Exploring MEDLINE Journals that Publish Randomized Controlled Trial Results, Trial Registry Unique Identifiers, and Journal Policies

Kristina Elliott,
Associate Fellow, 2014-2015

February 25, 2015

Project Sponsors:

Deborah A. Zarin, MD

Tony Tse, PhD

Contents

Background

Clinical trial registration fosters transparency in medical research. In the current atmosphere, published clinical trial results are subject to selective reporting bias, where positive outcomes are published at a higher rate than negative or inconclusive results.[1,2] This creates an incomplete picture of the research evidence, leading to issues such as the conduct of redundant trials that unnecessarily put human volunteers at risk and the application of incomplete information into patient care.

To mitigate this bias, the International Committee of Medical Journal Editors (ICMJE) set forth a policy in 2004 recommending all medical journal editors require prospective registration of clinical trials as a condition for publication.[3] The ICMJE accepts registration in World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP) primary registries and ClinicalTrials.gov, maintained by the National Library of Medicine (NLM). Prospective and complete registration of a trial in one of these registries satisfies the requirements of the ICMJE registration policy, and published articles resulting from trials are advised to list registry numbers at the end of the paper abstract.

Since the policy was first proposed, editors from both ICMJE member and nonmember journals have publicly adopted registration policies.[4] The ICMJE policy went into effect in July 2005, and registration in the months following increased sharply.[5] However, mixed compliance with the policy's full requirements has been found and there remain problems with the quality of registration even among registered trials.[6,7]Despite broad support for trial registration among key stakeholders, some biomedical journal editors and trialists are unsure of its benefits.[8,9]

Nearly 10 years after the ICMJE policy went into effect, questions still remain about whether or not trial registration reduces bias,[10] and many published trials still lack registration.[11] Although journals

---

[1] Ross JS. Confronting Bias. *JAMA Intern Med*. 2015;175(2):307. doi:10.1001/jamainternmed.2014.6933.
[2] Malički M, Marušić A, OPEN (to Overcome failure to Publish nEgative fiNdings) Consortium. Is there a solution to publication bias? Researchers call for changes in dissemination of clinical research results. *J Clin Epidemiol*. 2014;67(10):1103-1110. doi:10.1016/j.jclinepi.2014.06.002.
[3] DeAngelis CD, Drazen JM, Frizelle FA, et al. Clinical trial registration: A statement from the International Committee of Medical Journal Editors. *JAMA*. 2004;292(11):1363-1364. doi:10.1001/jama.292.11.1363.
[4] Journals following the ICMJE recommendations. International Committee of Medical Journal Editors Web site. http://www.icmje.org/journals-following-the-icmje-recommendations. Accessed February 23, 2015.
[5] Zarin DA, Tse T, Ide NC. Trial registration at ClinicalTrials.gov between May and October 2005. *N Engl J Med*. 2005;353:2779-2787. doi: 10.1056/NEJMsa053234.
[6] Viergever R, Karam G, Reis A, Ghersi D. The quality of registration of clinical trials: Still a problem. *PLoS One*. 2014. doi: 10.1371/journal.pone.0084727.
[7] Ross J, Mulvey GK, Hines EM, Nissen SE, Krumholz HM. Trial publication after registration in ClinicalTrials.gov: A cross-sectional analysis. *PLoS One*. 2009. doi: 10.1371/journal.pmed.1000144.
[8] Wager E, Williams P. "Hardly worth the effort"? Medical journals' policies and their editors' and publishers' views on trial registration and publication bias: Quantitative and qualitative study. *BMJ*. 2013;347:f5248. doi: http://dx.doi.org/10.1136/bmj.f5248.
[9] Reveiz L, Krleza-Jerić K, Chan AW, de Aguiar S. Do trialists endorse clinical trial registration? Survey of a Pubmed sample. *Trials*. 2007;8(30). doi:10.1186/1745-6215-8-30.
[10] Emdin C, Odutayo A, Hsiao A, Shakir M, Hopewell S, Phil D, Rahimi K, Altman DG. Association of cardiovascular trial registration with positive study findings: Epidemiological study of randomized trials (ESOER). *JAMA Intern Med*. 2015;175(2):304-307.doi:10.1001/jamainternmed.2014.6924.

have adopted policies requiring trial registration as a prerequisite for publication, compliance is still unclear.

This exploratory cross-sectional analysis examines the extent of trial registration policy adoption and implementation across MEDLINE-indexed journals that published results of at least one randomized controlled trial in 2013.

Introduction

This fall project's objective was to understand factors that influence clinical trial registration through pilot quantitative and qualitative analyses on a set of MEDLINE journal citations for the results of randomized controlled trials (RCTs). The quantitative analysis characterized the sample for the frequency of trial registration across journal titles, stratified by impact factor (≥10 and <10) and subject category type (general and specialized). The qualitative analysis explored the degree of stated registration policy compliance among subsets of journals categorized in the quantitative analysis. Analysis was performed on a set of MEDLINE citations retrieved from PubMed after creation and refinement of a comprehensive and specific original search strategy.

Quantitative analysis was completed to characterize the results set in two ways. A comparison of citations with and without MEDLINE-indexed trial registry unique identifiers (or "registry numbers") was completed to attain an estimated rate of compliance (percentage registered) among the journals. Given the significant range of this rate, assessing the journals' rate of compliance against their impact factors and subject category type was then used to identify trends and areas for qualitative analysis.

Qualitative analysis built upon the quantitative analysis and was used to explore individual journal policies in greater depth. The quantitative analysis had limitations that hindered drawing conclusions based solely on the analysis. Journals have different trial registration policies, and some have none at all. Conclusions about policy enforcement therefore required a qualitative approach that examined individual journal policies. The qualitative analysis was highly exploratory and used to develop preliminary findings on individual policies.

Methods

Search strategy

The Associate iterated a series of increasingly comprehensive and highly specific PubMed search strategies to retrieve a set of MEDLINE citations for journal articles published in 2013 (print and e-pub) that (1) report the results of registered and unregistered clinical trials and (2) are subject to the medical journal editors' stated registration policies (see Appendix A). Each subsequent refinement of the search strategy reduced the percentage of citations not subject to registration policies in the retrieved set (based on manual review of a subset), and eventually resulted in a more even distribution between citations with and without registry numbers (see Table 1).

[11] van de Wetering FT, Scholten RJPM, Harin T, Clarke M, Hooft L. Trial registration numbers are underreported in biomedical publications. *PLoS One*. 2012. doi: 10.1371/journal.pone.0049599.

Table 1 ***Results of Iterative Search Strategies***

Search Strategy	Total Number of Citations	Total Number of Citations with Registry Number (% Total)	Total Number of Citations without Registry Number (% Total) (Estimated % False Positives [FPs])
Iteration 1. Publication Type: All Phase Trials, Controlled, RCT, OR Clinical Trial	39,849	6,288 (16%)	33,561 (84%) (30% FPs)
Iteration 2. Publication Type: Only RCTs	25,831	5,022 (19%)	20,809 (81%) (47.5% FPs)
Iteration 3. Refined Strategy based on COCHRANE	11,020	5,318 (48%)	5,702 (52%) (5% FPs)

To distinguish between citations of registered and unregistered trials, MEDLINE-indexed registry numbers from the 16 World Health Organization (WHO) primary registries[12] and ClinicalTrials.gov, each of which is recognized by the ICMJE registration policy, served as searchable proxies for "successful" trial registration in this project in lieu of time-intensive manual review of PubMed citations. These registry numbers are discoverable through PubMed's Secondary Source ID ([si]) field, which carries a variety of databank names [13] including trial registries and their registry numbers[14] as listed on the last line of the abstract.

The initial iteration of the search strategy joined all related terms under the Clinical Trial Publication Type MeSH tree except for Observational Studies (not covered by the ICMJE policy) and Multicenter Studies (applies to studies not covered by ICMJE policy such as epidemiological studies) either with ("AND") or without ("NOT") registry Secondary Source IDs as a pair of queries (see Appendix A, First Search Strategy). This search, however, proved unreliable; randomized manual review of 40 retrieved citations without a registry number by the Associate indicated that many fall outside the scope of the ICMJE policy, such as nonhuman studies, protocol descriptions without trial results, meta- and secondary analyses, and other non-interventional studies. These nonrelevant results were considered "false positive" (FP) results and this process was used to evaluate the precision of each subsequent iteration (see Table 1*, Estimated % False Positives [FPs]*).

[12] WHO International Clinical Trial Registry Platform (ICTRP) Primary Registries. World Health Organization Web site. http://www.who.int/ictrp/network/primary/en. Accessed February 23, 2015.
[13] MEDLINE databank sources. U.S. National Library of Medicine Web site. http://www.nlm.nih.gov/bsd/medline_databank_source.html. Accessed February 23, 2015.
[14] Clinical trial registry numbers in MEDLINE/PubMed records. U.S. National Library of Medicine Web site. http://www.nlm.nih.gov/bsd/policy/clin_trials.html. Accessed February 23, 2015.

The project sponsors agreed to limit the search to the Randomized Controlled Trial (RCT) Publication Type because the ICMJE policy unambiguously applies to citations accurately indexed as RCTs. The remainder of the search strategy remained the same (see Appendix A, Second Search Strategy). The manual review of the search results for FPs revealed decreased precision with citations that either (1) fell outside the scope of the ICMJE policy (e.g., a "vignette experiment"[15]) or (2) did not appear to be RCTs as defined by the ClinicalTrials.gov data element definition, the MeSH term scope note, or the MEDLINE indexing policy (e.g., trials with nonrandomized interventions).

To develop a final search strategy with a high degree of specificity to capture exclusively RCT citations whether or not indexed as a Randomized Controlled Trial Publication Type by NLM indexers in MEDLINE, the Associate conducted a literature review to identify published search strategies for finding RCTs in PubMed (see Appendix B). In addition to numerous search strategies,[16,17,18] the review also yielded published reports of problems in retrieving RCTs using Publication Type searching in PubMed, attributed mainly to limitations in the indexing process including the short time spent on indexing an article and indexers' differing views on what constitutes an RCT.[19] The literature also showed that Publication Type searching was not highly sensitive,[18] retrieving a high proportion of nonrelevant results. The Associate found the literature about searching PubMed for RCTs to use in systematic reviews to be most useful and built upon the Cochrane Highly Sensitive Search Strategy for Identifying Randomized Trials in MEDLINE[20] to create a new, comprehensive, and specific search strategy that did not significantly sacrifice sensitivity (see Appendix A, Final Search Strategy based on COCHRANE). This strategy incorporated title/abstract text search for terms common in published RCTs, such as "randomized," "randomly," and variants. Manual review of the new results found decreased retrieval of nonrelevant results that were outside the scope of the ICMJE policy.

---

[15] Mulder R, Pouwelse M, Lodewijkx H, Bolman C. Workplace mobbing and bystanders' helping behaviour towards victims: The role of gender, perceived responsibility and anticipated stigma by association. *Int J Psychol*. 2014;49(4):304-12. doi: 10.1002/ijop.12018.

[16] Royle PL, Waugh NR. Making literature searches easier: A rapid and sensitive search filter for retrieving randomized controlled trials from PubMed. *Diabet Med*. 2007;24(3):308-11. doi: 10.1111/j.1464-5491.2007.02046.x.

[17] Haynes RB, McKibbon KA, Wilczynski NL, Walter SD, Werre SR, Hedges Team. Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: Analytical survey. *BMJ*. 2005;330(7501):1179. doi: http://dx.doi.org/10.1136/bmj.38446.498542.8F.

[18] McKibbon KA, Wilczynski NL, Haynes RB, Hedges Team. Retrieving randomized controlled trials from Medline: A comparison of 38 published search filters. *Health Info Libr J*. 2009;(3):187-202. doi: 10.1111/j.1471-1842.2008.00827.x.

[19] Wieland LS, Robinson KA, Dickersin K. Understanding why evidence from randomised clinical trials may not be retrieved from Medline: comparison of indexed and non-indexed records. *BMJ*. 2012;344:d7501. doi: 10.1136/bmj.d7501.

[20] The Cochrane highly sensitive search strategies for identifying randomized trials in MEDLINE. Cochrane Handbook for Systematic Reviews of Interventions. http://handbook.cochrane.org/chapter_6/6_4_11_1_the_cochrane_highly_sensitive_search_strategies_for.htm. Accessed February 23, 2015.

Quantitative analysis

The two sets of citations (those with and without registry numbers as proxies for registered RCTs and unregistered RCTs, respectively) retrieved using the final search strategy were downloaded from PubMed as Comma Separated Values-formatted files (.CSV) and imported into Excel for analysis. The total percentage of citations with registry numbers for each journal was calculated. This results set was subdivided several times throughout the rest of the study for analysis. Each subdivided set was calculated for average number of RCT citations; average percentage registered; mode number of citations; median number of citations; and range of citations.

To analyze these data by journal impact factor and subject category, the journals were matched to their 2013 impact factors and medical subject categories downloaded from Journal Citation Reports[21] (JCR). All journal titles in the sample of citations and listed as "Medicine, General & Internal" on JCR were categorized as "general" subject category journals; the remaining journal titles were categorized as "specialized" subject category journals. Because some non-general journal titles were associated with more than one specialty subject category, duplicates were removed from the list before analysis.

These two sets of general and specialized journals were further categorized by impact factor as "high impact" (i.e., 2013 JCR impact factor of 10 or higher) and "low impact" (i.e., 2013 JCR impact factor lower than 10) based on a previously reported study.[22] This left four sets of journals that would be used in the qualitative analysis: a set of high-impact/general subject category journals; high-impact/specialized journals; low-impact/general journals; and a set of low-impact/specialized journals.

The average number of RCT citations per journal was calculated for each of these four sets. For all sets except the high-impact/general journals, five journals at or above this average were randomly selected for further exploration in the qualitative analysis. There were only six journals in the high-impact/general set, and excluding journals below the average number of RCT citations would have left fewer than five for analysis; therefore, high-impact/general journals were randomized without the exclusion process. In the low-impact/specialized subject set, the journal *Trials* was excluded from randomization due to the high number of published protocols for RCTs, which are not covered by the ICMJE policy.

Qualitative analysis

The websites of the 20 randomly selected journal titles were reviewed for a statement of their trial registration policies. Journal titles that form the ICMJE and nonmembers that stated their adoption of the ICMJE policy were noted. All other journal titles were individually examined for registration policies through website searching and reading author guidelines. Any requirements related to trial registration were noted. The percentage of registered RCTs by journal were matched to the noted policy and used to assess degree of policy compliance.

---

[21] Thomson Reuters Journal Citation Reports. Thomson Reuters InCites Web site. http://about.jcr.incites.thomsonreuters.com. Accessed February 23, 2015.

[22] Becker JE, Krumholz HM, Ben-Josef G, Ross JS. Reporting of results in ClinicalTrials.gov and high-impact journals. *JAMA*. 2014;311(10):1063-1065. doi:10.1001/jama.2013.285634.

Results

Quantitative analysis

The final search strategy retrieved a total of 11,020 citations with RCTs in 1,590 unique journal titles published in 2013 (mean number of citations/journal title: 6.90; mode: 1; median: 3; range: 1 to 363) (see Figure 1). This total included 810 journals with at least one RCT citation that included a registry number, and 1,422 journals publishing at least one citation without a registry number. There was an overlap of 642 journals between these, representing journals that have published at least one RCT citation with and one citation without a registry number.

Of the total, 5,318 citations (48%) listed registry numbers (mean: 3.34; mode: 0; median: 1; range: 0 to 359), and 5,702 (52%) did not list registry numbers (mean: 3.59; mode: 1; median: 2; range: 0 to 82).

JCR impact factors were matched for only 843 of the 1,590 (53.02%) journal titles in the total sample. Since manual review of citations from the journal *Trials* showed it mainly published protocols and therefore fell outside the scope of this study, *Trials* was excluded from further analysis. This left the total of journals at 842. The number of journals and average percentage of registered RCTs per impact factor and subject category varied (see Table 2).

Table 2 ***Journal Titles, RCT Citations, and Citations with Registry Number by Impact Factor and Subject Category***

2013 JCR Impact Factor	JCR Subject Category	Total Number of Journals	Average Number of RCTs per Journal in 2013	Average Percent with Registry Numbers
Low (<10)	General	47	5.83	40.05%
High (≥10)	General	6	51.67	89.97%
Low (<10)	Specialized	770	7.50	29.96%
High (≥10)	Specialized	19	29.84	67.64%
All	All	842	8.23	31.82%

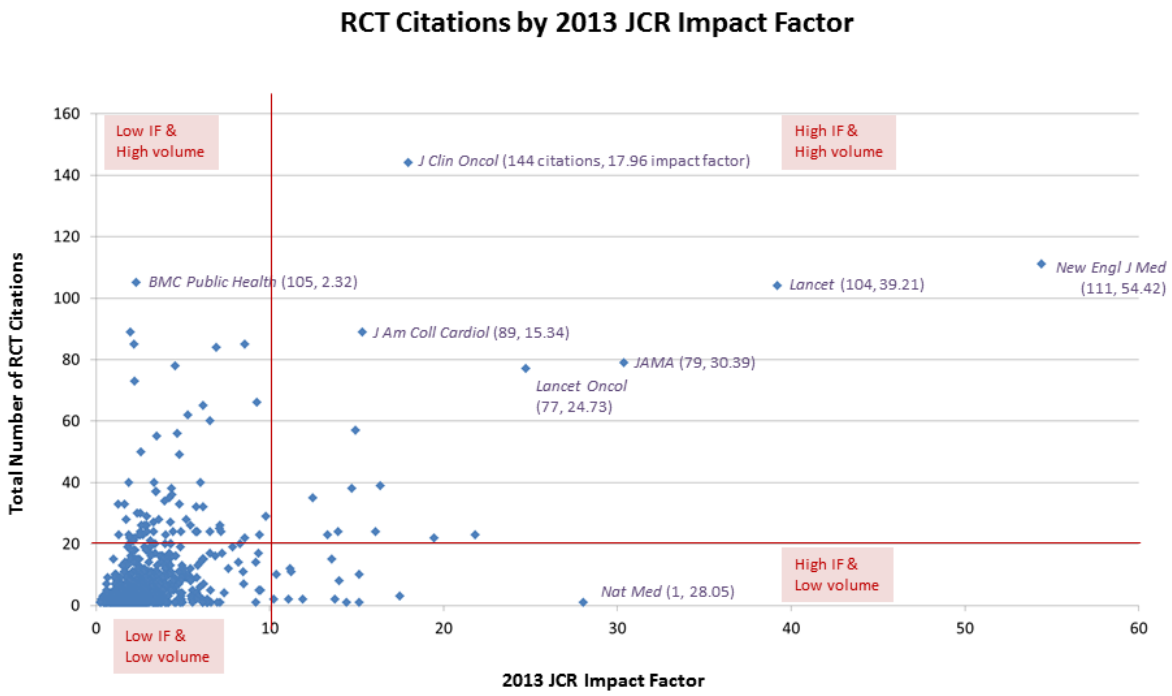
While low-impact journals published the majority of citations in the set matched to JCR, high-impact journals published more RCT results per journal (see Table 3).

Table 3 ***Characteristics of JCR Impact Factor/Category RCT Citations***

2013 JCR Impact Factor/Category	Number of RCT Citations	Mode Number of RCT Citations	Median Number of RCT Citations
Low/Specialized	5776 (83.00%)	1	3
High/Specialized	567 (8.20%)	2	22
Low/General	274 (4.00%)	1	2
High/General	310 (4.48%)	24	39
Total	6927 (100%)	1	3

Overall, high-impact journals were found to publish more RCTs on average in 2013 than low-impact journals (see Fig. 1).

Figure 1



Qualitative analysis

All 20 journals examined in the qualitative analysis had a stated clinical trial registration policy. Most of the journals (*n*=15 or 75%) followed the ICMJE policy, but only six were full members of the ICMJE. The other journals stated a nonspecific registration policy, such as only stating a journal should include a registry number at the end of an abstract.[23] The explicitness of the policy requirements varied across impact factor and subject category, and within each of the impact factor/subject category sets (see Fig. 2).

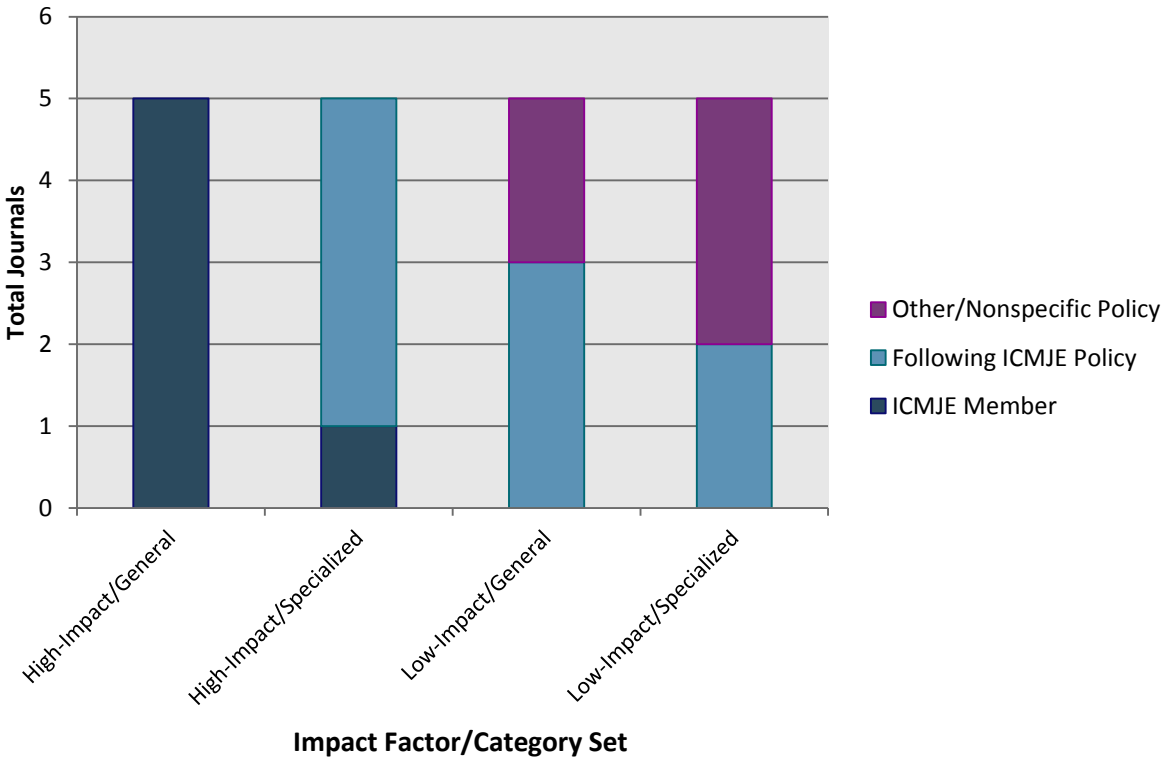
Of the high-impact/general journals, all followed the ICMJE policy. Only the high-impact/general journals were heterogeneous, with all these journals being full members of the ICMJE. The high-impact/specialized journals had one ICMJE member, with the other four listed as following the policy as nonmembers.[24]

[23] International Journal of Clinical Practice Author Guidelines. International Journal of Clinical Practice Web site. http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1742-1241/homepage/ForAuthors.html. Accessed February 25, 2015.

[24] Journals Following the ICMJE Recommendations. International Committee of Medical Journals Editors Web site. http://www.icmje.org/journals-following-the-icmje-recommendations. Accessed February 23, 2015.

Low impact factor journals were found to have less specific policies and none had full membership in the ICMJE. Of the five low-impact/specialized subject journals, two followed the ICMJE policy as nonmembers. Three of the five low-impact/general subject journals followed the ICMJE policy.

Figure 2 ***Policies by Impact Factor/Subject Category Set***



Of the 898 total RCT citations examined in the qualitative analysis, 357 (40%) were from high-impact/general subject journals. High-impact/specialized subject journals published 296 citations (33%). The remainder were distributed between low-impact/specialized subject journals (*n*=165 or 18%) and low-impact/general subject journals (*n*=80 or 9%).

As reported for the quantitative analysis results, the high impact factor journals used in the qualitative analysis also published the most RCTs on average in 2013 (71.4 RCTs), with high-impact/specialized subject journals second (59.2 RCTs). Low-impact/specialized subject journals published an average of 33 RCTs, and low-impact/specialized subject journals published an average of 16 RCTs.

When compared to the impact factor/category sets they were randomly selected from, the journals used for qualitative analysis had higher percentages of registered citations per journal. Journals in the high-impact/general subject set had the highest average percentage of citations with a registry number (92.7%) followed by low-impact/general subject journals (73.73%) high-impact/specialized subject journals (71.2%), and low-impact/specialized journals (60.88%).

Discussion

This fall project used quantitative and qualitative approaches to explore and understand factors that influence clinical trial registration among MEDLINE journals publishing the results of randomized controlled trials (RCTs). Quantitative analysis showed that trial registration could be increased, especially among low-impact journals. Qualitative analysis found that journal policies requiring registration as a prerequisite for publication have been widely adopted, but do not correspond to high rates of compliance. While the ICMJE statement encourages policy adoption to mitigate publication bias, this study found preliminary evidence that journal editors are not complying with their own policies.

Only a small number of journals were found to comply with trial registration policies by listing registration numbers in published articles. The final iteration of the search strategy found about half (*n*=5,702 or 51.74%) of the citations for articles reporting the results of RCTs in 2013 did not include a registry number. Citations that did include a registry number were published by only about a third (*n*=810 or 36.29%) of the journals in the overall sample. According to these results, a minority of journals seem to be publishing registry numbers in half of articles publishing the results of RCTs in 2013.

There was not a strict dichotomy among journals publishing results with registry numbers. Journals publishing both articles with and without a registry number made up 28.76% (*n*=642) of the total. This implies journals are not consistent with requiring registry numbers in articles.

These findings are preliminary and limited. Some registered trials can still be missed during MEDLINE indexing, especially when registry number is not listed at the end of an abstract, even when it is listed within the full text of the article. Additionally, the number of RCTs published per journals was found to range greatly, with the majority of journals publishing only one RCT citation (mode=1). More detailed analyses, including manual review of full text articles for registry numbers and searching WHO ICTRP and ClinicalTrials.gov registries for registered trials published without registry numbers, are necessary to confirm the findings of this exploratory project and account for these variations.

Examining journals by impact factor provided further clarity. These data suggest journal impact factor is more likely than subject category to be associated with the degree to which registry numbers are published in MEDLINE RCT citations. The 25 high impact factor journals were found to be more likely to publish at least one RCT result citation with a registry number (*n*=22 or 92%) than the 817 low-impact journals (*n*=438 or 53.61%) in 2013 (i.e., a differential of 38.39%). When compared by subject category, the 53 general category journals were only somewhat more likely to publish RCT citations with registry numbers (*n*=34 or 64.15%) than the 789 specialized category journals (*n*=427 or 54.12%).

A confounding factor is that many ICMJE member journals are also high-impact and therefore more likely to follow the ICMJE registration policy. For example, *New England Journal of Medicine* had the highest impact factor (54.42) and published the most registered RCT citations in its subject category with 110. Even after removal of the 9 ICMJE member journal titles from the high-impact set, the average percentage of citations with a registry number was 69.12%.

Although low impact factor journals constituted the vast majority of journals in the sample, high-impact journals were found to publish more RCT citations on average (30.54 MEDLINE citations per journal, or 12.66% of all citations) than low-impact journals (7.41 citations per journal).

The results of the qualitative analysis are consistent. All 10 high impact factor journals had explicit registration policies, while only half (*n*=10 or 50%) of the low-impact journals had such policies. Low adoption of registration policies by medical journal editors of low-impact journals has been shown in previous research, including among low-impact journals that publish few trial results. [8]

The final search strategy developed for this project may be useful for other applications at the NLM. Single-term Randomized Controlled Trial Publication Type searching in MEDLINE/PubMed has been shown to balance recall and precision,[25] with manual review of results needed to select relevant results. The timeframe and scope of this exploratory study did not allow for extensive manual review in order to select a study set of relevant RCT citations; therefore, the search strategy aimed for higher precision (i.e., a higher proportion of relevant results).

To limit the reduction in recall that follows an increase in precision, the final search strategy incorporated the COCHRANE Highly Sensitive Search Strategy for identifying randomized trials in MEDLINE but filtered out search terms found in the review of nonrelevant citations (e.g., "*meta-analysis*" or "*prospective*" in title and abstract). Limited review of the final search strategy results found that, when compared to a small set of the COCHRANE strategy results, retrieval of these nonrelevant citations was decreased and precision indeed appeared to have increased.

Limitations

Overall, the depth of the analysis in this project was limited by the timeframe required of a fall Associate project. Although it wasn't possible in this project, a larger selection of journal policies to review could provide better insight into policy compliance.

Some published results of RCTs were excluded in the final iteration of the search. This search was conducted only on MEDLINE indexed journals, limiting the scope of this project, as RCTs are also published in nonindexed journals. The search was also limited to English language citations and citations with an abstract.

This study used MEDLINE-indexed registry numbers as a proxy for evaluating the degree of registration policy compliance on a large scale, but this can only examine compliance with the registration requirement. Citations with registry numbers do not necessarily comply with all journal registration policy requirements; listing of registry numbers in a published article is only one of the ICMJE's required practices. For example, the ICMJE also requires prospective registration (defined as "at

[25] McKibbon KA, Wilczynski NL, Haynes RB; Hedges Team. Retrieving randomized controlled trials from Medline: A comparison of 38 published search filters. *Health Info Libr J*. 2009(3):187-202. doi: 10.1111/j.1471-1842.2008.00827.x.

or before the time of first patient enrollment") and submission of complete registration data.[26] Previous research has found published registered trials that have not met these requirements.[27] This study can only assess the degree of compliance with the first step of the registration process and one policy requirement.

Additionally, although registry numbers are indexed for PubMed, some registered trials can still be indexed without a number due to error. These Secondary Source IDs are added during the data creation phase of the indexing process, and citations that do not follow a structured abstract format and clearly list a registry number at the end of an abstract may not have the ID added. The search string does target citations that have language indicative of a trial registry number in the abstract text (i.e., *"trial registration"[Title/Abstract]*), but registry numbers can still appear without searchable, signifying text in the abstract or in the nonsearchable full text of the article.

Recommendations

This project is exploratory, and there are improvements that could be made to the search process to refine the results and limit nonrelevant citations outside the scope of this study. Published protocols continue to appear in the results set, and a new search strategy that eliminates this type of publication would be valuable. Additionally, greater manual review of search results can inform search strategy improvement.

Due to the large number of results, manual review was not feasible in the time period required for this project. However, manual review is the only way to gain a sense of the completeness and accuracy of search results. The Associate recommends determining a reasonable, representative number of search results for review before changing the current search strategy.

Although impact factor is a useful tool for analysis in this project, there were difficulties in matching JCR's journal abbreviations with the NLM's. Manual matching is required, since the Associate prioritized matching high impact titles. For more complete analysis, manual matching for all titles is required.

The use of MEDLINE/PubMed searching to compare large sets of results is novel, with traditional use of PubMed designed to ultimately rely on user selection of relevant results. If comparison of PubMed results as aggregate data shows this method to be a reliable measure of trends in publishing and indexing, the NLM can use an iteration of this study's final search strategy to review and improve the efficiency and accuracy of clinical trial indexing. This can inform more consistent indexing of trials, including updates in indexing policy and the Clinical Trials MeSH tree. Consistent indexing of RCTs would also make ongoing monitoring of registration compliance more feasible for researchers and

[26] Clinical Trial Registration. International Committee of Medical Journal Editors Web site. http://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html. Accessed February 24, 2015.

[27] Huić M, Marušić M, Marušić A. Completeness and changes in registered data and reporting bias of randomized controlled trials in ICMJE journals after trial registration policy. *PLoS One*. 2011. doi: 10.1371/journal.pone.0025258.

policymakers outside the NLM. The results of this project and the expertise of both the ClinicalTrials.gov and indexing staff should continue to be shared and consulted to support this goal.

Acknowledgments

I thank Dr. Deborah Zarin for her project sponsorship, continuous support of this project, and recognition of my dedicated effort. I would like to especially acknowledge Dr. Tony Tse for his unceasing support, curiosity, and investment in not only this project but my personal and professional growth.

Thank you to Lou Knecht, who is a true model Associate and knows how to make things happen. Also thank you to Susan Schmidt, Olga Printseva, Mary Ann Hantakas, and Melissa Yorks for their help and interest in my search strategy.

To those who know how to motivate, inspire, and encourage me – thank you to Dr. Kathel Dunn, Dr. Wanda Whitney, Maureen Madden, Lori Harris, Erin Foster, and Ariel Deardorff.

Appendix A

Iterative Search Strategies

First Search Strategy

ALL registered clinical trials EXCEPT Multicenter Studies AND Observational Studies IN 2013 IN Medline:

(Clinical Trial, Phase I[pt] OR Clinical Trial, Phase II[pt] OR Clinical Trial, Phase III[pt] OR Clinical Trial, Phase IV[pt] OR Controlled Clinical Trial[pt] OR Randomized Controlled Trial[pt] OR Clinical Trial[pt:noexp]) AND medline[sb] AND 2013:2013[dp] AND (ANZCTR[si] OR ReBec[si] OR ChiCTR[si] OR CRiS[si] OR CTRI[si] OR RPCEC[si] OR EudraCT[si] OR DRKS[si] OR IRCT[si] OR JPRN[si] OR NTR[si] OR PACTR[si] OR SLCTR[si] OR clinicaltrials.gov[si] OR ISRCTN[si])

ALL clinical trials EXCEPT Multicenter Studies AND Observational Studies WITHOUT a registry ID IN 2013 IN Medline:

(Clinical Trial, Phase I[pt] OR Clinical Trial, Phase II[pt] OR Clinical Trial, Phase III[pt] OR Clinical Trial, Phase IV[pt] OR Controlled Clinical Trial[pt] OR Randomized Controlled Trial[pt] OR Clinical Trial[pt:noexp]) AND medline[sb] AND 2013:2013[dp] NOT (ANZCTR[si] OR ReBec[si] OR ChiCTR[si] OR CRiS[si] OR CTRI[si] OR RPCEC[si] OR EudraCT[si] OR DRKS[si] OR IRCT[si] OR JPRN[si] OR NTR[si] OR PACTR[si] OR SLCTR[si] OR clinicaltrials.gov[si] OR ISRCTN[si])

Second Search Strategy

ALL RCT publication type *with* a registry number with humans in 2013 in MEDLINE:

Randomized Controlled Trial[pt] AND medline[sb] AND 2013[PDAT]:2013[PDAT] AND (ANZCTR[si] OR ReBec[si] OR ChiCTR[si] OR CRiS[si] OR CTRI[si] OR RPCEC[si] OR EudraCT[si] OR DRKS[si] OR IRCT[si] OR JPRN[si] OR NTR[si] OR PACTR[si] OR SLCTR[si] OR clinicaltrials.gov[si] OR ISRCTN[si] OR TCTR[si]) AND "humans"[MeSH Terms]

ALL RCT publication type *without* a registry number with humans in 2013 in MEDLINE:

Randomized Controlled Trial[pt] AND medline[sb] AND 2013[PDAT] : 2013[PDAT] AND "humans"[MeSH Terms] NOT (ANZCTR[si] OR ReBec[si] OR ChiCTR[si] OR CRiS[si] OR CTRI[si] OR RPCEC[si] OR EudraCT[si] OR DRKS[si] OR IRCT[si] OR JPRN[si] OR NTR[si] OR PACTR[si] OR SLCTR[si] OR clinicaltrials.gov[si] OR ISRCTN[si] OR TCTR[si])

Final Search Strategy based on COCHRANE[28]

ALL RCTs (based on publication type and title/abstract searching) *with* a registry number with humans in 2013 in MEDLINE:

(randomized controlled trial[pt] OR controlled clinical trial[pt] AND randomized[Title/Abstract] OR randomised[Title/Abstract] OR placebo[Title/Abstract] OR "clinical trials as topic"[MeSH Terms:noexp] OR randomly[Title/Abstract] OR trial[Ti] NOT “meta-analysis”[Title/Abstract] NOT retrospective[Title/Abstract]) AND medline[sb] AND (2013:2013[dp]) AND "humans"[MeSH Terms] AND hasabstract[text] AND English[lang] AND (ANZCTR[si] OR ReBec[si] OR ChiCTR[si] OR CRiS[si] OR CTRI[si] OR RPCEC[si] OR EudraCT[si] OR DRKS[si] OR IRCT[si] OR JPRN[si] OR NTR[si] OR PACTR[si] OR SLCTR[si] OR clinicaltrials.gov[si] OR ISRCTN[si] OR TCTR[si] OR “trial registration”[Title/Abstract] OR “trial registry”[Title/Abstract]) NOT “observational study”[Ti] NOT “prospective”[Ti]

[28] The Cochrane highly sensitive search strategies for identifying randomized trials in MEDLINE. Cochrane Handbook for Systematic Reviews of Interventions. http://handbook.cochrane.org/chapter_6/6_4_11_1_the_cochrane_highly_sensitive_search_strategies_for.htm

ALL RCTs (based on publication type and title/abstract searching) *without* a registry number with humans in 2013 in MEDLINE:

(randomized controlled trial[pt] OR controlled clinical trial[pt]) AND (randomized[Title] OR randomised[Title] OR placebo[Title] OR randomly[Title] OR trial[Ti] NOT retrospective[Title/Abstract]) AND medline[sb] AND (2013:2013[dp]) AND "humans"[MeSH Terms] AND hasabstract[text] AND English[lang] NOT (ANZCTR[si] OR ReBec[si] OR ChiCTR[si] OR CRiS[si] OR CTRI[si] OR RPCEC[si] OR EudraCT[si] OR DRKS[si] OR IRCT[si] OR JPRN[si] OR NTR[si] OR PACTR[si] OR SLCTR[si] OR clinicaltrials.gov[si] OR ISRCTN[si] OR TCTR[si] OR “trial registration”[Title/Abstract] OR “trial registry”[Title/Abstract]) NOT “meta-analysis”[Title/Abstract] NOT “observational study”[Ti] NOT “prospective”[Ti]

Appendix B

Literature Review of Searching for Randomized Controlled Trials (RCTs) in PubMed

Search Strategy

The Associate searched PubMed for relevant articles on how to find and retrieve Randomized Controlled Trials (RCTs) in MEDLINE/PubMed.

Search strategies used:

"information storage and retrieval"[MeSH Terms] AND "medline"[MeSH Terms] AND "randomized controlled trials as topic"[MeSH Terms]

"abstracting and indexing as topic"[MeSH Terms] AND "medline"[MeSH Terms] AND "randomized controlled trials as topic"[MeSH Terms]

Findings

A number of strategies for searching MEDLINE for RCTs have been proposed. Search strategies are required for retrieval of RCTs due to several factors, including indexers' view of what constitutes an RCT and the short time spent on indexing.[29]

Efforts have been made to clarify published RCT abstract structure and items, most prominently through the Consolidated Standards of Reporting Trials (CONSORT) Statement for Abstracts guidelines.[30] A clear, structured abstract may help indexers more efficiently and accurately assess whether a published study is an RCT. Journals and papers that strictly adhere to the CONSORT guidelines can also make MEDLINE search simpler because terminology is standardized, and can therefore be integrated into the search strategy.

Narrowing search results to only RCTs is beneficial in systematic reviews[31] and much of the literature on MEDLINE searching for RCTs comes from this field.

Annotated List of References

Glanville JM, Lefebvre C, Miles JN, Camosso-Stefinovic J. How to identify randomized controlled trials in MEDLINE: Ten years on. *J Med Libr Assoc*. 2006;94(2):130-6. Erratum in: *J Med Libr Assoc*. 2006;94(3):354.

[29] Wieland LS, Robinson KA, Dickersin K. Understanding why evidence from randomised clinical trials may not be retrieved from Medline: Comparison of indexed and non-indexed records. *BMJ*. 2012;344:d7501. doi: 10.1136/bmj.d7501.

[30] Ghimire S, Kyung E, Kang W, Kim E. Assessment of adherence to the CONSORT statement for quality of reports on randomized controlled trial abstracts from four high-impact general medical journals. *Trials*. 2012 Jun 7;13:77. doi: 10.1186/1745-6215-13-77.

[31] Zhang L, Ajiferuke I, Sampson M. Optimizing search strategies to identify randomized controlled trials in MEDLINE. *BMC Med Res Methodol*. 2006;6(23).

Authors compared MEDLINE citations indexed with the Randomized Controlled Trial Publication Type term from years where a Cochrane-NLM RCT retagging collaboration took place and years where the collaboration had not yet taken place. They found the publication type search on retagged citations was sensitive and precise, and recommended this simple search strategy for busy researchers. Publication type searching on citations that had not been retagged needed a modified strategy for good sensitivity and precision. The authors also state that researchers requiring high sensitivity, such as systematic reviewers and those conducting meta-analyses, should add additional search terms to the publication type search strategy. They recommend a modified Cochrane Highly Sensitive Search Strategy for this.

McKibbon KA, Wilczynski NL, Haynes RB; Hedges Team. Retrieving randomized controlled trials from Medline: A comparison of 38 published search filters. *Health Info Libr J*. 2009 Sep;26(3):187-202. doi: 10.1111/j.1471-1842.2008.00827.x.

The authors examined 38 search filters used for identifying RCTs in MEDLINE/PubMed. They found these search filters ranged greatly in sensitivity, specificity, and precision. The authors also note that when searches are fine-tuned for one of these measures, it is usually at the sacrifice of other measures. Therefore, creating a search filter requires balancing these measures based on the unique requirements of the search.

Wieland LS, Robinson KA, Dickersin K. Understanding why evidence from randomised clinical trials may not be retrieved from Medline: Comparison of indexed and non-indexed records. *BMJ*. 2012;344:d7501. doi: 10.1136/bmj.d7501.

Authors found a large number of RCT citations not indexed using the Randomized Controlled Trial Publication Type in MEDLINE. They suggest that research looking for RCTs in MEDLINE not rely solely on publication type searching.