2018

Evaluating Shepard's, KeyCite, and Bcite for Case Validation Accuracy

Paul Hellyer
College of William & Mary, phellyer@wm.edu

Evaluating Shepard's, KeyCite, and BCite
for Case Validation Accuracy[*]

Paul Hellyer[**]

This study evaluates and compares how accurately three legal citators (Shepard's, KeyCite, and BCite) identify negative treatment of case law, based on a review of 357 citing relationships that at least one citator labeled as negative. In this sample, Shepard's and KeyCite missed or mislabeled about one-third of negative citing relationships, while BCite missed or mislabeled over two-thirds. The citators' relative performance is less clear when examining the most serious citator errors, examples of which can be found in all three citators.

Table of Contents

Introduction

¶1 Lexis, Westlaw, and Bloomberg Law are among the most trusted legal databases,[1] and their citators are one of their most crucial features: Shepard's on Lexis Advance, KeyCite on Westlaw, and BCite on Bloomberg Law. Legal citators have a critical role in validating case law by identifying negative treatment from other cases. Most importantly, citators should indicate whether a case has been cited and overruled by a higher court. Without a reliable citator, lawyers and judges

---

1. *See, e.g.*, Kendall F. Svengalis, Legal Information Buyer's Guide & Reference Manual 157 (2017 ed.); Olufunmilayo B. Arewa, *Open Access in a Closed Universe: Lexis, Westlaw, Law Schools, and the Legal Information Market*, 10 Lewis & Clark L. Rev. 797, 799, 829 (2006); William R. Mills, *The Decline and Fall of the Dominant Paradigm: Trustworthiness of Case Reports in the Digital Age*, 53 N.Y.L.S. L. Rev. 917, 927–28 (2008–09).

might rely on invalid case law.[2] Pronouncements from citators can even sway courts and influence the development of law.[3] In short, reliable citators are essential to the legal profession.

¶2 In their advertisements, Lexis, Westlaw, and Bloomberg Law boast about the reliability and convenience of their citators. Lexis claims to follow "rigorous quality controls," and asserts that Shepard's is "the best way to be confident your case is built on the best authority."[4] Westlaw calls KeyCite "[t]he industry's most accurate, up-to-the-minute citation service," and invites users to "[u]se it to instantly verify whether a case, statute, regulation, or administrative decision is good law."[5] Advertising for BCite is less exuberant but confident, with statements such as "*Bloomberg Law* makes it easy to verify that your cases remain good law with . . . BCite."[6]

¶3 This study evaluates how accurately Shepard's, KeyCite, and BCite identify negative treatment of case law. The results will likely disappoint anyone who relies on these citators. I found highly inconsistent results and egregious mistakes. In this study, BCite's statistical performance is the lowest by a wide margin, but the citators' relative performance is less clear when specific citations are examined qualitatively. The results for all three citators are troubling. I begin this article with a review and analysis of previous comparison studies, followed by a discussion of my methodology, statistical results, and a discussion of specific mistakes. As far as I am aware, this is the largest statistical comparison study of citator performance for case validation, the first statistical comparison study involving BCite, and the first time both KeyCite and Shepard's have been statistically compared with a third citator.

Previous Citator Comparison Studies

¶4 Shepard's has a long history dating back to 1873, and has been available online since the early 1980s when both Westlaw and Lexis offered it.[7] In the 1990s, Lexis gained control of Shepard's; in response, Westlaw developed KeyCite.[8] In the years following KeyCite's 1997 debut, several law librarians conducted and published comparison tests of the two competing citators.

¶5 Fred Shapiro published the first major comparison study in the April 1998 issue of *Legal Information Alert*.[9] Shapiro gathered a random sample of 421 federal and state cases decided in January 1996, and compared the KeyCite and Shepard's

---

2. On the importance of citators, see, for example, *Gosnell v. Rentokil, Inc.*, 175 F.R.D. 508, 510 n.1 (N.D. Ill. 1997) ("It is really inexcusable for any lawyer to fail, as a matter of routine, to Shepardize all cited cases."); Judith D. Fischer, *Bareheaded and Barefaced Counsel: Courts React to Unprofessionalism in Lawyers' Papers*, 31 Suffolk U. L. Rev. 1, 9–11 (1997); Ellie Margolis, *Surfin' Safari—Why Competent Lawyers Should Research on the Web*, 10 Yale J.L. & Tech. 82, 92 (2007).

3. Joel Heller, *Subsequent History Omitted*, 5 Calif. L. Rev. Circuit 375, 378 (2014).

4. *Shepard's Citation Service*, LexisNexis, https://www.lexisnexis.com/en-us/products/lexis-advance/shepards.page [https://perma.cc/36WP-8NWR].

5. *KeyCite on Westlaw Next*, Thomson Reuters (2012), http://info.legalsolutions.thomson-reuters.com/pdf/wln2/l-356347.pdf [https://perma.cc/Z4ZH-HY9Y].

6. *Litigation on Bloomberg Law*, Bloomberg BNA, https://www.bna.com/litigation-m57982084299/ [https://perma.cc/AY4R-9XQX].

7. Laura C. Dabney, *Citators: Past, Present, and Future*, 27 Legal Reference Servs. Q. 165, 166, 169 (2008).

8. *Id.* at 171–75.

9. Fred R. Shapiro, *KeyCite and Shepard's—Coverage and Currency of Citations to Recent Cases: A Comparative Study*, Legal Info. Alert, Apr. 1998, at 1.

report for each case. He found that KeyCite returned over 50% more citing references than Shepard's, mainly due to better coverage of citations from unreported cases and law review articles.[10] He also found that KeyCite had an edge in currency.[11] Shapiro did not evaluate the accuracy of either citator. *Legal Information Alert* published a response from Lexis claiming that Shapiro did not use comparable data in his citation count comparison. Lexis claimed that Shepard's listed only citing references that had been analyzed by its editors, whereas most of KeyCite's citing references were merely computer-generated, with no analysis. Lexis also claimed that KeyCite's apparent edge in currency was due to KeyCite's practice of listing citing references before they had been analyzed.[12]

¶6 In 1999, *Legal Reference Services Quarterly* published a second comparison study.[13] Elizabeth McKenzie discussed the competing claims and counterclaims made by Lexis and Westlaw about their respective citators, and then presented the results of her own study. McKenzie selected a dozen federal and state cases, and compared the Shepard's and KeyCite reports for each. Like Shapiro, McKenzie found that KeyCite retrieved more citing references from unreported cases and law review articles, and she also found that KeyCite added more headnote numbers to the citing references.[14] She praised KeyCite's innovative "Table of Authorities" and depth of treatment features, and predicted that Shepard's would have to improve in response to KeyCite.[15]

¶7 In 2000, William L. Taylor published a comparison study in *Law Library Journal*,[16] which appears to be the most significant statistical comparison study of KeyCite and Shepard's published prior to this article. Taylor began by evaluating the completeness and currency of the two citators to see if Shepard's had closed the gap in citation retrieval. Comparing the Shepard's and KeyCite reports for a random sample of 459 state and federal cases decided in April 1997, Taylor found that, as of September 1999, there was no longer a significant difference between the citators in the number of citing references they retrieved.[17] To test currency, Taylor retrieved a set of U.S. Courts of Appeals decisions decided just two workdays earlier, and identified eighty-seven case citations in these decisions. He then tested how long it took for these new case citations to appear in the Shepard's and KeyCite reports for the cited cases. He found that all eighty-seven citations appeared in both citators within two workdays of the date of the citing opinion, but that the analyses sometimes took longer. Taylor found that Shepard's was somewhat faster at adding analyses for unrelated cases, but he was uncertain about the validity of his results since he looked at only forty-five citing relationships with analysis in unrelated cases.[18]

---

10. *Id*. at 3.
11. *Id*. at 14.
12. *Shepard's Response to the Shapiro Comparative Study of Shepard's and KeyCite*, Legal Info. Alert, Apr. 1998, at 4.
13. Elizabeth M. McKenzie, *Comparing KeyCite with Shepard's Online*, Legal Reference Servs. Q., 1999 No. 3, at 85.
14. *Id*. at 92–96.
15. *Id*. at 98–99.
16. William L. Taylor, *Comparing KeyCite and Shepard's for Completeness, Currency, and Accuracy*, 92 Law Libr. J. 127, 2000 Law Libr. J. 13.
17. *Id*. at 129, ¶ 9. Taylor attributed the difference between his results and Shapiro's to recent changes that Shepard's had made.
18. *Id*. at 130–32, ¶¶ 12–20.

¶8 Taylor went a step further than Shapiro and McKenzie by comparing the citators' case validation accuracy as well, albeit in a limited way. He looked at 146 citing relationships that at least one of the citators labeled as negative.[19] Out of these 146, 77 were labeled as negative by both citators, 36 were labeled as negative only by Shepard's, and 33 were labeled as negative only by KeyCite. If we assume that all the negative labels were correct, then Shepard's missed 23% of the negative references and KeyCite missed 25%. But why assume all the negative labels were correct? Taylor explained this as follows:

> It is possible that some of these negative analyses are unique to one system because they are mistakes. Looking at the text of each citing opinion, I did find some that I thought were incorrectly identified as negative, but I have decided not to interpose my own judgment in this very subjective area.[20]

¶9 I agree with Taylor that this area can sometimes be subjective, so I can understand his reluctance to use his own judgment. But his reticence has drawbacks. Each time a citator creates a false negative label, Taylor's method gives it credit for its mistake, while penalizing the other citator for not making the same mistake. It's possible that one citator has a greater tendency to create false negative labels, which would invalidate the comparison. It's also plausible that some of the discrepancies between the citators resulted from reasonable differences of opinion in interpreting ambiguous treatment, but Taylor did not make any attempt to separate reasonable differences of opinion from clear mistakes. For these reasons, it's possible that the real error rates for one or both citators are much lower than Taylor's estimate.

¶10 On the other hand, Taylor might have significantly undercounted the mistakes in one or both citators. Taylor only considered whether treatment was labeled as negative; he did not consider conflicts between the descriptive phrases that Shepard's and KeyCite applied. If Shepard's says a case was "overruled," while KeyCite says it was merely "distinguished," it's unlikely that they could both be considered correct. Second, Taylor did not count any negative treatment that was missed by both Shepard's and KeyCite. If Taylor had been able to compare results from a third citator, he might have identified additional errors.

¶11 Lexis and Westlaw both published responses to Taylor's study and, not surprisingly, their criticism focused on Taylor's method of determining accuracy. In an appendix to his article, Taylor listed the citing relationships that had been labeled as negative by one citator, but not the other. Both Lexis and Westlaw made good use of this list. Jane Morris, writing on behalf of Lexis, offered two examples of "errors" attributed to Shepard's, which were in fact false negative labels generated by KeyCite.[21] Not to be outdone, Westlaw commissioned an independent panel of experts to review all the citing relationships in Taylor's appendix and determine which citator deserved the blame for each discrepancy. Dan Dabney, writing on behalf of Westlaw, announced that the panel had found that in this sample, KeyCite

19. *Id.* at 133, ¶ 23. Taylor looked at two different samples of citing relationships; I'm combining his numbers here.

20. *Id.* at 133 n.17, ¶ 22 n.17.

21. Jane W. Morris, *A Response to Taylor's Comparison of Shepard's and KeyCite*, 92 Law Libr. J. 143, 143–144, 2000 Law Libr. J. 14, ¶¶ 4–6.

had a slight edge in accuracy over Shepard's: thirty-five correct entries for KeyCite, versus thirty-three for Shepard's.[22]

¶12 Westlaw's use of an independent panel of experts is intriguing. Dabney revealed that the panel members frequently disagreed with each other: they reached unanimous agreement only about half the time, and were split three to two on eighteen out of sixty-eight citing relationships.[23] But Dabney's article left many of the details unclear, making it hard to determine why the panel members disagreed so frequently. Apparently, Westlaw asked the panel members to give a simple yes or no answer on whether each citation should be labeled as negative.[24] Dabney did not say whether the panel members were able to confer with each other, nor if they were given any guidance. It's unclear whether each panel member came up with their own working definition of "negative treatment." But Dabney's conclusion is clear enough—he believes that citation analysis is a subjective process. He wrote:

> Perhaps the most salient point to take from this exercise is that one should not rely too much on the history judgments supplied by the editors for either system. Each system missed history tags that were correctly identified by the other, and each applied some tags that were properly omitted by the other.
>
> But, while both systems could undoubtedly do somewhat better in applying history tags, it is not to be expected that they can always anticipate the judgments of individual legal researchers. If individual researchers do not agree among themselves, there is no way that any system can always agree with all of them. Thus, while history tags are of immense assistance to the legal researcher, they should not be relied upon exclusively. Careful researchers should also examine citations that do not have negative tags.[25]

¶13 Westlaw's advertising for KeyCite does not reflect Dabney's advice.[26]

¶14 In the seventeen years since Taylor's article was published, there have been no comparable follow-up studies that I'm aware of, although some of the later citator commentaries have been interesting.[27] Perhaps the most significant develop-

---

22. Dan Dabney, *Another Response to Taylor's Comparison of KeyCite and Shepard's*, 92 Law Libr. J. 381, 383, 2000 Law Libr. J. 33, tbl. 1. Careful readers may notice a slight discrepancy between Taylor's article and Dabney's, which results from an error in Taylor's article. In the main text of his article, Taylor wrote that there were a total of sixty-nine negative labels that were unique to either KeyCite or Shepard's: thirty-three for KeyCite, and thirty-six for Shepard's. Taylor, *supra* note 16, at 133, ¶ 23. But Dabney referred to only sixty-eight unique negative labels. This is because Taylor listed only sixty-eight unique negative labels in his appendix: thirty-two for KeyCite, and thirty-six for Shepard's. Taylor either omitted a citing relationship from his appendix or miscounted when writing his article.

23. Dabney, *supra* note 22, at 384, tbl. 2.

24. Dabney did not say this explicitly, but he did not list anything other than simple binary responses.

25. Dabney, *supra* note 22, at 384–85, ¶¶ 19–20.

26. *See KeyCite on WestlawNext, supra* note 5.

27. In 2003, Alan Wolf and Lynn Wishart examined an inherent shortcoming in both Shepard's and KeyCite: they miss important new legal authorities that don't cite the target case. Alan Wolf & Lynn Wishart, *Shepard's and KeyCite Are Flawed (Or Maybe It's You)*, N.Y. St. B.J., Sept. 2003, at 24. This is an important problem often overlooked by novice researchers. For example, if a state legislature passes a new statute that supersedes an existing case, the statute itself won't appear in the case's citator report because the statute doesn't cite the case. Although we can't expect a citator to pick up non-citations, the databases must take some blame for their misleading advertising. Contrary to what their advertising suggests, Shepard's, KeyCite, and BCite *don't* tell you if a case is still good law—they merely tell you about the sources that cited it. More recently, in 2013 Susan Nevelow Mart found that Shepard's outperformed KeyCite in identifying relevant headnotes. Susan Nevelow Mart, *The Case for Curation: The Relevance of Digest and Citator Results in Westlaw and Lexis*, 32 Legal Reference Servs. Q. 13, 41–42 (2013). In 2016, Aaron Kirschenfeld compared the descriptive phrases that KeyCite and

ment since Taylor's article is the appearance of Bloomberg Law and its citator BCite, which launched in late 2009.[28] I am not aware of any significant comparison tests involving BCite.[29] Part of this article's purpose is to determine whether BCite is comparable to Shepard's and KeyCite for case validation, but perhaps more importantly, the addition of a third citator allows for a more thorough assessment of all three citators.

¶15 Designing a citator accuracy test is a difficult task. Taylor tried a minimalist approach, for which he was criticized. But Taylor's critics ran into problems of their own. Westlaw went to the trouble of commissioning an independent study, only to find that their panel members frequently disagreed with each other. Westlaw's Dan Dabney seemed ready to give up on the idea of comparing accuracy, instead offering his conclusion that citation analysis is a matter of opinion.

¶16 Based on my own experience with the study presented here, I concede that citation analysis is partly subjective. There are two key reasons for this. First, courts sometimes use ambiguous language when discussing other cases, sometimes deliberately. This creates ambiguity in a citator's underlying data, and ambiguity will always be open to different interpretations. Second, different researchers may have different ideas on what "negative treatment" means, not to mention more specific terms like "distinguished" or "criticized."

¶17 Nonetheless, I believe it's possible to design a reasonably objective comparison of citator accuracy. In the vast majority of the citing relationships I saw, the courts expressed themselves clearly, and if the underlying data are reasonably clear, there's no reason why citators and citator comparisons can't also be reasonably clear and accurate. To illustrate, let's consider again the independent study that Westlaw commissioned in response to Taylor's article. At first glance, it may seem that the panel members often disagreed on what the citing cases said. But some of those apparent conflicts may have been due to flaws in the way Westlaw designed the comparison test.

¶18 Westlaw apparently asked each panelist for a simple yes or no answer on whether each citing relationship was negative. But Westlaw could have offered a third option—it could have allowed the panelists to say that a citing relationship was ambiguous. It could be that in many of the apparent conflicts, the panelists would have agreed with each other that the citing relationship was ambiguous. If Westlaw had excluded these ambiguous relationships from the results, no doubt they would have found a higher rate of agreement on the remaining relationships. Alternatively, Westlaw could have excluded citing relationships where the panelists split three to two, and based its comparison on the remaining citing relationships. Either way, it's possible to create a more objective comparison test by filtering out

---

Shepard's applied to negative citing references. Aaron S. Kirschenfeld, *Yellow Flag Fever: Describing Negative Legal Precedent in Citators*, 108 Law Libr. J. 77, 2016 Law Libr. J. 4. Due to a small sample size, he did not come to any conclusions about which citator was more accurate. *Id.* at 91–92, ¶ 40.

28. Robert J. Ambrogi, *Can It Be a Contender? Bloomberg Law*, Or. St. Bar Bull., Feb./Mar. 2010, at 15.

29. I am aware of two very limited comparison tests involving BCite, Shepard's, and KeyCite. Neither produced any statistically valid results. Carole A. Levitt & Mark Rosch, *Are All Citator Services Created Equal?*, Internet for Lawyers (2012), http://www.netforlawyers.com/content/caselaw-citators-comparison-Google%20Scholar-Fastcase-Casemaker-LexisNexis-WestlawNext-Bloomberg-0059 [https://perma.cc/97K8-WW8Q]; Michael Robak, *The Bloomberg Citator: A First Look at BLAW's Citation Function*, AALL Spectrum, July 2009, at 24.

unreliable data. Some citing relationships are clear while others are hopelessly muddled; we shouldn't treat both types of citing relationships the same way in a citator comparison test.

¶19 Furthermore, it's not clear that Westlaw supplied the panel members with any definition of "negative treatment" or that Westlaw allowed the panel members to confer with each other on a definition. Even when a citing relationship is clear and two researchers read it the same way, they may apply conflicting labels if they don't agree on what the labels mean. In any comparison test, it's essential for all the testers to share clear and consistent understandings of what the labels mean, and to disclose these understandings when publishing their results.

¶20 Finally, it's not clear that the conflicts between Westlaw's panel members resulted from differences of opinion, rather than simple mistakes. Even when a citing relationship is clear, it may be hard to spot—sometimes I found a key phrase buried in a footnote, or sometimes the relevant discussion came a few paragraphs after the citation. If the panel members had been allowed to confer with each other at the end of the process, they might have been able to resolve many of their conflicts simply by pointing out the relevant passages to each other.

¶21 In the end, I expect it would be rare for two expert researchers to view the same citing relationship, point out the relevant passages to each other, agree that the language is unambiguous, and still disagree on what the citing court meant to say. There is more room for disagreement on how a citing relationship should be *labeled*, but a citator can and should explain any labels to end users through detailed definitions. In a comparison test, labels that are different but comparable can both be considered correct.

¶22 I've tried to improve on the accuracy of previous studies, but in the end no evaluation will be entirely free from subjectivity. So that readers may judge my work for themselves, I offer a detailed description of my methodology and many examples of the errors I perceived in the citators.

Methodology: Designing a Citator Comparison Test

¶23 In this study, I evaluate how accurately and thoroughly the citators identify negative citing references, from the perspective of a typical user. My "typical user" perspective centers on the needs of practicing lawyers, but it does not involve a particular jurisdiction or a particular legal problem. The question is whether a citing reference would negatively impact a lawyer's reliance on a case in *any* jurisdiction for *any* purpose. If a citator fails to identify any such negative treatment, I count this failure against the citator, regardless of whether the citator followed its own procedures. This explains my basic approach.[30]

¶24 As a starting point, I used Lexis to retrieve all seventy-three published decisions of the Ninth Circuit Court of Appeals from January 1984.[31] I selected the

---

30. It might be useful to separately evaluate how well the citators follow their own procedures, but my knowledge of the citators' behind-the-scenes work is incomplete. Only the database providers could fully and accurately evaluate how well the citators follow their internal procedures.

31. To retrieve these cases, I used the search query *date = 1/1984* in Lexis's database of Ninth Circuit Court of Appeals cases, and then used the filter option for reported cases. This yielded seventy-six cases, three of which were false hits. Two of the false hits were cases that the court withdrew prior to final publication (Levi Strauss & Co. v. Blue Bell, Inc., 732 F.2d 676 (9th Cir. 1984), and Piatt v.

Ninth Circuit because of its reputation for a high reversal rate.[32] In its October 1984 term, the U.S. Supreme Court reviewed twenty-eight Ninth Circuit cases and reversed them twenty-seven times.[33] I didn't design this article to evaluate the Ninth Circuit, but in any case my choice did not disappoint: as of mid-2017, according to the citators, these seventy-three Ninth Circuit cases have garnered a total of 357 negative citing references from other courts. These citing references are the basis for this comparison study.

¶25 Before going any further, I would like to stress that this study is *not* limited to results from January 1984, nor is it limited to Ninth Circuit cases. The seventy-three Ninth Circuit cases are merely a starting point. The relevant treatment comes from the *citing* cases, which date from 1984 to 2017, and come from a variety of courts. Federal cases predominate among the citing cases, which might skew the results somewhat if one or more citators handle federal cases with greater or lesser care than state cases. Otherwise, the results presented here should be representative of the citators' overall validation performance.

¶26 After I had identified the seventy-three Ninth Circuit cases using Lexis, I skimmed or read portions of each case so that I understood its subject matter, and downloaded its Shepard's, KeyCite, and BCite reports. The 357 citing cases I examined include each citing case that at least one of the three citators identified as negative, not including negative treatment from administrative decisions. All the citing cases I examined were from domestic courts (state, federal, and territorial), including both published and unpublished opinions. All three citators place citing cases into separate categories for "history" (i.e., other cases from the same course of litigation) and unrelated cases, but I make no such distinction and treat all negative case law treatment the same way, regardless of what category it comes from.[34]

¶27 For BCite, I counted as negative any citing relationships labeled as "distinguished," "criticized," "superseded by statute," "prior overruling," "overruled in part," or "overruled." For KeyCite, I counted as negative any citing relationships on the "Negative Treatment" tab. For Shepard's, I counted as negative any citing relationships with a red, orange, or yellow square, as well as citing relationships labeled as "Among Conflicting Authorities Noted In" (ACAN).[35] Throughout this paper, I refer to all of these indicators as "negative." For each citing relationship that I stud-

---

MacDougall, 723 F.2d 721 (9th Cir. 1984)), and another was a case listed twice by Lexis, the second time with a different docket number and no Shepard's results (EEOC v. Borden's, Inc., 724 F.2d 1390 (9th Cir. 1984)). I used only the first entry for Borden's (the one with Shepard's results).

32. *See, e.g.*, Marybeth Herald, *Reversed, Vacated, and Split: The Supreme Court, the Ninth Circuit, and the Congress*, 77 Or. L. Rev. 405 (1998); Diarmuid F. O'Scannlain, *A Decade of Reversal: The Ninth Circuit's Record in the Supreme Court Since October Term 2000*, 14 Lewis & Clark L. Rev. 1557 (2010); Kevin M. Scott, *Supreme Court Reversals of the Ninth Circuit*, 48 Ariz. L. Rev. 341 (2006); Stephen J. Wermiel, *Exploring the Myths About the Ninth Circuit*, 48 Ariz. L. Rev. 355 (2006).

33. Wermiel, *supra* note 32, at 357.

34. BCite further separates unrelated cases into two tabs, "Case Analysis" and "Citing Documents." *See Product Help & Walkthrough: BCite*, Bloomberg Law, https://www.bna.com/bcite/ [https://perma.cc/W4N6-ESQ5]. The cases on the "Citing Documents" tab have no treatment symbols or descriptions, and it's not clear if they are analyzed at all. In this study, I make no distinction between the two tabs. I count all failures to appropriately label a case, regardless of whether the case is on the "Case Analysis" or "Citing Documents" tab.

35. Shepard's considers ACAN to be neutral treatment, but the ACAN label is comparable to a negative label on KeyCite, so I decided to include ACAN labels in this study for purposes of comparison. *See infra* ¶¶ 42–44.

ied, I recorded the descriptive word or phrase that each citator applied (for example, "distinguished by" or "disagreement recognized by").[36]

¶28 I gathered the raw data for this study between April 25 and June 6, 2017. For each of the seventy-three Ninth Circuit cases, I always retrieved reports from all three citators on the same day, so that no citator would receive an unfair advantage due to timing. Prior to the publication of this article, I shared my findings with the database providers, and in response they have already made changes to their citator reports. I don't take into account any changes to the reports after I downloaded them, so this article does not necessarily reflect the citators' current content.

¶29 For each of the 357 citing relationships that I examined, I retrieved the citing case, read the relevant passages, and made my own judgment as to how the citing case treated the cited Ninth Circuit case.[37] If I felt that reasonable people might disagree on whether a citing relationship was negative, I recorded it as ambiguous. In these situations, I did not treat any citator as right or wrong. For example, in *Fyffe v. Heckler*,[38] a disability benefits case, a U.S. District Court cited the Ninth Circuit's *Kail v. Heckler*[39] in an ambiguous way. The district court wrote:

> Fyffe argues that his objectively identified psychological disorder is "severe" as defined in the regulations. He then argues that the existence of this severe non-exertional impairment absolutely precludes a directed denial of disability pursuant to the grids [vocational guidelines]. There is some support for Fyffe's argument. *See, e.g.*, *Kail v. Heckler*, 722 F.2d 1496, 1498 (9th Cir.1984); *McCoy v. Schweiker*, *supra* (where non-exertional impairments are present, the grids may serve only as guidelines); *Roberts v. Schweiker*, 667 F.2d 1143 (4th Cir.1981) (recourse must be had to evidence other than the grids). Here, however, Fyffe's real argument is not with the ALJ's reliance on the grids, but rather is with the ALJ's finding that Fyffe's non-exertional impairment is not so severe as to preclude sedentary, unskilled work.[40]

The court did not cite *Kail* again, and ruled against the plaintiff. How did the court treat *Kail*? It seems clear that the court distinguished *McCoy* and *Roberts*, and by implication it seems to have distinguished *Kail* as well, but it's hard to be certain because the district court didn't include any parenthetical for *Kail*. The district court said nothing explicit about *Kail*, except that it provided "some support for Fyffe's argument." According to Shepard's, *Fyffe* distinguished *Kail*, but according to KeyCite and BCite, *Kail* did not receive any negative treatment here. I tend to agree with Shepard's, but I recognize some room for reasonable difference of opinion. I marked this citing relationship as ambiguous, so that none of the citators are credited or penalized.[41]

---

36. I used abbreviations in my data, such as "D" for "distinguished by" and "DR" for "disagreement recognized by."

37. To improve the accuracy of my results, I reviewed all the citing cases a second time, a few months after my initial review. I recorded my own judgments again without looking at my initial judgments, and then compared my initial set of judgments with my second set of judgments. I reviewed for a third time the citing relationships where my judgment was inconsistent and produced a corrected set of results.

38. 580 F. Supp. 310 (D. Ariz. 1984).

39. 722 F.2d 1496 (9th Cir. 1984).

40. *Fyffe*, 580 F. Supp. at 314.

41. Another example of ambiguous treatment can be seen in *Ferguson v. Park City Mobile Homes*, No. 89 C 1909, 1989 U.S. Dist. LEXIS 11010 (N.D. Ill. Sept. 15, 1989). In *Ferguson*, a U.S. District Court in Illinois agreed with the Ninth Circuit's decision in *Brothers v. First Leasing*, 724 F.2d 789 (9th Cir.

¶30 For each citing relationship that was not ambiguous, I decided whether the relationship was negative. If a citing relationship was negative, but a citator failed to identify it as such, I marked this as a failure. These failures can occur in one of two ways: (i) failing to put a negative label on a negative case, or (ii) omitting a negative case altogether. In my records, I distinguished between the two types of situations. I sometimes refer to the first type of situation as "mislabeling," but to be more precise, some of these cases have no analysis label at all.[42] I also tracked negative labels that citators applied to citing relationships that were in fact positive or neutral; in the results section of this article, I report these instances separately.

¶31 My understanding of "negative" treatment means any treatment that invalidates the cited case in any jurisdiction, casts doubt on its validity in any jurisdiction, criticizes its reasoning, and/or limits its application. It's not uncommon for a court to treat a cited case both positively and negatively in the same opinion. In my view, any negative treatment needs to be labeled as such, regardless of whether it is mixed with positive treatment. Ideally, a citator should include both positive and negative labels for the same citing case where appropriate, with specific pin cites. Shepard's has this capability, but KeyCite and BCite do not. If a citing case gives a mix of positive and negative treatment, I expect a citator to apply a negative label, but in order to treat all three citators comparably, I overlook any failure to provide a positive label in addition to the negative label.[43]

¶32 In a few instances, I noticed that a citator omitted a positive or ambiguous citing case from its list of citing cases. Positive omissions would come to my attention only if another citator mislabeled the positive case as negative. Although I recorded these positive and ambiguous omissions when I found them, I did not count these omissions as errors since I'm not judging the citators' overall recall performance—I'm judging only their performance in identifying negative treatment.

1984), that a lease could be a form of credit covered by the Equal Credit Opportunity Act. But at the same time, the District Court disagreed with part of the Ninth Circuit's reasoning. *Ferguson*, 1989 U.S. Dist. LEXIS 11010, at *7 ("While this Court does not follow *Brothers*'s analysis—the Ninth Circuit reasoned that the ECOA incorporates the proscriptions of the TILA as amended by the CLA—it agrees, based on what has been presented to the Court, that the ECOA can apply to this transaction."). In effect, the District Court merely concurred with the result reached by the Ninth Circuit. In the KeyCite report for *Brothers*, Westlaw labeled *Ferguson* as "not followed by." Shepard's did not label *Ferguson* as negative, while BCite missed *Ferguson* altogether. I tend to agree that *Ferguson* is negative, although I would have described it as "criticized by." But it could be argued that declining to follow a line of reasoning is not the same as criticism, and that the departure in reasoning is outweighed by the agreed result. Seeing room for reasonable difference of opinion, I marked this citing relationship as ambiguous.

42. It's a worse mistake to put a distinctly positive label on a negative case than no analysis label at all, but the different citators don't have comparable ways of handling non-negative cases. KeyCite doesn't use positive or neutral analysis labels at all. In this article, I treat citing references that have no analysis label as being comparable to citing references with a positive or neutral label.

43. For example, in *Marchese v. Shearson Hayden Stone, Inc.*, 734 F.2d 414 (9th Cir. 1984), the Ninth Circuit cited its earlier opinion in *Amaro v. Continental Can Co.*, 724 F.2d 747 (9th Cir. 1984). The *Marchese* court applied *Amaro* to one of the plaintiff's claims, 734 F.2d at 421, but held that *Amaro* was distinguishable with respect to another claim, *id.* at 423. Although Shepard's can label a citing case as both negative and positive, the Shepard's report for *Amaro* labeled *Marchese* only as "distinguished by." KeyCite and BCite listed *Marchese* as a citing case, but did not label it as negative treatment. I marked Shepard's as correct, and KeyCite and BCite as incorrect.

¶33 All three citators do more than just mark certain citing references as negative. They also apply descriptive words or phrases to specify the type of negative treatment, such as "distinguished by" or "overruled." These descriptions are important, as some forms of negative treatment are more severe than others, and users may choose to examine only certain types of negative treatment. I evaluated the descriptive labels that the citators applied to negative citing relationships, and counted any that were plainly incorrect.[44] I allowed for some difference of opinion, so that two different descriptive labels applied by two different citators might both be considered correct. For example, in *Chance Management, Inc. v. South Dakota*,[45] the Eighth Circuit said the following about a Ninth Circuit opinion: "We are not entirely persuaded by the reasoning in *Western Oil and Gas*; however, even if we were to agree, this case is different."[46] The court went on to describe how its case was different, but it made no critical remarks about the Ninth Circuit's reasoning. Shepard's labeled this citing relationship as "distinguished by," while KeyCite labeled it as "called into doubt by." I think that Shepard's label is more accurate, but some users might feel that the phrase "we are not entirely persuaded by the reasoning" warrants the more negative label applied by KeyCite. Allowing for reasonable difference of opinion, I marked both labels as correct. BCite didn't label this as negative, which I counted as a failure.

¶34 Sometimes different parts of a citing case engage in different types of negative treatment. This creates a potential problem in the underlying data, but not necessarily any ambiguity. Ideally, a citator should accurately describe all the applicable types of negative treatment. Shepard's has this capability, but I didn't notice any citing cases in KeyCite or BCite that had more than one negative descriptive label. At a minimum, citators need to alert users to the most serious type of negative treatment. Where more than one type of negative treatment is involved, I accept as correct any descriptive phrase that correctly describes the most serious type. For example, in *Williams v. Baltimore County*, the U.S. District Court in Maryland both criticized and distinguished a Ninth Circuit opinion.[47] Shepard's handled this the ideal way, by applying both "criticized by" and "distinguished by" labels. KeyCite and BCite labeled *Williams* as distinguishing only, which I counted as failures. If KeyCite or BCite had labeled *Williams* as "criticized by," but not "distinguished by," I would have marked them as correct.

¶35 Each citator has its own standardized list of descriptive terms and phrases that it applies to citing cases.[48] Many terms and phrases are common to all three

---

44. For my statistical comparison, I didn't evaluate KeyCite's "most negative treatment" label, which is designed to highlight the single most negative reference for a case. *See* Kim Ellenberg, *Westlaw Next Tip of the Week: Checking Cases with KeyCite*, Legal Solutions Blog (June 11, 2012), http://blog.legalsolutions.thomsonreuters.com/legal-research/reference-attorney-tips/westlaw-next-tip-of-the-week-checking-cases-with-keycite/ [https://perma.cc/7ZLT-4GGJ]. I didn't mark any "most negative" labels as wrong partly because Shepard's and BCite lack a comparable feature, and partly because a mistake in a "most negative" label usually results from mistakes in other labels in the same KeyCite report. Counting "most negative" labels as wrong would subject KeyCite to a form of double-counting not applied to Shepard's or BCite.

45. 97 F.3d 1107 (8th Cir. 1996).

46. *Id.* at 1114 (citing Western Oil & Gas Ass'n v. Cory, 726 F.2d 1340 (9th Cir. 1984)).

47. No. 13-03445, 2016 U.S. Dist. LEXIS 32124, at *40–42 (D. Md. March 11, 2016) (citing Wrighten v. Metro. Hosps., Inc., 726 F.2d 1346 (9th Cir. 1984)).

48. LexisNexis, Shepard's *Signal Indicators and Treatments* 2 (2016), https://www.lexisnexis.com/pdf/lexis-advance/Shepards-Signal-Indicators-and-analysis-phrases-Final.pdf [https://perma.cc/Y2CH-9H8J]; Thomson Reuters, *KeyCite on Thomson Reuters Westlaw* 2, 4 (2016), https://

citators, but there are some differences. Minor differences are easily harmonized. For example, Shepard's uses the phrase "questioned by," while KeyCite uses the comparable phrase "called into doubt by." Wherever the phrase "questioned by" is correct, I also consider the phrase "called into doubt by" to be correct. But sometimes a label in one citator has no counterpart in another. If a citator is unable to correctly identify relevant negative treatment because it has no appropriate label it can apply, I count this as a failure.

¶36 For example, in *In re Schwartz*, a bankruptcy court in the Third Circuit declined to apply a Ninth Circuit decision (*Matthews*) because it directly conflicted with a Third Circuit case (*Pristas*).[49] The District Court merely stated: "Although *Mathews* [sic] and the numerous other cases provide a tenable basis for the debtors' stance, they are in direct conflict with *Pristas*. We are, of course, bound by *Pristas*."[50] KeyCite correctly labeled this treatment as "not followed," while Shepard's mislabeled it as "criticized by" and BCite mislabeled it as positive treatment. Shepard's and BCite do not appear to have any label equivalent to "not followed." It's possible that the editor at Shepard's chose the label "criticized by" because it was the best available label in Shepard's. But the *Schwartz* court did not criticize the Ninth Circuit, according to any reasonable understanding of the term "criticized."[51] On the contrary, the *Schwartz* court acknowledged that *Matthews* offers a "tenable basis" and that it aligns with "numerous other cases." From the perspective of the user, the Shepard's label is wrong. It makes no difference to the user whether Shepard's mistake resulted from an individual editor's decision or from a broader design problem in Shepard's list of descriptive terms and phrases. Likewise, BCite's lack of any negative label here is a validation failure from the user's perspective. For a lawyer in the Third Circuit, *In re Schwartz* indicates that *Matthews* is not good law, yet BCite applies no negative label.

¶37 BCite's list of descriptive terms and phrases is significantly shorter than the lists offered by KeyCite and Shepard's, which puts BCite at some disadvantage in this comparison study.[52] I do not automatically count these missing labels against BCite, but rather I verify in each instance whether the labels applied by the other citators are accurate and relevant to the typical user. For example, if KeyCite applied a "not followed" label to a citing reference that *did* follow the cited case, I would mark KeyCite's label as incorrect and credit BCite for having no negative label. If KeyCite applied a "not followed" label to an ambiguous citing reference, no

---

lscontent.westlaw.com/images/content/KeyCite%20on%20Westlaw.pdf [https://perma.cc/KR66-C3AY]; *Bloomberg Law Citator*, https://www.bna.com/btaccitator/ [https://perma.cc/4L47-6EBF]. I found a few additional terms and phrases in Shepard's and KeyCite results that are not included in their published lists, but they don't vary widely.

49. 52 B.R. 314, 316–17 (Bankr. E.D. Pa. 1985) (citing Pristas v. Landaus of Plymouth, Inc., 742 F.2d 797 (3rd Cir. 1984), and Matthews v. Transamerica Fin. Servs., 724 F.2d 798 (9th Cir. 1984)).

50. *Id.* at 317.

51. In a user guide, Lexis explains its use of the term "criticized," but the explanation reflects the ordinary meaning of the word. *See* LexisNexis, Shepard's *Signal Indicators and Treatments*, *supra* note 48. I found no instances in which a citator's user guide conflicted with common sense.

52. The only labels in BCite that might be considered negative are "distinguished," "criticized," "superseded by statute," "prior overruling," "overruled in part," and "overruled." See *Bloomberg Law Citator*, *supra* note 48.

citator would be credited or penalized. In their comments to me, Bloomberg Law asserted that I am unfairly penalizing BCite for having a shorter list of labels.[53]

¶38 The term "distinguished" appears to be the most common negative treatment label in all three citators, and requires some clarification. As I understand the term in this context, a "distinguishing" case reaches a different outcome or applies a different rule of law as compared to the cited case because of factual differences (procedural or substantive) between the two cases.[54] This is a form of negative treatment because it may limit the scope of the cited case. By contrast, a citing case that points out a factual difference in the cited case, but nonetheless applies the same rule of law and reaches the same outcome, is *not* what I consider a "distinguishing" case. On the contrary, it's a positive reference that reaffirms the cited case and perhaps even extends it.

¶39 A citation in *United States v. Gleave*[55] illustrates this distinction. *Gleave* pointed out differences between its fact pattern and the Ninth Circuit's *United States v. Snowadzki*,[56] yet *Gleave* did not distinguish *Snowadzki*. Both cases found that an informant was not acting as a government agent under a Fourth Amendment analysis, so that evidence seized by the informant could be admitted. *Gleave* described the factual differences as follows:

> In this case, the facts indicating government involvement in the alleged burglary of Knoll's office are less than in *Snowadzki*. Unlike *Snowadzki*, in this case there is no evidence that the government knew of a potential seizure before it took place and no rewards to private parties were mentioned.[57]

Here, *Gleave* is pointing out that its defendant has an even weaker case than the defendant in *Snowadzki*, who also lost on this point. *Gleave* is in no way a negative reference for *Snowadzki*, yet KeyCite labeled it as "distinguished by." I marked KeyCite's label as incorrect. Shepard's and BCite treated *Gleave* as a positive reference, which I marked as correct.

¶40 Some readers may have a broader view of the term "distinguished," but in the context of citators it would be problematic to group together both positive and negative treatment under the same "distinguished" label. Labels should aid users in differentiating between types of treatment. Moreover, labeling positive cases as "distinguished" would conflict with the structure of Shepard's and KeyCite, and is inconsistent with the data I gathered for BCite. Shepard's assigns any distinguishing treatment a yellow caution sign (which indicates possible negative treatment), and KeyCite puts any distinguishing treatment on the "Negative Treatment" tab. From the user's perspective, *Gleave*'s appearance in a negative category would seem to be a mistake. In BCite, the negative nature of the "distinguished" label is less clear to

---

53. See *infra* ¶ 75.
54. KeyCite's published definition of "distinguished" matches my own, while the definitions published by Shepard's and BCite are consistent with mine, but open to interpretation. *See* Thomson Reuters, *KeyCite on Thomson Reuters Westlaw, supra* note 48 ("Distinguished: There is a difference in facts, procedural posture, or law between the two cases that compels the citing court to reach a different result than the cited case"); LexisNexis, Shepard's *Signal Indicators and Treatments, supra* note 48 ("Distinguished by: the citing case differs from the case you are Shepardizing, either involving dissimilar facts or requiring a different application of the law"); *Bloomberg Law Citator, supra* note 48 ("Distinguished: . . . one or more courts differentiate this opinion on the law or the facts").
55. 786 F. Supp. 258 (W.D.N.Y. 1992).
56. 723 F.2d 1427 (9th Cir. 1984).
57. *Gleave*, 786 F. Supp. at 290 n.20.

the user.[58] In their comments to me, Bloomberg Law stated that their "distinguished" label is not necessarily negative.[59] However, BCite's actual practice was to apply no distinguishing label to any of the nine citing references like *Gleave* in my dataset. Although BCite's policy might pose a problem for future comparison tests, I regard it as a moot point in this comparison test.

¶41 Another important distinction is that "distinguished" doesn't mean the same as "distinguishable." Any case is distinguishable under the right circumstances. The purpose of a citator is not to tell us that a case is distinguishable, but rather to tell us which cases have distinguished it. This may seem like an obvious point, but the citators sometimes disregard it. For example, in *Vic Wertz Distributing Co. v. Teamsters Local 1038*,[60] the court noted that *George Day Construction Co. v. United Brotherhood of Carpenters*[61] is "distinguishable from the typical case," and then in the next paragraph, it stated that its own case is like *George Day*—in other words, not like the typical case.[62] Curiously, both BCite and KeyCite labeled *Vic Wertz* as distinguishing *George Day*. But surely users would expect the "distinguished by" label to refer to the relationship between *Vic Wertz* and *George Day*— not to the relationship between *George Day* and a nameless "typical case."[63] I marked BCite and KeyCite as incorrect.[64] Shepard's labeled *Vic Wertz* as a positive reference, which I marked as correct.

¶42 Shepard's "Among Conflicting Authorities Noted In" (ACAN) label presents a puzzle. Shepard's considers ACAN to be neutral treatment, but KeyCite has a similar descriptive phrase called "Disagreement Recognized By" (DR), which it treats as negative. BCite does not have a comparable label. Both Shepard's and KeyCite apply these labels to cases that say the cited case is in conflict with other cases. For example, in *Burds v. Union Pacific Corp.*,[65] the Eighth Circuit noted a circuit split involving the Ninth Circuit's opinion in *Amaro v. Continental Can Co.*[66] The Eighth Circuit said that *Amaro* conflicted with opinions from the Seventh and Eleventh Circuits, but the *Burds* court did not decide which side to take.[67] Shepard's labeled this as ACAN (neutral treatment), while KeyCite labeled it as DR (negative

---

58. In BCite, cases that have been distinguished can't have a positive composite analysis label. From this, I inferred that the "distinguished" label in BCite is distinct from positive treatment. But unlike Shepard's and KeyCite, BCite's interface does not explicitly place distinguished references in a negative category.

59. Letter from Darby J. Green, Commercial Director of Litigation & Bankruptcy, Bloomberg Law, to author (July 10, 2018) (on file with author).

60. 898 F.2d 1136 (6th Cir. 1990).

61. 722 F.2d 1471 (9th Cir. 1984).

62. *Vic Wertz Distributing Co.*, 898 F.2d at 1140.

63. A citator might use a label such as "distinguishing recognized by," similar to the label "overruling recognized by," to cover situations where a court refers to a relationship between the cited case and some other case. It's not clear whether such a label should be considered negative or whether it would be useful to users. But I have never seen such a label in any of the citators.

64. On this point, I don't believe the database providers and I have any policy disagreement. In response to my findings, Bloomberg Law and Westlaw have already corrected their citator reports for *George Day*. Also, Bloomberg Law confirmed to me that their "distinguished" labels are supposed to refer to the relationship between the citing case and the cited case. Letter from Darby J. Green, *supra* note 59.

65. 223 F.3d 814 (8th Cir. 2000).

66. 724 F.2d 747 (9th Cir. 1984).

67. *Burds*, 223 F.3d at 817.

treatment).[68] I believe it would be better to label this as negative treatment, since it would certainly be negative for anyone in the Seventh or Eleventh Circuits. If *Burds* were the only citing reference to indicate that *Amaro* is not good law in the Seventh or Eleventh Circuits, users from those circuits might be disserved by Shepard's neutral label.[69] But it would be unfair to say that Shepard's ACAN label is wrong, and it would work for users provided they were paying close attention.

¶43 A tougher puzzle is presented when a citing case notes a conflict, but decides to follow the cited case. For example, in *De Pace v. Matsushita Electric Corp. of America*,[70] a U.S. District Court in New York cited *Amaro* and noted the same circuit split with the Seventh and Eleventh Circuits, but sided with the Ninth Circuit's view.[71] Shepard's labeled this as ACAN (neutral treatment), while KeyCite merely said that *De Pace* "discussed" *Amaro*. From the perspective of users in the Ninth or Second Circuits, *De Pace* appears to be positive treatment, but from the perspective of users in the Seventh or Eleventh Circuits, *De Pace* tells them that *Amaro* is not good law. Here, I think Shepard's neutral label is best, but I see room for reasonable differences of opinion on whether this citing relationship should be marked as positive, neutral, or negative. (Neither KeyCite nor BCite has an appropriate neutral label that could be applied here.)

¶44 Citing references that Shepard's labels as ACAN are difficult to compare to KeyCite and BCite. My solution was to include in my study all citing references that had the ACAN label, even though Shepard's considers this to be neutral. (In this article, whenever I refer to relationships that are labeled as "negative," that includes relationships that Shepard's labeled as ACAN.) In situations like *Burds*, where the citing case did not decide whether to follow the cited case, I marked the ACAN label as correct, as well as the DR label from KeyCite. If a citator failed to apply a negative label in these situations, I counted this as a failure. In situations like *De Pace* where a citing case notes a conflict but follows the cited case, I marked the citing relationship as ambiguous, so that no citator is credited or penalized.

¶45 Negative citing references that appear in dissenting opinions also present some difficulties, as the three citators handle these types of citations differently. Shepard's puts them in a neutral category called "cited in dissenting opinion" and never labels these citations as negative.[72] But BCite and KeyCite apply negative labels to dissenting opinions in at least some situations. In my dataset, BCite labeled four citations from dissenting opinions as negative, and KeyCite labeled a fifth one as negative. As I stated earlier, I'm not judging the citators according to how well they follow their internal procedures, so it doesn't matter for my purposes that Shepard's withheld these negative labels according to a deliberate policy. If the negative treatment is clearly relevant to a typical user, it should be labeled as negative—I count anything less as a failure.

---

68. BCite labeled this as positive with no descriptive phrase other than "cited in."

69. It's not unusual for negative treatment to be a matter of perspective, and the citators generally label as negative any treatment that will be negative for some users, but not for others. For example, if the Eighth Circuit had expressly disagreed with *Amaro*, the treatment would obviously be labeled as negative, even though it presents no problem for users practicing in the Ninth Circuit.

70. 257 F. Supp. 2d 543 (E.D.N.Y. 2003).

71. *Id.* at 557–60.

72. Letter from Liz Christman, Senior Director of Case Law and Shepard's Solutions, LexisNexis, to author (June 13, 2018) (on file with author).

¶46 But is negative treatment clearly relevant if it comes from a dissenting opinion? Not necessarily. If a U.S. Supreme Court majority opinion cites a Ninth Circuit case favorably, but the dissenting opinion criticizes the Ninth Circuit, I certainly wouldn't mark a citator as wrong for listing the Supreme Court case as a positive reference. As any first-year law student knows, it's the majority opinion that counts. But what if a dissenting opinion cites a case and points out that it has been treated negatively by another case? My dataset includes two examples of this. One is *Enlow v. Salem-Keizer Yellow Cab Co.*,[73] a Ninth Circuit panel decision in which Judge Ferguson wrote a separate opinion concurring in part and dissenting in part. Ferguson cited a Ninth Circuit case, *EEOC v. Borden's, Inc.*, and noted that it had been overruled on other grounds by the U.S. Supreme Court.[74] The majority opinion didn't cite *Borden's*. Users would certainly want to know that *Borden's* has been overruled, regardless of who points out this fact.[75] BCite handled this perfectly by describing the treatment as "prior overruling in" and labeling *Enlow* as a concurring/dissenting opinion. The KeyCite and Shepard's reports for *Borden's* didn't indicate any negative treatment from *Enlow*, which I counted as failures. To make matters worse, the Supreme Court case itself is incorrectly listed as a positive reference in the Shepard's report for *Borden's*. For Shepard's, omitting the negative treatment in *Enlow* compounded a critical mistake. Nothing in *Borden's* Shepard's report indicates that it was overruled by the Supreme Court.

¶47 I credited KeyCite for applying an "overruling recognized by" label to another dissenting opinion similar to *Enlow*, which BCite and Shepard's failed to label as negative.[76] The other three negative citations that BCite identified in dissenting opinions were distinguishing references. Some users might feel that distinguishing references in dissenting opinions are too insignificant to justify negative labels, but I tend to think that negative labels are useful here, provided that a citator follows BCite's practice of clearly identifying the sources. Since I can see room for reasonable difference of opinion, I hesitate to mark KeyCite and Shepard's as wrong for not labeling these references as negative. Instead, I marked these three citations as ambiguous.

Statistical Results

¶48 Before I started my independent analysis of the citators' results, I could see a problem: the three citators are rarely in agreement. I looked at 357 citing relationships that have at least one negative label from a citator. Out of these, all three citators agree that there was negative treatment only 53 times. This means that in 85% of these citing relationships, the three citators do not agree on whether there was negative treatment. Even when they all agree there was negative treatment, their

---

73. 371 F.3d 645 (9th Cir. 2004).

74. *Id.* at 656 (citing EEOC v. Borden's Inc., 724 F.2d 1390 (9th Cir. 1984)).

75. There might be some concern that the dissenting opinion mischaracterized the Supreme Court's treatment of *Borden's*, but that concern is present whenever the "overruling recognized by" label is applied. Citators aren't designed to verify whether negative treatment is justified; they merely point it out so that the user may investigate further. In any case, Judge Ferguson was correct in pointing out that *Borden's* had been overruled.

76. United States v. Keene, 933 F.2d 711, 723 (9th Cir. 1991) (Alarcon, J., dissenting) (citing Pacific Mut. Life Ins. Co. v. American Guar. Life Ins. Co., 722 F.2d 1498 (9th Cir. 1984)).

Table 1

Citator Performance in Identifying 309 Negative Citing Relationships

	Relationships omitted or mislabeled as positive or neutral	Relationships labeled as negative, but incorrectly described	Overall failure to correctly identify
Shepard's	85 (28%)	18 (6%)	103 (33%)
KeyCite	105 (34%)	11 (4%)	116 (38%)
BCite	211 (68%)	11 (4%)	222 (72%)

descriptions often conflict. The three databases substantively agree on the type of negative treatment in only 40 of these citing relationships,[77] which means that in this sample, they all agree with one another only 11% of the time.

¶49 On the surface, even without any further examination, these numbers are troubling. They mean that at least one citator is making a lot of mistakes, or (viewed more charitably) that the citators are very likely to have differences of opinion. Even the latter explanation is worrisome, because all three citators present their results as fact—none of the citators labeled any of the citing relationships as ambiguous or open to interpretation. This means that when you cite a case that has negative treatment, the results you get depend mainly on which citator you happen to be using.

¶50 The results of my independent analysis are just as troubling. Out of 357 citing relationships that at least one of the citators marked as having negative treatment, I found that 21 actually had no negative treatment and 27 were ambiguous, leaving 309 citing relationships in which there was clearly negative treatment. Out of these 309 negative citing relationships, Shepard's failed to mark 85 (28%) as negative and applied incorrect descriptive phrases to another 18 (6%);[78] overall, Shepard's failed to correctly identify 103 (33%) of the negative references. KeyCite failed to mark 105 (34%) as negative and applied incorrect descriptive phrases to another 11 (4%); overall, KeyCite failed to correctly identify 116 (38%) of the negative references. BCite failed to mark 211 (68%) as negative and applied incorrect descriptive phrases to another 11 (4%); overall, BCite failed to correctly identify 222 (72%) of the negative references.

¶51 Most of the negative relationships that the citators missed were mislabeled, rather than omitted altogether. Of the 85 negative relationships that Shepard's missed, 78 were mislabeled, while only 7 were omitted altogether. Of the 105 negative relationships that KeyCite missed, 99 were mislabeled, while only 6 were omitted altogether. BCite had a more significant problem with omissions. Of the 211 negative relationships that BCite missed, 178 were mislabeled, while 33 were omitted altogether. All omissions in all three citators were unpublished cases, which

---

77. By "substantively agree," I don't mean that the descriptions were necessarily identical, only that there was no clear substantive difference. For example, if Shepard's described a citation as "distinguished," while KeyCite described it as "declined to extend," I would consider them to be in substantial agreement.

78. An incorrect descriptive label would be one that indicates the wrong type of negative treatment. For example, a citing reference that is labeled "distinguished by" when it should say "overruled by."

Table 2

Incorrect Negative Labels

	Relationships labeled as negative	Relationships that were actually positive or neutral	Relationships that were negative, but incorrectly described	Total incorrect negative labels
Shepard's	235	11 (5%)	18 (8%)	29 (12%)
KeyCite	218	14 (6%)	11 (5%)	25 (11%)
BCite	102	4 (4%)	11 (11%)	15 (15%)

would be less important to users.[79] For all three citators, the significant problems occur in the editorial analysis process, after the initial process of identifying the citing cases.

¶52 I also calculated how accurately the three citators described the negative treatment that they did identify. Excluding the ambiguous references, Shepard's listed 235 references as negative; of these, 11 (5%) were actually positive or neutral, while another 18 (8%) were negative but not correctly described. Overall, 12% of the negative labels in Shepard's were incorrect. KeyCite listed 218 references as negative; of these, 14 (6%) were actually positive or neutral, while another 11 (5%) were negative but not correctly described. Overall, 11% of the negative labels in KeyCite were incorrect. BCite listed 102 references as negative; of these, 4 (4%) were actually positive or neutral, while another 11 (11%) were negative but not correctly described. Overall, 15% of the negative labels in BCite were incorrect.

¶53 The citators are much more likely to label a negative reference as positive than label a positive reference as negative. Unfortunately, the first type of situation is more serious, since it may lead users to cite invalid case law. Users can't be expected to examine every citing reference for themselves. When they see a citing reference marked as positive, they're not likely to investigate further. The second type of situation is more likely to be a mere inconvenience. A user who sees a false negative reference will probably investigate further and make their own judgment, or skip over the cited case and use another case.

¶54 As discussed earlier in this article, William L. Taylor's comparison study rested on the assumption that no positive citing relationships were mislabeled as negative (an assumption for which he was much criticized).[80] My results partly vindicate Taylor's approach, and the performance rates I found for Shepard's and KeyCite are fairly close to what Taylor found. I found that Shepard's missed 28% of the negative references, compared to Taylor's 23%. I found that KeyCite missed 34% of the negative references, compared to Taylor's 25%. The different performance rates I found are due in part to my comparison with a third citator, as BCite sometimes spotted negative treatment that both Shepard's and KeyCite missed. Other explanations for the difference in my results include the larger sample size

---

79. A citator could have omitted a citing case for one of two reasons: (i) the citing case wasn't in the database or (ii) the citator failed to spot the citation in the text of the citing case. I didn't determine which of these two reasons explain each omission, but the fact that all the omissions involve unpublished cases suggests that the first reason explains most of the omissions.

80. *See supra* ¶¶ 8–11.

in this study, and the much broader time frame covered here. This study examines citing relationships dating from 1984 to 2017, compared to Taylor's much narrower time frame of 1997 to 1999.[81]

¶55 Neither Taylor's study nor mine is designed to catch all failures in identifying negative treatment. Because I looked only at citing relationships that at least one of the three citators labeled as negative, I would not have caught any negative treatment that all three citators missed. Given the citators' poor overall performance, it's likely that many negative references were missed by all three.[82] The performance rates presented here may underestimate the extent of the problem.

¶56 Although my statistics suggest that BCite underperforms in case validation as compared to Shepard's and KeyCite, readers should keep in mind that these statistics treat all failures as equal. The reality is more complex, as discussed in the next section.

Examples of Citator Errors

¶57 The statistics tell only part of the story, since not all citator errors are of equal importance. Some are trivial. If a citator misses a brief distinguishing reference from an unpublished trial court case, users are not likely to suffer any harm. Moreover, a citator might reasonably focus its limited resources on labeling the more significant references, at the expense of less important ones. But the mistakes I found are not necessarily trivial. Most were significant, and some were astonishing. In this section of the article, I describe some of the most blatant and potentially harmful mistakes.[83]

¶58 Perhaps no form of negative treatment is more important than an overruling from the U.S. Supreme Court. In my dataset, there are four of these overrulings, all directed at published opinions of the Ninth Circuit Court of Appeals. Incredibly, Shepard's mislabeled three of them, while KeyCite mislabeled two and BCite mislabeled one. These four citing relationships are obviously not a statistically valid sample size, but the point is that serious mistakes of this kind should be exceedingly rare. The fact that Shepard's mislabeled three out of four Supreme Court overrulings is evidence enough of a serious problem. Each of these mistakes is worth discussing in detail.

¶59 First, in *Public Employees Retirement Systems of Ohio v. Betts*, the Supreme Court construed a section of the Age Discrimination in Employment Act that exempts certain age-based employee benefit plans.[84] At one point in the opinion, the Court considered whether the exemption should be limited to plans that have a cost justification for age-based differentials, a position that had been urged by the Equal Employment Opportunity Commission (EEOC). The Court acknowledged two Ninth Circuit cases that had recognized the cost justification requirement,

---

81. Taylor, *supra* note 16, at 129, ¶ 8. It seems plausible that the citators have had some ups and downs in their performance rates over the years, but from the user's perspective, it's their cumulative performance that counts.

82. One such failure came to light during my review. See *infra* ¶ 65.

83. Because the database providers have already made changes to their citator reports in response to a draft of this article, the errors discussed here may no longer exist online. The errors are drawn from the citator reports I downloaded in 2017.

84. 492 U.S. 158 (1989) (construing 29 U.S.C. § 623(f)(2)).

including *EEOC v. Borden's*.[85] In the next three paragraphs, the Court attacked the position taken by the EEOC and the Ninth Circuit. The Court wrote that "[t]here are a number of difficulties with this explanation for the cost-justification requirement," and said the EEOC's position was "quite difficult to believe."[86] The Court further wrote that the EEOC's position was "weakened further" by regulatory authority, and finally concluded that the "cost-justification requirement is contrary to the plain language of the statute and is invalid."[87] The Supreme Court's rejection of the *Borden's* holding is perfectly clear. KeyCite labeled *Betts* as disagreeing with *Borden's*, and BCite labeled it (more appropriately) as overruling *Borden's*.[88] But the Shepard's report for *Borden's* inexplicably labeled *Betts* as "cited by," indicating no negative treatment whatsoever from *Betts*.

¶60 Following the *Betts* decision, at least eight published federal court opinions have explicitly stated that *Betts* overruled *Borden's*.[89] KeyCite or BCite acknowledged these statements with the "overruling recognized by" or "prior overruling in" label, but the Shepard's report for *Borden's* doesn't clearly acknowledge any of these statements. The best that Shepard's could manage was to say that *Borden's* had been "questioned" by only two other courts. The incorrect Shepard's report for *Borden's* resulted not just from a single mistake, but from a series of failures.

¶61 Next, in *Peacock v. Thomas*, the Supreme Court resolved a circuit split about ancillary jurisdiction involving the Ninth Circuit's *Blackburn Truck Lines, Inc. v. Francis*.[90] At the outset of its opinion, the Supreme Court said it had granted *certiorari* "to resolve a conflict among the Courts of Appeals."[91] In a footnote at the end of this sentence, the Court cited *Blackburn* and cases from the Third and Seventh Circuits as being aligned with the Fourth Circuit case under review, in opposition to cases from the Fifth and Tenth Circuits.[92] The Supreme Court reversed the Fourth Circuit.[93] Although the Court did not cite *Blackburn* again, the effect on *Blackburn* is clear. The Supreme Court declared that it was resolving a circuit split involving *Blackburn*, and *Blackburn* wound up on the losing side. BCite correctly labeled *Peacock* as overruling *Blackburn*, but Shepard's and KeyCite both fell short. Shepard's described *Peacock*'s treatment of *Blackburn* as "validity questioned by," and KeyCite described it as "called into doubt by."[94] Shepard's and KeyCite's deci-

---

85. *Id.* at 173 (citing EEOC v. Borden's, 724 F.2d 1390 (9th Cir. 1984)).

86. *Id.* at 173–74.

87. *Id.* at 174–75.

88. I marked both KeyCite and BCite as correct, although KeyCite's label is less than ideal. *Betts* did more than just "disagree" with *Borden's*; it invalidated part of *Borden's*. KeyCite's label was on the borderline of what I consider to be reasonably correct.

89. Enlow v. Salem-Keizer Yellow Cab. Co., 371 F.3d 645, 656 (9th Cir. 2004) (Ferguson, J., dissenting); Cooney v. Union Pac. R.R. Co., 258 F.3d 731, 734–35, 735 n.2 (8th Cir. 2001); Johnson v. New York, 49 F.3d 75, 78 (2d Cir. 1995); Rebar v. Marsh, 959 F.2d 216, 218 n.5 (11th Cir. 1992); AARP v. Farmers Grp., Inc., 943 F.2d 996, 1004 (9th Cir. 1991); Mitchell v. Mobil Oil Corp., 896 F.2d 463, 470 (10th Cir. 1990); EEOC v. Newport Mesa Unified Sch. Dist., 893 F. Supp. 927, 930 (C.D. Cal. 1995); EEOC v. California Micro Devices Corp., 869 F. Supp. 767, 771–72 (D. Ariz. 1994).

90. 516 U.S. 349 (1996) (citing Blackburn Truck Lines, Inc. v. Francis, 723 F.2d 730 (9th Cir. 1984)).

91. *Id.* at 352.

92. *Id.* at 352 n.2.

93. *Id.* at 360.

94. I gave Shepard's and KeyCite credit for correctly identifying *Peacock* as a negative reference, but I marked their descriptive phrases as incorrect.

sion to use these weaker descriptive phrases means that both citators display only yellow or orange warning signs for *Blackburn*, as opposed to the red warning signs that should have been displayed.[95] Curiously, both Shepard's and KeyCite noted that other cases had acknowledged *Peacock*'s invalidation of *Blackburn*, but apparently this didn't prompt anyone at Shepard's or KeyCite to reconsider their description of *Peacock* itself. In the KeyCite report for *Blackburn*, two published opinions from U.S. Courts of Appeals are labeled as "abrogation recognized by," and they were both referring to *Peacock*'s abrogation of *Blackburn*.[96] Shepard's marked one of these as "validity questioned by," but failed to identify any negative treatment in the other.[97] But Shepard's did say that *Blackburn* was "overruled in part as stated in" an unpublished U.S. District Court case, which again was referring to *Peacock*'s overruling of *Blackburn*.[98]

¶62 KeyCite's initial description of *Blackburn*'s overruling led to an even worse mistake in *Blackburn*'s KeyCite report. KeyCite has a "most negative" highlighting feature designed to highlight the single most negative citing case in a KeyCite report. The "most negative" case is listed first on the "Negative Treatment" tab, where it receives a prominent red label; it also appears directly on the "Document" tab where users view the full text of *Blackburn*.[99] According to KeyCite, *Blackburn*'s most negative treatment came from the Fourth Circuit in *Thomas v. Peacock*,[100] the very same case that the Supreme Court grouped with *Blackburn* and reversed. Although *Blackburn* is on the losing side of a circuit split resolved by the Supreme Court, KeyCite directs its users' focus away from the Supreme Court ruling to highlight some negative treatment from a circuit on the same side as *Blackburn*.[101] It's easy to imagine a busy lawyer who looks only at the "most negative" case, sees that it isn't very serious, and decides to rely on *Blackburn* without pulling up the rest of the KeyCite report.

¶63 The last in this trio of Supreme Court cases is *Republic National Bank of Miami v. United States*,[102] which also resolved a circuit split against a Ninth Circuit case. In *Republic National Bank*, the treatment was not as clear as it had been in *Peacock*, because of small differences in the Supreme Court's phrasing. The Court

---

95. Shepard's displayed an orange "Q" sign for *Blackburn* as opposed to a red stop sign; KeyCite displayed a yellow flag as opposed to a red flag.

96. Ellis v. All Steel Constr., Inc., 389 F.3d 1031, 1035 (10th Cir. 2004); Futura Dev. of P.R., Inc. v. Puerto Rico, 144 F.3d 7, 11 (1st Cir. 1998).

97. Shepard's described the negative treatment in *Futura Development* as "validity questioned by," even though *Futura* made it clear that the Supreme Court had ruled against *Blackburn*. I gave Shepard's credit for identifying *Futura* as a negative reference, but marked its descriptive phrase as incorrect. BCite failed to identify either *Ellis* or *Futura* as negative treatment for *Blackburn*.

98. 21st Century Fin. Servs., LLC v. Manchester Fin. Bank, No. 15-cv-1848 BTM (BGS), 2016 U.S. Dist. LEXIS 66822, at *4 n.1 (S.D. Cal. May 19, 2016). KeyCite and BCite also correctly labeled this case. Note the discrepancy between Shepard's correct description of *21st Century Financial* ("overruled in part as stated in") and its incorrect description of *Futura Development* ("validity questioned by").

99. For a description of KeyCite's "most negative" feature, see Ellenberg, *supra* note 44.

100. 39 F.3d 493 (4th Cir. 1994).

101. KeyCite was correct to point out that the Fourth Circuit disagreed with part of *Blackburn*, but it's absurd to suggest that this is more serious than an overruling by the U.S. Supreme Court. I didn't count the "most negative" label as a separate error in my statistical comparison (see *supra* note 44), but it does help to illustrate why KeyCite's description of *Peacock* is wrong. Users need accurate descriptive labels so they can see the relative importance of different citing relationships.

102. 506 U.S. 80 (1992).

in *Republic National Bank* considered the limits of *in rem* jurisdiction, and said it granted *certiorari* "[i]n view of inconsistency and apparent uncertainty among the Courts of Appeals."[103] Although this sentence could have been clearer, I interpret it to mean that the Supreme Court is resolving a circuit split. The Court is hedging somewhat with the word "apparent," but note that this word qualifies "uncertainty," not "inconsistency." A footnote in this sentence gives examples of the circuit split, but it doesn't separate the cases as neatly as the Court did in *Peacock*. Here's the text of the footnote, in its entirety:

> Compare *United States v. One Lot of $25,721.00 in Currency*, 938 F.2d 1417 (CA1 1991); *United States v. Aiello*, 912 F.2d 4 (CA2 1990), cert. denied, 498 U.S. 1048 (1991); *United States v. $95,945.18, United States Currency*, 913 F.2d 1106 (CA4 1990), with *United States v. Cadillac Sedan Deville, 1983*, 933 F.2d 1010 (CA6 1991) (appeal dism'd); *United States v. Tit's Cocktail Lounge*, 873 F.2d 141 (CA7 1989); *United States v. $29,959.00 U. S. Currency*, 931 F.2d 549 (CA9 1991); and the Court of Appeals' opinion in the present case. Compare also *United States v. $57,480.05 United States Currency and Other Coins*, 722 F.2d 1457 (CA9 1984), with *United States v. Aiello*, 912 F.2d at 7, and *United States v. $95,945.18, United States Currency*, 913 F.2d at 1110, n. 4.[104]

¶64 The Supreme Court reversed. The question for this study is how the Court treated the Ninth Circuit's *$57,480.05 United States Currency and Other Coins*,[105] which it cited only in this footnote. The wording in the footnote requires some thought, but in the end its meaning is clear. In the first sentence, *Aiello* and *$95,945.18* are contrasted with the "Court of Appeals' opinion in the present case." In the second sentence, *$57,480.05* is contrasted with *Aiello* and *$95,945.18*. This places *$57,480.05* on the same side as the case under direct review, which was reversed. This interpretation can be confirmed by comparing the Ninth Circuit's holding in *$57,480.05* with the Supreme Court's holding in *Republic National Bank*.[106] The Ninth Circuit held that federal courts lose jurisdiction over *in rem* cases if they lose control over the *in rem* currency, and further held that if *in rem* currency is transferred to the United States Treasury, federal courts have no power to order the Treasury to return it.[107] In *Republic National Bank*, the Supreme Court rejected both aspects of this rule. The Supreme Court held that federal courts do *not* need continuing control over *in rem* currency to retain *in rem* jurisdiction,[108] and further held that courts *can* order the Treasury to return *in rem* currency.[109] The treatment in *Republic National Bank* is a good example of something that's easy to overlook, but not ambiguous after close examination. Catching it requires a sharp eye. Shepard's and BCite failed to identify any negative treatment from *Republic National Bank* in their reports for *$57,480.05*; KeyCite's report for

---

103. *Id.* at 84.

104. *Id.* at 84 n.3.

105. 722 F.2d 1457 (9th Cir. 1984).

106. Ordinarily, I expect the citators to make a judgment based only on the text of the citing case, without consulting the text of the cited case. If the text of the citing case doesn't reveal any negative treatment, I don't expect citators to independently consider conflicts between the two cases. But here, the text in *Republic National* is certainly enough to signal negative treatment; a comparison of the two cases merely confirms how the treatment should be described.

107. *$57,480.05 United States Currency*, 722 F.2d at 1458–59.

108. *Republic National Bank*, 506 U.S. at 84–85.

109. *Id.* at 95–96 (opinion of Rehnquist, J.).

$57,480.05 came closer by labeling *Republic National Bank* as "called into doubt by."[110]

¶65 The three Supreme Court cases discussed above are part of my dataset, because each of them was labeled as a negative reference by at least one citator. During my review, I discovered yet another mishandled Supreme Court case that *should* be part of my dataset. In *Alexander v. Sandoval*,[111] the Supreme Court rejected a holding in *Larry P. v. Riles*,[112] one of the Ninth Circuit cases that I used as a starting point for this study. All three citators included *Alexander* as a citing case for *Larry P.*; the reason why *Alexander* doesn't appear in my dataset is that none of the three citators labeled *Alexander* as a negative reference. The majority opinion in *Alexander* didn't cite *Larry P.*, but Justice Stevens, in a dissent joined by three other justices, cited *Larry P.* as an example of a case that the majority was rejecting.[113] *Alexander* illustrates why citations in dissenting opinions need to be analyzed.[114] Any lawyer evaluating *Larry P.*'s validity would want to know about Justice Stevens' citation. After *Alexander* was decided, it took twelve years for another court to cite *Larry P.* and point out that it had been invalidated by the Supreme Court.[115] Shepard's and KeyCite appropriately labeled this other case, but BCite missed it.[116] Nothing else in the citator reports indicates that *Larry P.* is inconsistent with Supreme Court precedent.

¶66 *En banc* decisions in U.S. Courts of Appeals that reverse prior panel decisions are another critically important form of negative treatment. In *Wallace v. Christensen*,[117] the Ninth Circuit sitting *en banc* reviewed a decision of the United States Parole Commission. As the court explained, "[p]revious cases in this circuit have referred to an 'abuse of discretion' standard as the basis for review of Commission decisions, and have implicitly assumed the existence of jurisdiction to conduct that analysis."[118] Immediately after this sentence, the court cited several Ninth Circuit panel decisions, including *Roth v. U.S. Parole Commission*[119] and *Torres-Macias v. U.S. Parole Commission*.[120] The court said that it "took this case en banc to reconsider this assumption and evaluate our standard of review."[121] The court disagreed with its prior rulings, and held that federal courts have no power to review the Commission's use of discretion even under an abuse of discretion standard, but that courts may consider whether the Commission has acted outside its statutory authority or violated the Constitution.[122] So far, the court's treatment of *Roth* and *Torres-Macias* is similar to the Supreme Court's treatment of *Blackburn* and *$57,480.05*. The court announced that it was reviewing a holding from prior cases

---

110. I gave KeyCite credit for recognizing *Republic National Bank* as a negative reference, but I marked its descriptive phrase as incorrect.

111. 532 U.S. 275 (2001).

112. 793 F.2d 969 (9th Cir. 1984).

113. *Alexander*, 532 U.S. at 295–96 n.1 (Stevens, J., dissenting).

114. For further discussion of negative treatment in dissenting opinions, see *supra* ¶¶ 45–47.

115. Landegger v. Cohen, 5 F. Supp. 3d 1278, 1288 n.12 (D. Colo. 2013).

116. In their reports for *Larry P.*, Shepard's labeled *Landegger* as "abrogated in part as stated in," KeyCite labeled it as "abrogation recognized by," and BCite labeled it as a positive reference.

117. 802 F.2d 1539 (9th Cir. 1986).

118. *Id.* at 1542.

119. 724 F.2d 836 (9th Cir. 1984).

120. 730 F.2d 1214 (9th Cir. 1984).

121. *Wallace*, 802 F.2d at 1542.

122. *Id.* at 1550–52.

472

cited by name, then declared a rule of law that conflicts with the holding in those prior decisions. Readers who are more sympathetic to the citators might feel that this sort of treatment just isn't clear enough: they may feel that the reviewing court needs to connect the overruling more directly to the cited cases. In *Wallace*, the court did just that. At the end of its opinion, it said: "We overrule our previous decisions only to the extent that we have held that Commission decisions granting or denying parole are subject to judicial review for an abuse of discretion."[123] This was followed by a footnote citing cases including *Roth* and *Torres-Macias*.[124] In the reports for *Roth* and *Torres-Macias*, Shepard's and KeyCite correctly labeled *Wallace* as an overruling case, but BCite labeled *Wallace* as a positive case.

¶67 Most of the citing relationships I studied are less important than the ones discussed so far. But some of these relatively unimportant relationships involve especially blatant mistakes that cast further doubt on the citators' overall quality control. For example, in *In re Di Noto*, the Ninth Circuit Bankruptcy Appellate Panel spent three paragraphs and a footnote discussing *In re Comer*, referring to it by name six times.[125] The Panel distinguished *Comer* in three different ways, writing that "*Comer* is not applicable to our case,"[126] then in the next paragraph writing "a second reason for distinguishing *Comer* . . . ,"[127] and then in a footnote writing "there is also a third possible basis for distinguishing *Comer*."[128] Although Shepard's and BCite correctly labeled *In re Di Noto* as a distinguishing reference, KeyCite somehow failed to identify any negative treatment here.

¶68 In another mistake equally strange but going in the other direction, Shepard's indicated that *In re Peterson*[129] distinguished *United States v. Zolla*.[130] In fact, *Peterson* didn't cite *Zolla* at all.[131]

¶69 In *Roldan v. Rocette*, the Second Circuit spent over a full page in the *Federal Reporter* heaping criticism on a line of Ninth Circuit cases, including *Thorsteinsson v. INS*.[132] *Thorsteinsson* came in for particular scorn, with the Second Circuit writing that "[t]he critical flaw in the *Mendez* rule is especially evident in *Thorsteinsson*," followed by a full paragraph explaining why.[133] Despite this and three other citing cases that clearly gave *Thorsteinsson* negative treatment,[134] BCite gave *Thor-*

---

123. *Id.* at 1554.

124. *Id.* at 1554 n.10.

125. *In re* Di Noto, 46 B.R. 489, 491 (B.A.P. 9th Cir. 1984) (citing *In re* Comer, 723 F.2d 737 (9th Cir. 1984)).

126. *Id.* at 491.

127. *Id.*

128. *Id.* at 491 n.1.

129. 152 B.R. 329 (D. Wyo. 1993).

130. 724 F.2d 808 (9th Cir. 1984).

131. KeyCite and BCite didn't include *Peterson* as a citing case for *Zolla*, so I marked KeyCite and BCite as correct.

132. Roldan v. Rocette, 984 F.2d 85, 89–90 (2d Cir. 1993) (citing Thorsteinsson v. INS, 724 F.2d 1365 (9th Cir. 1984)).

133. *Id.* at 90.

134. The three other cases are *Santiago-Rodriguez v. Holder*, 657 F.3d 820, 831–33 (9th Cir. 2011) (distinguishing *Thorsteinsson*), *Baez v. INS*, 41 F.3d 19, 22–23 (1st Cir. 1994) (rejecting the Ninth Circuit's *Mendez* line of cases, including *Thorsteinsson*), and *Arreaza-Cruz v. INS*, 39 F.3d 909, 912 (9th Cir. 1994) (distinguishing *Thorsteinsson*).

steinsson a green plus sign indicating that *Thorsteinsson* hadn't received negative treatment from any case.[135]

¶70 BCite also bungled the report for *United States v. Beckett*,[136] which was discussed at length, featured in a block quotation, and rejected by the Eleventh Circuit in *United States v. Stone*,[137] and rejected another time by the Sixth Circuit in *United States v. Mari* with language including "we register our disagreement with the Ninth Circuit," and "[*Beckett*] is contrary to the Supreme Court's later decision . . . as well as illogical."[138] BCite missed the negative treatment in *Stone* and *Mari*, and managed to identify only one of two other cases that distinguished *Beckett*.[139]

¶71 I counted a total of 470 failures in the citators' validation of seventy-three Ninth Circuit cases. These failures can't be readily dismissed as insignificant. Although some failures are trivial, others are harmful; and while some failures are hard to see, others are blatant. The bottom line is that users can make no safe assumptions about what their citators will do for them.

Responses from the Database Providers

¶72 Thomson Reuters (publisher of Westlaw), LexisNexis, and Bloomberg Law reviewed a draft version of this article and my dataset prior to publication. All three companies offered some comments, and LexisNexis and Bloomberg Law also provided lists of citing references they disputed.[140] In response, I made some revisions to the text and agreed to recode four citing references,[141] but many disagreements remained.

¶73 Thomson Reuters conceded that I had identified some incorrect determinations in KeyCite, but said they disagreed with most of my results and conclusions. They did not provide examples. They wrote that the "complex nature of legal research, and analyzing judicial opinions specifically, does not lend itself to an 'objectively correct' interpretation," and claimed that most of their labels involve "subjective determinations."[142] They further wrote that "Thomson Reuters stands behind KeyCite. It is the most complete, accurate and up to the minute citator available in the market."[143]

¶74 LexisNexis and Bloomberg Law offered more specific comments. They both protested that I sometimes marked as incorrect citing references that had been

---

135. KeyCite correctly identified all of *Thorsteinsson*'s negative treatment, and Shepard's correctly identified all but *Santiago-Rodriguez*.

136. 724 F.2d 855 (9th Cir. 1984).

137. 9 F.3d 934, 940 (11th Cir. 1993).

138. 47 F.3d 782, 787 (6th Cir. 1995).

139. The distinguishing reference that BCite identified was from *United States v. Defazio*, 899 F.2d 626, 636 (7th Cir. 1990). The distinguishing reference that BCite missed was from *United States v. Mang Sun Wong*, 884 F.2d 1537, 1542–43 (2d Cir. 1989). Shepard's and KeyCite correctly identified all the negative treatment for *Beckett*.

140. All correspondence is on file with the author. Lexis listed forty citing references in dispute, while Bloomberg Law listed twenty-nine.

141. I agreed to recode as ambiguous three citing references that were somewhat unclear, and for another citing reference I agreed to credit BCite instead of Shepard's and KeyCite due to a reading error on my part.

142. Letter from Leann Blanchfield, Vice President for Product Development, Thomson Reuters, to author (June 10, 2018) (on file with author).

143. *Id.*

handled correctly according to their own internal procedures. Naturally, the results presented here would be more favorable to all three citators if I judged each citator according to its own procedures, but as I explained previously, I judge the citators according to their end results from the perspective of a typical user.

¶75 Bloomberg Law's key objection not only challenges my methodology, but perhaps questions the overall relevance of this article. As noted earlier, BCite uses a relatively short list of negative labels.[144] For many types of negative treatment (such as "called into doubt" or "disagreement recognized by"), BCite simply has no label that can be applied. This issue accounts for roughly one-fifth of the missing negative labels I counted in BCite. In their comments to me, Bloomberg Law insisted that this design feature is not a weakness and should not have been counted against them. They explained that they made a deliberate decision to use a short list of negative labels in response to market research showing that most users don't want as many negative labels as Shepard's and KeyCite offer.[145] Some users may agree with Bloomberg's decision, but in any case it's important for users to be aware of this significant difference between the citators.[146]

¶76 LexisNexis defended Shepard's policy of withholding negative labels from what it calls "citing convention negatives." An example explains the general idea. In *AARP v. Farmers Group, Inc.*, the Ninth Circuit briefly cited one of its earlier decisions as follows: "*Cf. E.E.O.C. v. Borden's Inc.*, 724 F.2d 1390, 1396–97 (9th Cir.1984), *overruled on other grounds, Betts*, 492 U.S. at 173–75."[147] In its report for *Borden's*, Shepard's did not apply an "overruled as stated in" label to *AARP* because its negative treatment of *Borden's* can be attributed to a "citing convention." I was unable to determine how Shepard's defines a "citing convention," but the idea is that these citing convention negatives are too inconsequential to note.[148] According to LexisNexis, this policy accounts for twelve of the missing labels I counted in Shepard's.[149] I decided not to exclude these missing labels from my statistics. After receiving LexisNexis's explanation and viewing many examples, I still don't fully understand how Shepard's applies this policy, so I don't believe it would be clear to the typical user. Moreover, these "citing convention negatives" are not necessarily

---

144. *See supra* ¶ 37.

145. Letter from Darby J. Green, *supra* note 59. In my view, a truncated list of negative labels is not an improvement. I realize that users are often uninterested in certain types of negative labels, but in KeyCite and Shepard's they can choose to focus on whatever labels they feel are important. If some users find KeyCite and Shepard's confusing, BCite could have offered improved customization so that individual users could easily choose the level of detail they prefer.

146. BCite's more limited approach is not immediately clear to the user. Although Bloomberg publishes a complete list of its BCite labels, it doesn't indicate that the list is complete. On the contrary, it presents the list as "a description of the main components," which suggests there may be additional components not listed. *Bloomberg Law Citator, supra* note 48. Knowing that Westlaw and Lexis publish similar lists of citator labels that are not complete, I assumed at first that BCite's list was not complete. Moreover, a typical user is not likely to compare the lists from the different citators, determine what's missing from BCite, and realize how this impacts the validation process.

147. 943 F.2d 996, 1004 (9th Cir. 1991).

148. As LexisNexis explains, "our policy is based on the rationale that including citing convention references artificially inflates negative citing references and results in researchers having to go check cases to review these allegedly negative references unnecessarily." Letter from Liz Christman, *supra* note 72.

149. *Id*. I can't verify that number since I don't understand the boundaries of the policy.

inconsequential. I have already discussed the serious deficiencies in the Shepard's report for *Borden's*,[150] and the missing label for *AARP* is part of the problem.[151]

¶77 LexisNexis also pointed out its policy of withholding editorial analysis labels from citing references in unpublished California Court of Appeal decisions.[152] This accounts for five of the missing negative labels I counted in Shepard's.[153] According to LexisNexis, the rationale for its policy "is that California follows a strict rule of not citing unpublished opinions as precedential authority."[154] Again, I decided not to exclude these missing labels from my statistics. From the user's perspective, Shepard's policy looks like a mistake. For example, in its report for *Javor v. United States*,[155] Shepard's lists seven citing references from unpublished California Court of Appeal cases. These look like any other citing cases in a Shepard's report. All seven have a blue square with a "cited by" label (which indicates no negative treatment), but a user who views these cases would discover that at least three of them are in fact negative. Of course, these are minor problems since these citing cases are of little importance,[156] but I don't separate major and minor problems in my statistics.

¶78 Finally, there were additional, specific citing references that Bloomberg Law and/or LexisNexis argued were simply ambiguous. Deciding which citing references to code as ambiguous is perhaps the most subjective part of this study, but lines must be drawn somewhere. Some of the references we disagreed on were admittedly close to the line,[157] while many others struck me as perfectly clear.[158] I admit that different researchers might choose to code somewhat more or somewhat

---

150. *Supra* ¶¶ 46, 59–60.

151. When a citing convention negative notes an express overruling, the information is usually redundant. But redundancy can be a good way of mitigating errors like the one that occurred in *Borden's* Shepard's report. In other situations, citing convention negatives might point out implicit overrulings that would not otherwise be noted in a citator report.

152. E-mail from Liz Christman to author (June 13, 2018, 18:34 EST) (on file with author).

153. A sixth unpublished California Court of Appeal decision was omitted altogether from a Shepard's report.

154. E-mail from Liz Christman, *supra* note 152.

155. 724 F.2d 831 (9th Cir. 1984).

156. These citing cases are unimportant, but not irrelevant. In this study, the typical user is assessing the validity of published Ninth Circuit cases. The question is whether the Ninth Circuit cases should be cited, not whether any unpublished California cases should be cited. Viewed in this context, unpublished California cases should have the same, limited relevance as any other unpublished cases.

157. For example, *Sheley v. Dugger*, 833 F.2d 1420 (11th Cir. 1987), cited two cases called *Toussaint*. First, it cited *Toussaint v. McCarthy*, 801 F.2d 1080 (9th Cir. 1986), which is not part of my dataset. *Sheley*, 833 F.2d at 1425. *Sheley* referred to this case with the shorthand reference "*Toussaint*" and clearly distinguished it. *Id.* at 1426–27. Immediately after this discussion, *Sheley* cited *Toussaint v. Yockey*, 722 F.2d 1490 (9th Cir. 1984), which is part of my dataset. *Id.* at 1427. Although most readers would infer that the two *Toussaints* are related cases, *Sheley* doesn't say this explicitly. KeyCite and Shepard's noted no distinguishing treatment here for *Toussaint v. Yockey*, but BCite did. I marked BCite as incorrect. I thought it was clear enough that the two *Toussaints* should not be treated as one and the same, but some readers might see ambiguity here.

158. For example, almost the entire opinion in *Berryman v. Wong*, 2010 U.S. Dist. LEXIS 9910 (E.D. Cal. Jan. 15, 2010), was about the applicability of *Javor v. United States*, 724 F.2d 831 (9th Cir. 1984). The court extensively discussed the factual differences between the two cases and concluded that "[t]he presumed prejudice standard under *Javor* does not apply." *Berryman*, 2010 U.S. Dist. LEXIS, at *18. Shepard's and BCite failed to apply a "distinguished" label here, which LexisNexis argued was just a "judgment call." Letter from Liz Christman, *supra* note 72. Bloomberg Law did not dispute this reference.

fewer citing references as ambiguous, but I doubt these differences would result in much change to the overall outcome.

Conclusion

¶79 This study evaluates one key aspect of legal citators: their performance in flagging negative citing references. Other important aspects such as currency, overall retrieval of citing references, labeling of positive references, ease of use, customer satisfaction, and cost are not covered here. This study is not an assessment of any citator's overall merit. I do not claim that any citator is "best" or "worst."

¶80 I strove to make this study as fair and objective as possible, but any evaluation of the citators' case validation performance will rest partly on personal opinion. I've mitigated the role of personal opinion by attempting to exclude ambiguous citing relationships from my statistics, disclosing my methodology, offering many examples of what I consider to be errors, and giving the benefit of the doubt to the citators on close questions. Some readers may remain unconvinced. But it's not only my own judgment that conflicts with the citators—for almost 90% of the negative labels reviewed here, at least one citator conflicts with another.[159] If these discrepancies can't be objectively resolved, then the citators themselves are based on nothing more than idiosyncratic opinion. If true, this would undermine the citators' usefulness. The citators can be reliable or they can be idiosyncratic, but they can't be both.

¶81 I believe that most citing relationships are clear and can be objectively described, that labels in citators can be right or wrong, and that all three citators can and should do better. I don't expect perfection. But surely there is room for improvement, especially in the most important citing relationships, such as citations from the U.S. Supreme Court. In the meantime, users may need to reconsider the trust they place in citators, and law librarians may need to rethink how they discuss citators with their patrons. Citators will always be an essential part of the legal research process, but researchers need to be aware of the citators' shortcomings. Relying only on a citator's treatment symbols is a risky strategy for case validation.

---

159. *See supra* ¶ 48.