



7-2019

## Search Results Outliers Among MEDLINE Platforms

Christopher Sean Burns

*University of Kentucky*, [sean.burns@uky.edu](mailto:sean.burns@uky.edu)

Robert M. Shapiro II

*University of Kentucky*, [shapiro.rm@uky.edu](mailto:shapiro.rm@uky.edu)

Tyler Nix

*University of Michigan - Ann Arbor*

Jeffrey T. Huber

*University of Kentucky*, [jeffrey.huber@uky.edu](mailto:jeffrey.huber@uky.edu)

Follow this and additional works at: [https://uknowledge.uky.edu/slis\\_facpub](https://uknowledge.uky.edu/slis_facpub)



Part of the [Health Sciences and Medical Librarianship Commons](#)

**Right click to open a feedback form in a new tab to let us know how this document benefits you.**

---

### Repository Citation

Burns, Christopher Sean; Shapiro, Robert M. II; Nix, Tyler; and Huber, Jeffrey T., "Search Results Outliers Among MEDLINE Platforms" (2019). *Information Science Faculty Publications*. 63.

[https://uknowledge.uky.edu/slis\\_facpub/63](https://uknowledge.uky.edu/slis_facpub/63)

This Article is brought to you for free and open access by the Information Science at UKnowledge. It has been accepted for inclusion in Information Science Faculty Publications by an authorized administrator of UKnowledge. For more information, please contact [UKnowledge@lsv.uky.edu](mailto:UKnowledge@lsv.uky.edu).

---

## Search Results Outliers Among MEDLINE Platforms

### Abstract

**Objective:** Hypothetically, content in MEDLINE records is consistent across multiple platforms. Though platforms have different interfaces and requirements for query syntax, results should be similar when the syntax is controlled for across the platforms. The authors investigated how search result counts varied when searching records among five MEDLINE platforms.

**Methods:** We created 29 sets of search queries targeting various metadata fields and operators. Within search sets, we adapted 5 distinct, compatible queries to search 5 MEDLINE platforms (PubMed, ProQuest, EBSCOhost, Web of Science, and Ovid), totaling 145 final queries. The 5 queries were designed to be logically and semantically equivalent and were modified only to match platform syntax requirements. We analyzed the result counts and compared PubMed's MEDLINE result counts to result counts from the other platforms. We identified outliers by measuring the result count deviations using modified z-scores centered around PubMed's MEDLINE results.

**Results:** Web of Science and ProQuest searches were the most likely to deviate from the equivalent PubMed searches. EBSCOhost and Ovid were less likely to deviate from PubMed searches. Ovid's results were the most consistent with PubMed's but appeared to apply an indexing algorithm that resulted in lower retrieval sets among equivalent searches in PubMed. Web of Science exhibited problems with exploding or not exploding Medical Subject Headings (MeSH) terms.

**Conclusion:** Platform enhancements among interfaces affect record retrieval and challenge the expectation that MEDLINE platforms should, by default, be treated as MEDLINE. Substantial inconsistencies in search result counts, as demonstrated here, should raise concerns about the impact of platform-specific influences on search results.

### Keywords

Information Retrieval, Search Queries, Medical Subject Headings (MeSH), MEDLINE

### Disciplines

Health Sciences and Medical Librarianship | Library and Information Science

### Notes/Citation Information

Published in *Journal of the Medical Library Association*, v. 107, no. 3, p. 364-373.

Copyright © 2019 Christopher Sean Burns, Robert M. Shapiro II, Tyler Nix, Jeffrey T. Huber

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

# Search results outliers among MEDLINE platforms

Christopher Sean Burns; Robert M. Shapiro II; Tyler Nix; Jeffrey T. Huber

*See end of article for authors' affiliations.*

**Objective:** Hypothetically, content in MEDLINE records is consistent across multiple platforms. Though platforms have different interfaces and requirements for query syntax, results should be similar when the syntax is controlled for across the platforms. The authors investigated how search result counts varied when searching records among five MEDLINE platforms.

**Methods:** We created 29 sets of search queries targeting various metadata fields and operators. Within search sets, we adapted 5 distinct, compatible queries to search 5 MEDLINE platforms (PubMed, ProQuest, EBSCOhost, Web of Science, and Ovid), totaling 145 final queries. The 5 queries were designed to be logically and semantically equivalent and were modified only to match platform syntax requirements. We analyzed the result counts and compared PubMed's MEDLINE result counts to result counts from the other platforms. We identified outliers by measuring the result count deviations using modified z-scores centered around PubMed's MEDLINE results.

**Results:** Web of Science and ProQuest searches were the most likely to deviate from the equivalent PubMed searches. EBSCOhost and Ovid were less likely to deviate from PubMed searches. Ovid's results were the most consistent with PubMed's but appeared to apply an indexing algorithm that resulted in lower retrieval sets among equivalent searches in PubMed. Web of Science exhibited problems with exploding or not exploding Medical Subject Headings (MeSH) terms.

**Conclusion:** Platform enhancements among interfaces affect record retrieval and challenge the expectation that MEDLINE platforms should, by default, be treated as MEDLINE. Substantial inconsistencies in search result counts, as demonstrated here, should raise concerns about the impact of platform-specific influences on search results.



This article has been approved for the Medical Library Association's Independent Reading Program  
<<http://www.mlanet.org/page/independent-reading-program>>.



See end of article for supplemental content.

## INTRODUCTION

The replication and reproduction of research, or lack thereof, is a perennial problem among research communities [1–3]. For systematic reviews and other research that relies on citation or bibliographic records, the evaluation of scientific rigor is partly based on the reproducibility of search strategies. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Guidelines and the Cochrane Handbook for Systematic Reviews and Interventions are examples of how scholars recognize the need for

systematic reporting of methods and the organization of review research [4, 5].

Differences in search interfaces, article indexing, and retrieval algorithms also impact reproducibility and replication, which are important aspects of the scientific process, evidence-based medicine, and the creation of systematic reviews [6–12]. Even if search strategies are methodical and well documented, searches might not be reproducible because many platforms are proprietary products, and thus the code, algorithms, and, in general, the software that drives these products are not available for public

review. Consequently, one can only speculate how such systems work by inference from use; for example, by comparing them to similar products [13, 14].

Although the National Library of Medicine (NLM) maintains the MEDLINE records and provides free (i.e., federally subsidized) access to them through PubMed, they also license these records to database vendors (hereafter, “platforms”). Furthermore, although these platforms operate with the same MEDLINE data, each platform applies its own indexing technologies and its own search interface, and it is possible that these alterations influence different search behaviors and retrieval sets [15, 16].

Some studies used queries that were designed to study reproducibility across platforms by comparing recall and precision for retrieval sets across platforms [17–19]. However, different query syntax across platforms has been highlighted as an important problem itself [20, 21]. One small study, for example, compared search queries and results among different interfaces to the CINAHL database and reported reproducible search strategies except for queries that contained subject-keyword terms [22]. Another study reported that different interfaces to the same underlying database or set of records produced different search results and noted that the practical implications of missing a single record from a literature review could skew results or alter the focus of a study [23]. A third study found that PubMed retrieved more records than Ovid’s MEDLINE, but this study did not include MEDLINE subset results in PubMed [24]. A reply to this study suggested that the differences could be explained by basic problems with bibliographic and MEDLINE searching and concluded that “database and search interface providers should agree on common standards in terminology and search semantics and soon make their professional tools as useful as they are intended to” [25].

The purpose of this study was to document how different MEDLINE platforms influenced search result counts (presumably based on the same MEDLINE data file) by creating equivalent, structured, and straightforward queries to search across these platforms (i.e., by controlling for query syntax). The authors asked the research question: how much do search result counts among

MEDLINE platforms vary after controlling for search query syntax?

## METHODS

We examined five MEDLINE platforms by creating twenty-nine sets of search queries for each platform and comparing search count results. The platforms were PubMed’s MEDLINE subset, ProQuest’s MEDLINE, EBSCOhost’s MEDLINE, Web of Science’s MEDLINE, and Ovid’s MEDLINE, hereafter simply referred to by their main platform name (e.g., PubMed, Ovid).

Our queries were organized into 29 sets, with each set containing 5 equivalent queries, 1 per platform, and numbered sequentially (s01, s02...s29), for a total of 145 searches. Two authors collected the counts for all platforms by running the queries in the platforms and recording the total records returned in a spreadsheet. PubMed search counts were recorded on results sorted by *most recent* since PubMed alters the search query, and thus the search results, when sorting by *best match* [26]. The other MEDLINE platforms do not alter search records or counts based on sorting parameters.

Each of the 29 search sets targeted various search operators and metadata fields. For example, Table 1 reports an example set of queries and search result counts for search set s09 (composed of a single MeSH term appearing on a single branch of the MeSH tree, exploded, and combined with a keyword and date limit). All 145 queries, search logic descriptions, and search count results are provided in supplemental Appendix A. Some of our queries were limited by publication dates so that we could limit the influence of records that have been newly added and reduce deviations based on updates to PubMed and then updates to the other platforms.

Table 1 represents how the five queries per set were designed to be semantically and logically equivalent and were modified only to match the syntax required by each platform. In another example, our first set of queries (s01) compared the same all-field keyword search (e.g., “neoplasms”[All] AND medline[SB] in PubMed) across these five platforms, and our second set of queries (s02) compared the same single MeSH term (single

**Table 1** Example search queries and results for search set s09

	Search set s09	Result counts
PubMed	"neoplasms"[MH] AND "immune"[ALL] AND 1950:2015[DP]	72,297
ProQuest	MESH.EXPLODE("neoplasms") AND NOFT("immune") AND YR(1950-2015)	72,641
EBSCOhost	MH("neoplasms+") AND TX("immune") AND YR 1950-2015	72,987
Web of Science	MH:exp=("neoplasms") AND TS=("immune") AND PY=(1950-2015)	14,711
Ovid	1. EXP neoplasms/ AND immune.AF 2. limit 1 to YR=1950-2015	71,594

branch, no explode) searches across platforms (supplemental Appendix A). The remaining queries were constructed to explore other permutations of simple searches, including searches with single MeSH terms on single and multiple branches as well as other field searches like journal titles, author names, and date limits. Queries were constructed to specifically search the MEDLINE subset of each platform when it was not the default. For example, PubMed queries that did not contain MeSH terms included the limiter "medline[sb]", and all Ovid queries were run in the "mesz" segment, which includes only documents with MEDLINE status and omits epub ahead of print, in-process, and other non-indexed records contained in the "ppez" segment.

The queries were not designed to mimic end user usage nor were they designed to examine database coverage. Rather, they were designed to explore search result counts stemming from basic query syntax and differences in search field indexing. That is, our goal was to understand baseline deviations and to detect outliers to help understand whether reproducing queries across MEDLINE platforms is hindered by the platforms. All searches were created and pilot-tested in the summer of 2018. The results reported here are from searches conducted in October 2018.

To answer our research question, our analysis is based on a comparison of search result counts and modified z-scores ( $m_i$ ) for the result counts in each search set. The modified z-score is a version of the standard z-score and is likewise interpreted and applicable in locating deviations; however, it is more robust against outliers [27]. Generally, the standard z-score is compared to the mean (or the center of the data), but we centered our scores around the PubMed result counts from each search set to highlight search result counts that deviate from

those of PubMed. In particular, we defined search result outliers as any modified z-score that deviated more than  $\pm 3.5$  from PubMed, as recommended by Iglewicz and Hoaglin [27]. In addition to the z-score, we highlighted search count differentials (result counts as compared to those of PubMed) for all searches, as reported in the table in supplemental Appendix B. Even if results do not deviate from PubMed by  $\pm 3.5$  standardized points, differences in counts help highlight deviations across MEDLINE platforms.

The analysis was conducted in the R programming language with additional software libraries [28–34]. Code and data for this analysis are provided in supplemental Appendixes C and D.

## RESULTS

Overall, we found that most searches resulted in retrieval differences among MEDLINE platforms and that some platforms deviated from PubMed more than others. In general, ProQuest and EBSCOhost exhibited similar patterns of search result count deviations from PubMed, but ProQuest deviated from PubMed more substantially, with three search queries classified as outliers. Web of Science exhibited the most idiosyncratic search result count deviations from PubMed searches, with five search queries returning substantially different counts. Although Ovid's search result counts showed fewer and less exaggerated deviations, it consistently returned fewer records than PubMed, even for publications restricted by publication date range 1950–2015. This deviation suggested that there was an important difference between PubMed and Ovid in how they indexed their records. By fixing the publication dates to a range, ongoing updates to the database content should have had less influence on these differences.

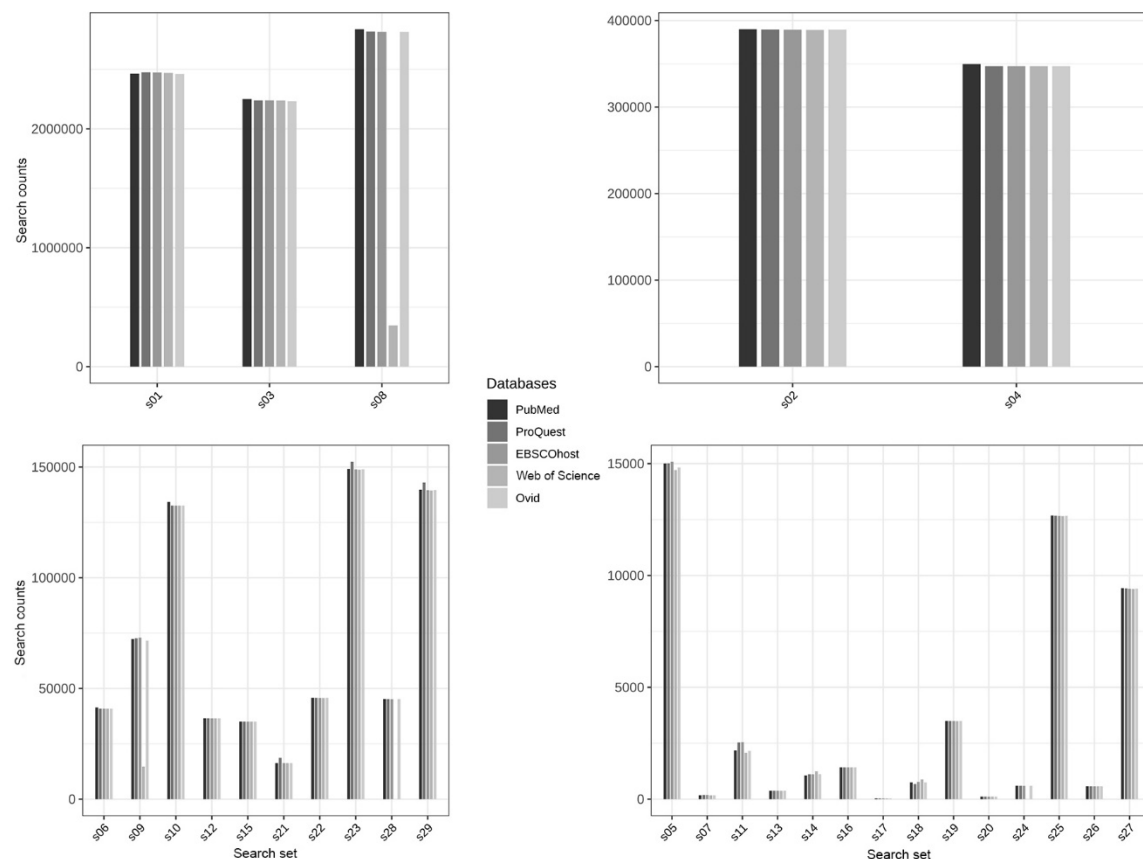
Figure 1 shows the total records returned for each set of queries. The figure is faceted into 4 plots by the magnitude of search result counts. On the surface, most searches in each set appear to be consistent with the others. However, there are a few obvious inconsistencies in the results. For example, the Web of Science search returned only 20% of the search records that PubMed returned for the equivalent query in search set s09 (Table 1) and only 12% of the search records that PubMed returned in search set s08 (Table 1). In both searches, the queries exploded the MeSH term “Neoplasms,” indicating a problem with how Web of Science explodes terms.

To derive the search result differentials, we subtracted each query’s total number of search count results from the PubMed search count results in the respective set to analyze how far each search deviated from PubMed. We also examined the search differentials using a modified z-score, which allowed us to zoom in on the discrepancies. (The table in supplemental Appendix B reports the

differentials.) For example, in search set s10, PubMed returned 134,217 records with publication dates limited from 1950–2015 for a search against the MeSH term “Dementia,” exploded. The other 4 platforms returned between 1,618 to 1,627 fewer records.

These deviations in this set were fairly consistent across the 4 platforms and statistically small, per the z-scores (indicated in parentheses in supplemental Appendix B). However, search set s23 also queried for “Dementia” (exploded) but did not limit results by publication date. Here, PubMed returned 149,146 total records, and the other 4 platforms returned a more varied number of results. In this case, ProQuest returned the greatest differential and was a statistical outlier, by retrieving 3,266 more records than the equivalent PubMed search. The remaining 3 platforms returned fewer records than the PubMed search, although they were closer, ranging from 167 to 348 fewer records.

**Figure 1** Total search result counts for each of the 29 search sets



The four plots are organized by the magnitude of results.

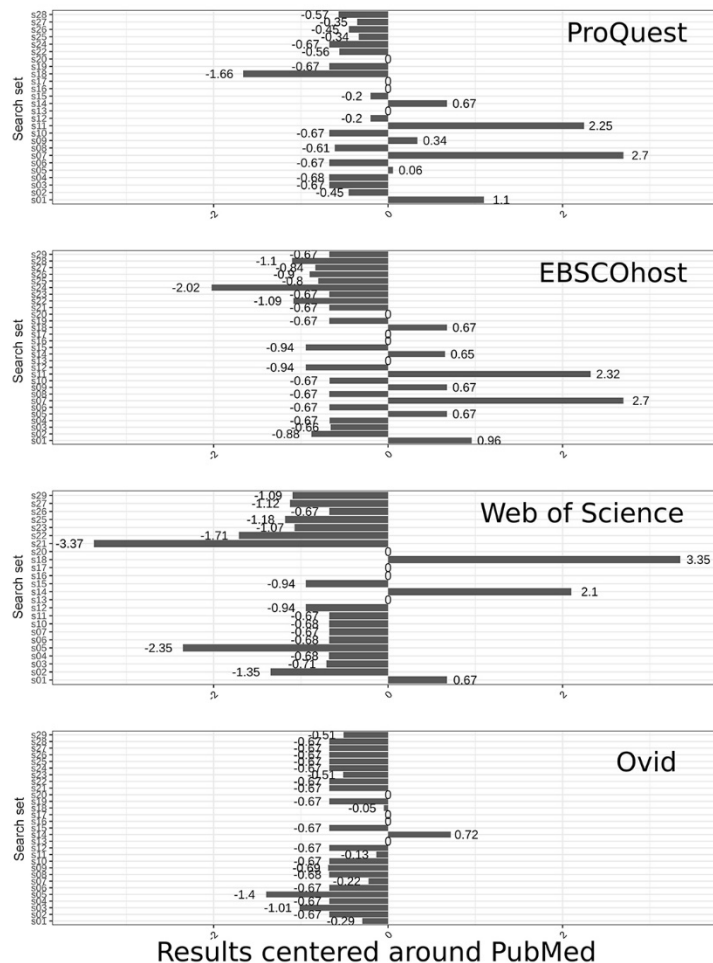
Figures 2 and 3 present the z-scores and highlight the deviations for all searches compared to PubMed. Figure 2 includes the search sets within  $\pm 3.5$  deviations from PubMed, and Figure 3 includes deviations outside that range that are, therefore, classified as outliers. In both figures, PubMed results are represented by the center, that is, 0 deviations.

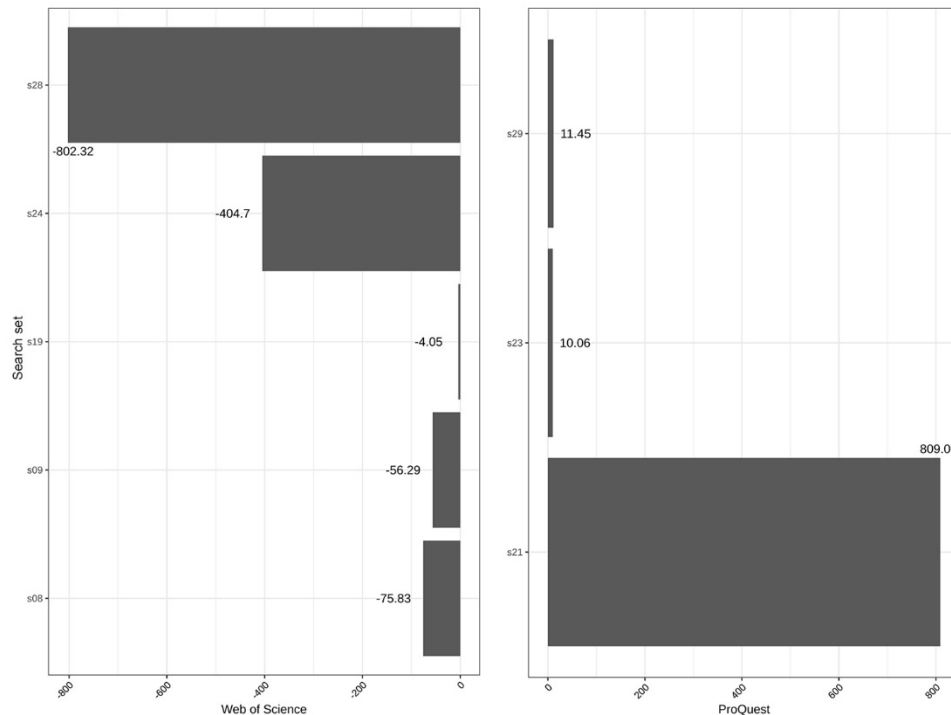
Figure 2 highlights the substantial inconsistencies between PubMed and the 4 platforms and low consistency in the deviation across the 4 platforms themselves. For example, search s07 (a single MeSH term, "Dementia," not exploded, with an additional keyword, "immune," and a date restriction) shows that ProQuest and EBSCOhost returned results equivalent to an average of 2.7 more records than PubMed (Figure 2; supplemental Appendix B). However, that same search returned fewer average records for Ovid and

Web of Science. Two searches (s16 and s17) consistently retrieved the same number of results across all platforms. A third search (s13) retrieved the same results across PubMed, ProQuest, EBSCOhost, and Ovid but not Web of Science, and a fourth search (s20) was consistent across PubMed, ProQuest, EBSCOhost, and Web of Science but not Ovid. In searches s13 and s20, respectively, Web of Science was only 2 results below the other platforms, and Ovid retrieved only 1 fewer result.

Figure 3 shows the outliers, defined as search result counts beyond  $\pm 3.5$  standard deviations away from PubMed. Only Web of Science and ProQuest had search result count outliers, with each for different search sets. All Web of Science outliers included results that returned fewer records than PubMed, and all ProQuest outliers included results that returned more records than PubMed.

**Figure 2** Deviations per platform from PubMed's MEDLINE, excluding outlier searches



**Figure 3** Outlier search results in ProQuest and Web of Science

Numbers represent modified z-scores. A score outside of  $\pm 3.5$  is considered an outlier.

Two of the high outliers for ProQuest searches included at least 1 MeSH term that appeared on multiple branches and that were exploded (s23,  $m_i=10.06$ ; s29,  $m_i=11.45$ ) (Figure 3). In both searches, ProQuest returned thousands more results than PubMed. Web of Science (s23,  $m_i=-1.07$ ; s29,  $m_i=-1.09$ ) also deviated by more than 1 standard deviation from PubMed in these 2 searches, but in the opposite direction, returning fewer records. EBSCOhost (s23,  $m_i=-0.67$ ; s29,  $m_i=-0.67$ ) and Ovid (s23,  $m_i=-0.51$ ; s29,  $m_i=-0.51$ ) also returned fewer results, but these results were much closer to PubMed's. Similar differences are seen in search s21, in which ProQuest retrieved thousands more results than PubMed ( $m_i=809.06$ ), although the other 3 platforms retrieved fewer results than the PubMed baseline. This search examined the equivalent of "All Fields" across the platforms combined with 2 journal titles.

As stated, Web of Science result counts deviated most often from PubMed searches. In the 15 Web of Science searches that deviated from the equivalent PubMed searches by at least 1 standard deviation, 5 of those searches were extreme outliers (s08, s09, s19, s24, and s28; supplemental Appendix A). The

first 2 searches (s08,  $m_i=-75.83$ ; s09,  $m_i=-56.29$ ) highlighted issues with how Web of Science exploded MeSH terms. The third search (s19,  $m_i=-4.05$ ) returned only 6 fewer records than PubMed but is considered an outlier relative to how closely the other 3 platforms matched PubMed's results. The fourth search (s24,  $m_i=-404.07$ ) returned 0 records even though PubMed retrieved 600 records, and the other 3 platforms returned approximately the same. The fifth search (s28,  $m_i=-802.32$ ) returned 0 records, compared to over 45,000 records retrieved on the other 4 platforms, with a query that included 2 MeSH terms.

Author name searches were problematic across the platforms except when they were attached to MeSH terms, which seemed to help disambiguate the names (s17 and s18; supplemental Appendix A). In a search for a single author name only, Ovid (s18,  $m_i=-0.05$ ) returned results that were nearly equal with PubMed results. However, by increasing magnitude, EBSCOhost ( $m_i=0.67$ ) returned more results, ProQuest ( $m_i=-1.66$ ) returned fewer results, and Web of Science ( $m_i=3.35$ ) returned more than PubMed. When the author name was attached to 2 MeSH terms (s17; supplemental Appendix A), all 4



platforms returned the same number of results as PubMed.

We found that very specific search queries were more likely to produce more consistent results across all five platforms. In addition to the search query described above that included MeSH terms and a single author name (s17; supplemental Appendix A), there were searches that resulted in perfect or nearly perfect agreement among all platforms (s13, s16, and s20; supplemental Appendix A). The first of these searches (s13) included two MeSH terms and a title keyword and exploded the second MeSH term. The second of these searches (s16) included two MeSH terms (one not exploded and one exploded) joined by a Boolean NOT and searched against one journal title. The third of these searches (s20) included a title term search against two journal titles.

Likewise, four other searches produced fewer records than PubMed but near consistent results among each other (s4, s6, s10, and s26). These were also very specific searches, including only MeSH terms. In addition, the first three of these searches were limited by publication dates. However, including only specific terms did not guarantee consistent results across all platforms. In particular, Web of Science often deviated from the others when only MeSH terms were included in the query. The deviations were likely the result of how Web of Science explodes terms.

## DISCUSSION

In this research, we constructed queries across five MEDLINE platforms in order to understand how search result counts vary after controlling for necessary differences in search query syntax across platforms. Hypothetically, content in the MEDLINE platforms is consistent across platforms because each uses MEDLINE records created by NLM. However, this assumption has lacked thorough scientific testing, which can be problematic, especially if studies combine multiple MEDLINE platforms under a single "MEDLINE" category [35]. Although one might expect some variation in search results across platforms since search interfaces and syntax are vendor-specific; in general, search results should be similar, if not identical, for queries that are equivalent.

It appears, however, that no MEDLINE platform can be a substitute for another MEDLINE platform, which is problematic if researchers, clinicians, and health information professionals do not have access to all of them and, thus, do not have the ability to cross-reference searches and de-duplicate search records when they conduct literature searches. The inability to substitute one MEDLINE platform for another can be caused by various interventions by platform vendors (possibly including data ingest workflows, term indexing and retrieval algorithms, and interface features) that affect record retrieval. Hence, our results challenge the expectation that all MEDLINE platforms produce equivalent results and that they should be treated as MEDLINE. The inconsistencies seen here across platforms should raise concerns about the impact of vendor-specific indexing algorithms. It appears that the features provided by the proprietary platforms have a significant impact on the retrieved results of even basic queries. This, in turn, affects the replication and reproducibility of search query development and, possibly, the conclusions drawn from those literature sets.

Practically speaking, the queries that returned the most similar result counts to their equivalent PubMed searches were multifaceted and included either MeSH terms or a title keyword and then were combined with another field, such as a journal title, author name, or journal search (e.g., s13, s16, s17, s20). However, deviations were not generally consistent across platforms nor in relation to specific query elements (e.g., specific metadata combinations). As such, there appear to be no ready solutions for mitigating inconsistencies in search results across platforms as an end user. Because perhaps few users have access to all MEDLINE platforms, this could be problematic, since, as noted, even one missing study can skew or alter scientific or clinical conclusions [23].

Although Ovid produced the most consistent results with PubMed, there were still differences in search result counts. In all those cases where Ovid and PubMed differed, Ovid returned fewer results (without de-duplicating). We were able to rule out that these differences were solely the result in lag time between MEDLINE updates, that is, the time between when PubMed is updated and the licensed platforms are updated, because the Ovid search

counts were lower even for those queries that were limited by publication dates (1950–2015).

Without knowing what has been left out of these search results, it would be difficult to know how those results might impact clinical care, especially because MEDLINE has been deemed an important source for practice and where even one record can have important consequences in treatment [35, 36]. As such, future studies should include research questions related to understanding the contents of retrieved sets in order to understand how the bibliographic records are influencing retrieval across the platforms.

Lag time between updates of the MEDLINE file across the platforms also could not explain differences in results for ProQuest and EBSCOhost. The higher counts in ProQuest and in EBSCOhost suggested that their indexing algorithms were more sensitive and defaulted to more inclusive retrieval sets. This claim was supported by ProQuest's highest outlier, which included a keyword search against 2 specific journal titles ( $s_{21}$ ,  $m_i=179.42$ ; supplemental Appendix A). Given the variances observed across platforms, it is important to understand under what conditions queries across MEDLINE platforms might be more sensitive.

One limitation of our study is that it is only a snapshot at one moment. Therefore, future studies could examine longitudinal changes in how these systems respond to basic searches to increase understanding of the effects that vendor-specific algorithms have on search result counts, because it could be that such algorithms are modified over time. Additional lines of research include examining how retrieval of non-indexed and in-process citations in PubMed's MEDLINE subset differentiate from comparable databases or subsets.

Also, as noted earlier, this study examined baseline differentiation for permutations of simple searches. However, searches documented in the literature for, among other things, systematic reviews should also be compared across platforms. Such studies could help researchers understand the maximum differentiation that these systems might exhibit since the queries documented in these studies are generally complex.

We also used PubMed search counts as the point of reference, and we did this because NLM is

responsible for both MEDLINE and the PubMed interface. However, other platforms could function as a point of reference and doing so might be useful in explaining differences in indexing, Boolean logic, and other aspects of searching. Lastly, although analyzing baseline deviation using search counts helps illustrate fundamental differences among MEDLINE platforms, future research could examine and compare the content of records that are returned to better understand the source of these deviations.

## FUNDING

The first author received support for this work with the 2018 Summer Faculty Research Fellowship Program of the College of Communication and Information at the University of Kentucky. This funding source had no involvement in the study design, collection, analysis, or interpretation of the data.

## REFERENCES

1. Amrhein V, Korner-Nievergelt F, Roth T. The earth is flat ( $p>0.05$ ): significance thresholds and the crisis of unreplicable research. *Peer J*. 2017 Jul 7;5:e3544.
2. Baker M. 1,500 scientists lift the lid on reproducibility. *Nature*. 2016 May 26;533(7604):452–4.
3. Open Science Collaboration. Estimating the reproducibility of psychological science. *Science*. 2015 Aug 28;349(6251):aac4716. DOI: <http://dx.doi.org/10.1126/science.aac4716>.
4. Moher D, Liberati A, Tetzlaff J, Altman DG, Group TP. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. *PLOS Med*. 2009 Jul 21;6(7):e1000097. DOI: <http://dx.doi.org/10.1371/journal.pmed.1000097>.
5. Cochrane. Cochrane handbook for systematic reviews of interventions [Internet]. Cochrane [cited 6 May 2019]. <<https://training.cochrane.org/handbook>>.
6. Buchanan S, Salako A. Evaluating the usability and usefulness of a digital library. *Libr Rev*. 2009 Oct 9;58(9):638–51. DOI: <http://dx.doi.org/10.1108/00242530910997928>.
7. Edwards A, Kelly D, Azzopardi L. The impact of query interface design on stress, workload and performance. In: Hanbury A, Kazai G, Rauber A, Fuhr N, eds. *Advances in information retrieval*. Springer International Publishing; 2015. p. 691–702. (Lecture Notes in Computer Science).
8. Goodman SN, Fanelli D, Ioannidis JPA. What does research reproducibility mean? *Sci Transl Med*. 2016 Jun 1;8(341):341ps12. DOI: <http://dx.doi.org/10.1126/scitranslmed.aaf5027>.

DOI: [dx.doi.org/10.5195/jmla.2019.622](https://doi.org/10.5195/jmla.2019.622)

9. Ho GJ, Liew SM, Ng CJ, Shunmugam RH, Glasziou P. Development of a search strategy for an evidence based retrieval service. *PLOS One*. 2016 Dec 9;11(12):e0167170. DOI: <http://dx.doi.org/10.1371/journal.pone.0167170>.
10. Peng RD. Reproducible research and biostatistics. *Biostatistics*. 2009 Jul;10(3):405–8. DOI: <http://dx.doi.org/10.1093/biostatistics/kxp014>.
11. Toews LC. Compliance of systematic reviews in veterinary journals with Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) literature search reporting guidelines. *J Med Libr Assoc*. 2017 Jul;105(3):233–9. DOI: <http://dx.doi.org/10.5195/jmla.2017.246>.
12. Lam MT, McDiarmid M. Increasing number of databases searched in systematic reviews and meta-analyses between 1994 and 2014. *J Med Libr Assoc*. 2016 Oct;104(4):284–9. DOI: <http://dx.doi.org/10.5195/jmla.2016.141>.
13. Bethel A, Rogers M. A checklist to assess database-hosting platforms for designing and running searches for systematic reviews. *Health Inf Libr J*. 2014 Mar;31(1):43–53.
14. Ahmadi M, Sarabi RE, Orak RJ, Bahaadinbeigy K. Information retrieval in telemedicine: a comparative study on bibliographic databases. *Acta Inform Med*. 2015 Jun;23(3):172–6. DOI: <http://dx.doi.org/10.5455/aim.2015.23.172-176>.
15. Younger P, Boddy K. When is a search not a search? a comparison of searching the AMED complementary health database via EBSCOhost, OVID and DIALOG. *Health Inf Libr J*. 2009 Jun;26(2):126–35. DOI: <http://dx.doi.org/10.1111/j.1471-1842.2008.00785.x>.
16. De Groote SL. PubMed, Internet Grateful Med, and Ovid. *Med Ref Serv Q*. 2000 Winter;19(4):1–13. DOI: [http://dx.doi.org/10.1300/J115v19n04\\_01](http://dx.doi.org/10.1300/J115v19n04_01).
17. Bramer WM, Giustini D, Kramer BM, Anderson P. The comparative recall of Google Scholar versus PubMed in identical searches for biomedical systematic reviews: a review of searches used in systematic reviews. *Syst Rev*. 2013 Dec 23;2:115.
18. Haase A, Follmann M, Skipka G, Kirchner H. Developing search strategies for clinical practice guidelines in SUMSearch and Google Scholar and assessing their retrieval performance. *BMC Med Res Methodol*. 2007 Jun 30;7:28.
19. Nourbakhsh E, Nugent R, Wang H, Cevik C, Nugent K. Medical literature searches: a comparison of PubMed and Google Scholar. *Health Inf Libr J*. 2012 Sep;29(3):214–22.
20. Craven J, Jefferies J, Kendrick J, Nicholls D, Boynton J, Frankish R. A comparison of searching the Cochrane library databases via CRD, Ovid and Wiley: implications for systematic searching and information services. *Health Inf Libr J*. 2014 Mar;31(1):54–63.
21. Dunikowski LG. EMBASE and MEDLINE searches. *Can Fam Physician*. 2005 Sep 10;51(9):1191. (Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1479462/>. [cited 17 Jan 2019].)
22. Allison MM. Comparison of CINAHL® via EBSCOhost®, OVID®, and ProQuest®. *J Electron Resour Med Libr*. 2006;3(1):31–50.
23. Boddy K, Younger P. What a difference an interface makes: just how reliable are your search results? *Focus Altern Complement Ther*. 2009;14(1):5–7. DOI: <http://dx.doi.org/10.1111/j.2042-7166.2009.tb01854.x>.
24. Katchamart W, Faulkner A, Feldman B, Tomlinson G, Bombardier C. PubMed had a higher sensitivity than Ovid-MEDLINE in the search for systematic reviews. *J Clin Epidemiol*. 2011 Jul;64(7):805–7. DOI: <http://dx.doi.org/10.1016/j.jclinepi.2010.06.004>.
25. Boeker M, Vach W, Motschall E. Semantically equivalent PubMed and Ovid-MEDLINE queries: different retrieval results because of database subset inclusion. *J Clin Epidemiology*. 2012 Aug;65(8):915–6. DOI: <http://dx.doi.org/10.1016/j.jclinepi.2012.01.015>.
26. Collins M. Updated algorithm for the PubMed best match sort order. *NLM Tech Bull [Internet]*. 2017 Jan–Feb;(414):e3 [cited 17 Jan 2019]. <[https://www.nlm.nih.gov/pubs/techbull/jf17/jf17\\_p\\_m\\_best\\_match\\_sort.html](https://www.nlm.nih.gov/pubs/techbull/jf17/jf17_p_m_best_match_sort.html)>.
27. Iglewicz B, Hoaglin DC. How to detect and handle outliers. Milwaukee, WI: ASQC Quality Press; 1993.
28. R Foundation. The R project for statistical computing [Internet]. The Foundation [cited 17 Jan 2019]. <<https://www.r-project.org/>>.
29. Wickham H, François R, Henry L, Müller K, RStudio. dplyr: a grammar of data manipulation [Internet]. 2018 [cited 17 Jan 2019]. <<https://CRAN.R-project.org/package=dplyr>>.
30. Wickham H. ggplot2: elegant graphics for data analysis. New York, NY: Springer; 2009.
31. Auguie B, Antonov A. gridExtra: miscellaneous functions for “grid” graphics [Internet]. 2017 [cited 17 Jan 2019]. <<https://CRAN.R-project.org/package=gridExtra>>.
32. Makiyama K. magicfor: magic functions to obtain results from for loops [Internet]. 2016 [cited 17 Jan 2019]. <<https://CRAN.R-project.org/package=magicfor>>.
33. Wickham H. Reshaping data with the reshape package. *J Stat Softw*. 2007 Nov 13;21(1):1–20. DOI: <http://dx.doi.org/10.18637/jss.v021.i12>.
34. Dahl DB, Scott D, Roosen C, Magnusson A, Swinton J, Shah A, Henningsen A, Puetz B, Pfaff B, Agostinelli C, Loehnert C, Mitchell D, Whiting D, da Rosa F, Gay G, Schulz G, Fellows I, Laake J, Walker J, Yan J, Andronic L, Loecher M, Gubri M, Stigler M, Castelo R, Falcon S, Edwards S, Garbade S, Ligges U. xtable: export tables to LaTeX or HTML [Internet]. 2018 [cited 17 Jan 2019]. <<https://CRAN.R-project.org/package=xtable>>.
35. Dunn K, Marshall JG, Wells AL, Backus JEB. Examining the role of MEDLINE as a patient care information resource: an analysis of data from the Value of Libraries study. *J Med Libr Assoc*. 2017 Oct;105(4):336–46. DOI: <http://dx.doi.org/10.5195/jmla.2017.87>.
36. Ogilvie RI. The death of a volunteer research subject: lessons to be learned. *CMAJ*. 2001 Nov 13;165(10):1335–7.

**SUPPLEMENTAL FILES**

- **Appendix A:** [All queries, search logic descriptions, and search count results](#)
- **Appendix B:** [Search differentials \(search count and modified z-score differences\) with PubMed as the reference point for each search set](#)
- **Appendix C:** [R programming code used to analyze the data](#)
- **Appendix D:** [Data set formatted for use in the R programming language](#)

**AUTHORS' AFFILIATIONS**

**Christopher Sean Burns**, [sean.burns@uky.edu](mailto:sean.burns@uky.edu), <http://orcid.org/0000-0001-8695-3643>, Associate Professor, School of Information Science, University of Kentucky, Lexington, KY

**Robert M. Shapiro II**, [shapiro.rm@uky.edu](mailto:shapiro.rm@uky.edu), <http://orcid.org/0000-0003-4556-702X>, Assistant Professor, School of Information Science, University of Kentucky, Lexington, KY

**Tyler Nix**, [tnix@umich.edu](mailto:tnix@umich.edu), <http://orcid.org/0000-0002-0503-386X>, Informationist, Taubman Health Sciences Library, University of Michigan, Ann Arbor, MI

**Jeffrey T. Huber**, [jeffrey.huber@uky.edu](mailto:jeffrey.huber@uky.edu), <http://orcid.org/0000-0002-3317-0482>, Professor, School of Information Science, University of Kentucky, Lexington, KY

*Received October 2018; accepted March 2019*



Articles in this journal are licensed under a [Creative Commons Attribution 4.0 International License](#).



This journal is published by the [University Library System](#) of the [University of Pittsburgh](#) as part of its [D-Scribe Digital Publishing Program](#) and is cosponsored by the [University of Pittsburgh Press](#).

ISSN 1558-9439 (Online)