified so that events that terminate claims, parties, or both, are better recorded. Another question for the Judicial Conference is whether it would make sense to provide more widespread scholarly access to the information stored in PACER to help develop better tools for measuring not just class actions, but also other forms of litigation. The CM/ECF system is funded by PACER fee revenue, so either Congress would have to increase funding to the judiciary or increases in access would have to be done in ways that do not cannibalize the PACER revenue stream. Although this is a topic for another day, we believe that would be possible.