Citation Databases for Legal Scholarship

John R. Beatty
University at Buffalo School of Law, jrbeatty@buffalo.edu

This is an Accepted Manuscript of an article published by Taylor & Francis in Legal Reference Services Quarterly on March 19, 2020, available online: https://www.tandfonline.com/doi/full/10.1080/0270319X.2020.1738198.

Citation Databases for Legal Scholarship

John R. Beatty [a]*

[a] *Charles B. Sears Law Library, University at Buffalo, State University of New York, Buffalo, NY, USA*

John Beatty, Charles B. Sears Law Library, O'Brian Hall, Buffalo, NY 14260, jrbeatty@buffalo.edu, ORCID ID: 0000-0002-4940-8483

John R. Beatty is the Faculty Scholarship Outreach Librarian at the Charles B. Sears Law Library at the University at Buffalo.

This is an Accepted Manuscript of an article published by Taylor & Francis in Legal Reference Services Quarterly on March 19, 2020, available online: https://www.tandfonline.com/doi/full/10.1080/0270319X.2020.1738198.

Citation Databases for Legal Scholarship

Traditional citation sources, such as Web of Science, index limited numbers of law journals. Consequently, although not designed for generating scholarship citation metrics, many law scholarship citation studies use law-specific databases like Westlaw or LexisNexis to gather citations. This article compares citation metrics derived from Web of Science and Westlaw to metrics derived from Google Scholar and HeinOnline's citation tools. The study finds that HeinOnline and Westlaw generate higher metrics than Web of Science, and Google Scholar generates higher metrics than both. However, metrics from all four sources are highly correlated, so rankings generated from any may be very similar.

Keywords: citation databases; rankings; citation analysis; research metrics; bibliometrics; unique citations; citation overlap; h-index

Introduction

Impact is an important part of the scholarship life cycle. Scholars contribute to the conversation in their research areas through their scholarship. Although an imperfect measurement, citation metrics are often used to measure the impact of a single publication, an individual scholar, a faculty, or a journal on its subject area. Although there are newer metrics that use more immediately available measurements, such as downloads or media mentions, citation metrics remain the predominant measurement.

The reasons to treat citation metrics with skepticism are many. Metrics, like the Journal Impact Factor (JIF), are routinely abused. Publishers advertise the JIFs of their journals to entice authors to submit. Administrators assume that the JIF is a measure of a typical citation rate, but many articles have a lower citation rate for reasons that have little to do with article quality.[1] Indeed,

[1] Philip Campbell, *Escape from the Impact Factor*, 8 ETHICS SCI. & ENVTL. POL. 5, 5–6 (2008).

the JIF of a particular journal says nothing about the quality of a specific article in the journal.[2] Reference standards may fail to include articles published in multidisciplinary journals, which in some cases may not capture "a considerable portion" of the relevant literature.[3] Further, metrics are also subject to manipulation. Recent studies show that the standard bibliometrics have become targets, and scholarly publishing has been altered to better conform to the target metrics.[4]

Another reason to distrust citation metrics is the questionable accuracy and completeness of the source of citation data.[5] Most sources contain only a subset of journals. In many cases, interdisciplinary publications are not well covered. Books and conference proceedings are missing from others. Within a database, citations may not be correctly captured by the algorithms that match a citation to its source publication.[6]

In the sciences, two sources reign supreme: Web of Science and Scopus. Web of Science is the

---

[2] Berenika M. Webster, *Principles to Guide Reliable and Ethical Research Evaluation Using Metric-Based Indicators of Impact*, 18 PERFORMANCE & MEASUREMENT METRICS 5 (2017).

[3] Lutz Bornmann et al., *Citation Counts for Research Evaluation: Standards of Good Practice for Analyzing Bibliometric Data and Presenting and Interpreting Results*, 8 ETHICS SCI. & ENVTL. POL. 93, 98 (2008).

[4] Michael Fire & Carlos Guestrin, *Over-Optimization of Academic Publishing Metrics: Observing Goodhart's Law in Action*, 8 GIGASCIENCE 2 (2019), https://academic.oup.com/gigascience/article/8/6/giz053/5506490 (last visited June 4, 2019).

[5] *See* María del Carmen Calatrava Moreno, Thomas Auzinger & Hannes Werthner, *On the Uncertainty of Interdisciplinarity Measurements due to Incomplete Bibliographic Data*, 107 SCIENTOMETRICS 213, 214 (2016).

[6] Juan Gorraiz & Christian Schloegl, *A Bibliometric Analysis of Pharmacology and Pharmacy Journals: Scopus Versus Web of Science*, 34 J. INFO. SCI. 715, 721 (2008).

older, more established product. Scopus is newer, with slightly different coverage. Although both are used heavily in the hard sciences, their inadequacy in other areas has been noted by a number of researchers.[7] Google Scholar is a newer, freely available source of citation data that is often more complete in subject areas where Web of Science and Scopus have little content. Finally, some subject areas, like law and nursing, have databases that can be used to derive citation metrics, despite not being dedicated citation databases themselves.

Citation studies in law are similar to other studies in the social sciences in that the subjects are covered by Web of Science and Scopus, but not well. But law also has its unique challenges. The legal scholarship publishing regime is unlike most other disciplines. There, the main avenue of publication is not in peer-reviewed journals published by large academic journal publishers or scholarly societies, but in legal journals published by law schools and run by student editors.

Competition is fierce to get published in the top journals. Authors submit pieces to multiple journals and use publication offers from one journal as leverage to get the article published in a higher-ranked journal. The top journals are largely those published by the law schools ranked highest in the U.S. News & World Report rankings.[8] Those top journals largely feature articles written by professors at those same top-ranked schools.[9] This leaves a lot of quality scholarship

[7] *See, e.g.,* Anne-Wil Harzing & Satu Alakangas, *Google Scholar, Scopus and the Web of Science: A Longitudinal and Cross-Disciplinary Comparison*, 106 SCIENTOMETRICS 787, 788 (2016).

[8] A.H. Yoon, *Editorial Bias in Legal Academia*, 5 J. LEGAL ANALYSIS 309, 315 (2013).

[9] Lawprofblawg & Darren Bush, *Law Reviews, Citation Counts, and Twitter (Oh My!): Behind the Curtains of the Law Professor's Search for Meaning*, 50 LOY. U. CHI. L.J. 327, 335 (2018).

to lower-ranked journals.[10] Furthermore, just about every ABA-approved law school publishes at least one journal, and most publish more than one, so there are many outlets for publication. Because so much legal scholarship is published, much of it is never cited.[11]

Due to the manner in which articles are selected for the top journals, coverage matters. There are hundreds of legal journals are published in the United States. Web of Science and Scopus selec only a small portion of them. This approach may be inadequate for assessing the impact of a scholar or faculty because many citations or articles published in non-indexed journals will be missed. Moreover, both have selected journals with low citation rates published by large commercial publishers over more highly cited journals published by law schools. Consequently, law-specific citation sources are often used to assess legal scholarship. But because of the lack of interdisciplinary coverage in those sources, the use of law-specific sources may disadvantage scholars who publish in other disciplines.

This article seeks to assess the differences in coverage of legal scholarship between several of the most-used citation databases and the effects of those differences on the most heavily used citation metrics.

Previous Studies

Several studies have examined the differences in citations and citation metrics, such as between Web of Science and Google Scholar, and between Web of Science, Scopus, and Google Scholar.

[10] Dennis J. Callahan & Neal Devins, *Law Review Article Placement: Benefit or Beauty Prize?*, 56 J. LEGAL EDUC. 374, 385 (2006).

[11] Jeffery L. Harrison & Amy R. Mashburn, *Citations, Justifications, and the Troubled State of Legal Scholarship: An Empirical Study*, 3 TEX. A&M L. REV. 45, 79 (2015).

A few have also examined Web of Science and Scopus, and at least one study covered all three sources, along with a single disciplinary database. Because of the difficulty of cleaning citation data, especially that from Google Scholar, studies are often confined to a small number of scholars in one discipline or institution. Others study only the citations to a specific set of journals. Most of these studies analyze only a small subset of basic citation metrics: number of citations, number of papers, and h-index. The h-index is a single number that "measures the broad impact of an individual's work" while avoiding the problems of other indices and also allowing direct author comparisons.[12] An author has "index h if h of his or her N papers have at least h citations each and the other $(N - h)$ papers have $\leq h$ citations each.[13]

Results of Previous Studies

Mingers and Lipitakis compared Web of Science and Google Scholar in the fields of business and management by examining citations to the scholarship of all academics at three U.K. business schools.[14] They found that only 45% of the scholarship appeared in journal articles, with the remaining 55% in conference papers, books, and other sources not covered by Web of Science.[15] Google Scholar returned 89% of journal articles and 66% of all publications, while Web of Science returned only 48% of journal articles.[16] Google Scholar found considerably

---

[12] ANA ANDRÉS, MEASURING ACADEMIC RESEARCH 62 (2009).

[13] *Id.*

[14] John Mingers & Evangelia A.E.C.G. Lipitakis, *Counting the Citations: A Comparison of Web of Science and Google Scholar in the Field of Business and Management*, 85 SCIENTOMETRICS 613, 614 (2010).

[15] *Id.* at 618.

[16] *Id.*

more citations as well, with 14.7 citations per paper to Web of Science's 8.4.[17]

Bar-Ilan compared the h-indexes generated by Web of Science, Scopus, and Google Scholar for a list of forty heavily cited Israeli researchers.[18] That study found that there was no significant difference in the h-indexes generated by Web of Science and Scopus.[19] Google Scholar results were similar for many of the researchers, but there was some discipline-specific bias.[20] Many of the researchers had similar h-indexes (within 30% higher or lower), but most of the mathematicians and computer scientists had an h-index over 30% higher on Google Scholar, and all three of the high-energy physicists had h-indexes more than 30% lower.[21]

Franceschet compared Web of Science and Google Scholar citations for a group of Italian computer scientists.[22] He examined the statistical correlation between the two sources for a number of metrics and found good correlation for citation-based metrics and moderate correlation for paper-based metrics, with h-type indexes spread throughout the group.[23]

Adriaanse and Rensleigh, in a total citation study covering five years of nine South African

---

[17] *Id.* at 621.

[18] Judit Bar-Ilan, *Which H-Index?—A Comparison of WoS, Scopus and Google Scholar*, 74 SCIENTOMETRICS 257, 262 (2008).

[19] *Id.* at 265.

[20] *Id.* at 267.

[21] *Id.*

[22] Massimo Franceschet, *A Comparison of Bibliometric Indicators for Computer Science Scholars and Journals on Web of Science and Google Scholar*, 83 SCIENTOMETRICS 243, 245 (2010).

[23] *Id.* at 251.

environmental sciences journals, found a 62.5% overlap between the three sources.[24] Web of Science had the largest number of unique citations, with Google Scholar not far behind, and Scopus having very few.[25]

Meho and Yang, in a study of library and information science faculty, found a 58% overlap between Web of Science and Google Scholar, with Scopus having almost two-thirds of the total unique citations.[26] The authors also found that Google Scholar identified 53% more citations than Scopus and Web of Science combined, and that its unique citations were 48% of the total number of citations.[27]

Vanclay, asserting that the h-index is robust, and that the great majority of errors in citation databases are in the "long tails" that do not greatly affect the h-index, compared his own publication record in Web of Science and Google Scholar.[28] After correcting obvious errors in the data, he calculated his h-index using the data from each source.[29] Using Google Scholar as the data source instead of Web of Science increased his h-index by one, from twelve to

---

[24] Leslie S. Adriaanse & Chris Rensleigh, *Web of Science, Scopus and Google Scholar: A Content Comprehensiveness Comparison*, 31 ELECTRONIC LIBR. 727, 734 (2013).

[25] *Id.*

[26] Lokman I. Meho & Kiduk Yang, *Impact of Data Sources on Citation Counts and Rankings of LIS Faculty: Web of Science Versus Scopus and Google Scholar*, 58 J. AM. SOC'Y INFO. SCI. & TECH. 2105, 2113–14 (2007).

[27] *Id.* at 2115–16.

[28] Jerome K. Vanclay, *On the Robustness of the H-Index*, 58 J. AM. SOC'Y INFO. SCI. & TECH. 1547, 1547–48 (2007).

[29] *Id.*

thirteen.[30]

De Groote and Raszewski examined the h-indexes of nursing researchers in Web of Science, Scopus, Google Scholar, and a sole disciplinary database.[31] They found a strong correlation between h-indexes derived from each source.[32] They also generated h-indexes from citations aggregated from multiple sources and found a strong correlation between the aggregated metrics and single-source metrics.[33] In addition, they found significant overlap between citations in the various databases.[34] Unique citations ranged from 6.5% (Web of Science) to 37.5% (Google Scholar) in a sample of thirty articles.[35]

Martín-Martín et al. found, in a comparison of citations across 252 subject categories, a strong correlation in citation counts between Scopus and Google Scholar, and between Web of Science and Google Scholar.[36]

Harzing and Alakangas, looking at 146 senior academics in five broad disciplinary areas, found that Scopus had more citations than Web of Science in all areas except the sciences, while Google Scholar had more in all areas than either of the other two.[37] Similarly, they found that h-

---

[30] *Id.*

[31] Sandra L. De Groote & Rebecca Raszewski, *Coverage of Google Scholar, Scopus, and Web of Science: A Case Study of the H-Index in Nursing*, 60 NURSING OUTLOOK 391, 394 (2012).

[32] *Id.* at 397.

[33] *Id.*

[34] *Id.* at 398.

[35] *Id.*

[36] Alberto Martín-Martín et al., *Google Scholar, Web of Science, and Scopus: A Systematic Comparison of Citations in 252 Subject Categories*, 12 J. INFORMETRICS 1160, 1172–73 (2018).

[37] Harzing and Alakangas, *supra* note 7, at 796.

indexes were slightly higher in Scopus than Web of Science, and higher still in Google Scholar.[38] The increase in h-index from Scopus to Google Scholar was slight in the sciences and life sciences, a bit larger in engineering (33%), and considerably higher in the social sciences (79%) and humanities (286%).[39]

In a study of 340 soil researchers, Minasny et al. found that, although there was a large difference in the number of citations, number of publications, and h-indexes between Web of Science, Scopus, and Google Scholar, there was a high correlation between those metrics from each.[40]

Bakkalbasi et al. looked at the overlap in citations in oncology and condensed matter physics between Google Scholar, Scopus, and Web of Science. They found that, both the degree of overlap for each database, and the database with the highest number of unique citations, was different for each discipline.[41]

While it does not appear that there is a comparable study in law, there have been two recent scholarly rankings of law faculties, with each using a separate citation source. The Leiter-Sisk rankings, first published as a ranking of twenty-five law faculties by Brian Leiter in 2007 (and

---

[38] *Id.* at 797.

[39] *Id.*

[40] Budiman Minasny et al., *Citations and the H Index of Soil Researchers and Journals in the Web of Science, Scopus, and Google Scholar*, 1 PEERJ e183, 4, 6 (2013).

[41] Nisa Bakkalbasi et al., *Three Options for Citation Tracking: Google Scholar, Scopus and Web of Science*, 3:7 BIOMEDICAL DIGITAL LIBR. 1, 6–7 (2006), https://bio-diglib.biomedcentral.com/articles/10.1186/1742-5581-3-7 (last visited June 13, 2019).

updated in 2010), used author searches in Westlaw to find citations.[42] Leiter's rankings were extended to seventy schools in 2010 by Gregory Sisk.[43] Sisk has updated the rankings every three years, most recently in 2018.[44] Paul Heald and Ted Sichelman, in a 2019 ranking of one hundred law faculties, used citation data from HeinOnline.[45] Although both the methodologies and data sources of the two studies were different, their final rankings were highly correlated (0.88).

Methodologies of Previous Studies

Mingers and Lipitakis used known-publications lists for each researcher and looked up each publication individually in Web of Science and Google Scholar.[46] Franceschet used name searches in each source, but studied a group of computer scientists in his own department in order to check results against known-publications lists.[47] Meho and Yang examined library and information science citations by studying the full output of the fifteen faculty members in their

---

[42] Brian Leiter, *Top 35 Law Faculties Based on Scholarly Impact, 2007* (2007), http://www.leiterrankings.com/faculty/2007faculty_impact.shtml (last visited Jan. 10, 2020); Brian Leiter, *Top 25 Law Faculties in Scholarly Impact, 2005–2009* (2010), http://www.leiterrankings.com/new/2010_scholarlyimpact.shtml (last visited Jan. 10, 2020).

[43] Gregory C. Sisk et al., *Scholarly Impact of Law School Faculties: Extending the Leiter Rankings to the Top 70 (2010)*, (U. St. Thomas Legal Res. Stud. Paper No. 10-24, 2010), http://www.ssrn.com/abstract=1674764 (last visited Jan. 10, 2020).

[44] Gregory C. Sisk et al., *Scholarly Impact of Law School Faculties in 2018: Updating the Leiter Score Ranking for the Top Third*, 15 U. ST. THOMAS L.J. 95 (2018).

[45] Paul J. Heald & Ted Sichelman, *Ranking the Academic Impact of 100 American Law Schools*, 60 JURIMETRICS J. 1, 4 (2019).

[46] Mingers and Lipitakis, *supra* note 14, at 618.

[47] Franceschet, *supra* note 22, at 247, 249.

department.[48] They used the cited-author search in Web of Science and known-item title searches in Scopus and Google Scholar.[49] De Groote and Raszewski looked at the h-indexes of thirty members of one nursing faculty as well as the total number of citations and the overlap between databases for a subset of articles.[50]

Sisk, as well as Heald and Sichelman, used faculty lists for the law faculties studied.[51] Sisk compiled a preliminary list and asked deans at each school to confirm that the list contained all tenured faculty, with a 97% response rate.[52] Heald and Sichelman compiled an initial list of traditional tenured and tenure-track faculty using the *AALS Directory of Law Teachers 2015–2016*, and they then asked deans at each school to confirm that the lists were correct, with a 60% response rate.[53] The lists were further corrected using information from HeinOnline.[54] Sisk used targeted author searches to find citations in Westlaw.[55] In contrast, Heald and Sichelman obtained raw citation counts for each author and law review in HeinOnline's database.[56] Like other citation databases, those citation counts are tied to the articles in the HeinOnline database.

---

[48] Meho and Yang, *supra* note 26, at 2110.

[49] *Id.*

[50] De Groote and Raszewski, *supra* note 31, at 393–94.

[51] Sisk et al., *supra* note 44, at 108–09; Heald and Sichelman, *supra* note 45, at 7–8.

[52] Sisk et al., *supra* note 44, at 109.

[53] Heald and Sichelman, *supra* note 45, at 8.

[54] *Id.*

[55] Sisk et al., *supra* note 44, at 109–110.

[56] Heald and Sichelman, *supra* note 45, at 9.

Because subject searches are difficult to compare between databases, Bakkalbasi et al. selected a random sample of journal articles from a random sample of journals in two disciplines, then gathered citations using known-item searches.[57] Martín-Martín et al. used a sample of papers taken from Google Scholar's Classic Papers and analyzed them using several different subject classification schemes.[58] Minasny et al. gathered researchers who listed interests in various fields of soil research in their Google Scholar author profiles and compiled publication lists for them using Web of Science, Google Scholar, and Scopus.[59]

Methodology

For this study, ten authors were chosen from the University at Buffalo School of Law faculty. Every fifth faculty member was selected from the full faculty list. If the chosen faculty member had five or fewer publications, the next faculty member on the list was selected instead. The resulting set of authors covered faculty members with publication histories spanning from twelve to fifty-two years and quantitively, from nine to eighty-one publications. A number of research areas were included, with little overlap between authors. Table 1 shows the publication areas of each of the ten chosen authors.

(INSERT TABLE 1)

Citations were gathered from four sources: Westlaw, Web of Science, HeinOnline, and Google Scholar. The citations from each source were matched to determine which citations were unique to each source, along with the degree of overlap between the four sources.

[57] Bakkalbasi et al., *supra* note 41.

[58] Martín-Martín et al., *supra* note 36, at 1162, 1172–74.

[59] Minasny et al., *supra* note 40, at 3.

Sources

Web of Science

Web of Science[60] is the oldest citation index. It is a selective database that purports to include the journals with the biggest citation impact in each discipline.[61] Its coverage of disciplines outside of the hard sciences, however, is not as robust as its marketing materials make it seem. Various studies have noted its lack of coverage in several areas, particularly when compared with Google Scholar.[62] Web of Science's coverage of law is especially troubling.[63] Washington and Lee University School of Law publishes an annual law journal impact ranking that is widely used by law schools.[64] Web of Science only indexes about three-quarters of Washington and Lee's top fifty journals. Additionally, about half of the journals it indexes are specialty journals published by large academic journal publishers, most of which rank below five hundred in Washington and

---

[60] WEB OF SCIENCE, http://webofknowledge.com (last visited Sept. 12, 2019).

[61] Clarivate, *Web of Science Core Collection*, WEB OF SCIENCE, https://clarivate.com/webofsciencegroup/solutions/webofscience-core-collection-editorial-selection-process/ (last visited Sept. 11, 2019).

[62] Meho and Yang, *supra* note 26, at 2105–06; Mingers and Lipitakis, *supra* note 14, at 615, 624–25; Franceschet, *supra* note 22, at 256–57.

[63] *See, e.g.*, Andrew T. Hayashi & Gregory Mitchell, *Maintaining Scholarly Integrity in the Age of Bibliometrics*, __ J. LEGAL EDUCATION __, at 8, n.21 (forthcoming 2020), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3419054 (last visited Feb. 25, 2020) (noting the lack of coverage of American legal sources as a probable reason for its nonuse in studies of those sources).

[64] Washington and Lee University Law Library, *W&L Law Journal Rankings*, WASH. & LEE SCH. OF L., https://managementtools4.wlu.edu/LawJournals/ (last visited Feb 25, 2020).

Lee's list.[65]

Westlaw

Westlaw[66] is, along with LexisNexis[67], one of the two largest legal information platforms. Both are aimed at practicing attorneys, but they contain a large amount of secondary material, including an almost complete collection of U.S. law reviews. Although neither offers a dedicated citation index, Westlaw and Lexis are often used for law citation studies because their coverage is more complete than Web of Science.[68] Additionally, both platforms include a citation report feature for primary legal materials (KeyCite and Shepard's, respectively), which includes secondary materials like law reviews. Washington and Lee's law journal rankings are researched

[65] As of the 2018 W&L Law Journal Rankings (released in September 2019), only the top three hundred U. S. law journals and the top fifty law journals published outside of the United States are ranked. The remaining journals are marked "NR" for every data category. *See* Washington and Lee University Law Library, *About & How to Use*, WASH. & LEE SCH. OF L., https://managementtools4.wlu.edu/LawJournals/Default2.aspx (last visited Sept. 11, 2019). The comparison discussed here was made between the 2017 Journal Citation Reports and the 2017 W&L Law Journal Rankings.

[66] WESTLAW, https://westlaw.com (last visited Sept. 12, 2019).

[67] Welcome to LexisNexis Legal & Professional, https://www.lexisnexis.com/en-us/home.page (last visited Sept. 12, 2019).

[68] *See, e.g.,* James M. Donovan & Carol A. Watson, *Citation Advantage of Open Access Legal Scholarship*, 103 LAW LIBR. J. 553 (2011); James M. Donovan, Carol A. Watson & Caroline Osborne, *The Open Access Advantage for American Law Reviews*, 2015–03A EDISON 1 (2014); John P. Joergensen, *Second Tier Law Reviews, Lexis, and Westlaw: A Pattern of Increasing Use*, 21 LEGAL REFERENCE SERVS. Q. 43 (2002); Sisk et al., *supra* note 41; Karen L. Wallace & M. Sara Lowe, *HeinOnline and Law Review Citation Patterns*, 103 L. LIBR. J. 55 (2011).

using Westlaw.[69] Westlaw is also used in this study due to the comparative ease with which a researcher can export both search results and KeyCite reports.

Google Scholar

Google Scholar is Google's entry into the scholarly search landscape.[70] Its interface does not allow easy extraction of citation metrics, but the Publish or Perish tool, developed by Anne-Wil Harzing, enables the use of Google Scholar as a back end database for citation studies.[71] Publish or Perish allows a researcher to search Google Scholar, then easily clean up the results to remove duplicates and false positives. Once the results are cleaned, Publish or Perish will calculate a number of popular citation metrics.

HeinOnline

HeinOnline is a database of legal materials, which includes an extensive collection of full-text law review articles, American Law Institute (ALI) materials, state bar association publications, and foreign materials.[72] Although primarily a research database, HeinOnline offers article and case citation counts for works in the database. Article results are culled from its own database, while case citations are available through a partnership with Fastcase.[73] Citation information was

---

[69] Washington and Lee University Law Library, *Ranking Methodology*, WASH. & LEE SCH. OF L., https://managementtools4.wlu.edu/LawJournals/Default3.aspx (last visited Sept. 12, 2019).

[70] Google Scholar, GOOGLE.COM, https://scholar.google.com/ (last visited Sept. 12, 2019).

[71] Anne-Wil Harzing, *Publish or Perish*, HARZING.COM (Feb. 6, 2016), https://harzing.com/resources/publish-or-perish (last visited Sept. 12, 2019).

[72] HEINONLINE, https://home.heinonline.org/ (last visited Sept. 12, 2019).

[73] *Fastcase | HeinOnline*, HEINONLINE.ORG, https://home.heinonline.org/content/Fastcase/ (last visited Sept. 12, 2019).

first added to HeinOnline in 2009, when the first version of its ScholarCheck ranking was launched.[74] Five years later, it introduced author profiles, which gather all of an author's works together on one page with citation counts.[75]

In late 2017, HeinOnline greatly expanded its functionality by assembling the citation counts for all of an institution's authors in a single comma-separated values (CSV) file that can be downloaded by users.[76] HeinOnline has not yet been used widely for legal citation studies. Heald and Sichelman appear to have published the first such study.[77] In February 2019, shortly after the data was first gathered for the present study, U.S. News & World Report announced that it would be compiling a scholarly ranking of U.S. law faculties using citation data from HeinOnline.[78] Its choice of HeinOnline as a citation data source may be related to early results from Heald and Sichelman.[79]

---

[74] Shane Marmion, *Most Cited Authors in HeinOnline—Hein's ScholarCheck*, HEINONLINE BLOG (Jan. 30, 2009), https://home.heinonline.org/blog/2009/01/most-cited-authors-in-heinonline-heins-scholarcheck/ (last visited Sept. 12, 2019).

[75] Shannon Furtak, *New Feature: Author Profile Pages*, HEINONLINE BLOG (Dec. 31, 2014), https://home.heinonline.org/blog/2014/12/new-feature-author-profile-pages/ (last visited Sept. 12, 2019).

[76] Lauren Mattiuzzo, *NEW: Citation Data Extraction Is Now Available for Author Profile Pages*, HEINONLINE BLOG (Nov. 17, 2017), https://home.heinonline.org/blog/2017/11/new-citation-data-extraction-is-now-available-for-author-profile-pages/ (last visited Sept. 12, 2019).

[77] Heald and Sichelman, *supra* note 45.

[78] Robert Morse, *U.S. News Considers Evaluating Law School Scholarly Impact*, U.S. NEWS & WORLD REP. (Feb. 13, 2019), https://www.usnews.com/education/blogs/college-rankings-blog/articles/2019-02-13/us-news-considers-evaluating-law-school-scholarly-impact (last visited Sept. 12, 2019).

[79] Heald and Sichelman, *supra* note 45, at 6.

Since the announcement, Hein has worked to improve its matching algorithms in order to increase the accuracy of its citation counts.[80] Hein has also quietly added more metrics, including the h-index, to the author profiles and CSV downloads. Apart from the partnership with U.S. News, HeinOnline is intriguing as a citation platform because of the similarity of its secondary source coverage to that of Westlaw and LexisNexis. If HeinOnline's citation capabilities become more robust and easy to use, it could replace both as the go-to database for legal citation metrics.

Data Collection and Deduplication

All data for this study was gathered twice—first, over a five-day period early in February 2019; and then again over a seven-day period in November 2019. This method was used, both for verification of results and because Hein spent the summer of 2019 improving its metrics infrastructure following the announcement of its partnership with U.S. News & World Report.[81]

Two sets of data were gathered for each author. One set was a list of the author's works from each of the four citation sources. The second set was a list of citing references to those works. The methods for obtaining the citations from each source varied widely. Collection began with a known publications list for each author.

---

[80] Lauren Mattiuzzo, *U.S. News Considers Using Hein's ScholarCheck to Evaluate Law School Scholarly Impact*, HEINONLINE BLOG (Feb. 25, 2019), https://home.heinonline.org/blog/2019/02/u-s-news-considers-using-heins-scholarcheck-to-evaluate-law-school-scholarly-impact/ (last visited Sept. 12, 2019); Lauren Mattiuzzo, *U.S. News Law School Scholarly Impact: An Update from HeinOnline*, HEINONLINE BLOG (Apr. 22, 2019), https://home.heinonline.org/blog/2019/04/u-s-news-law-school-scholarly-impact-an-update-from-heinonline/ (last visited Sept. 12, 2019).

[81] *The U.S. News Scholarly Impact Project*, HEINONLINE.ORG (2019), https://help.heinonline.org/wp-content/uploads/USNewsFlyer.pdf (last visited Dec. 4, 2019).

Author Metrics

Publish or Perish was used to obtain the author metrics from Google Scholar. Each author was searched by name, limited to citations beginning in the year of the author's first publication. Results were checked against the known-publications lists. Any nonmatching publications were removed from the list, and duplicates were combined into a single entry. Publish or Perish automatically calculates metrics based on the search results and recalculates them as the list is updated. Once the publication lists were cleaned, the metrics for all authors were exported as a CSV file.

Web of Science was similarly straightforward. An author search was performed in the Web of Science Core Collection for each author using the "Author Search" tab. Web of Science has recently started assembling works for individual author profiles. Additionally, as discussed further below, its exclusive use of works in its database for the citation counts in the profiles means that its citation counts are incomplete. In addition, the author profiles only include a few basic calculated metrics. Web of Science, however, offers the ability to create a search result based on an author profile. Instead of the traditional method of finding an author's works by searching on all known name variations and refining the search to exclude works by similarly named authors, a search result can be created with one click. Not all of the name-disambiguation is complete, however, so a few of the search results required further refining. These search results were exported from Web of Science and into Publish or Perish. Like the Google Scholar results, the metrics for all authors were exported from Publish or Perish to a single CSV file.

HeinOnline contains author profiles, similar to those offered by Web of Science. These include basic metrics, however, Hein does not provide an export function. Westlaw offers no author profiles or metrics at all. A set of import templates for Publish or Perish was created by exporting the Web of Science searches and then deleting the citation counts. Each template was updated by hand to include the citation counts from HeinOnline and Westlaw. HeinOnline citation counts were taken straight from the author profiles. Westlaw citation counts were gleaned from KeyCite reports for each article. As with Web of Science, this approach necessarily meant that the citation counts did not include every citation available in the database. The completed templates were accordingly loaded into Publish or Perish, and then the metrics were exported.

Citing References

As with the author searches, the methods for pulling the citations from each source varied. Data cleaning and deduplication likewise required different methods for each source. The publications lists were used both for generating searches, if applicable, and for checking results. For each source, results were exported using whatever method was feasible. Citing references were combined into one spreadsheet for each source. Each spreadsheet was checked manually for duplicates. First, duplicate entries were highlighted in the title column and also in the citation column where one was available. The spreadsheets were sorted by these two columns, and duplicate entries were checked and removed by hand.

Once the spreadsheets were checked for duplicate entries, they were loaded as tables into a Microsoft Access database, where they were combined into a single table. That single table was then checked for duplicate entries. If any author was cited more than once in the same work, only one citation was retained. This step was taken because Web of Science's deduplication filtering removes any work that appears more than once in an author's full citation report, even if that work cites multiple works written by the same author. All of the other results were filtered this way by hand to ensure the comparison of like results from each database.

The method for collecting citations from Westlaw depended on whether there was a KeyCite report available for a particular article. Each article title was searched. If the article was included in Westlaw, the KeyCite report was exported as a CSV file. If the article was not included in Westlaw, a title search was performed in Westlaw's "Secondary Sources" database, and the search results were saved as a CSV file. All these results were combined into a single Excel spreadsheet. A second set of searches was performed for each author, starting with the year that author's first article was published. These searches were intended to capture references that were included in Westlaw, but were absent from the KeyCite reports.

The Westlaw results were filtered for duplicates in two stages. First, the spreadsheet containing the work searches and KeyCite reports was checked for duplicate entries. Then, the author-search spreadsheet was checked for duplicates. After each was checked and cleaned, the two spreadsheets were imported into the Access database and combined. The two spreadsheets were combined using a query that joined the tables and filtered out entries from the author-searches table that duplicated titles in the work-searches table. This new table was then checked manually for duplicates, and any remaining duplicates were removed. Once the duplicates were removed, there were 712 additional citations from the author searches that were not included in the work searches and KeyCite reports.

Gathering results from Web of Science was slightly more straightforward. A cited-reference search was performed for each author, once again starting with the year of the author's first published work. Each initial search was targeted, using all known variations of the author's name, including initials and full names. The resulting cited-reference index was checked against the author's known-publications list, and all matching publications were selected. The final results were exported to a CSV file. As previously noted, Web of Science automatically filters duplicates from cited-reference searches, so no further duplicate checking was performed.

Using a cited-reference search allows a researcher to find citations to items that are not contained in Web of Science databases. Like other citation services, Web of Science scans each item in its database for citations and attempts to match each citation to an item in its database. The result of these matches is the citation count shown by each work in a search result. A cited-reference search also captures the unmatched citations in the database. It will also find citations not properly attributed to a particular author, similar to performing a separate author search in Westlaw. Paging through the large set of search results for each author is time consuming, but more thorough than relying on Web of Science's built-in tools. The cited-reference searches returned more than twice the number of citations as the author searches for five of the ten authors.

Exporting results from HeinOnline was a bit more difficult. Citations are accessed from HeinOnline's author profiles. There are several ways to reach an author profile. The easiest is through an author's affiliated institution. Clicking on the "Author Profiles" button at the top of the Law Journal Library screen opens a list of institutions. Clicking on an institution will pull up the list of affiliated authors. Clicking on an author's name will retrieve that author's profile, which includes a list of their articles indexed in HeinOnline. On the right side of each article's information is a list of the number of articles, cases, and ALI documents in HeinOnline that cite that article. Clicking on the number of citing articles will open a list of those articles. Similarly, clicking on the number of cases or ALI documents will open a list of citing cases or ALI documents, respectively. Direct export of the list is not possible. The entire list can, however, be exported as a set of bookmarks. For this, a MyHein account is required. Once logged in, the user can export an entire citation list as MyHein bookmarks.

Each citation list was exported to a set of bookmarks named after the author. Once the bookmarks were exported for each article by each author, the citing references were exported from the "Saved Bookmarks" page. The export list was compiled by scrolling through the list of bookmarks for each author and selecting all valid citations. Once all of an author's citing references were checked, the list was exported as a CSV file. This export method only works for journal articles. Cases were exported separately by e-mail and the results loaded into a spreadsheet by hand. Hein's system will not export large numbers of citations, so the results for the two authors with the largest number had to be transferred in parts.

Results were exported from Google Scholar using Publish or Perish. Each author was searched, and the results were checked against the known-publications lists. Any nonmatching publications were removed from the list. The cleaned publication lists were used to generate a citing-works search. Those results were exported to Excel spreadsheets and then combined into a single spreadsheet.

That spreadsheet required extensive cleaning. As multiple commentators have noted, Google Scholar's results often contain duplicates.[82] Unlike a controlled database, Google Scholar gathers results from all over the internet. A typical law review article, for example, may be available in multiple repositories; there may be a preprint on SSRN, a version on the publishing journal's

---

[82] *See, e.g.*, Adriaanse & Rensleigh, *supra*, note 24, at 732; Meho & Yang, *supra*, note 26, at 2111; Mingers & Lipitakis, *supra*, note 14, at 618.

website, and another in the institutional repository of the author's school. Additionally, Google Scholar now indexes articles in HeinOnline, so if the article is included in HeinOnline, every one of those other copies is a potential duplicate.

The full results list was cleaned by hand using OpenRefine[83] to remove punctuation and change all words in the title field to lowercase. OpenRefine's text analysis tools were used to find duplicate articles with slightly different titles and change them to match. This made duplication filtering in Excel simpler. The full spreadsheet was checked again by hand. Duplicate entries were kept in the following order of priority based on the source: publisher's website, HeinOnline, JSTOR, law review website or repository, institutional repository of the author's school, and SSRN.

In total, there were 6,567 good entries, 919 duplicates, and 329 incorrect or unverifiable citations. The incorrect citations were a mixture of garbled citations; citations to materials other than scholarship; citing documents that did not cite to one of the authors in the study; and other errors. A citation was unverifiable where a copy of the citing document could not be found. Over 15% of the citations needed to be removed before the results could be used. Original documents were not exhaustively checked for citations to works by the ten authors. However, suspect documents were spot-checked and removed if no citation could be found.

Once all of the results were imported into the Access database, they were combined into one table using a series of SQL queries. Results from Westlaw, HeinOnline, and Web of Science were matched by citation. As each source was added to the combined table, it was checked by

[83] OPENREFINE, https://openrefine.org/ (last visited Dec. 4, 2019). OpenRefine is an open-source application for data wrangling.

hand for any missing matches. The Google Scholar results did not include citations, so the entries were matched by title. Once all of the results were in one table, the table was sorted a number of different ways (by title, source publication, and citation). Entries were matched, and remaining duplicates were removed by hand.

Coding

Once deduplication was finished, the results were coded. Unfortunately, none of the databases reliably categorize work types. Web of Science results feature a simple categorization that includes several work types, such as journal articles, books, and series. The February results contained only two types of results—journal articles and books. When checked, 12% of those labels were incorrect. The November results also contained series results. All article and book results were correct. Series results were largely book series, but they also included a few journals.

Westlaw uses a system that categorizes cited works into a number of types, including several for primary law sources (e.g., cases, statutes, and trial court orders), and several for secondary sources, including a catch-all category called "Other Secondary Source." Unfortunately, the type of category returned in Westlaw depends on the search. KeyCite results will return the specific type of source for both primary and secondary sources. Other searches, however, will return only a general "Secondary Source" category for all types of secondary sources.

HeinOnline returned no document types. Most results were journal articles, along with a few case opinions and some practitioner materials.

Google Scholar's document types were not very useful, with one exception. The type for most documents in Google Scholar is "PDF," "BOOK," or "CITATION." "Citation" results are those where the search tool retrieves something that looks like a citation, but the document it references is either not in the Google Scholar database, or the matching algorithm was not able to match the citation to the article. Consequently, there is no link to the citing article for any citation results. Because they can be unreliable, Publish or Perish offers an option to filter out the "Citation" results. This option was unnecessary for this study, however, because many of the results for this group of authors were correct. An initial round of spot-checking found a number of good matches to results from Westlaw and HeinOnline. Instead, the unique citation results were checked manually, and those that were incorrect or unverifiable were removed from the data set. Book results in Google Scholar are harvested directly from the Google Books database, so citations marked "BOOK" were generally correct. The results, however, were incomplete; there were book entries that were associated with another document type, or with none at all.

With the possible exception of Westlaw, the provided classifications do not precisely capture the available types of documents present in the search results. Accordingly, results from all four sources were hand-coded into one of thirteen document types. The bulk of the results fit into a few categories that are fairly standard in citation metrics and in institutional repositories. "Book" is any material published in a book, whether as a single volume, multiple volumes, or a chapter. "Journal" is any piece published in a professional or academic journal, including law reviews. The "Conference Paper" category includes any material published in conference proceedings, whether in print or online. "Report" includes white papers, reports published by non-profit groups, and similar documents. "Working Papers," in this context, are works published on SSRN, or as part of a university or non-profit organization's paper series, but have not been published in a journal as of the date the citations were gathered.

Student work is broken into two categories: Thesis and “Student Paper.” The Thesis category includes all theses and dissertations submitted as a final work for a degree, whether a master’s, doctorate, or honors undergraduate degree. Theses and dissertations are grouped together for two reasons—first, because of the differences between the definitions of each in the United States and elsewhere; and second, because they are not consistently coded at each hosting site. “Student Paper” refers to all other student work, including senior theses, non-final master’s theses, coursework, and undergraduate paper awards.

All remaining documents were categorized using Westlaw document types. Court documents were separated into three categories based on the type of court: “Case”, “Administrative Decision”, or “Trial Court Order.” Briefs are also available in Westlaw, but were not collected for this study. Three more categories were used for secondary sources not fitting into one of the already listed categories: “ALR,” “Bar Journal,” and “Other Secondary Source.” ALR refers to the *American Law Reports*, a series of focused topical reports. Bar Journals are professional journals published by bar associations that generally contain shorter articles aimed at practitioners. In Westlaw, “Other Secondary Source” encompasses various practitioner resources, such as treatises, practice guides, continuing legal education materials, and more. In this study, the group was expanded to include presentations, magazines, newsletters, and newspapers.

Results

As expected, the number of citations returned from each of the four citation sources varied widely. Google Scholar returned the most, over three-quarters of the total returned by all four sources. Westlaw returned half of the citations, HeinOnline about a third, and Web of Science the fewest, less than a quarter of the total. Unique results were similar. Google Scholar again had far more than the others, but it was HeinOnline that had the fewest. Table 2 shows the total number of citations returned by each source, as well as the number of unique citations that each returned.

(INSERT TABLE 2)

In general, the greatest similarity in coverage between Westlaw, HeinOnline, and Google Scholar is for journals, especially law reviews, so it was surprising that Westlaw's unique citations were mostly journals (75%), with 85% of those being law reviews. Most of the remaining unique Westlaw citations represented secondary practitioner materials, such as treatises, loose-leaf services, bar journals, and ALRs. At first look, it seems likely that this difference in unique citations stems from the storage method used in the databases. Westlaw stores text. HeinOnline is built on PDF files, mostly scanned journals that have had optical character recognition (OCR) applied. Google Scholar gets much of its law journal content from HeinOnline. The 648 law review citations not returned by Google Scholar and HeinOnline could have been missed due to poor results from the OCR process. However, author searches were used to supplement the Westlaw results. This process gave a comparable result to the Web of Science cited-reference search results. However, it does not offer a direct comparison with HeinOnline and Google Scholar, which do not have this extra search capability. When the author search results are removed from the Westlaw search results, it loses half of its unique citations, including 501 law review citations.

Almost all (89%) of the unique citations from HeinOnline represented journal articles. The remainder were seven bar journal articles, seven ALI drafts, and four cases. Over half (52%) of the 156 unique journal citations were for articles published before 1993, which seems to be the start of full law review coverage in Westlaw. Most of the remaining journal citations should have been in Westlaw; these may have been missed by Westlaw's algorithm, but caught by HeinOnline.

Almost two-thirds (60%) of the unique Web of Science citations were for journal articles. Over half (52%) of the unique journal results were from journals published by six large academic publishers: Cambridge University Press, Elsevier, Oxford University Press, Sage Publications, Taylor & Francis, and Wiley. Most of the journal citations were from interdisciplinary journals, or journals outside the field of law. But around 20% were from journals available in some of the other three sources, including *Law & Society Review*, *Law & Social Inquiry*, and the *Journal of Legal Education*. The remaining 40% of unique Web of Science results were from books.

The majority of Google Scholar's unique citations were almost evenly split between journals (35%) and books (40%). There were also theses and dissertations (15%) and working papers (4.5%). The remaining citations (4%) were for conference papers, reports, student work, and other sources. Almost a quarter (22%) of the journal citations were for law review articles. Almost 13% of these were published in a language other than English. The remaining journals represented a hodgepodge of law titles from major and small publishers, along with titles that focus on other disciplines, including anthropology, social work, architecture and planning, history, public policy, and sociology. Of the unique law review citations, about three-quarters (77%) were drawn from HeinOnline. Most of the rest were culled from SSRN, JSTOR, or law school institutional repositories.

Table 3 shows the overlap between sources. HeinOnline had significant overlap with both

Westlaw (84%) and Google Scholar (82%). Web of Science had almost as large an overlap (77%) with Google Scholar. Google Scholar's overlap with Westlaw was a little lower (66%).

(INSERT TABLE 3)

Although it appears that Google Scholar, once the results are cleaned of chaff and duplicates, is the best source for citations, the reality is more complicated. First, the proportion of citations returned between the sources varied wildly between authors. Although Google Scholar most often returned the largest number of results, it did not always. And even when it did, the difference between it and the next source was not always large. Table 4 shows the total number of citations returned by each citation source for all ten authors. Westlaw returned the most citations for three authors: B, E, and H. It was a close second to Google Scholar for A and D, but Westlaw returned less than half of the total citations for four authors.

(INSERT TABLE 4)

HeinOnline did not fare nearly so well, coming in third or fourth for every author. It returned roughly half of the total citations for only three authors: A, B, and D. As discussed further below, this is likely because those authors write about topics that are most appropriate for law reviews. Accordingly, most of those authors' publications appear in law reviews, and those are mostly cited in other law reviews. For the remaining authors, HeinOnline generally returned between 20% and 40% of the total citations for each author, with the two exceptions of C (about 13%) and J (less than 5%). Author C's main avenues of publication are interdisciplinary journals and books, which are not covered well in any of the citation databases except Google Scholar. Author J's main form of publication is books.

Web of Science, unsurprisingly, did not fare well, returning at most about 30% of the total citations to any one author, and frequently returning 20% or less. In one extreme example, Web of Science returned only one citation for Author E. Author E rarely publishes in law reviews, instead publishing in dedicated tax journals, which are not indexed by Web of Science, and only sparsely by Google Scholar and HeinOnline. For the other authors, Web of Science favored those who publish in interdisciplinary journals and books.

Some of the variations in results are probably due to the different types of documents indexed in the four sources. Table 5 shows the number of citations of each document type returned by each source. The dominant type of document returned was journal articles, overwhelmingly so for every source except Google Scholar. Only the Google Scholar and Web of Science results contained books. Westlaw and HeinOnline both returned a small number of cases. Westlaw also returned a number of practitioner materials, a little under 5% of its total citations, but more than every other non-journal document type combined. Google Scholar returned a number of unique (or almost unique) document types, including working papers, newsletters, reports, conference papers, presentations, theses and dissertations, and other student work. When the Westlaw author search results were removed, it had 604 fewer journals, 86 fewer “Other Secondary Sources,” and 4 fewer bar journals.

(INSERT TABLE 5)

The usefulness of citations from most of these document types for citation counting can be, and often is, debated. Setting aside arguments for or against counting citations from dissertations, case opinions, bar journals, and the like, the more concerning issue for a researcher (and an author) is the number of citations to journal articles each source missed.

Table 6 shows each source's coverage of journals published by the top fifteen publishers of citing journals. Student-run law reviews are not included in this table, but are covered separately below. The coverage of these journals, which includes law and other journals published by large and small journal publishers, university presses, associations, and law schools, is generally good to very good in Google Scholar (83%), not as good in Web of Science (46%), and very poor in Westlaw (23%) and HeinOnline (11%).

(INSERT TABLE 6)

There are exceptions, however. Westlaw had the best coverage for the publications of the Environmental Law Institute and Indiana University Press. Westlaw also demonstrated unexpectedly good coverage of Wiley and Cambridge University Press, though this was largely due to the high number of citations from two journals: *Law & Society Review* and *Law & Social Inquiry*. Westlaw also returned about half or more of the total citations for law-focused publishers, such as the Association of American Law Schools, Duke University School of Law, the American Association of Law Libraries, and its own parent company, Thomson Reuters.

HeinOnline, in particular, performed very poorly in returning results from most of the aforementioned publishers. Specifically, it returned only 9% of all the Wiley citations. Westlaw was next with 40%. HeinOnline's noticeably poor showing for Author C is likely a result of it missing almost two hundred citations to that author's work from both *Law & Society Review* and *Law & Social Inquiry*. In contrast, Web of Science returned more than half of the citations for each of the top ten citing-journal publishers. However, it returned no results at all for the vast majority of the remaining publishers.

(INSERT TABLE 7)

The results for student-run law reviews were somewhat uneven. With the exception of Web of Science, this should not have happened. Web of Science does not cover the complete set of U.S. law reviews, so it returned fewer citations than the other three sources. The other three, however, should have returned more similar results. HeinOnline includes or indexes the entire runs of all, or almost all, of the citing law reviews. Google Scholar uses HeinOnline's database for the bulk of its law-review indexing. Westlaw does not include the full runs of all law journals, but it is mostly complete from the mid-1990s forward. For articles published prior to that time, Westlaw's coverage depends on the journal. Some, like the *Harvard Law Review*, have full coverage from the publication's inception. Others have a period of selected coverage before full coverage begins later. Table 7 shows the number of citations returned by each source from the top twenty citing law reviews.

Remarkably, even though it does not fully cover all journals, Westlaw returned more citations to almost every law review than the other sources. Of the 233 journals that cited the sample five times or more, Westlaw returned all of the citations from eighty-eight of the journals. None of the other citation sources came close. Google Scholar returned the second most with twenty-five. HeinOnline returned fourteen, while Web of Science returned only one.

In the top thirty citing journals, Google Scholar returned more citations than Westlaw three times. The only source to return every citation in any of those thirty journals was Westlaw, which returned all of the citations from the *Columbia Law Review*, *Emory Law Journal*, *Vanderbilt Law Review*, and *Iowa Law Review*.

Because HeinOnline indexes the complete runs of most of these journals, Hein and Google Scholar should have returned all of the citations from each journal. But neither returned all of the citations for any journal, and the citation counts rarely matched each other. This discrepancy is likely the result of the different methods each source uses to find citations. Westlaw, Google Scholar, and Web of Science appear to use an algorithm to scan source material for citations. In contrast, HeinOnline relies on specialized searches.

In Westlaw, articles appear to be scanned for citations, and then when citations are found, the system creates links to those works that are also available in the database. Any cases in the database will then appear in the "Table of Authorities" for the article. Links to the work from any other type of document in the database will appear in a "Citing References" list. The system, however, does not flag every citation. If part of a citation is missing or incorrect, the Westlaw system may miss it. If a citation is missed, there is no link in the text, and it won't appear in the "Citing References" list. These may be identified, however, in an author or title search.

Unlike Westlaw, Web of Science is primarily an index, rather than a full-text source, particularly where law journals are concerned. For each article in a covered title, a citing-references list is available. Where the cited reference is also in a covered journal, the citing references are linked to the cited article's database entry. Where the cited reference is not in a covered journal, the algorithm attempts to assemble all of the citations in discrete database entries. However, Web of Science does not cope well with pinpoint cites, often listing citations to different pages of the same work as multiple works. It also does a poor job of matching the correct authors to works in journals it does not index.

Unlike Web of Science, Westlaw, and Google Scholar, HeinOnline does not appear to have a separate citations database. Each article contains a small "ScholarCheck" widget with up to three numbers. When clicked on, the widget expands to show the number of cases, articles, and ALI documents that cite the article. Clicking on one of the numbers generates a real-time search in the database for the citation in several formats. The author profiles are generated similarly. Opening an author profile shows a list of articles that appears to be returned by a live search in the system. This method, like Westlaw's, misses any citation that is not in a format that the system recognizes, or which contains certain typographical errors. And because the system is entirely based on articles in its own database, it will only return citations to the articles in that database, even when the author profile system is used. This shortcoming has been noted by critics of Heald and Sichelman's study, as well as the anticipated study by U.S. News & World Report.[84] There appears to be no separate reference search performed on the articles in the database, so no outside citations are captured.

Google Scholar appears to work similarly to Web of Science. Documents in the database are scanned for citations. If there is a citation to another document in the database, they are matched. If the cited document is not in the database, it is nevertheless tracked. These results (labeled "CITATION") can be unreliable. In this study, 44% of the "Citation" results were removed because the citing document could not be found, the citing document was found but did not actually cite one of the authors in the sample, or the citing document was not a scholarly publication. Another problem, as discussed above, is the presence of duplicates. Google Scholar does not filter articles present in multiple sources. Searches performed directly on Google Scholar return all of the sources. Publish or Perish likewise does not attempt to filter duplicates, so they must be manually identified and removed.

---

[84] Gregory Sisk, *Measuring Law Faculty Scholarly Impact by Citations: Reliable and Valid for Collective Faculty Ranking*, 60 JURIMETRICS J. 41, 48 (2019); Hayashi and Mitchell, *supra* note 60, at 9.

Google Scholar appears to use HeinOnline as its primary source for law review articles. The vast majority of law review articles returned in the Google Scholar searches list HeinOnline as the source and link to the full-text article in Hein if the user has a subscription. Consequently, the difference between the citation counts for law review articles in the two databases is likely to be the result of the different methods they utilize to match citations to cited articles. Of the two, Google Scholar (72%) returns more than HeinOnline (63%), but Westlaw's 86% return is well ahead of both until the author search results are removed. Once the author search results are removed, Westlaw loses much of its advantage, returning 81% of the total results, compared to 80% from Google Scholar and 69% from HeinOnline. But the more concerning problem is that without those searches, there are almost 600 journal citations to the sample authors that are not returned by any of the four sources.

These results highlight the overall concern with reliance on any citation database. All four of these sources use some type of algorithm for detecting and matching citations, but none of them get close to achieving a perfect score, even in matching up citations to works that are available in full text within.

(INSERT FIGURE 1)

Additionally, subject coverage is drastically different between the four sources. Although the articles in this study were not coded by subject, the authors' research interests had little overlap, so the differences in the coverage for each individual author is instructive. While the total numbers would suggest that Google Scholar is always the best choice for coverage purposes, this is not always true. Google Scholar returned the largest number of citations for seven authors, but Westlaw outperformed it for the remaining three. In one case (Author E), Google Scholar returned less than 40% of his total citations, while Westlaw returned 84%.

In general, Westlaw performed better than both HeinOnline and Web of Science. It returned the highest number of citations for three authors (B, E, and H), and at least half of the citations for three more authors (A, D and F). For the remaining four authors, Westlaw placed second for all but Author J, where it was third.

HeinOnline only retrieved more than half of the citations for one author (A). For all but two of the authors (C and J), it retrieved the third-highest number of citations. HeinOnline performed surprisingly worse than Westlaw, considering the overlap in coverage between the two databases. This large difference is likely attributable to two things: first, the extra coverage of non-law journals in Westlaw (particularly Wiley journals); and second, the greater efficiency of Westlaw's algorithms for finding citations in full-text materials.

Web of Science fared particularly poorly, returning less than 30% of the citations for each author. It managed a third-place finish twice (C and J), both times for authors who publish largely in interdisciplinary journals and books.

The largest disparity in the present study's results was seen for authors C, E, and J. In all three cases, one database vastly outperformed the other three. Author E's publications focus almost exclusively on the topic of taxation. His articles are primarily published in specialty tax journals, and he contributes to one of the most well-known multivolume tax treatises. Most of his articles appear in specialty tax journals, which are included in Westlaw, but which are not widely available on platforms outside of those catering to tax practitioners. Accordingly, none of the other three sources returned even half of Author E's total citations.

Author C has a long career and an equally long publications list, but most of his work appears in books and interdisciplinary journals. Consequently, most of the citations to his publications are found in interdisciplinary journals, books, and non-law journals. Coverage of all of those types of resources is poor in Westlaw and HeinOnline. The disparity between the performances of Westlaw and HeinOnline here is largely because of Westlaw's more extensive coverage of three specific journals: *Law & Social Inquiry*, *Law & Society Review*, and *Annual Review of Law & Social Science*. Author C has many citations from those journals carried by Westlaw, but not HeinOnline. Web of Science's comparatively strong showing in this instance is due to its inclusion of journals in disciplines that are not covered by the two law databases. But Google Scholar, with access to the major journal publishers and many more books than Web of Science, is the clear winner here.

Author J is a newer faculty member with fewer publications, but with a publication pattern and citation pattern similar to that of Author C.

Authors A, B, and H publish in different subject areas, but had similar results in the present study. All had the most results from Westlaw, but with a less drastic difference in results from Google Scholar and HeinOnline than that seen for Author E. Like Author E, all three professors publish in more focused areas of law (intellectual property, bankruptcy, and environmental law). Those areas, however, are less specialized than tax, and those specialty publications are more widely available.

(INSERT TABLE 8)

In most similar studies, metrics derived from Google Scholar citation counts were larger than those from Web of Science. This study is no exception. The Google Scholar metrics are also larger than those from Westlaw and HeinOnline, although not by as much. Table 8 shows the means of three citation metrics derived from the citation results. Papers ("pap") is a count of the number of papers found by each source for each author. Citations ("cit") is a count of the total number of citations to all of the papers found by each source. H-index ("h") is the h-index derived from the citations to each paper for a particular author. Google Scholar ("gs") returned the largest number of papers, with Westlaw ("wl") second, HeinOnline ("hol") third, and Web of Science ("wos") last. Number of citations and h-index followed this pattern.

(INSERT TABLE 9)

Table 9 shows the ratios between the mean number of papers, mean number of citations, and mean h-indexes returned by the four sources. On average, Google Scholar returned 2.6 times as many papers as Web of Science. Westlaw and HeinOnline's results were more similar. Google Scholar only returned 1.35 times as many papers as Westlaw, and 1.83 as many as HeinOnline. This result for Web of Science is set in between that of Franceschet, who found 5.48 times as many papers in Google Scholar as in Web of Science, and Minasny et al., who found 2.33 times as many.[85]

Citation numbers were also higher in Google Scholar, which returned seven times as many citations as Web of Science. This result was again similar to what Franceschet found (7.76), and far more than that discovered by Minasny et al.[86] Again, Westlaw and HeinOnline were much closer at 1.73 and 2.39, respectively. These ratios are similar to what Minasny et al. found (1.87)

[85] Franceschet, *supra* note 22, at 251; Minasny et al., *supra* note 40, at 5.
[86] *Id.*

between Google Scholar and Web of Science.[87] The greatest disparities were observed between Web of Science and the other databases. Google Scholar returned almost seven times as many citations, and HeinOnline and Westlaw returned close to three and four times as many, respectively.

The h-index results reflect similar differences between the four databases. The greatest difference in the h-index was seen between Google Scholar and Web of Science at a 2.84 ratio. This is almost identical to what Franceschet found with computer science researchers (2.86), and slightly below what Harzing and Alakangas found for social science researchers (2.24).[88] Again, Westlaw and HeinOnline's results were closer to those of Google Scholar.. The Google Scholar-Westlaw ratio for the average h-index was 1.21 in this study, better than the performance of Web of Science compared to Google Scholar for most of the other published studies. Harzing and Alakangas noted better performance from Web of Science for science and life science researchers.[89] In the present study, the Google Scholar-HeinOnline ratio was 1.39, similar to what Harzing and Alakangas found for Web of Science for engineering researchers.[90]

(INSERT TABLE 10)

Spearman's rank correlation was calculated for the three metrics using R version 3.6.0.[91] Table 9

[87] Minasny et al., *supra* note 40, at 5.

[88] Franceschet, *supra* note 22, at 251; Harzing and Alakangas, *supra* note 7, at 797.

[89] Harzing and Alakangas, *supra* note 7, at 797.

[90] *Id.*

[91] R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria 2019), https://www.R-project.org/ (last visited Feb. 25, 2020).

shows the correlations for the three variables between the four sources. Surprisingly, there was excellent correlation between all four databases for the h-index. Correlation between the numbers of papers was very similar, and the number of citations was slightly lower, but still very high. This finding is similar to that of Minasny et al., who used a sample of 340 soil researchers.[92] Franceschet's sample of thirteen computer scientists found a much lower correlation, which is what was expected here.[93] In Francheschet's study, the correlation between the number of citations from Web of Science and Google Scholar was high, at 0.92.[94] But the correlation between the number of papers (0.69) and the h-index (0.65) was much lower.[95] These differences might be attributed to several factors, including sample size, changes to Web of Science and Google Scholar since 2010, and differences in coverage for legal scholarship and computer science in Web of Science and Google Scholar.

High correlation between Westlaw and HeinOnline was expected due to the similarity of coverage. Heald and Sichelman found high correlation between the law faculty rankings compiled from Westlaw data by Sisk and their own ranking compiled from HeinOnline data. The correlation between the two was 0.88, which is a little lower than the 0.94 and 0.95 for the metrics compiled for the present study.[96] Those two rankings, however, were compiled using different methodologies and faculty lists, while here the comparison between metrics was made using data compiled from different sources using as similar a methodology as possible.

[92] Minasny et al., *supra* note 40, at 6.

[93] Franceschet, *supra* note 22, at 252.

[94] *Id.*

[95] *Id.*

[96] Heald and Sichelman, *supra* note 45, at 31.

Although the high correlation between HeinOnline and Google Scholar was unexpected, it is unsurprising, given that Google Scholar is now indexing HeinOnline content for a large portion of its law journal articles. However, because sample size is a possible factor, further study is warranted.

Conclusion

There is still no ideal citation database for legal scholarship. Each of the major sources of citation data has multiple shortcomings. Web of Science has the best tools for generating metrics, but has the least content, with many heavily cited law journals being omitted. Google Scholar has the most content, but it was not designed to be a citation database. Publish or Perish is an excellent front-end tool that will generate most commonly-used metrics. However, it relies on data scraping and could be blocked by Google at any time. Google Scholar also has serious data quality issues, and its results require manual cleaning to remove duplicates and spurious entries. Westlaw has a useful amount of content, though it is light in terms of its inclusion of interdisciplinary journals. Westlaw was likewise not designed to be a citation-generating database. Westlaw's raw search results need to be exported and manipulated with external software.

HeinOnline does have incorporated citation tools. However, these tools are new and still rudimentary. The ScholarCheck metric is proprietary and so not particularly useful. Like Westlaw, searches need to be exported and manipulated with external software to generate standard citation metrics. Publish or Perish's import filter can be used for this purpose, provided that the results spreadsheets are in a compatible format. And every database has problems with name disambiguation, and algorithms that do not match citations with articles contained within.

Differences in content notwithstanding, the correlation between the three studied metrics from each database is very high, indicating that rankings from any of them should be very similar. This observation is undercut somewhat by the small sample size. Further study, perhaps of multiple and complete law school faculties, is warranted. These initial results, however, are promising. If further studies have similar results, then it may be possible to generate useful comparisons with only one or two of the sources studied here. This is particularly true if it becomes easier to clean Google Scholar results, or if HeinOnline or Westlaw improve their name disambiguation, coverage, and matching process

Table 1. Subject areas

Author	Subjects
A	Intellectual Property, Legal History, Media Law
B	Bankruptcy, Torts, Commercial Law
C	Torts, Law and Society, Asian Legal Cultures, Disability Law, Law and Anthropology
D	Election Law, Constitutional Law
E	Tax
F	Civil Rights Law, Insurance Law, Law and Economics, Labor and Employment Law, Health Law, Critical Legal Studies
G	Civil Rights Law, Legal Theory, Critical Legal Studies, Law and Gender
H	Environmental Law, Land Use Law, Property Law, Federal Indian Law, Administrative Law
I	Critical Legal Studies, Law and Society, Corporate Law, Law and Economics, Legal History
J	Comparative Law, Criminal Law, Law and Anthropology, Contracts

Table 2. Number of citations returned by each citation source

Source	**Total Citations**	**Percentage of Total Citations**	**Unique Citations**	**Unique Percentage of Total**
Westlaw	4221	50%	1010	12%
Web of Science	1889	22%	277	3%
HeinOnline	2802	33%	174	2%
Google Scholar	6567	78%	2915	35%
All Databases	8444	100%		

Table 3. Overlap between citation sources

	Westlaw	**Web of Science**	**HeinOnline**	**Google Scholar**
Westlaw (n=4221)		972	2356	2802
Web of Science (n=1889)	972		638	1466
Hein (n=2802)	2356	638		2312
Google Scholar (n=6567)	2802	1466	2312	

Table 4. Total citations returned for each author

Author	Total	Westlaw	Web of Science	HeinOnline	Google Scholar
A	373	241	65	189	259
B	137	123	21	67	85
C	2108	647	611	265	1801
D	1905	1334	371	916	1440
E	85	71	1	19	30
F	795	400	139	278	638
G	615	232	122	158	518
H	401	260	81	135	240
I	1863	886	428	770	1422
J	162	27	50	5	134

Table 5. Citations returned by type of publication

Type	Total	Westlaw	Web of Science	HeinOnline	Google Scholar
Administrative Decision	4	4	0	0	0
ALR	3	3	0	0	0
Bar Journal	21	14	0	8	0
Book	1401	0	213	0	1292
Case	44	40	0	39	0
Conference Paper	59	0	1	0	59
Journal	6039	3932	1675	2747	4579
Other Secondary Source	246	227	0	8	13
Report	25	0	0	0	25
Student Paper	30	0	0	0	30
Thesis	436	0	0	0	436
Trial Court Order	1	1	0	0	0
Working Paper	133	0	0	0	133

Table 6. Citations returned by publishers

Publisher	**Total**	**Westlaw**	**Web of Science**	**Hein**	**Google Scholar**
Wiley	310	123	181	29	256
Cambridge University Press	260	116	147	38	217
Taylor & Francis	173	4	87	6	157
Sage Publishing	129	2	79	4	117
Oxford University Press	86	23	49	23	60
AALS	66	35	50	39	46
Elsevier	58	0	44	0	50
Non-English	55	0	0	2	55
Springer	43	1	22	0	41
Annual Reviews	39	16	33	0	38
Emerald Group Publishing Ltd.	33	0	18	0	32
DeGruyter	26	10	3	7	21
Duke University School of Law	26	17	5	17	21
University of Chicago Press	26	9	20	6	18
Environmental Law Institute	24	24	0	0	9

Table 7. Citations returned by law review title

Source	**Total**	**Westlaw**	**Web of Science**	**HeinOnline**	**Google Scholar**
Buffalo Law Review	82	66	49	44	56
Harvard Law Review	68	59	36	29	37
Michigan Law Review	62	49	36	36	53
Rutgers Law Journal	59	54	0	50	39
Yale Law Journal	57	48	38	31	51
Texas Law Review	48	40	28	27	34
Northwestern University Law Review	40	31	26	30	26
Albany Law Review	38	37	0	29	32
Wisconsin Law Review	38	34	18	29	29
William and Mary Law Review	37	32	0	29	31
Fordham Law Review	36	29	22	21	24
Connecticut Law Review	35	33	0	22	26
Univ. of Pennsylvania Law Review	34	31	22	24	32
Stanford Law Review	34	33	17	20	27
Columbia Law Review	33	33	30	30	32
Indiana Law Journal	32	29	25	16	26
Georgetown Law Journal	32	31	22	19	26
UCLA Law Review	32	30	22	20	24
Virginia Law Review	31	30	21	24	25
Temple Law Review	30	29	8	26	28

Table 8. Citation metrics derived from the four sources

source	pap	cit	h
gs	41.9	909.9	12.5
wl	31	527.1	10.3
hol	22.9	380.4	9
wos	16.1	134.2	4.4

Table 9. Ratios between metrics derived from the four citation sources

	pap	**cit**	**h**
gs.wos	2.64	7.03	2.84
gs.wl	1.35	1.73	1.21
gs.hol	1.83	2.39	1.39
wl.gs	0.74	0.58	0.82
wl.wos	1.95	4.07	2.34
wl.hol	1.35	1.39	1.14
hol.gs	0.55	0.42	0.72
hol.wos	1.44	2.94	2.05
hol.wl	0.74	0.72	0.87

Table 10. Correlations between citation sources

Variable	**By variable**	**Spearman coefficient rank**	**Prob > [p]**
GS *h*-index	WoS *h*-index	0.97	2.45E-06
GS *h*-index	WL *h*-index	0.96	1.04E-05
GS *h*-index	HOL *h*-index	0.87	0.001086985
WL *h*-index	WoS *h*-index	0.97	5.71E-06
WL *h*-index	HOL *h*-index	0.94	3.80E-05
HOL *h*-index	WoS *h*-index	0.91	0.000250988
GS no. citations	WoS no. citations	0.92	0.000204472
GS no. citations	WL no. citations	0.92	0.000204472
GS no. citations	HOL no. citations	0.85	0.001636803
WL no. citations	WoS no. citations	0.94	5.48405E-05
WL no. citations	HOL no. citations	0.94	5.48E-05
HOL no. citations	WoS no. citations	0.90	0.000343612
GS no. papers	WoS no. papers	0.89	0.000566965
GS no. papers	WL no. papers	0.97	5.25E-06
GS no. papers	HOL no. papers	0.92	0.000204472
WL no. papers	WoS no. papers	0.90	0.00032182
WL no. papers	HOL no. papers	0.95	1.79E-05
HOL no. papers	WoS no. papers	0.96	7.62E-06

Figure 1. Percentage of Total Citations per Author by Database