2016

Making Doctrinal Work More Rigorous: Lessons from Systematic Reviews

William Baude

Adam S. Chilton

Anup Malani

Making Doctrinal Work More Rigorous: Lessons from Systematic Reviews

Will Baude,[*] Adam Chilton,[†] and Anup Malani[‡]

July 18, 2016

Legal scholars, lawyers, and judges frequently make positive claims about the state of legal doctrine. Yet despite the profligate citation norms of legal writing, these claims are often supported in a somewhat imprecise way – where the exact evidence is unclear or difficult for others to probe or falsify. In response to similar issues, other disciplines have developed methodological standards for conducting "systematic reviews" that summarizes the state of knowledge on a given subject. In this essay we argue that methods for performing systematic reviews that are specifically tailored to legal analysis should be developed. We purpose a simple four-step process that could be used whenever someone is trying to make objective claims about the state of legal doctrine and illustrate the value of method by applying it to doctrinal claims that have been made in recent legal scholarship.

[*] Neubauer Family Assistant Professor of Law, University of Chicago (baude@uchicago.edu).

[†] Assistant Professor of Law, University of Chicago (adamchilton@uchicago.edu).

[‡] Lee and Brena Freeman Professor of Law, University of Chicago (amalani@uchicago.edu).

We would like to thank participants in the *University of Chicago Law Review's* symposium on "Developing Best Practices for Legal Analysis" for helpful feedback on this paper. We thank Cooper Tang, Kelly Holt, and Rob Golan-Vilella for research assistance.

INTRODUCTION

Legal scholars, advocates, and judges commonly make positive claims about the state of legal doctrine. For example, a legal scholar might claim that there is a trend in recent federal court decisions to allow a particular pretrial procedure or a judge might claim that most courts endorse a given legal proposition. These claims, however, are frequently made without a systematic demonstration of supporting evidence. When this occurs, it not only makes it difficult for the reader to evaluate the validity of the claim, but it also may impede future legal analysis and allow for either conscious or unconscious bias.

In response to analogous concerns about their literature, other disciplines have developed rigorous methods for conducting "systematic reviews." A systematic review is a method of summarizing the results of prior literature on a research question. Typically employed in the medical and psychological sciences, but increasingly being used in the social sciences, systematic reviews have four key lessons for doctrinal work. First, the researcher should state clearly the question she is attempting to answer as this affects the manner in which she conducts her review. Second, the research must justify and be transparent about how she defines and obtains the sample for her review. Third, the research ought to explain any weighting that is applied to the cases in the sample. Fourth, the researcher needs to justify and be transparent about the manner in which she analyzes the sample cases she reviews. Together these steps can prevent bias from case-selection and improve the legitimacy of conclusions drawn from the review.

In this essay we argue that legal scholars, lawyers, and judges should conduct a four step systematic review when they are making positive claims about the state of legal doctrine. In Part I we survey the lack of systematic support for positive claims and explain the benefits of greater methodological rigor. In Part II we discuss systematic reviews in other fields and their applicability to legal analysis. In Part III we propose a four-step process for conducting systematic review of legal doctrine and then demonstrate its use by applying it to recently made doctrinal claims. Finally, we conclude discussing the areas that we believe would benefit most from the application of our proposed method.

I. THE CASE FOR INCREASED RIGOR

We begin by surveying unsystematic claims about the state of legal doctrine, then go on to explain why, even if the claims are true, there are still benefits to more systematic review.

A. Examples

Lawyers regularly make claims about the law, and in particular about case law. Indeed, it might be one of the research tasks that they are most frequently paid to do. And while much legal scholarship is more normative, claims about the law are still prevalent. For example, a civil procedure scholar may argue that a particular rule for class action cases is the increasingly prevailing view in federal courts, or a public law scholar may discuss the administrability problems created by a trend in state constitutional law. Yet those scholars might point to only two or three cases as evidence of the trend, and with no information about the universe from which they were chosen.

These are not just hypothetical examples—both are from recent law review articles. We stress that in each case, the authors may well be right. Indeed, we have no particular reason to doubt that these experts in their field are wrong. And by describing these examples, we do not mean to criticize them for failing to adhere to an existing standard of proof or citation (which is why we do not name them here). In fact, our argument is that these examples are not unique, and instead illustrative of a broad pattern.

To get a better sense of what kind of evidence is provided to establish legal claims, we reviewed every article published in the last completed volume of ten top law reviews.[4] For each article, we had a research assistant read the abstract and record any claim about the state of legal doctrine.[5] The research assistant then read the article and recorded the evidence that was provided as support for those claims. Finally, we coded the support provided for the doctrinal claim into one of three categories: citing to 1 case or less for support; citing to multiple cases as support; or conducting some form of a systematic review (that is, define the entire set of cases that was relevant to the claim and the evidence to support it).

The results of this research are presented in Table 1. Our analysis suggested that roughly 50% (69 of 139) of articles included a claim about the state of legal doctrine in the abstract. Of these 69 articles, only about 20% (14 of 69) provided any form of systematic review to support the doctrinal claim. The rest of the articles provided string cites to cases (and occasionally, academic articles as well), but did not explain how they identified the universe of cases or whether they are representative.

This strikes us as suboptimal. The norms of citation in legal academia ought to be designed to give non-expert readers a chance to test those claims, and a sense of how much confidence those claims deserve. Again, we do not fault anybody for failing to adhere to a norm that does not yet exist. But our suggestion is that it would be good for legal academia to *develop* a standard that helps legal analysts more rigorously see and more persuasively show what the law is.

[4] We set out to analyze the flagship law reviews of the 10 highest ranked schools in the 2015 U.S. News and World Report ranking of law schools. Because the flagship journals of two schools—the *University of Pennsylvania Law Review* and *University of Virginia Law Review*—did not consistently have abstracts for their articles, we skipped these schools in our analysis and moved to the next schools on the list.

[5] We focused on doctrinal claims made in the abstract because our goal was to identify doctrinal claims that were central to the articles argument.

Table 1: Support for Doctrinal Claims in Recent Volume of 10 Major Law Reviews

Journal	Articles in the Volume	Doctrinal Claims in Abstract	Support for the Doctrinal Claim		
			≤ 1 Case	Multiple Cases	Systematic Review
California Law Review (vol. 115)	14	4	1	3	0
Columbia Law Review (vol. 103)	10	5	1	3	1
Duke Law Journal (vol. 65)	16	8	0	6	2
Harvard Law Review (vol. 129)	7	3	0	3	0
Michigan Law Review (vol. 114)	18	12	0	10	2
Northwestern U. Law Review (vol. 110)	11	9	0	6	3
NYU Law Review (vol. 90)	17	10	0	9	1
Stanford Law Review (vol. 67)	13	4	2	1	1
Univ. of Chicago Law Review (vol. 82)	17	7	0	4	3
Yale Law Journal (vol. 124)	16	7	0	6	1
Total	**139**	**69**	**4**	**51**	**14**

Although this analysis focused on legal scholarship, we also see the same problems in more formal academic output, the Restatements of Law published by the American Law Institute. The Restatements have long been an important and widely-cited resource in American law,[6] and a recent volume has been given "the highest praise" for its "clear and careful exposition of the law."[7]

But that very same volume has proven controversial in the courts. In a recent Supreme Court case, the Justices divided over whether to accept a special master's decision that had relied heavily on the Third Restatement of Restitution and Unjust Enrichment.[8] The majority adopted the master's recommendation, repeatedly citing the Restatement,[9] while the dissent complained that the Restatement "lacks support in the law," would "alter the doctrinal landscape of contract law," and had not been relied on by courts.[10]

[6] Caleb Nelson, *The Persistence of General Law*, 106 COLUM. L. REV. 503, 510 n.35 (2006) ("[C]ourts continue to treat the Restatements as presumptively accurate summaries of general American jurisprudence."). *See also* Bennett Boskey, *The American Law Institute A Glimpse at Its Future*, 12 GREEN BAG 2d 255, 258 (2009).

[7] Ben Kremer, *Book Review*, 35 MELB. U. L. REV. 1197, 1215 (2011) (praising Restatement (Third) of Restitution and Unjust Enrichment (2011)); Lionel Smith, *Book Review*, 57 MCGILL L.J. 629, 633 (2012) (same).

[8] Kansas v. Nebraska, 135 S.Ct. 1042 (2015).

[9] *Id.* at 1056-57, 1058.

[10] *Id.* at 1068-69 (Thomas, J., concurring in part and dissenting in part) (internal quotation marks omitted).

Justice Scalia wrote separately to criticize the Restatement even more pointedly. "[M]odern Restatements," he said, "must be used with caution."[11] They "have abandoned the mission of describing the law," and contain provisions "that should be given no weight whatever as to the current state of the law." Hence, he concluded, "it cannot safely be assumed, without further inquiry, that a Restatement provision describes rather than revises current law."[12]

The power of these criticisms is exacerbated by the vague methodological ambitions of the Restatements. The Restatements purport to "reflect the law as it presently stands *or* might appropriately be stated by a court."[13] To the extent this methodology is descriptive, the restatements tell us little about how to test their validity or know when to trust them. But some Restatement reporters are trying to do better: For instance, Oren Bar-Gill, Omri Ben-Shahar and Florencia Murotta-Wurgler are writing a new Restatement of the Law, Consumer Contracts using principles analogous to the ones we discuss here.[14] This is a great step. Our goal is to assist this kind of endeavor and encourage more of them.[15] We hope that these principles of systemic analysis can help.

B. The Value of a More Rigorous Approach

Even if a given claim about legal doctrine is correct, that there are benefits to establishing the claim in a more rigorous way. We will briefly mention five.[16]

First, a more rigorous demonstration of evidence makes it easier for readers to evaluate whether the ultimate claims are true or false. When less comprehensive support is provided, readers instead rely on their outside knowledge or rely on the author's credibility as evidence for the validity of the claim. Expecting readers to rely on these proxies is problematic because not everyone will have the same outside knowledge or view of the author's reputation. Using reputation as a proxy also invites ad hominem attacks on the author's credibility.

Second, a more rigorous demonstration of evidence makes it easier for readers to assess how much uncertainty is associated with a given claim. For example, it may be true that courts

---

[11] Ironically, the one original Restatement that Justice Scalia cited as an example of trustworthy craft – the First Restatement of Conflict of Laws – is one that had been singled out for opprobrium by a recent officer at the American Law Institute. *See* Boskey, *supra* note 6, at 257 ("[T]he judiciary and the bar welcomed the help of most of the Restatement First (possibly excepting the Restatement of the Conflict of Laws, for which the ideologically-imprisoned Professor Joseph H. Beale had been the reporter).").

[12] Kansas v. Nebraska, 135 S.Ct. at 1064 (Scalia, J., concurring in part and dissenting in part).

[13] https://www.ali.org/publications/frequently-asked-questions/#differ (emphasis added).

[14] *See generally* Oren Bar-Gill, Omri Ben-Shahar & Florencia Murotta-Wurgler, *Searching for the Common Law: The Empirical Approach of the Consumer Contracts Restatement*, 84 U. CHI. L. REV. (forthcoming).

[15] *Cf.* Jay M. Feinman, *The Restatement of the Law of Liability Insurance As A Restatement: An Introduction to the Issue*, 68 RUTGERS U.L. REV. 1, 8 (2015) (noting a recent ALI discussion reacting to Justice Scalia's criticisms).

[16] These benefits largely parallel the arguments that have been used to motivate the transparency and replication movement that has been taking place in the social sciences. *See generally* Lee Epstein & Gary King, *The Rules of Inference*, 69 U. CHI. L. REV. 1, 38-54 (2002).

generally agree on a point of law, but valuable to know how many cases have disagreed. Similarly, it is valuable to know whether a trend has been shown only in certain courts, or in certain years. This evidence helps a reader understand the degree of uncertainty associated with a claim and also know the scope conditions of when that claim is valid.

Third, providing more complete support for claims can reduce error. Even authors who are fairly confident in their knowledge make mistakes. When authors undertake to demonstrate their work, they will be less likely to make a mistakenly false statement. This logic has been one of the reasons that quantitative researchers are increasingly expected to provide their data and code. Simply put, the original researcher will be more careful when they know it will be easier for future researchers to double check their work.

Fourth, more complete documentation of support increases general progress in the field. Both common law legal reasoning and research are social enterprises in that they build on the work from the past. When authors do not document the support for their claim, however, people trying to answer the same question in the future have to recreate their work. Because research is a social enterprise, research norms should support this kind of documentation, just as journals and funding agencies increasingly require empirical researchers to publish their data.

Fifth, providing such demonstrations can help to reduce actual or perceived bias. A large body of scholarship has studied the role that political ideology has on legal decision making. This literature has consistently found that the political views of judges predict their decisions,[17] and more recently has even found the political views of law professors predict the conclusions they reach in their scholarship.[18] One way to help reduce the risk or perception of bias is to provide the evidence that the claim is based on.

II. SYSTEMATIC REVIEWS

In this Part, we discuss the history and justifications for systematic review, explain the steps of systematic reviews, and discuss why or why not it might be an appropriate model for doctrinal work. The last step is the most critical, as systematic review is not a perfect fit for doctrinal work and so only steps that are profitably imported into analysis of case law should be recommended.

A. History and Justification

The sciences, especially the biological and psychological sciences, have long recognized the need for a methodology to synthesize the results of prior research on a scientific question. An individual study may have a limited sample and thus statistical power to answer a research question. Moreover, its specific conclusions may be bound by the specific circumstances in which it was conducted. By contrast, a review could aggregate the data and contexts from

[17] *See, e.g.*, Cass R. Sunstein & Thomas J. Miles, *Depoliticizing Administrative Law*, 58 DUKE L.J. 2193, 2197 (2009).

[18] Adam Chilton & Eric Posner, *An Empirical Study of Political Bias in Legal Scholarship*, 44 J. LEGAL. STUD. 277 (2015).

multiple studies to yield both a more precise and generalizable study.[19] The intellectual challenge of finding a method to combine results from multiple studies has long attracted the attention of leading statisticians, including Karl Pearson and Sir Ronald Fisher in the early twentieth century.[20] A famous early example is Pearson's effort to synthesize a number of studies that examined the value of enteric fever inoculation in 1904.[21]

Demand for a method for synthesizing studies was initially limited, however, because there were simply too few medical studies conducted to be synthesized and because medical practice was informal and decentralized. As reliable research designs developed—especially the randomized controlled trial—and computing power increased, more and more primary research was conducted.[22] Moreover, in the 1970s, a movement emerged that argued that medical practice should be driven by research evidence and not physicians' idiosyncratic personal experiences or hunches.[23]

One of the principle products of the evidence-based medicine movement is the Cochrane Collaboration, which promotes the development of a rigorous methodology for synthesis, also known as "systematic reviews," and hosts an online database of reviews of prior research.[24] The need to define best practices for systematic reviews is now embraced widely in the medical literature, which has generated consensus statements on how such reviews ought to be conducted.[25]

The primary alternative methodology to the systematic review is the narrative review. A narrative review is mainly qualitative, critical examination of the prior literature on a subject. The main criticism of this methodology—and thus the justification for systematic reviews—is that the authors have discretion to select which medical studies they review and how they interpret the studies they select. This discretion can lead to confirmation bias—authors select articles that

---

[19] *See* Amit X. Garg, et al., *Systematic Review and Meta-analysis: When One Study Is Just not Enough*, 3 CLINICAL JOURNAL OF THE AMERICAN SOCIETY OF NEPHROLOGY 253, 254 (2008); J.C. Valentine, T.D. Pigott, & T. Lau, *Systematic reviewing and meta-analysis*, in J. Wright, e.d., INTERNATIONAL ENCYCLOPEDIA OF THE SOCIAL AND BEHAVIORAL SCIENCES, 2ND ED, 906 (2015).

[20] *See* RJS Simpson & Karl Pearson, *Report on certain enteric fever inoculation statistics*, 3 BRITISH MEDICAL JOURNAL 1243 (1904) (an early effort to combine results from different sources); Ronald Aylmer Fisher, STATISTICAL METHODS FOR RESEARCH WORKERS 99 (1925) ("…it sometimes happens that although few or [no statistical tests] can be claimed individually as significant, yet the aggregate gives an impression that the probabilities are lower than would have been obtained by chance").

[21] Pearson, *supra* note 20, at 1235.

[22] Valentine et al., *supra* note 19, at 908.

[23] Archibald Cochrane, EFFECTIVENESS AND EFFICIENCY: RANDOM REFLECTIONS ON HEALTH SERVICES (1972).

[24] *See* The Cochrane Collaboration, http://www.cochrane.org (accessed April 17, 2016).

[25] *See, e.g.*, David Moher, et al., *Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement*, 354 THE LANCET (1999); D. F. Stroup, et al., *Meta-analysis of observational studies in epidemiology: A proposal for reporting*, 283 JAMA (2000) (for observational studies); David Moher, et al., *Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement*, 151 ANNALS OF INTERNAL MEDICINE 1299, 1300 (2009).

tend to reinforce the author's priors.[26] Moreover, the narrative review does little to address the problem of publication bias, which is the tendency for papers with less interesting results—usually results showing no effect, also known as null results—not to be published.[27] This omission leads to overestimates of correlations, which often means the reviews will conclude treatments have effects even when they actually may not.[28]

B. Steps in a Systematic Review

Systematic reviews address these biases with four basic steps. First, a review's author clearly defines the question she seeks to answer.[29] For example, what is the value of bariatric surgery for reducing obesity?[30] This helps ensure that the author stays on target when searching for relevant literature. Although it may be too obvious to need stating, a major cause of bias is an author answering a different question than the one that motivated a review.[31]

Second, the author conducts an exhaustive search for relevant studies. In order for readers to judge how well the search was done, the author should be explicit about the databases searched, the search terms used, and any inclusion or exclusion criteria applied.[32] The latter are criteria that determine whether a study falls within the ambit of their search or is to be dropped because it does not.[33] Disclosures about the search methods also allow the reader to judge the potential for bias in the review[34] and the development of "best practices" for search. The

---

[26] Julia H. Littell, *Evidence-based or biased? The quality of published reviews of evidence-based practices*, 30 CHILDREN AND YOUTH SERVICES REVIEW 1299, 1300 (2008); Garg et al., *supra* note 19, at 253.

[27] *See, e.g.*, Phillipa J. Easterbrook, et al., *Publication bias in clinical research*, 337 THE LANCET (1991); Jerome M Stern & R John Simes, *Publication bias: evidence of delayed publication in a cohort study of clinical research projects*, 315 BMJ (1997).

[28] LM Schmidt & PC Gotzsche, *Of mites and men: reference bias in narrative review articles: a systematic review*, 54 THE JOURNAL OF FAMILY PRACTICE (2005) (finding narrative reviews are overly positive in their assessments of treatments relative to systematic reviews and clinical trials).

[29] *See, e.g.*, Khalid S. Khan, et al., *Five steps to conducting a systematic review*, 96 JOURNAL OF THE ROYAL SOCIETY OF MEDICINE 118, at 118, 119 (2003); Denise O'Connor, Sally Green and Julian PT Higgins, *Defining the review question and developing criteria for including studies*, in Julian PT Higgins and Sally Green, eds. COCHRANE HANDBOOK FOR SYSTEMATIC REVIEWS OF INTERVENTIONS: THE COCHRANE BOOK SERIES 83 (The Cochrane Collaboration 2008); Valentine et al., *supra* note 19, at 910.

[30] H. Buchwald, et al., *Bariatric surgery: A systematic review and meta-analysis*, 292 JAMA 1724 (2004).

[31] Mark Crowther, et al., *Systematic review and meta-analysis methodology*, 116 BLOOD 3140, 3141 (2010).

[32] *See, e.g.*, Khan et al., *supra* note 29, at 118, 119-20; Carol Lefebvre, Eric Manheimer and Julie Glanville on behalf of the Cochrane Information Retrieval Methods Group, Searching for studies, in Higgins & Green, *supra* note 29, at 95; Julian PT Higgins and Jonathan J Deeks, *Selecting studies and collecting data*, in Higgins & Green, *supra* note 29, at 151.

[33] H. C. Van Spall, et al., *Eligibility criteria of randomized controlled trials published in high-impact general medical journals: A systematic sampling review*, 297 JAMA (2007).

[34] Garg et al., *supra* note 19, at 253.

literature search step is crucial since an importance source of confirmation bias is the omission of relevant studies that may agree with the authors' prior beliefs about the correct answer to their research question.[35]

Third, the author appraises the quality of the studies that she has gathered.[36] This is different than exclusion criteria, which are typically based on explicit criteria like whether the studies look at the right treatment, the target patient population, the intended outcome, etc. The quality appraisal looks instead at things like the methodology employed in the study (e.g., was an observational study or a randomized controlled trial[37] or was double, single or not blinded[38]). This step is used to increase the weight of methodologically sound studies in the author's subsequent synthesis of the evidence across studies.

Finally, the author synthesized the results of the different studies that survive. The author should be explicit about the methodology she uses to synthesize the study.[39] For example, she may use a voting method in which she simply counts the number of studies that find positive impacts of a treatment and that do not and reports what the majority of studies find, perhaps with different votes for different classes of studies where classes are defined by the quality of the study.[40] She may be even more rigorous and extract the statistical results from each and combine them using meta-analysis, a quantitative methodology for combining summary statistics or even the data from multiple studies.[41] The author should also be explicit about how

---

[35] *See* Littell, *supra* note 26, at 1300. Garg et al., *supra* note 19, at 256-57, also argue that more comprehensive searches also reduce the risk of publication bias.

[36] *See, e.g.*, Khan et al., *supra* note 29, at 118, 120; Julian PT Higgins and Douglas G Altman on behalf of the Cochrane Statistical Methods Group and the Cochrane Bias Methods Group, *Assessing risk of bias in included studies*, in Higgins & Green, *supra* note 29, at 187.

[37] An observational study looks retrospectively at outcomes from treatments that patients chose while a randomized controlled trial randomly assigns patients to treatment to address selection bias. MIQUEL PORTA, et al., A DICTIONARY OF EPIDEMIOLOGY 203 (Oxford University Press, USA. 2014).

[38] A single blind of the study subject prevents the subject from changing his behavior in response to the treatment, including dropping out of the study. Such behavior introduces selection effects either due to unobservable behavior while on treatment or unravels the benefit of random assignment. A single blind of the investigator prevents the investigator from seeing what treatment the patient received in order to limit the measurement error wherein the investigator measurement of (especially subjective) outcomes reflects her priors about the value of a treatment. A double blind study blinds both the subject and the investigator. See Porta et al., *supra* note 37, at 27.

[39] *See, e.g.*, Khan et al., *supra* note 29, at 118, 121; Carol Lefebvre, Eric Manheimer and Julie Glanville on behalf of the Cochrane Information Retrieval Methods Group, Searching for studies, in Julian PT Higgins and Sally Green, eds., COCHRANE HANDBOOK FOR SYSTEMATIC REVIEWS OF INTERVENTIONS: THE COCHRANE BOOK SERIES at 95 (The Cochrane Collaboration 2008).

[40] *See* Valentine et al., *supra* note 19, at 909.

[41] The term meta-analysis was coined by G.V. Glass and M.L. Smith, *Meta-analysis of research on the relationship of class-size and achievement*, 1 EDUCATIONAL EVALUATION AND POLICY ANALYSIS 2 (1979). Details of how to conduct meta-analyses may be found in, e.g., Jonathan J Deeks, Julian PT Higgins and Douglas G Altman on behalf of the Cochrane Statistical Methods Group, Analysing data and

she thinks publication bias may affect the conclusions she is able to draw. Obviously, the better the method of synthesis the author employs, the better the review. However, being explicit about the method is almost as important as the method itself because it allows others to replicate the review author's work, ensuring that the review was not manipulated and increasing confidence in the review's conclusions.[42]

C. An Appropriate Model for Doctrinal Work?

Although much of the impetus for development of a methodology for systematic reviews comes from the biological and psychological sciences, it would seem to be of value to any field wherein there is a need for synthesizing the results from multiple inquiries into the same issue. One of the early converts to systematic reviews was the public policy literature, which set up the Campbell Collaboration to support and disseminate such reviews of policy interventions, especially in the fields of education, crime and justice, social welfare and international development.[43] Efforts have also been made to import the methodology to management science[44] and even software engineering.[45]

It would seem that legal research, especially doctrinal work, would be a natural candidate for application of systematic reviews. As noted above, many scholars make descriptive claims about the law, and that work may be vulnerable to conscious or unconscious bias because the author neglects cases that do not fit.[46] Readers of doctrinal work cannot assess any bias from this case selection process, and can compound the problem by citing uncritically the conclusions of the doctrinal analysis in their own legal analysis.

The mere need to synthesize prior work, however, is not sufficient for justifying the wholesale importation of the methodology of systematic reviews. There are important difference between the medical sciences, for which the approach was developed, and doctrinal analysis. First, medical studies are quantitative while legal cases are qualitative. It is more difficult to aggregate or combine qualitative research. Second, medical studies have positive aims—figure out whether a treatment works or not—while legal analysis often embeds normative aims (e.g. like arguing that one rule is better than another).[47]

---

undertaking meta-analyses, in Julian PT Higgins and Sally Green, eds., COCHRANE HANDBOOK FOR SYSTEMATIC REVIEWS OF INTERVENTIONS: THE COCHRANE BOOK SERIES at 243 (The Cochrane Collaboration 2008).

[42] *See* Valentine et al., *supra* note 19, at 911.

[43] *See* Campbell Collaboration, *About us*, http://www.campbellcollaboration.org/about_us/index.php (accessed on April 17, 2016).

[44] David Tranfield, et al., *Towards a methodology for developing evidence-informed management knowledge by means of systematic review*, 14 BRITISH JOURNAL OF MANAGEMENT 207 (2003).

[45] Jorge Biolchini, et al., *Systematic review in software engineering*, 679 SYSTEM ENGINEERING AND COMPUTER SCIENCE DEPARTMENT COPPE/UFRI, TECHNICAL REPORT ES (2005).

[46] *See supra* Part I.B.

[47] Similar arguments were made against the importation of systematic reviews into management science. *See* Tranfield, *supra* note 44, at 212.

These differences justify caution when translating elements of systematic reviews to doctrinal work, but do not necessarily justify ignoring entirely the lessons of the methodology. The fact that prior cases are qualitative does not at present prevent lawyers and legal academics from drawing conclusions from prior cases about what courts are likely to do in future cases. The lesson we should learn from systematic reviews is that even when conducting qualitative synthesis, an author should be clear about which cases made her sample. This will reduce the risk that the author draws incorrect conclusions because her qualitative synthesis ignored certain relevant cases, and allow future researchers to know where to expand on or replicate the author's claims. She should also be clear about the sorts of logical steps she took when conducting her qualitative synthesis (e.g., what cases she valued more because of the judge or because the context was more generalizable).

Likewise, the fact that legal work is often normative is not an argument against greater rigor during case selection and transparency about the nature of legal analysis. Indeed it is the opposite. Systematic review does not deny the need for normative work but rather clarifies the division between positive and normative by preventing an author from hiding normative analysis in work she is advertising as positive analysis. When an author wants to be normative, and this is appropriate, e.g., when she is writing a legal brief for a client, she should disclaim systematic review. When she seeks to be positive, she should embrace certain steps of systematic to lend her work legitimacy. Much scholarship and judicial writing aims to do both at the same time, and the steps of systemic review make it easier to tell which one the author is doing.

III. DEVELOPING A METHOD OF SYSTEMATIC REVIEW FOR LEGAL ANALYSIS

In this Part we first outline a process for how to conduct a systematic review of legal doctrine, and then provide an example of this process for a recent piece of legal scholarship.

A. A Four-Step Process for Conducting a Systematic Review of Legal Doctrine

We propose a four-step process for making claims about the state of legal doctrine: (1) clearly stating the legal question that is being answered; (2) defining the sample of cases that will be used; (3) explaining how the cases in the sample will be weighted; (4) conducting the analysis of the sample of cases and stating the conclusion. We briefly explain each of these four steps below.

1. Stating the Question

The first step in providing the evidence for a legal claim is defining the exact question that the subsequent analysis is trying to answer. There are two things to keep in mind at this stage.

First, the question should be precise. The idea of stating a legal question will obviously be familiar to anyone in the legal profession. Legal questions are asked during Socratic cold calls during law school, used to motivate legal memos, and guide many forms of legal briefs. These questions, however, are often asked in a fairly broad manner. The key when asking a legal question to motivate a systematic review of legal doctrine is to make sure to state a question that

is sufficiently precise as to guide the time frame, jurisdictions, and relevant universe of cases that will be used to answer the question.

Second, it is helpful to think about what evidence is required to establish a given claim. For example, if the question is how courts "typically" decide a particular type of case, answering the question requires knowing, say, the median way that courts have decided the case. Once again, knowing what evidence is required for the question helps to determine exactly what sample of cases are relevant and how to analyze them. Below we provide examples of common kinds of claims and the evidence they require.

- *Courts generally decide issue X in way Y*. This kind of claim can be thought of as calling for the median outcome, or "majority rule" for a given kind of cases. To establish this kind of claim it is both necessary to establish the universe of relevant cases and to classify the outcomes of those cases in some way.
- *Courts have increasingly decided issue X in way Y*. This kind of claim can be thought of as calling for the correlation of outcomes over time. To establish this kind of claim it is necessary to establish the universe of relevant cases, to classify the outcomes of those cases, and to make note of when those cases occurred.
- *There is a split in how courts decide issue X*. This kind of claim can be thought of as making a claim about the variance of outcomes. Depending on the scope of this claim may be necessary to establish the universe of relevant cases and to classify the outcomes of those cases.
- *Courts have frequently confronted issue X*. This can be thought of as a claim about the size of a given sample. Making this claim thus requires documenting the number of cases that meet the relevant criteria.
- *At least one court has decided issue X in way Y*. This can be thought of as a claim about the existence of a given phenomenon. To establish this claim, it is not necessary to establish the universe of cases. Instead it is simply necessary to find one case that meets a given criteria.

2. Defining the Sample of Cases

After a question has been clearly stated, the next step is to define the relevant sample of cases that were analyzed. There are also two major steps to this process.

First, it is important to establish what process was used to assemble the universe of cases. For example, one might say what courts one searched for cases from, and over what time period. This way it is possible for anyone else to understand exactly the universe of cases that was analyzed as support for a given doctrinal claim.

Second, it is important to state any inclusion or exclusion criteria that were applied to a sample of cases. For example, if the universe includes a large number of cases, it is important to say which cases were analyzed. In some cases, the entire sample of cases may be analyzed, but in others it might be a random sample. Alternatively, it may be the case that certain kinds of cases are excluded from the analysis because they are not relevant (e.g., all potentially express preemption cases in an inquiry into field preemption). All of these decisions should be clearly documented.

Finally, in an ideal world (or if a process like ours begins to become more commonplace) one might also hope that analysts would specifically document the *technology* of their search process. For instance, they might say what databases they searched, using what terms, on what dates. This is considered an important step of systematic reviews in the medical literature. But we suspect that there may be more reluctance and resistance to translating it into legal scholarship. This is likely partly for reasons of style and etiquette, but also because the legal research process is more heterogeneous than the research processes in other disciplines. Although it would be beneficial if scholars documented this part of their process as well, it is not as important as clear definition of the universe of cases.

3. Explaining the Weighting

Once the sample of cases is established, it is important to state how the cases in the sample will be weighted in the analysis. Just as it may not be appropriate to give all clinical studies equal weight during a systematic review of the medical literature on a given subject if the quality of the studies differs, it may not be appropriate to give all cases the same weight. For instance, it may be appropriate to weight cases differently based on whether they are: of different precedential status; more recent; cited more frequently or written by more frequently cited judges; or engage in more analysis on the relevant topic. Once again, the key is transparency. Legal analysis need not be the simple sum of equally weighted cases, but the weighting should be explained to readers.

4. Conducting the Analysis and Stating the Conclusion

The final step is analyzing the sample and answering the question posed. There are three pieces of information that should be provided about this process.

First, one should provide the criteria that were used to analyze the cases. This may be as simple saying "I counted any cases that mentioned issue X as relevant" or "I only counted cases as relevant if the central issue of the case was X."

Second, one should say how the cases were analyzed. For example, one approach may be to conduct a key word search over a set of cases, but another would be to carefully read all of the relevant cases.

Third, a conclusion should be stated that is not broader than what the evidence can support. For example, if the only federal district court opinions from 2010 to 2015 were analyzed during this process, the conclusion that follows is that "district court decisions between 2010 to 2015 handle issue X in way Y" and not "courts handle issue X in way Y". To be sure, scholarship frequently asks readers to make inferences from one set of data points to a broader one—the fact that a certain set of decisions handle issue X in way Y may be argued to imply that other courts do so as well. But once again, a clear analysis should make clear what claim is being made about the cases and what the requested inference is.

B. A Sample Review

We hope that this four-step process can serve as a relatively simple way to advance the rigor—and hence the credibility and transparency—of doctrinal analysis. In their own work, Oren Bar-Gill, Omri Ben-Shahar and Florencia Murotta-Wurgler are using a systematic review to

write a Restatement,[48] and we applaud the effort. We think similar methods can add to the value of legal scholarship, and will try to demonstrate with a concrete example.

One of us (Baude) previously published an article that investigated whether "originalism" is "our law," in part through a synthesis of Supreme Court opinions.[49] We think that the persuasiveness of that analysis might have been helped by the principles of systemic review. And so in the course of writing this article we decided to conduct a systemic review relevant to some of the claims in that article. Below, we describe the steps of that review and its results.

1. Stating the Question

One of the claims in the article was that the Supreme Court's cases, with no exceptions or relatively few exceptions, were consistent with what Baude described as "inclusive originalism."[50] More specifically, it claimed: "First, in cases where the Court acknowledges a conflict between original meaning or textual meaning and another source of constitutional meaning, the text and original meaning prevail. Second, across the larger run of cases that do not feature an explicit clash of methodologies, the Court never contradicts originalism."[51]

To check this claim more systematically, we examined a set of 280 Supreme Court cases[52] with the help of a research assistant, and answered the following questions for each case:

- Did the case decide a constitutional question?
- If so, did the Court either reject the original meaning or say that the original meaning would not matter to its analysis?

2. Defining the Sample

The previous article attempted to focus on Supreme Court cases that reflect our current positive law commitments, which include both modern cases and older decisions that continue to be recognized as "canonical."[53] For purposes of our review, we focused on a subset of these cases and used a media salience metric called the "NYT Measure": whether a case was listed on the front page of the New York Times.[54] We defined the sample to include all 280 cases decided between 1989 and 2009 (the most recent period available). We then excluded the 84 cases that did not decide a constitutional issue.

---

[48] *Supra* note 14.

[49] William Baude, *Is Originalism Our Law?*, 115 COLUM. L. REV. 2349, 2370-2386 (2015).

[50] *Id.* at 2391.

[51] *Id.* at 2371. This was not the only claim in the article, but it is the one most immediately susceptible to systematic review.

[52] *See infra* Part III.B.2 for how we defined that sample.

[53] Baude, *supra* note 57, at 2371, 2391.

[54] The metric was developed by Lee Epstein and Jeffrey Segal. *See* Lee Epstein & Jeffrey Segal, *Measuring Issue Salience*, 44 AM. J. POL. SCI. 66, 73-81 (2000). The cases are available at http://epstein.wustl.edu/research/salience.xls.

This is of course an incomplete sample, and we note that several important cases discussed in the article[55] did not appear in the dataset. But the metric is a "valid, reliable, and unbiased measure of salience,"[56] and therefore useful for a systematic review of salient cases.

3. Explaining the Weighting

Our narrow definition of the sample to focus only on salient cases means that almost all cases that discussed the original meaning of a constitutional provision could get equal weight. However, depending on the specific question, it could be appropriate to give older cases or cases whose reasoning is partly repudiated or contested less weight in the final analysis.

4. Conducting the Analysis

The results of our analysis are presented in Table 2. Of the 196 constitutional law cases in our sample, our systematic review revealed only one in which the Court seemed to say that the original meaning of the constitutional provision (known or not) did not matter: *Lawrence v. Texas*. It is worth noting that this case was discussed at length in the original article.[57]

Our review also uncovered eight other borderline cases: *Allegheny County v. Pittsburgh*[58] (since partly overruled by a 2014 decision[59]); *Planned Parenthood v. Casey* (also discussed at length in the article);[60] *BMW v. Gore*;[61] *Kelo v. City of New London*;[62] and a string of Eighth Amendment cases involving "evolving standards of decency."[63] Each of these cases probably does not reject inclusive originalism,[64] but presented a sufficiently close call that our review flagged them as unclear. This demonstrates an additional useful function of the review – identifying cases that might deserve further explanation—in addition to demonstrating one of the article's claims in a more systematic way.

[55] *See, e.g.*, Baude, *supra* note 57, at 2376 (discussing Crawford v. Washington, 541 U.S. 36, 60 (2004)).

[56] Epstein & Segal, *supra* note 54, at 72.

[57] Baude, *supra* note 57, at 2381-2382.

[58] 492 U.S. 573 (1989).

[59] Town of Greece, N.Y. v. Galloway, 134 S. Ct. 1811 (2014).

[60] Baude, *supra* note 57, at 2384.

[61] 517 U.S. 559 (1996).

[62] 545 U.S. 469 (2005).

[63] Hudson v. McMillian, 503 U.S. 1 (1992); Atkins v. Virginia, 536 U.S. 304 (2002); Roper v. Simmons, 543 U.S. 551 (2005); Kennedy v. Louisiana, 554 U.S. 407, as modified 554 U.S. 945 (2008).

[64] *See, e.g.*, Baude, *supra* note 57, at 2356 n.24 (discussing Eighth Amendment).

Table 2: Systematic Review of Originalism in Salient Supreme Court Cases

Period	NYT Cases	Decided Constitutional Question	Rejects Original Meaning	Borderline Case
1989-1991	73	54	0	1
1992-1994	46	28	0	2
1995-1997	37	28	0	1
1998-2000	38	25	0	0
2001-2003	38	31	1	1
2004-2006	26	18	0	2
2007-2009	22	12	0	1
Total	**280**	**196**	**1**	**8**

CONCLUSION

Although we believe that legal analysis could be improved if methodological standards for analyzing case law were developed, we acknowledge that our process has drawbacks. Most notably, documenting the steps we describe can consume time and space that could be spent on other things. Nor is systematic review appropriate for advocates making normative or proscriptive claims about what legal doctrine should be.

But we hope to convince others of the benefits of this framework when making positive claims about legal doctrine a central part of the analysis in law reviews, restatements, and judicial opinions. Law reviews articles provide research for lawyers, judges, and policymakers to rely on. They would be more useful—and perhaps more likely to be cited—if they provided all the evidence necessary to support their central claims. Systematic reviews could help the reporters of restatement alleviate the concern that they color their analysis to reach their desired conclusions. Systematic reviews could help courts by lending credibility and reducing any perception of bias about their decisions.

Even if many authors are reluctant to adopt these techniques directly, we believe their insights can be useful in other ways as well. For instance, for claims that are not central to an analysis, it still may be best to cite to secondary sources that *did* conduct a systematic review. This is because these sources would provide better evidence than articles that may have made the same claim while simply citing to other articles or legal materials. And when one is skeptically questioning a doctrinal claim that does not document its methodology, our framework may provide a useful point of departure—it can help critics and skeptics zero in on which part of an argument most needs to be supported and proven.

Finally, we emphasize that we recognize that there are many different ways to incorporate some of the insights of systematic reviews. We do not intend this essay to be the final statement on the matter, but instead hope to generate debate on how more rigorous methods can be incorporated into traditional legal analysis.