

THE USE AND RELIABILITY OF FEDERAL NATURE OF SUIT CODES*

Christina L. Boyd**
David A. Hoffman***

2017 MICH. ST. L. REV. 997

ABSTRACT

When filing a civil case in a federal district court, attorneys must identify one, and only one, of ninety issue area nature of suit (NOS) codes that best describes their case. While this may seem like a trivial moment in litigation, the selection of this single descriptor has significant implications for court statistics, empirical research findings, and the allocation of resources to federal courts, including judgeships. Despite the import of NOS codes, there is little within the process of choosing them to guarantee reliability in the selected NOS codes. To assess how reliable NOS codes are, we examine a database of nearly 2,500 federal civil complaints and the individual causes of action within those complaints. Our data reveal that for lawsuits like those involving employment discrimination and intellectual property, the selected NOS codes do a very good job of summarizing the legal content of the complaint. However, in other types of civil lawsuits, including many contract, tort, and real property cases, there is a great deal of inconsistency between the NOS codes and the complaint contents. The difficulty in reliably selecting an NOS code is particularly high as the number and variability of underlying legal claims rise. We conclude by recommending that federal courts adopt a modest revision to the NOS code selection strategy. Rather than the courts relying on attorneys to summarize their frequently complex lawsuits into a

* We thank Adam Candeub, Ryan Black, Ken Grady, Dan Linna, and the participants at the *Michigan State Law Review* symposium for the opportunity to participate and the feedback provided. We gratefully acknowledge the research assistance of Geoffrey Bauer, Keith Blackley, Matthew Canan, Antima Chakraborty, Melissa Jabour, Nicholas Mozal, Jacqueline Sievert, and Elizabeth Young. We also thank Zoran Obradovic and Kosta Ristovski for their work on related projects.

** Associate Professor, Department of Political Science, University of Georgia, CLBoyd@uga.edu.

*** Professor of Law, University of Pennsylvania Law School, DHoffman@law.upenn.edu.

single NOS code, filing attorneys should instead classify their individual causes of action. From there, the courts can automate the grouping of content-similar cases. The result will be NOS codes that much more accurately and reliably capture the nature of the suits.

TABLE OF CONTENTS

INTRODUCTION	998
I. THE UTILITY AND RELIABILITY OF NOS CODES	1000
A. Scholarship Using NOS Codes	1001
B. Court Administrators	1003
C. NOS Reliability	1006
II. EXAMINING NOS CODE SELECTION QUALITY	1008
A. Research Design & Data	1009
B. Analysis	1012
III. SUMMARY AND RECOMMENDATIONS	1024
APPENDIX A: NOS CODES AND CATEGORIES	1028
APPENDIX B: FULL REGRESSION RESULTS	1032

INTRODUCTION

Attorneys filing a complaint in a federal district court must complete, sign, and date a Civil Cover Sheet.¹ This one-page form contains basic case information—e.g., the name and counties of residence of the plaintiffs and defendants, attorney names, basis of jurisdiction, and the “nature of suit” (NOS) for the litigation. The filer must submit the completed form to the Clerk of the Court, who uses these data to populate the docket sheet for the lawsuit.²

Over one-third of the form is devoted to the selection of the case’s NOS code. The form lists ninety different NOS codes and indicates that attorneys should “Place an ‘X’ in One Box Only.”³ The back page of the Civil Cover Sheet further instructs attorneys to do the following when selecting a NOS code: “If there are multiple

1. Form JS 44 (Rev. 06/17), http://www.uscourts.gov/sites/default/files/js_044_1.pdf [<https://perma.cc/GY3L-QQZ8>].

2. See Margo Schlanger, *Inmate Litigation*, 116 HARV. L. REV. 1555, 1699 (2003). District court clerks are generally responsible for assigning the NOS code in prison condition and inmate cases (NOS codes 550 and 555). As Schlanger notes, the “directions to district court clerks on how to choose between 550 and 555 are extremely sketchy.” *Id.* Matters are complicated even further when the filing is by a nonindigent or done by an attorney representing the prisoner. *Id.*

3. Appendix A, *infra*, lists the ninety NOS codes and the thirteen categories that the NOS codes are grouped in.

nature of suit codes associated with the case, pick the nature of suit code that is most applicable.”⁴ For filing attorneys needing additional assistance in selecting their case’s NOS, the current Civil Cover Sheet provides a hyperlink to a document providing a one-sentence description of each of the ninety NOS codes.⁵ Some examples of these descriptions include:

- A case falling within NOS code 441, “Civil Rights: Voting,” is described as an “[a]ction filed under Civil Rights Act, 52 U.S.C. § 10101, and Voting Rights Act, 52 U.S.C. § 10301.”
- NOS code 240, “Real Property: Torts to Land,” is described as an “[a]ction alleging trespass to land, nuisance, contamination or other unlawful entry on or interference with real property possessed by another.”
- NOS 190, “Contract: Other Contract,” has the following as its description: “Action primarily based on rights and obligations under a contract not classifiable elsewhere under the specific natures of suit under ‘Contract.’”

For a handful of NOS categories, the description file provides a few further instructions to attorneys along with the description of the category.⁶ These include:

- NOS 195, “Contract: Contract Product Liability,” directs: “Actions primarily alleging personal injury or property damage caused by a defective product should be classified under the appropriate nature of suit code under ‘TORTS.’”
- NOS 320, “Torts/Personal Injury: Assault, Libel & Slander” says: “(Excludes a government employee).”
- NOS 370, “Personal Property: Other Fraud,” notes: “(Excludes any property that is not real property).”

4. The 2004 version of the Civil Cover Sheet instructed attorneys: “If the cause fits more than one nature of suit, select the most definitive.” Form JS 44 (Rev. 11/04), http://lawprofessors.typepad.com/mass_tort_litigation/files/js44-45.pdf [<https://perma.cc/8XGB-HUX2>].

5. The instructions for nature of suit selection currently end with “Click here for: Nature of Suit Code Descriptions,” linking interested attorneys to <http://www.uscourts.gov/forms/civil-forms/civil-cover-sheet> [<https://perma.cc/92M4-EZRW>]. The file, entitled “Civil Nature of Suit Code Descriptions,” includes basic definitions of each NOS code and is available at http://www.uscourts.gov/sites/default/files/js_044_code_descriptions.pdf [<https://perma.cc/XG2Y-MRDD>]. The 2004 version of the Civil Cover Sheet contained no hyperlink to or description of the different NOS codes. *See* Form JS 44 (revised 11/2004).

6. This coding advice assumes lawyers read it and follow it.

Codes generated from these sparse directions are the foundation for almost all quantitative analysis of the federal litigation system. Scholars seeking to understand the relationship between substance and procedure use NOS codes to select their datasets. So do court administrators looking to equalize workload between and among districts to justify new judgeship requests. A mighty edifice of practical learning rests on harried lawyers getting it mostly right—that is, effectively distilling the essence of their case to a reliable, replicable code. This Article asks, in essence, whether we are on firm footing. The bad news is that NOS codes in their current form are much too noisy to be reliable. The good news is that a complete and almost free solution is easily at hand.

After a review of the literature on the utility of NOS codes, in Part I we focus on a problem with the structure of NOS codes that is well-known but not well-explored: Lawsuits join multiple sorts of legal issues. In the best case, careful lawyers reliably pick particular codes to represent particular clusters of cases, which would mean that the resulting aggregations would be coherent representations of the underlying complaints. But given the reality of pleading practice, and the competing demands on lawyers' cognitive effort, we tend to doubt that such optimism is warranted. We test our intuition by exploiting a dataset of causes of action from ~2500 federal complaints. In Part II we describe our methodology and our findings. As we show, some NOS codes—like IP and employment discrimination—provide clear signals of the underlying complaints. Others, like contract and property, are almost incoherently noisy.

Part III explores the practical upshot of our findings. The solution is relatively obvious and cheap to administer: Require attorneys to select NOS codes for each cause of action, instead of one for each complaint, and then use cluster analysis to assign particular cases to administratively coherent groups.

I. THE UTILITY AND RELIABILITY OF NOS CODES

NOS codes are widely relied on by scholars, courts, and court administrators to draw conclusions about areas of law, recommend the allocation of resources among courts and judges, and advocate for policy reforms. Indeed, as Eisenberg and Schlanger put it, “for researchers seeking to identify all federal district court cases in a certain subject matter category,” the NOS code “is the easiest, and

perhaps the most reliable, method of doing so”⁷ In this Part, we first identify the consumers of the NOS code: scholars and court administrators. We then discuss the (admittedly sparse) literature on those codes’ reliability.

A. Scholarship Using NOS Codes

Scholars seeking to study particular issue areas commonly use NOS codes. These studies apply issue-specific theory, generate empirical results, and make recommendations and broad conclusions in the context of that issue area. It is worth examining a few representative examples in this area:⁸

- Galanter famously observed that trials are vanishing in federal district courts, with tort cases (defined by NOS codes) falling from 55% of all trials in 1962 to 23.4% of all trials by 2002.⁹
- Merz and Pace use NOS codes 820, 830, and 840 to conclude that patent lawsuit filings have increased, but copyright and trademark suits have not.¹⁰ The authors theorize that the difference may be driven by the “expensive nature of patent litigation, which would tend to dissuade court battles and likewise promote the settlement of suits.”¹¹
- Hadfield¹² discusses the changing distribution of federal district court cases from the 1970s to the 2000s, noting the sizable growth in tax and revenue cases on the federal civil docket over time, the slow growth in tort cases, and the steadiness of commercial cases. She concludes that “much of the growth in federal litigation . . . has come in the commercial sphere of the legal system, in particular areas where there has been little change in the underlying available causes of action.”¹³ Hadfield’s conclusions rest largely on trusting NOS coding since her classification of cases as “commercial,” “tax and revenue,” “tort,” or

7. Theodore Eisenberg & Margo Schlanger, *The Reliability of the Administrative Office of the U.S. Courts Database: An Initial Empirical Analysis*, 78 NOTRE DAME L. REV. 1455, 1463 (2003).

8. *See id.* at 1456-58 (providing a then-current list of studies using federal court data and examining specific issue areas via NOS codes).

9. Marc Galanter, *The Vanishing Trial: An Examination of Trials and Related Matters in Federal and State Courts*, 1 J. EMPIRICAL LEGAL STUD. 459, 466 (2004).

10. Jon F. Merz & Nicholas M. Pace, *Trends in Patent Litigation: The Apparent Influence of Strengthened Patents Attributable to the Court of Appeals for the Federal Circuit*, 76 J. PAT. & TRADEMARK OFF. SOC’Y 579, 589 (1994).

11. *Id.*

12. Gillian J. Hadfield, *Exploring Economic and Democratic Theories of Civil Litigation: Differences Between Individual and Organizational Litigants in the Disposition of Federal Civil Cases*, 57 STAN. L. REV. 1275 (2005).

13. *Id.* at 1290.

something else is entirely dependent on the underlying NOS codes for the cases.¹⁴

- Siegelman and Waldfogel report significantly different plaintiff win rates among contract, tort, and civil rights cases, with plaintiffs winning in nearly 73% of contract cases and less than 11% of civil rights cases.¹⁵ These results are important, the authors argue, “for understanding litigation more generally.”¹⁶
- Nielsen, Nelson, and Lancaster examine NOS code 442 in their study of civil employment discrimination.¹⁷ They observe that “[o]ne in five plaintiffs acts as his or her own lawyer” in this area and that these plaintiffs in pro se suits are “almost three times more likely to have their cases dismissed.”¹⁸ As Eisenberg and Schlanger report, pro se filers usually do not fill in a Civil Cover Sheet. As such, court clerks generally designate an NOS for pro se cases “based on their own understanding of a case’s subject matter.”¹⁹
- Hutton, Jiang, and Kumar examine how the political culture of a firm affects the likelihood of corporate litigation.²⁰ As a proxy for the likelihood of litigation, the authors “aggregate multiple same-type lawsuits filed against the firm in the same fiscal year into one observation for each type of litigation.”²¹ The types of lawsuits are defined by the authors using NOS codes—civil rights (442), labor (710, 720, 790), environmental (893), securities (850), and intellectual property (820, 830, 840).²²
- Boyd examines the likelihood of litigant appeals in civil rights (NOS codes 440, 442, 443, 444), business (NOS codes 190, 820, 830, 840), and personal injury tort (NOS codes 310, 320, 340, 340, 350) cases. She finds that civil rights cases are, on average, 13% more likely to be appealed than personal injury torts.²³

14. Hadfield fully details her categorization of NOS codes in the Appendix. *Id.* at 1323.

15. Peter Siegelman & Joel Waldfogel, *Toward a Taxonomy of Disputes: New Evidence Through the Prism of the Priest/Klein Model*, 28 J. LEGAL STUD. 101, 106 (1999). Siegelman and Waldfogel use NOS codes 190 (Contract); 310, 315, 340, 345, 350, 355, 360, and 365 (Tort); 440 and 442 (Civil Rights). *Id.*

16. *Id.* at 130.

17. Laura Beth Nielsen, Robert L. Nelson & Ryon Lancaster, *Individual Justice or Collective Legal Mobilization? Employment Discrimination Litigation in the Post Civil Rights United States*, 7 J. EMPIRICAL LEGAL STUD. 175, 175 (2010).

18. *Id.* at 188.

19. See Eisenberg & Schlanger, *supra* note 7, at 1463.

20. Irena Hutton, Danling Jiang & Alok Kumar, *Political Values, Culture, and Corporate Litigation*, 61 MGMT. SCI. 2905, 2905 (2015).

21. *Id.* at 2911.

22. *Id.*

23. Christina L. Boyd, *Litigant Status and Trial Court Appeal Mobilization*, 37 L. & POL’Y 294, 300-01, 312 (2015).

- Alexander, Eigen, and Rich argue that discrimination plaintiffs' attorneys have shifted employment discrimination litigation to Fair Labor Standards Act litigation.²⁴ Using NOS codes 442, 445, and 710, the authors find preliminary evidence of this in their examination of litigation patterns from 1977 to 2013.²⁵
- Olson finds that Civil Rights cases have a much higher rate of published opinions than contract cases (12% vs. 3%).²⁶
- Grossman uses the Civil Rights (Voting) code to help identify the number of cases filed against local instrumentalities.²⁷
- Sag identifies cases for a survey of all IP litigation filed between 1994 and 2014 by using the NOS codes for trademark, copyright, and patent.²⁸

These papers follow a conventional model. While some note that NOS codes might be unreliable for some purposes, they proceed to use the identifiers to first mark the limits of their data collection and to then draw conclusions (in the aggregate) about the disputes that they have found as representative of the identified universe of cases in that particular subject matter.

B. Court Administrators

Courts and court administrators also frequently use NOS codes in their work. While the use of NOS codes in this context varies, common uses include statistical reporting and resource allocation recommendations based on issue-area trends.

Much of the NOS-based statistical reporting comes from the Administrative Office of the U.S. Courts, the support agency within the federal judicial branch.²⁹ The AO publishes annual and biannual statistics on federal court cases, including the number and type of

24. Charlotte S. Alexander, Zev J. Eigen & Camille Gear Rich, *Post-Racial Hydraulics: The Hidden Dangers of the Universal Turn*, 91 N.Y.U. L. REV. 1, 2 (2016).

25. *Id.* at 55-58.

26. Susan M. Olson, *Studying Federal District Courts Through Published Cases: A Research Note*, 15 JUST. SYS. J. 782, 790 (1992).

27. Perry Grossman, *The Case for State Attorney General Enforcement of the Voting Rights Act Against Local Governments*, 50 U. MICH. J.L. REFORM 565, 590 n.118 (2017).

28. Matthew Sag, *IP Litigation in U.S. District Courts: 1994-2014*, 101 IOWA L. REV. 1065, 1071 (2016).

29. See *Judicial Administration*, U.S. CTS., <http://www.uscourts.gov/about-federal-courts/judicial-administration> [<https://perma.cc/MET3-M7F5>] (last visited Oct. 23, 2017).

cases filed, terminated, and pending each year, by circuit and district court.³⁰ These statistics are frequently broken down by NOS code or NOS category. For example, the annual judicial business report for 2015 details the number and percent of U.S. district court civil cases terminated during the previous twelve-month period.³¹ The data provided are broken down by NOS code and court action taken (e.g., case terminated before pretrial, during/after pretrial, or during/after trial) and, at times, indicate wide variation in the percent of cases by NOS code that reach trial.

The Federal Judicial Center (FJC)³² has developed case weights by NOS codes. The FJC describes the need for district court case weights like this:

Cases filed in the district courts require varying amounts of judicial work to process. At the time a case is filed, the best prediction of how much work will be required hinges on the nature of the case. Observers of the courts would agree, for example, that a judge is likely to spend more time processing a newly filed patent case than a newly filed student loan case. A number of case-specific factors can cause an individual patent or student loan case to depart from this pattern, but over a large number of cases, the general relationship holds true.³³

The resulting case weights vary considerably by NOS code. NOS code 190 (Other Contract Actions) receives a weight of 1.22, NOS 360 cases (Personal Injury torts) have a 0.90 weight, NOS 440 cases (Civil Rights: Other) have a 1.92 weight, NOS 442 cases (Civil

30. See *Statistical Tables for the Federal Judiciary*, U.S. CTS., <http://www.uscourts.gov/report-names/statistical-tables-federal-judiciary> [<https://perma.cc/F5WN-DDKL>] (last visited Oct. 23, 2017); see also *Judicial Business*, U.S. CTS., <http://www.uscourts.gov/statistics-reports/analysis-reports/judicial-business-united-states-courts> [<https://perma.cc/L745-ZJ4L>] (last visited Oct. 23, 2017).

31. 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, 2015*, U.S. CTS. (Sept. 30, 2015), http://www.uscourts.gov/sites/default/files/data_tables/C04Sep15.pdf [<https://perma.cc/6VRM-4D7Z>].

32. *About the FJC*, FED. JUD. CTR., <https://www.fjc.gov/about> [<https://perma.cc/4GK7-MZMQ>] (last visited Oct. 23, 2017).

33. Carol Krafka & Patricia Lombard, *2003-2004 District Court Case-Weighting Study: Final Report to the Subcommittee on Judicial Statistics of the Committee on Judicial Resources of the Judicial Conference of the United States*, FED. JUD. CTR. (Jan. 1, 2005), <https://www.fjc.gov/sites/default/files/2012/CaseWts0.pdf> [<https://perma.cc/Y7S7-8ARD>].

Rights: Employment) have a 1.67 weight, and NOS 791 cases (ERISA labor) have a 0.84 weight.³⁴

These NOS code-based case weights have been adopted by the Judicial Conference³⁵ and applied to make important resource allocation determinations about the federal district courts. For example, the Judicial Conference's determination that a district court needs additional judges is based on the court's current weighted filings per authorized judgeship.³⁶ Similarly, a district court is deemed to be in a state of judicial emergency if it has a district judge vacancy and has weighted filings over 600 per judgeship.³⁷

34. *Id.* at 5. Based on this weighting scheme, a case requiring an average amount of effort should receive a weight of 1. *See also* Philip Habel & Kevin Scott, *New Measures of Judges' Caseload for the Federal District Courts, 1964-2012*, 2 J.L. & CTS. 153, 153 (2014) (discussing how the calculation of a district court's caseload should incorporate various measures of judicial workload, including weighted filings, vacancies, and senior status judges).

35. *See, e.g., Report of the Proceedings of the Judicial Conference of the United States*, U.S. CTS. (Mar. 15, 2016), <http://www.uscourts.gov/sites/default/files/2016-030.pdf> [<https://perma.cc/MPN6-TQTA>] (adopting the FJC's 2016 "new district court case weights for each civil and criminal case type").

36. *See Federal Judgeships: General Accuracy of District and Appellate Judgeship Case-Related Workload Measures: Testimony Before the Subcomm. on the Courts, the Internet, and Intellectual Prop. of the H. Comm. on the Judiciary*, 108th Cong. (2003) (statement of William O. Jenkins, Jr., Director, Homeland Security and Justice Issues), <https://www.gpo.gov/fdsys/pkg/GAOREPORTS-GAO-03-937T/html/GAOREPORTS-GAO-03-937T.htm> [<https://perma.cc/J9A3-XDBN>] (noting that "[i]f the Conference determines that additional judgeships are needed, it transmits a request to Congress identifying the number, type (courts of appeals, district, or bankruptcy), and location of the judgeships it is requesting").

37. *See Judicial Emergency Definition*, U.S. CTS., <http://www.uscourts.gov/judges-judgeships/judicial-vacancies/judicial-emergencies/judicial-emergency-definition> [<https://perma.cc/RZ96-CEBD>] (last visited Oct. 23, 2017) (indicating in the definition of judicial emergency for a district court that "[i]n determining judgeship needs in the U.S. district courts, the Judicial Conference uses weighted filings as a means of accounting for differences in the time required for judges to resolve various types of civil and criminal actions. Rather than counting each case as a single case, weights are applied based on the nature of cases. For example, cases involving a defaulted student loan are counted as 0.16 for each case and antitrust cases are counted as 3.72 cases. The criminal weights are applied on a per-defendant basis. The total for 'weighted filings per judgeship' is the sum of all weights assigned to civil cases and criminal defendants, divided by the number of authorized judgeships").

C. NOS Reliability

The selected NOS is designed to describe what the filed lawsuit is about—i.e., a summary of its nature. This is surely easy to do for a complaint with just one legal claim. It is also relatively easy if the complaint has more than one cause of action but each of those causes of action center around the same issue area. But what are lawyers or litigants to do when their case, like so many others, involves multiple, discrete legal claims across differing issue areas? Is a complaint with two causes of action—a tort and a contract claim—best described as a tort or a contract case?

While they may not realize it, lawyers and litigants filling out Civil Cover Sheets are serving as coders for researchers, courts, and court administrators. But unlike research assistants, these coders have no training and no codebook. There are no rules provided that say that the NOS for a complaint with multiple causes of action should be coded based on the first conversation a lawyer had with his or her client about what happened, by the first cause of action listed, the most commonly occurring, or the most important cause of action in the complaint. There isn't even a requirement for internal consistency—i.e., that lawyers follow the same criteria each time they fill out a Civil Cover Sheet.

And, of course, there is little-to-no incentive for lawyers to care if they select the “right” NOS code even if we agree on what that might be. There is no punishment for improperly classifying a lawsuit's content or reward for selecting the “true” summary category. Lawyers may strategically pick codes to signal to a busy judge that a case is ripe for an aggressive (or passive) management approach. Even more cynically, in a profession where time is money, perhaps the only real NOS selection-related incentive is to select something quickly. This is not exactly the type of incentive structure that is likely to lead to high-quality, consistent coding. As Hadfield puts it, it is an “open question” as to whether the codes are reliable indicators.³⁸

How well does the NOS coding process line up with other data collection and coding efforts? As Epstein and Martin note, the task

38. Gillian K. Hadfield, *Judging Science: An Essay on the Unscientific Basis of Beliefs About the Impact of Legal Rules on Science and the Need for Better Data About Law*, 14 J.L. & POL'Y 137, 145 (2006) (“In addition, the nature of suit coding is based on whatever the plaintiff, her attorney or paralegal indicated as ‘the’ cause of action on the cover sheet she filled in when she filed the case. The reliability of this categorization for research purposes is an open question.”).

of coding data first entails “developing a precise scheme to account for the values of each variable.”³⁹ From there, coders should turn to “methodically and physically assigning all units a value for each variable.”⁴⁰ They then recommend that researchers “[e]stablish that the [v]alues of the [v]ariables are [m]utually [e]xclusive”⁴¹ since failing to do so can lead to confusion among coders.⁴² Instructions to coders, often via codebooks, should “minimize the need for interpretation.”⁴³ Epstein and Martin continue:

Human judgment should be removed as much as possible or, when judgment is necessary, the rules underlying the judgments should be clarified enough to make them wholly transparent to the coders and to others who will examine the study. Only by proceeding in this way can researchers help to ensure the production of reliable measures.⁴⁴

Reliability in coding is key. “A measure is reliable when it produces the same results regardless of who or what is actually doing the measuring.”⁴⁵ Given the above discussion about the process and instructions in place for selecting NOS codes, there can be little doubt that the quality of NOS coding is likely unreliable. The above-discussed instructions to lawyers to pick the “most applicable” NOS code are not transparent and do not remove the need for significant judgment calls on the part of the coding lawyer. As a result, two lawyers with similar complaints containing multiple, diverse legal claims may not code the NOS the same. In short, the NOS coding process goes against nearly everything that social scientists propagate.

Despite this, prior reviews on the quality of NOS codes are mixed. Eisenberg and Schlanger note that the NOS code “appears, from the limited research already done, to be highly accurate. (This too is unsurprising, because the AO [Administrative Office of the U.S. Courts] depends on the accuracy of reports on filings by case category code to allocate resources among courts.)”⁴⁶ However, others have expressed more skepticism.

39. LEE EPSTEIN & ANDREW D. MARTIN, AN INTRODUCTION TO EMPIRICAL LEGAL RESEARCH 95 (2014).

40. *Id.*

41. *Id.* at 105.

42. *Id.*

43. *Id.* at 112.

44. *Id.*

45. *Id.* at 48.

46. Eisenberg & Schlanger, *supra* note 7, at 1463.

Pardo, studying bankruptcy dockets, notes that parties often fail to select the appropriate NOS code in cases involving student loan dischargeability.⁴⁷ Similarly, in his empirical study of NOS-identified ADA Title II and III claims, Eagan found a 21% rate of miscoded NOS codes, noting that “attorneys often miscoded employment discrimination cases, general civil rights cases, or even personal injury cases as ADA Title II and III claims.”⁴⁸

An illuminating approach works backward from filed opinions. Sag takes this tack—he first collected a set of written opinions about copyright and identified 370 documents from 2000 to 2012. Of those, 80% had originally been coded as NOS 820 (copyright)—that is, 20% of the time attorneys filed a suit that resulted in a copyright opinion, it was not originally identified as a copyright case.⁴⁹ But, as he points out, “[m]any copyright cases are also trademark cases, contract cases, common law right of publicity cases, etc.”⁵⁰ Thus, attorneys coded the cases (ending up as copyright opinions) as “Contract, Cable/Sat TV, Other Statutory Actions, Insurance, Assault, Libel, & Slander, Other Personal Property . . . , Civil Rights, Fraud, Personal Injury and even some criminal filings.” Sag, noting the 80% success rate, shrugged: “It’s not bad.”⁵¹ Others might have a different view.

II. EXAMINING NOS CODE SELECTION QUALITY

As detailed above, there are many reasons to be concerned with the output of the NOS code selection process. But litigants, lawyers, and their staff may be better at classifying their lawsuits via NOS code selection than we would expect. As such, we now turn to an

47. Rafael I. Pardo, *The Undue Hardship Thicket: On Access to Justice, Procedural Noncompliance, and Pollutive Litigation in Bankruptcy*, 66 FLA. L. REV. 2101, 2128 (2014) (noting the history of searchability by NOS codes); *id.* at 2130 n.184 (noting that party selection on NOS code creates the possibility for error).

48. Jamie A. Eagan, *The Americans with Disabilities Act: An Empirical Look at U.S. District Court Litigation Involving Government Services and Public Accommodations Claims* 5-6 (Harv. Law. Sch., Working Paper, 2011), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1870601 [<https://perma.cc/9UVP-PAK2>].

49. See Matthew Sag, *Empirical Studies of Copyright Litigation: Nature of Suit Coding* 7 (Loyola Univ. Chi. Sch. of Law Pub. Law & Legal Theory, Research Paper No. 2013-017, 2013), <http://ssrn.com/abstract=2330256> [<https://perma.cc/E7VF-6LEH>].

50. *Id.* at 3.

51. *Id.* at 5.

empirical assessment of how well NOS codes summarize or capture the nature of filed civil cases. Rather than starting with opinions—an already heavily selected dataset⁵²—we will start with complaints. Those complaints are the best and most complete source for assessing what an initiated legal action is, objectively speaking, about.

A. Research Design & Data

Truly random selection of federal complaints remains nearly impossible: Since, for example, many complaints are not available electronically, paper complaints are archived around the country, and the traditional retrieval of a large sample of them (electronically or not) would be cost-prohibitive.⁵³ To gather our federal district court complaints, we instead turned to RECAP, a free digital archive of federal district court and bankruptcy case documents developed in 2008 by the Center for Information Technology Policy at Princeton University.⁵⁴ RECAP's repository is sourced through internet users of PACER (Public Access to Court Electronic Records), the federal judiciary's fee-based service for accessing electronic court records. The RECAP database now contains over 5 million federally filed documents, a number that represents approximately 1% of PACER's current library.⁵⁵

Within the RECAP electronic database, we identified approximately 80,000 electronically available civil complaints, from which we could retrieve unique identifying information like a case's

52. See David A. Hoffman, Alan J. Izenman & Jeffrey R. Lidicker, *Docketology, District Courts, and Doctrine*, 85 WASH. U. L. REV. 681, 681-82 (2007) (finding less than 5% of trial court orders are written into opinions).

53. *Electronic Public Access Fee Schedule*, PACER (Dec. 1, 2013) https://www.pacer.gov/documents/epa_feesched.pdf [https://perma.cc/CXW2-HFHH]. PACER currently charges \$0.10 per page to access court records. *Id.*

54. *RECAP Project – Turning PACER Around*, FREE L. PROJECT, <https://free.law/recap/> [https://perma.cc/8WHC-UXVT] (last visited Oct. 23, 2017).

55. *Id.* RECAP obtains electronic documents from federal courts when individuals install an extension into their Firefox internet browser which, after installed, transfers a copy of any file downloaded from PACER into the RECAP file sharing directory. *Id.* RECAP was seeded with several million documents in 2009 when Aaron Swartz, a 22-year-old Stanford dropout, entered a library at which the government had begun a free trial of PACER and managed to download around 20 percent of the entire PACER database at that time. John Schwartz, *An Effort to Upgrade a Court Archive System to Free and Easy*, N.Y. TIMES (Feb. 12, 2009), <http://www.nytimes.com/2009/02/13/us/13records.html> [https://perma.cc/3BJJ-8BXM]. This amounted to 19,856,160 pages of text. *Id.*

district name and docket number.⁵⁶ Our goal with these RECAP complaints was to build a dataset that somewhat resembles the population of civil complaints filed in federal courts. To do this, we selected a stratified sample of 2,500 complaints from the RECAP database based on an estimation of filed cases' NOS codes to reflect the overall distribution of NOS codes filed in the federal courts in 2007.⁵⁷ After the selection of our 2,500-complaint sample, we found two duplicate complaints (based on docket-number errors), leaving us with a final sample of 2,498 complaints.⁵⁸

The black bars in Figure 1 depict the NOS code distribution for all cases filed in federal district courts in 2007, as recorded by the Administrative Office. In the same figure in gray bars, we depict the same distribution of NOS codes for our resulting database of 2,498 complaints. Based on this comparison, we can see that nearly all major NOS categories are present in our newly created dataset, excepting prisoner petitions and social security cases. We can also see that as a result of this exclusion, several categories are moderately overrepresented in our data.

56. Our search of the RECAP database took place in late 2010.

57. Before the selection of our case sample, we excluded prisoner petitions and social security complaints as well as those complaints filed by a pro se plaintiff.

58. This results in a set of known non-random constraints in our database. We are biased toward more salient/interesting cases, and the users (who pulled the files) are biased in which districts they focused on.

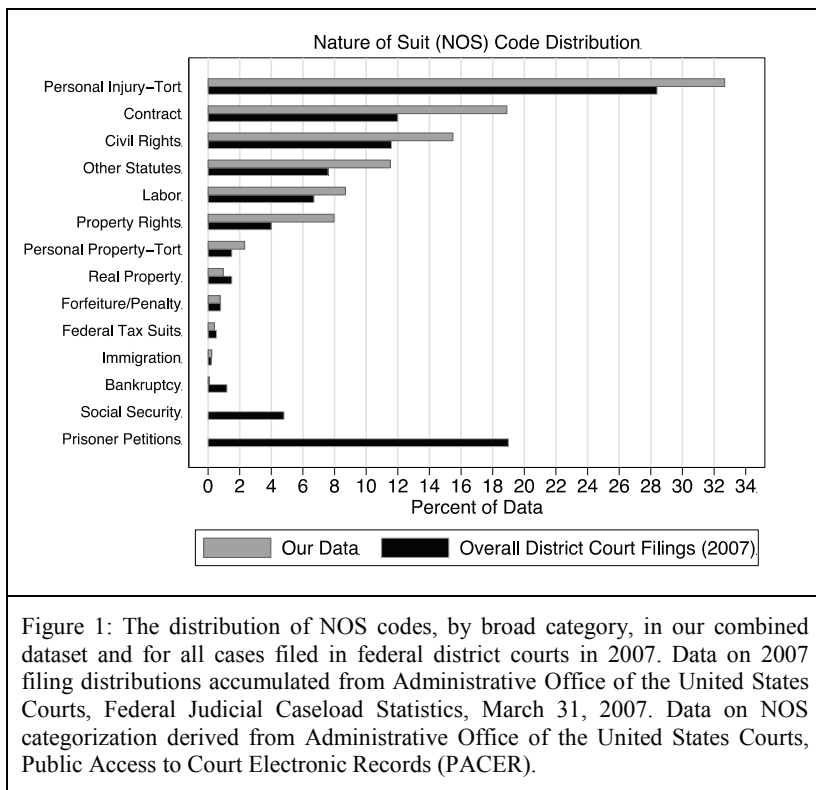


Figure 1: The distribution of NOS codes, by broad category, in our combined dataset and for all cases filed in federal district courts in 2007. Data on 2007 filing distributions accumulated from Administrative Office of the United States Courts, Federal Judicial Caseload Statistics, March 31, 2007. Data on NOS categorization derived from Administrative Office of the United States Courts, Public Access to Court Electronic Records (PACER).

With a set of federal complaints in hand, we then identified and categorized the causes of action within them. Full details of the categorization process are available in Boyd, Hoffman, Obradovic, and Ristovski’s work.⁵⁹ To briefly summarize, we first listed each cause of action. We developed a list of eighteen general categories of causes of action. We then assigned each of the causes of action to a category. That process ranged from easy text normalization (e.g., “Breach of Contract” and “Contract Breach” claims) to more complex coding (ensuring that all causes of action fit within only one category).

59. Christina L. Boyd et al., *Building a Taxonomy of Litigation: Clusters of Causes of Action in Federal Complaints*, 10 J. EMPIRICAL LEGAL STUD. 253, 259-61 (2013).

B. Analysis

Let's start by examining the basics about NOS codes and complaints in our data.

As discussed above, the easiest complaints for lawyers to code the NOS are likely to be those with just one cause of action (COA). In those cases there is no need to select between multiple legal claims to decide which best describes the lawsuit. How often does this happen? Not too frequently. Within our data, just under 17% of the complaints have only one COA. And indeed, the NOS selection here is quite good. We observe that in over 80% of these one-COA-only cases, the selected NOS matches our coding of the legal claim in the case. And, in many of the other 20% of the cases, the only claim listed in the complaint involves a claim for monetary or equitable damages, meaning that the selected NOS may describe the underlying claim even if the listed causes of action do not do a great job doing so.

While this is a positive sign for NOS selection quality, these 17% of the cases are surely the easiest candidates for success. What about the other 83% of the data? Figure 2 shows that there is wide distribution in the data with regard to the number of COAs per complaint. At just over 18%, the most common number of COAs is two. However, other higher frequencies are quite common, with nine COAs per complaint resting at 10% of our data. We would certainly expect that those nine COA complaints are much more difficult to capture in a single NOS than those with just one.

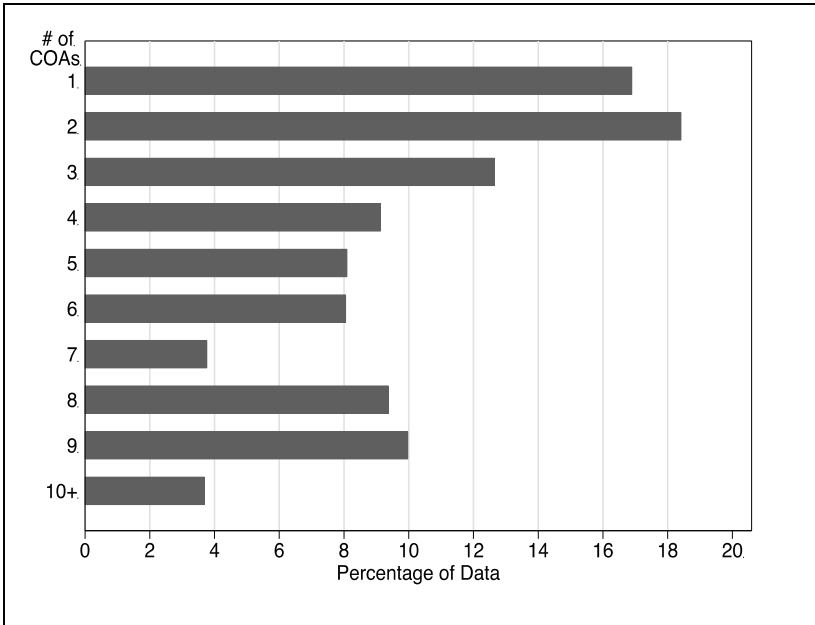


Figure 2: Distribution in the number of causes of action per complaint within the data.

One method of slicing the data seems to confirm this. Nearly all of our nine-COA cases are classified with a torts-related NOS code (90%). Within these, 52% of the underlying COAs are classified as tort COAs. That means, of course, that 48% are not. The other common COAs in these complaints involve legal claims better classified as resting in contract, fraud, intellectual property, civil rights, and damages. This diversity doesn't prove that these cases should not be described as "tort" cases, but it does seem to hint that *only* calling them "tort" cases is too simplistic.

Figure 3 delves further into this topic by plotting the number of COAs per complaint for common NOS codes within our data. The figures are plotted using box plots. Each box in the box plot shows the distribution of observations within the cluster. The 50th percentile (median) number of causes of action within each assigned NOS code is represented by the black line in the middle of each box; the 25th through 75th percentiles are indicated within the box; the 5th through 95th percentiles are presented with the whiskers; and outliers are represented with dots. For ease of figure interpretation,

we exclude all NOS codes that are present in fewer than ten cases in our data.

As we can see from the figure, certain NOS codes are frequently selected in the face of lots of COAs. These include 440 (Other Civil Rights), 160 (Stockholder Suits), 320 (Assault, Libel, & Slander), 365 (Personal Injury-Product Liability), 368 (Asbestos Personal Injury Product Liability), and 840 (Trademark). In a great many cases within our data, complaints classified using these NOS codes contain a variety of legal claims well beyond what is captured by the NOS code.

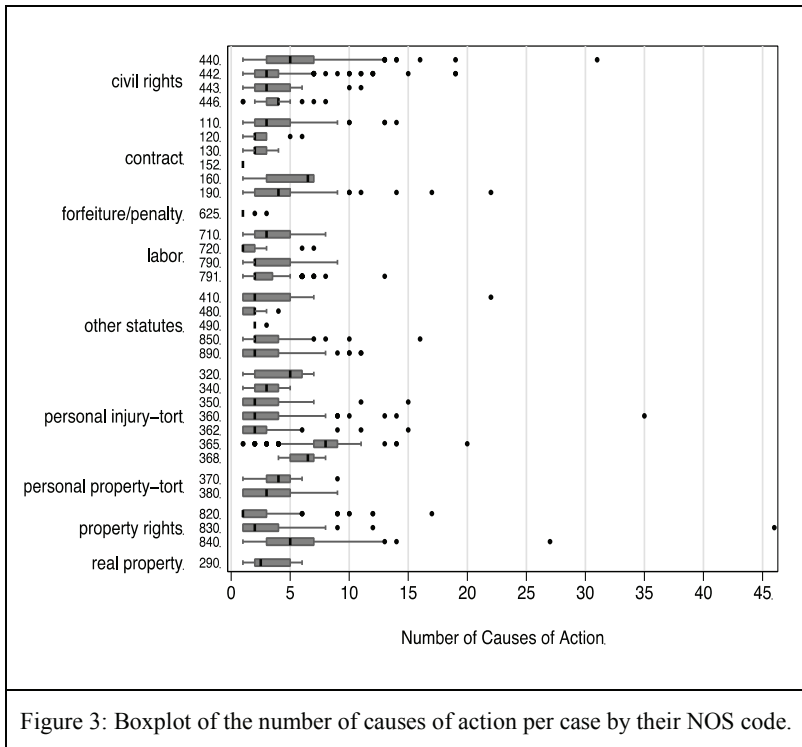
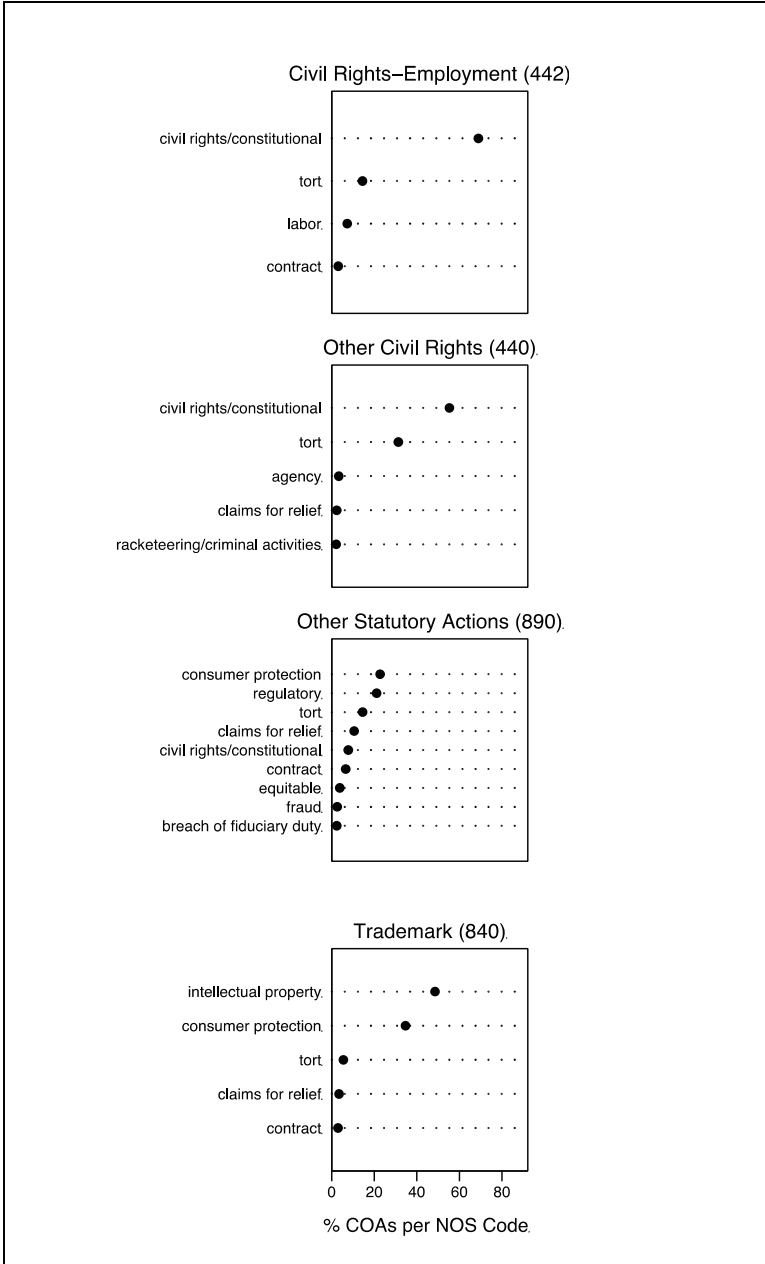


Figure 3: Boxplot of the number of causes of action per case by their NOS code.

One way to examine NOS selection quality is to focus on individual NOS codes and the COAs that are present in cases with those codes. In Figure 4 below, we illustrate this relationship descriptively between our causes of action and NOS codes. In particular, the figure illustrates, for each of the seven most frequently-occurring NOS codes in our data, the percent breakdown of causes of action within those NOS codes.

What do we learn about NOS code composition from Figure 4?

- NOS 442 (Civil Rights-Employment): The composition of COAs within this NOS code is the most homogeneous and predictable of the dataset. Nearly 70% of the COAs in NOS 442 cases are civil rights/constitutional. The next most common cause of action is tort at 15% of the category data.
- NOS 890 (Other Statutory Actions): This statutory-based catch-all NOS code has a combination of consumer protection (23%), regulatory (21%), tort (15%), relief (11%), civil rights (8%), and contract (7%) COAs, but no single COA dominates.
- NOS 840 (Trademark): The trademark NOS has just under 50% of its causes of action classified as intellectual property. Another huge percentage of NOS 840's COAs fall into the consumer protection category (35%).
- NOS 791 (Labor-ERISA): The majority of COAs in this area are related to labor (51%), but other common COAs include contract (18%), claims for relief (11%), enforcement (8%), and tort (5%).
- NOS 440 (Other Civil Rights): Here, civil rights/constitutional COAs are again in the majority (55%), but the degree of COA homogeneity within this category is much lower than it is for NOS 442. Tort COAs make up an additional 31% within this group.
- NOS 365 (Personal Injury-Product Liability): 52% of this category's COAs are classified as torts. Another 22% are contract, 14% fraud, and 9% claims for relief.
- NOS 190 (Other Contract): The COAs falling within this group are quite diverse and distributed, perhaps as the "other" in this NOS code's name suggests. 37% of the COAs are contract, 13% equitable contract, 12% tort, 10% fraud, 10% claims for relief, 7% consumer protection, and 3% breach of fiduciary duty.



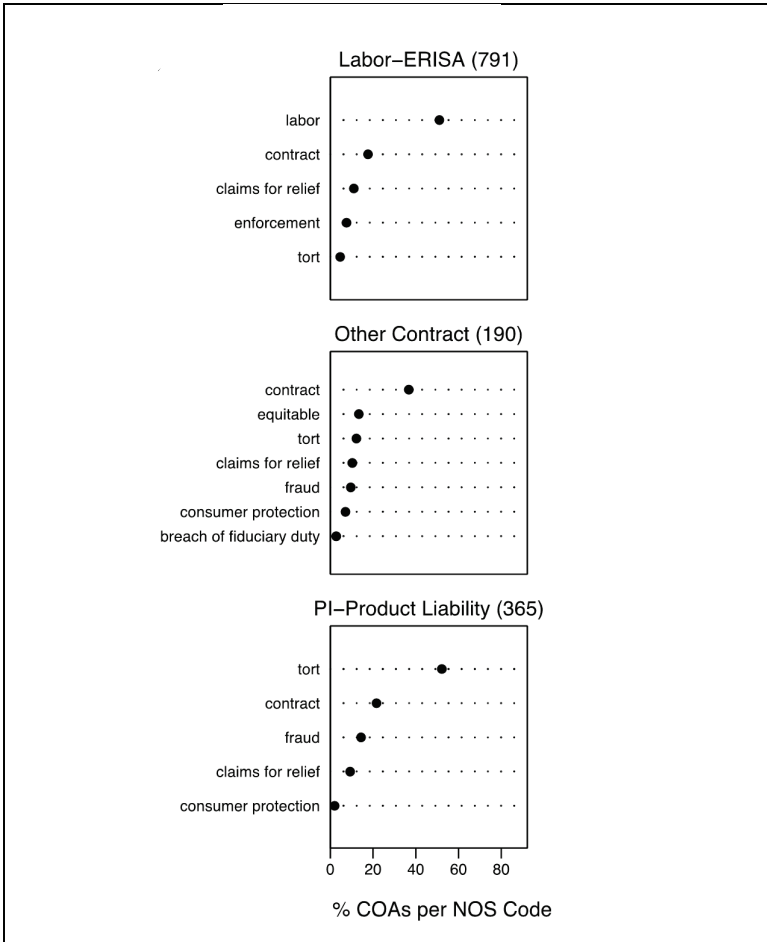


Figure 4: Dotplot of the causes of action present for the NOS codes in our data. The combination of causes of action per individual NOS code totals 100%. For purposes of exposition, the Figure focuses exclusively on the seven most commonly occurring NOS codes in our data and excludes causes of action that account for less than 2% of the causes of action within an NOS code in our data.

Figure 5 provides a very similar descriptive view of the data and the relationship between NOS codes and COA content in complaints, but it does so in the opposite direction. To do this, Figure 5 focuses on the broader NOS code categories that are most commonly selected when there is a specific COA in a complaint. As we can see, the story is familiar. For some COAs, like civil rights/constitutional law and IP, the choice of NOS category is nearly

always consistent. But for other COAs, there is much more distribution across NOS categories. For contract COAs, for example, a tort NOS category selection is much more likely than a contract one. For agency COAs, a civil rights or tort NOS is equally likely. The distribution spreads even more with consumer protection COAs, where NOS categories related to other statutes, property rights (intellectual), contract, and torts are all strong possibilities.

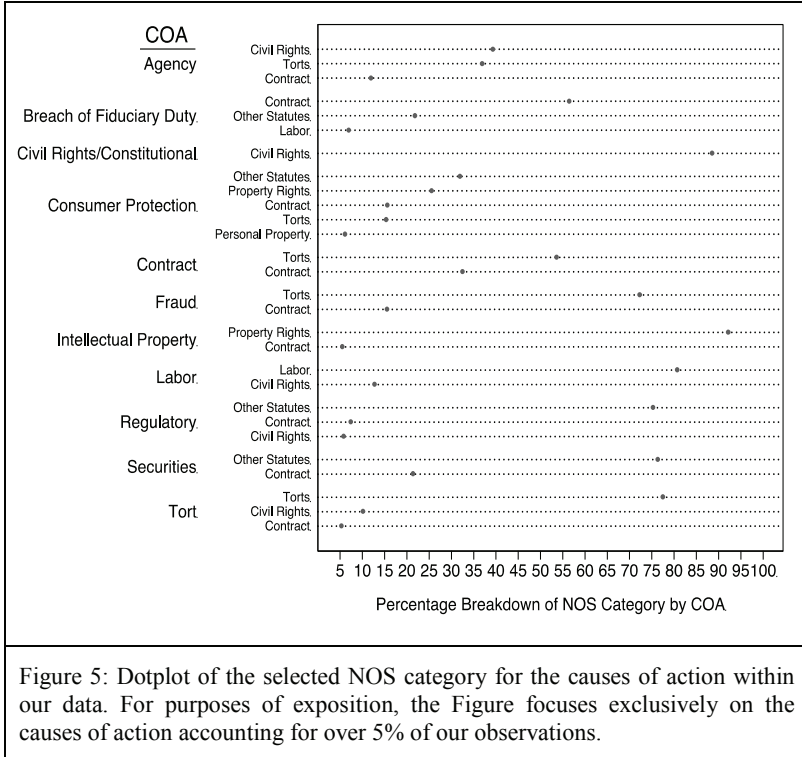


Figure 5: Dotplot of the selected NOS category for the causes of action within our data. For purposes of exposition, the Figure focuses exclusively on the causes of action accounting for over 5% of our observations.

While we now have a good sense from the descriptive data on this, to further and more systematically delve into the question of whether the presence of a COA leads to a particular NOS code selection, we turn to regression analysis of what COAs best explain the selection of a NOS category on the Civil Cover Sheet. To do this, we separately model logistic regressions for the major NOS categories within our data. This means we estimate seven models, one for each of the following NOS categories: Civil Rights, Torts, Contract, Labor, Other Statutory, Real Property, and Property Rights (Intellectual). The dependent variable in each model is dichotomous,

coded as 1 any time the NOS category of interest is selected in a case and 0 if it is not. The independent variables are each of the COA types in our data. These too are coded dichotomously, with COAs present in a case’s complaint coded as 1 and those not present coded as 0. The unit of analysis for this inquiry is the individual case complaint.

Since the statistical results of logistic regressions like these cannot be directly interpreted, we immediately focus on the substantive results of these models. Full regression results are reported in Appendix B. Table 1 reports the predicted probability that a case containing the listed COA in its complaint will have a NOS coded within the specific category. Cells marked “NS” indicate results that are not statistically significant—meaning that the estimates indicate that the probability of the NOS category being selected given that COA cannot be statistically distinguished from 0. Cells marked as “--” indicate that the COA perfectly predicts failure in that NOS, meaning that the COA never leads to that NOS being selected in the data. These COAs are excluded from the modeling in these cases.

Table 1: Predicted Probability of NOS Category Selection⁶⁰

COA	NOS: Civil Rights/ Con Law	NOS: Torts	NOS: Contract	NOS: Labor	NOS: Other Statutory	NOS: Real Property	NOS: Property Rights (Intellectual)
Agency	NS	NS	0.45	--	NS	NS	NS
Bad Faith	NS	--	0.94	NS	--	--	--
Breach of Fiduciary Duty	--	NS	0.67	NS	0.34	--	NS

60. The table displays the predicted probability that a particular NOS category is selected given the presence of a particular COA. Full regression results, from which these predicted probabilities are generated, are provided in Appendix B. NS indicates not statistically significant. The symbol “--” indicates that the COA perfectly predicts failure, meaning that the COA never leads to that NOS being selected in the data. Modeling excludes COAs related to relief and COAs that were classified as obscure/difficult to code.

Civil Rights/ Constitutional	0.90	0.001	0.02	0.01	0.04	NS	--
Consumer Protection	NS	0.01	0.17	0.02	0.61	NS	0.04
Contract	0.01	0.02	0.82	0.02	0.03	0.01	NS
Enforcement	NS	0.01	0.19	0.22	NS	NS	0.02
Equitable Contract	NS	0.01	0.46	NS	0.20	NS	NS
Fraud	0.01	0.19	0.05	NS	0.05	NS	NS
Intellectual Property	--	NS	0.05	--	NS	--	0.94
Labor	NS	NS	0.01	0.88	NS	--	--
Process	NS	NS	0.23	NS	0.51	--	--
Property	NS	NS	0.42	NS	NS	0.26	NS
Racketeering	NS	NS	0.21	NS	0.45	--	NS
Regulatory	NS	NS	0.10	NS	0.84	NS	NS
Securities	--	--	0.09	NS	0.90	--	--
Tax	--	--	--	--	0.62	--	--
Tort	0.02	0.85	0.03	0.01	0.05	NS	NS

Together, these regression results provide more interesting insights about the NOS selection process and how well it mirrors the content of civil complaints. As observed in our prior descriptive figures, things look quite good for civil rights and IP NOS selection. The presence of civil rights and IP COAs lead to over a 90% likelihood that the selected NOS category will be civil rights and property rights (intellectual), respectively. The results are also relatively strong for labor and tort NOS category selection (an above 85% likelihood of a COA issue-area match with the NOS category in each). But we also see a 19% likelihood that fraud COAs will get classified with a tort NOS.

Table 1's regression results indicate that contract, other statutory, and real property NOS categories are much more problematic—and for different reasons. The real property NOS category stands out for its lack of predictability. There, property COAs lead to only a 26% likelihood of the real property NOS being selected. None of the other COAs have a notable effect on the selection of a real property NOS.

For the contract NOS category, we see that the presence of a number of different COAs has a positive and meaningful effect. This includes areas that very naturally fit within contract—e.g., agency COAs (+0.45 probability of contract NOS selection), bad faith COAs (+0.94), breach of fiduciary duty (+0.67), equitable contract (+0.46), and contract (+0.82). But it also includes a high likelihood for many COAs that seem to more naturally fit in other NOS categories: consumer protection (+0.17), securities (+0.09), regulatory (+0.10), racketeering (+0.21), and intellectual property (+0.05).

The selection of the “other statutory” is also predicted by a diverse set of COAs. In some ways, this is to be expected given the broad catch-all content of NOS codes in this category, united simply by their statutory nature. We would certainly expect that high probabilities would be present for securities (+0.90), regulatory (+0.84), consumer protection (+0.61), and racketeering (+0.45) since each is explicitly listed under this category. But perhaps less expected are equitable contract (+0.20), tax (+0.62), breach of fiduciary duty (+0.34), civil rights (+0.04), tort (+0.05), and fraud (+0.05).

We now turn to an examination of what happens when two COAs are paired together in a complaint. When this pairing involves two COAs of the same type, this should make NOS selection easier. And when the pairing involves distinct COAs across different types of issue areas, we would expect the NOS selection process to be much more difficult. Figure 6 depicts the results of this COA pairing exercise for the most common COA pairings within our data. As it reveals, same-type pairings lead to a high degree of NOS selection success. Two civil rights/constitutional law COAs paired together in a complaint lead to the selection of a civil rights NOS code 90% of the time. Two paired IP COAs lead to a property rights (intellectual) NOS selection 96% of the time. And two paired tort COAs lead to a tort NOS selection 91% of the time. Interestingly, two paired contract COAs lead to a contract NOS selection only 35% of the time. More common with these pairings is the selection of a tort NOS code (in 54% of the relevant cases).

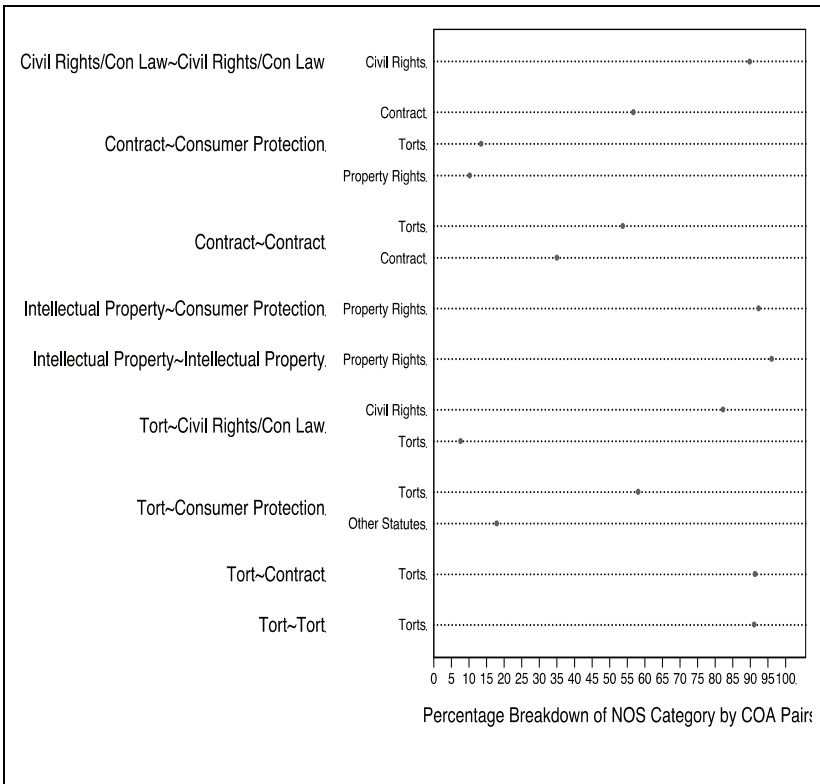


Figure 6: Dotplot of the percentage of observations within our data that lead to a particular NOS selection in which two COAs are paired together. The figure depicts only those COA pairings that compose more than 2% of our observations.

In the presumably more complicated arena where the paired COAs are distinct, the evidence in Figure 6 does seem to confirm that NOS selection varies more. Where contract and consumer protection COAs are paired together in complaints, lawyers vary in the NOS selection between contract (57%), tort (13%), and property rights (intellectual) (10%). With tort and consumer protection COA pairings, lawyers select tort NOS codes 58% of the time and other statutes 18% of the time.

However, with a few common cross-subject area COA pairings in the data, the choice of NOS category seems surprisingly easy given the uniformity in NOS selection. For example, in complaints with a paired tort and civil rights COA, 82% of lawyers select a civil rights NOS code. And in complaints with a tort and contract COA

together, attorneys select a tort NOS over 91% of the time. We see a similar pattern with IP and consumer protection COAs, where lawyers select a property rights (intellectual) NOS over 92% of cases. By contrast, these lawyers choose “other statutes” NOS codes (the more typical consumer protection NOS code) in only 1.2% of these paired cases.

As one final way to examine the NOS code selection process, we look to the first COA in a complaint. It is possible that attorneys plead with a particular COA as their primary COA and others are more “in the alternative.” If this is the case, we would expect the former COA to guide the choice of NOS code. To see whether this is the case within our data, we examine what percentage of the time the selected NOS category matches the complaint’s first listed cause of action. As Table 2 indicates, the first COA is a strong predictor in civil rights and IP NOS selection. Notably, however, these COAs were just as effective in NOS category selection when present at *any* point in a complaint.⁶¹ For other COAs, we see that the first COA listed in the complaint does not do a particularly good job of predicting NOS selection.

First Listed COA in Complaint	% NOS category matches first listed COA	% NOS category matches COA (any order in complaint) ⁶²
Civil Rights/Constitutional	91%	90%
Intellectual Property	93%	94%
Labor	82%	88%
Regulatory	79%	84%
Torts	77%	85%
Contract	51%	82%
Real Property	33%	26%

Table 2: Percent match between first listed COA in a complaint and selected NOS category

61. See *infra* Table 2, col. 3.

62. See *supra* Table 1.

III. SUMMARY AND RECOMMENDATIONS

For some sorts of suits, a coding system based on a lawyer identifying the predominate legal theory may result in an output that is both replicable and reliable. We found that with respect to particular NOS codes, that criterion was met by any reasonable standard. In particular, when attorneys use NOS codes 440 (Other Civil Rights), 442 (Employment), and 791 (ERISA), they signal that the underlying complaints contain such causes of action (and mostly only such actions). A single NOS code is, for such cases, literally good enough for government work.

For other codes and issue areas, NOS codes are a weaker signal. Of course, any normal coding system will result in Type I and Type II errors. We find evidence of both problems here. Type I error, in this context, would be an NOS code signaling the dominance of a legal problem, which the underlying complaints do not reflect. In our data, the clearest example of this error is the real property NOS codes, which predict underlying real property causes of action less than 30% of the time. That sort of mismatch suggests that attorneys simply do not understand the underlying category with sufficient precision to reliably code for its presence. They are, in effect, guessing.

Type II errors (*missing* an issue) are more systematic. For many of the issue areas we have described, an NOS code is at best a very noisy signal. It might reflect, in the case of complaints with multiple causes of action advancing different theories, the first-placed cause of action in the complaint (and that, in turn, likely relates to the cause of action the lawyer feels best about). Thus, causes of actions that are typically pleaded second in a case (as a back-up) will be rarely captured. Many state-based common law theories, attending complaints on the grounds of supplemental jurisdiction, fall into this trap.

Putting aside the ordering concern, NOS categories around contract (particularly 190) and other statutory causes seem to invite attorneys to shoehorn complaints of a variety of types. These NOS categories could not reliably be used to identify their underlying issue compositions. By contrast, though attorneys do mix in a variety of related causes of action to the intellectual property NOS codes, we find that 94% of the time such a code predicts at least one of the underlying intellectual property causes of action. (This is a slightly better mark than Matthew Sag's 80% prediction based on a reverse

engineering of copyright *opinions*.)⁶³ This suggests that the current level of granularity of the codes only sometimes matches reality: The intellectual property NOS codes (by hypothesis) are at the right level of detail, but the contract codes are not.

We think the solution to these problems of error and noise is relatively simple and easy to execute: The Judicial Conference should simply amend the federal Civil Cover Sheet to require filing attorneys to select NOS-type codes for *each cause of action in their complaint* rather than for whole cases. Given the universality of electronic filing, this would not be an excessively burdensome request of lawyers, nor would it create coding problems for the court clerks on the back-end as it might have in the days of paper dockets and manual data entry.

Populating each case for multiple codes solves several problems at once. For scholars, it would enable fine-grained selection based on causes of action, essentially eliminating Type II errors. This would give researchers some confidence that their nets are sieving the right sorts of problems and that their resulting analyses actually reflect how particular substantive issues are treated. It would also—as we suggested in a prior paper—permit researchers to test how litigation acts as a tournament for causes of action.⁶⁴ To the extent that particular causes of action are more likely than others to survive in similar contexts, we might be able to draw conclusions about the functioning of doctrine that our current gross and noisy perspective does not permit.

Cause-of-action-based NOS selection also has the potential to be of great use for the judiciary. As we have explored, the current annual statistical reporting and case weighting protocols rely on NOS codes. Using cause-of-action-based coding might permit the judiciary to instead use a *clusters-based* technique to assign cases. As we have previously explored, cluster analysis “aims to objectively group similar objects based on information found in the data.”⁶⁵ Using the same underlying complaint data as described here, we previously used spectral clustering “to classify and group” cases based on the “similarity of their individual causes of action.”⁶⁶ We found—at least with this set of complaints—eight clusters (or

63. See Sag, *supra* note 49, at 7.

64. See Boyd et al., *supra* note 59, at 272.

65. *Id.* at 261.

66. *Id.* at 262.

discrete groupings) of causes of action identified in the data.⁶⁷ We can roughly describe these groupings as:

- (1) Contract paired with quasi-contract;
- (2) Labor and ERISA;
- (3) Torts, contracts and fraud;
- (4) Securities Law;
- (5) IP and consumer protection;
- (6) Civil rights and state law associated torts;
- (7) Civil forfeiture; and
- (8) Regulatory actions.

As we showed, some of such clusters had more in common with each other than others—the first and third categories, for instance, are more alike to each other than they are to securities cases.⁶⁸ Whether these overlaps, and the underlying clusters, represent the present distribution of complaints is an open question (which could be answered with particularized NOS coding).

Developing a clusters-based approach would permit the administrative office to weigh cases based on their similarity to others filed with the same patterns of causes of action, permitting a more efficient allocation of resources than the noisy single-NOS system currently does. That is, a cluster analysis tells us with more precision what a particular complaint is going to look like (when compared to other similarly situated cases) because we can include significantly more information about the nature of the suit than a single identifier permits. We know vastly more about a case by assigning it to the civil rights cluster than we do by assigning it to either NOS 442 (Employment) or 440 (Other Civil rights), though those NOS codes happen to be very likely to be assigned to that grouping.

The Administrative Office of the U.S. Courts would be able—with a dataset of tens of millions of individual causes of action—to provide more certainty to this sort of analysis and assign a case with some precision to a group of cases with a similar pattern of underlying causes of action. This would allow for differentiation of cases that are entirely within one NOS code and those that cross codes/categories. It would also (potentially) permit a much more

67. *See id.* at 266-67.

68. *Id.* at 266.

efficient weighting algorithm for the assignment of workload across judges than the current gross system permits.

Though the change may seem to require more paperwork, requiring an attorney to select all of the applicable causes of action may, in fact, reduce cognitive effort—she need not *choose* at all which of many potential issues represents her case. Certainly it would eliminate the time and investment associated with strategic gaming. On the back end, a clusters-based analysis would require an initial investment to understand the relevant techniques but would otherwise require no more than a few lines of code in R.

In short, requiring particularized NOS coding is as close as we can imagine to a pareto-superior solution to a problem of real practical import.

APPENDIX A: NOS CODES AND CATEGORIES

CONTRACT	
110	Insurance
120	Marine
130	Miller Act
140	Negotiable Instrument
150	Recovery of Overpayment & Enforcement of Judgment
151	Medicare Act
152	Recovery of Defaulted Student Loans (Excl. Veterans)
153	Recovery of Overpayment of Veterans' Benefits
160	Stockholders' Suits
190	Other Contract
195	Contract Product Liability
196	Franchise
REAL PROPERTY	
210	Land Condemnation
220	Foreclosure
230	Rent Lease & Ejectment
240	Torts to Land
245	Tort Product Liability
290	All Other Real Property
TORTS: Personal Injury	
310	Airplane
315	Airplane Product Liability
320	Assault, Libel & Slander
330	Federal Employers' Liability
340	Marine
345	Marine Product Liability
350	Motor Vehicle
355	Motor Vehicle Product Liability
360	Other Personal Injury
362	Personal Injury - Medical Malpractice
365	Personal Injury - Product Liability

367	Health Care/Pharmaceutical Personal Injury Product Liability
368	Asbestos Personal Injury Product Liability
TORTS: Personal Property	
370	Other Fraud
371	Truth in Lending
380	Other Personal Property Damage
385	Property Damage Product Liability
CIVIL RIGHTS	
440	Other Civil Rights
441	Voting
442	Employment
443	Housing/Accommodations
444	Welfare
445	Americans w/Disabilities - Employment
446	Americans w/Disabilities - Other
PRISONER PETITIONS: Habeas Corpus	
463	Habeas Corpus - Alien Detainee
510	Motions to Vacate Sentence
530	General
535	Death Penalty
PRISONER PETITIONS: Other	
540	Mandamus & Other
550	Civil Rights
555	Prison Condition
560	Civil Detainee - Conditions of Confinement
FORFEITURE/PENALTY	
625	Drug Related Seizure of Property 21 USC 881
690	Other
LABOR	
710	Fair Labor Standards Act
720	Labor/Management Relations
730	Labor/Management Reporting & Disclosure Act
740	Railway Labor Act

790	Other Labor Litigation
791	Employee Retirement Income Security Act
IMMIGRATION	
462	Naturalization Application
465	Other Immigration Actions
BANKRUPTCY	
422	Appeal 28 USC 158
423	Withdrawal 28 USC 157
PROPERTY RIGHTS	
820	Copyrights
830	Patent
840	Trademark
SOCIAL SECURITY	
861	HIA (1395ff)
862	Black Lung (923)
863	DIWC/DIWW (405(g))
864	SSID Title XVI
865	RSI (405(g))
FEDERAL TAX SUITS	
870	Taxes (U.S. Plaintiff or Defendant)
871	IRS-Third Party 26 USC 7609
OTHER STATUTES	
375	False Claims Act
376	Qui Tam (31 USC 3729(a))
400	State Reapportionment
410	Antitrust
430	Banks and Banking
450	Commerce
460	Deportation
470	Racketeer Influenced and Corrupt Organizations
480	Consumer Credit
490	Cable/Sat TV
850	Securities/Commodities/Exchange
890	Other Statutory Actions

891	Agricultural Acts
893	Environmental Matters
895	Freedom of Information Act
896	Arbitration
899	Administrative Procedure Act/Review or Appeal of Agency Decision
950	Constitutionality of State Statutes

APPENDIX B: FULL REGRESSION RESULTS

COA	NOS: Civil Rights/ Con. Law	NOS: Torts	NOS: Contract	NOS: Labor	NOS: Other Statutory	NOS: Real Property	NOS: Property Rights (Intellectual)
Agency	0.362 (0.44)	0.613 (0.49)	1.076** (0.39)	--	-3.942** (1.20)	1.651 (1.14)	0.170 (0.67)
Bad Faith	0.742 (0.70)	--	4.020** (0.98)	0.877 (1.13)	--	--	--
Breach of Fiduciary Duty	--	-3.155** (0.83)	1.979** (0.65)	1.327** (0.63)	1.156* (0.65)	--	-1.751** (0.74)
Civil Rights/ Con. Law	6.256** (0.33)	-4.166** (0.37)	-2.828** (0.47)	- 1.889** (0.43)	-1.313** (0.28)	-1.587 (1.02)	--
Consumer Protection	-0.929* (0.51)	-1.544** (0.25)	-0.298 (0.20)	-0.356 (0.37)	2.247** (0.23)	-0.298 (0.66)	0.829** (0.39)
Contract	-1.008** (0.40)	-0.942** (0.23)	2.787** (0.20)	-0.430 (0.34)	-1.751** (0.27)	0.767 (0.54)	-1.565** (0.49)
Enforcement	-0.486 (1.41)	-1.144 (1.35)	-0.155 (0.34)	2.069** (0.40)	-2.442** (1.07)	0.015 (1.04)	0.276 (0.54)
Equitable Contract	-1.090* (0.62)	-2.197** (0.38)	1.134** (0.23)	0.031 (0.57)	0.403 (0.30)	0.574 (0.70)	-0.017 (0.61)
Fraud	-0.553 (0.45)	1.674** (0.22)	-1.568** (0.19)	-0.839* (0.47)	-1.101** (0.32)	-0.869 (0.72)	-0.790 (0.56)
Intellectual Property	--	-2.772** (0.77)	-1.704** (0.32)	--	-3.311** (0.57)	--	6.681** (0.38)
Labor	-0.455 (0.55)	-1.557** (0.53)	-3.535** (0.47)	5.361** (0.31)	-2.384** (0.65)	--	--
Process	-1.038 (1.60)	-0.523 (0.63)	0.059 (0.59)	-1.222* (0.71)	1.841** (0.70)	--	--
Property	0.952 (0.64)	-2.161** (0.77)	0.948** (0.46)	-0.524 (0.69)	-1.511 (0.96)	3.930** (0.50)	-0.673 (0.83)
Racketeering Criminal	-0.558 (0.54)	-0.134 (0.59)	-0.015 (0.49)	-1.100* (0.66)	1.616** (0.44)	--	0.378 (0.56)
Regulatory	-2.281** (0.46)	-3.426** (0.68)	-0.884** (0.39)	-0.091 (0.62)	3.493** (0.28)	0.750 (0.66)	-1.593 (1.88)
Securities	--	--	-1.079** (0.50)	-0.735 (1.03)	3.989** (0.54)	--	--
Tax	--	--	--	--	2.311** (0.66)	--	--
Tort	-0.138 (0.28)	4.880** (0.33)	-2.250** (0.21)	- 1.279** (0.38)	-1.136** (0.22)	-0.861 (0.63)	-1.826** (0.49)
Constant	-4.045** (0.31)	-3.146** (0.31)	-1.285** (0.19)	- 3.324** (0.28)	-1.813** (0.20)	-4.967** (0.37)	-3.985** (0.29)
Observations	2498	2498	2498	2498	2498	2498	2498

* p<0.10, ** p<0.05