Clinical Epidemiology

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Titles versus titles and abstracts for initial screening of articles for systematic reviews

This article was published in the following Dove Press journal:
Clinical Epidemiology
14 March 2013
Number of times this article has been viewed

Farrah J Mateen[1,2]
Jiwon Oh[1,2]
Ana I Tergas[1,3]
Neil H Bhayani[1,4]
Biren B Kamdar[1,5]

[1]Bloomberg School of Public Health, [2]Department of Neurology, [3]Division of Gynecologic Oncology, Johns Hopkins Hospital, Baltimore, MD, USA; [4]Department of Surgery, Howard University Hospital, Washington, DC, USA; [5]Division of Pulmonary and Critical Care Medicine, Johns Hopkins University, Baltimore, MD, USA

Background: There is no consensus on whether screening titles alone or titles and abstracts together is the preferable strategy for inclusion of articles in a systematic review.

Methods: Two methods of screening articles for inclusion in a systematic review were compared: titles first versus titles and abstracts simultaneously. Each citation found in MEDLINE or Embase was reviewed by two physician reviewers for prespecified criteria: the citation included (1) primary data; (2) the exposure of interest; and (3) the outcome of interest.

Results: There were 2965 unique citations. The titles first strategy resulted in an immediate rejection of 2558 (86%) of the records after reading the title alone, requiring review of 239 titles and abstracts, and subsequently 176 full text articles. The simultaneous titles and abstracts review led to rejection of 2782 citations (94%) and review of 183 full text articles. Interreviewer agreement to include an article for full text review using the titles-first screening strategy was 89%–94% (kappa = 0.54) and 96%–97% (kappa = 0.56) for titles and abstracts combined. The final systematic review included 13 articles, all of which were identified by both screening strategies (yield 100%, burden 114%). Precision was higher in the titles and abstracts method (7.1% versus 3.2%) but recall was the same (100% versus 100%), leading to a higher F-measure for the titles and abstracts approach (0.1327 versus 0.0619).

Conclusion: Screening via a titles-first approach may be more efficient than screening titles and abstracts together.

Keywords: meta-analysis, research methods, epidemiology, systematic review

Introduction

Systematic reviews, which summarize all of the available evidence on a topic, are increasingly necessary for clinical and health policy decision making. Information presented in a systematic review can come from observational studies (eg, questions of incidence and prevalence of a condition, etiology of a disease) and/or clinical trials (eg, questions of effectiveness and safety of an intervention). Recently, the Institute of Medicine published standards for performing systematic reviews for comparative effectiveness research;[1] although these standards mainly focus on systematic reviews addressing questions of intervention effectiveness, many of the items are applicable across all topics.

Because systematic reviews are informative for policy making, many groups wish to or need to conduct them, for example before deciding to undertake a new study or when developing clinical practice guidelines. However, systematic reviews require substantial resources in excess of what the investigative team may be able to commit. An unpublished 2005 report from the United Kingdom estimates the cost of

Correspondence: Farrah J Mateen
Department of International Health,
Bloomberg School of Public Health,
The Johns Hopkins University,
Room E5547, 615
N Wolfe Street, Baltimore,
MD 21205, USA
Tel +1 410 935 5181
Email fmateen@jhsph.edu

http://dx.doi.org/10.2147/CLEP.S43118

a systematic review to range from £17,000 to £80,000 with a time commitment between 6 and 18 months (Mugford et al, unpublished data, 2005). Cutting corners to minimize these financial and time investments may lead to lower quality systematic reviews and meta analyses that fail to identify all relevant studies.[2] In contrast, there is increasing recognition of unfinished and unpublished projects that could be beneficial if finalized.[2] Thus, a method to reduce the time and cost of a systematic review, without compromising its quality, would be helpful to those performing reviews and help avoid "wastage" in medical research.[2]

Among the steps in performing a systematic review, screening and selection of citations for inclusion in the review accounts for a large proportion of the time investment, making it an obvious target for time reduction strategies. While there are currently no data regarding the most effective strategy for screening citations before full text review,[3,4] the Institute of Medicine recommends a simultaneous title and abstract screening approach.[1]

Our objective was to compare two methods of performing an initial screening of citations obtained from searches of commonly used medical bibliographic databases. First, we performed a two-stage method whereby titles alone were screened, followed by screening of titles and abstracts of those not rejected by title alone ("titles first"). Second, we performed the traditional screening method of examining the title and abstract together ("titles and abstracts"). Our overall goal was to assess the numbers of citations reviewed at each step and determine whether each strategy ultimately yielded all relevant full text articles for the systematic review.

Methods

A five-member team of physicians performed a systematic review examining the association between breast cancer risk and night shift work exposure ("light at night").[5] Physician disciplines included pulmonary medicine, general surgery, neurology, and obstetrics and gynecology. Each reviewer had completed at least 1 year of graduate coursework in biostatistics and epidemiology. No member received payment or other incentives to participate in the systematic review. All screening occurred over a 4-week period as part of a graduate course on systematic review methods.

The review team worked with a librarian specialized in systematic reviews to construct search strategies for MEDLINE and Embase (see Supplementary material for search strategy). To accomplish screening of all retrieved citations by two independent reviewers, team members (denoted A through E) were assigned into five reviewer pairs (ie, AB, BC, CD, DE, and AE) with each reviewer assigned to two partners. The citation list was divided equally among the five reviewer pairs. The screening process assessed whether the citation: (1) presented data from an original research study; (2) focused on the exposure of interest, "light at night;" and (3) captured the outcome of interest, "breast carcinoma."

To determine whether the full text should be retrieved for a given citation, the two independent reviewers marked each citation using a "yes," "no," or "unknown" (unsure whether yes or no) designation. We first performed the screening using titles first then titles with abstracts and then redid the entire process using the traditional screening method (screening titles and abstracts simultaneously). For both methods, the reviewers evaluated the same set of citations.

When both reviewers marked "yes" for either screening approach, the citation was forwarded for further review. When both reviewers marked "no," the record was discarded from further review (Figure 1). Citations that received at least one "yes" or "unknown" were carried forward to the next round of review. Reviewers were unaware of their partners' decisions until after the screening process was completed. Full-text articles were not reviewed until completion of both screening processes.

Reviewer classifications ("yes," "unknown," or "no") using each of the two screening approaches were entered using drop-down menus in the Microsoft Office Excel® spreadsheet (Microsoft Corporation, Redmond, WA, USA). Full-text articles were then reviewed for potential eligibility. The total number of records screened, number classified as "no" "no" (discard), number retained for the next step, number of full-text articles screened, and number of articles ultimately eligible for the systematic review and meta-analysis were recorded at each step (Figure 1). Agreement was defined as the proportion of the total number of citations classified identically, "yes" "yes" or "no" "no," by the reviewer pair. Interreviewer agreement for each method was assessed using a kappa statistic.

The sensitivity of the two screening approaches was measured as the proportion of articles eligible for the review correctly identified by each screening method. The titles first versus titles and abstracts methods were also compared using previously reported metrics used in the evaluation of

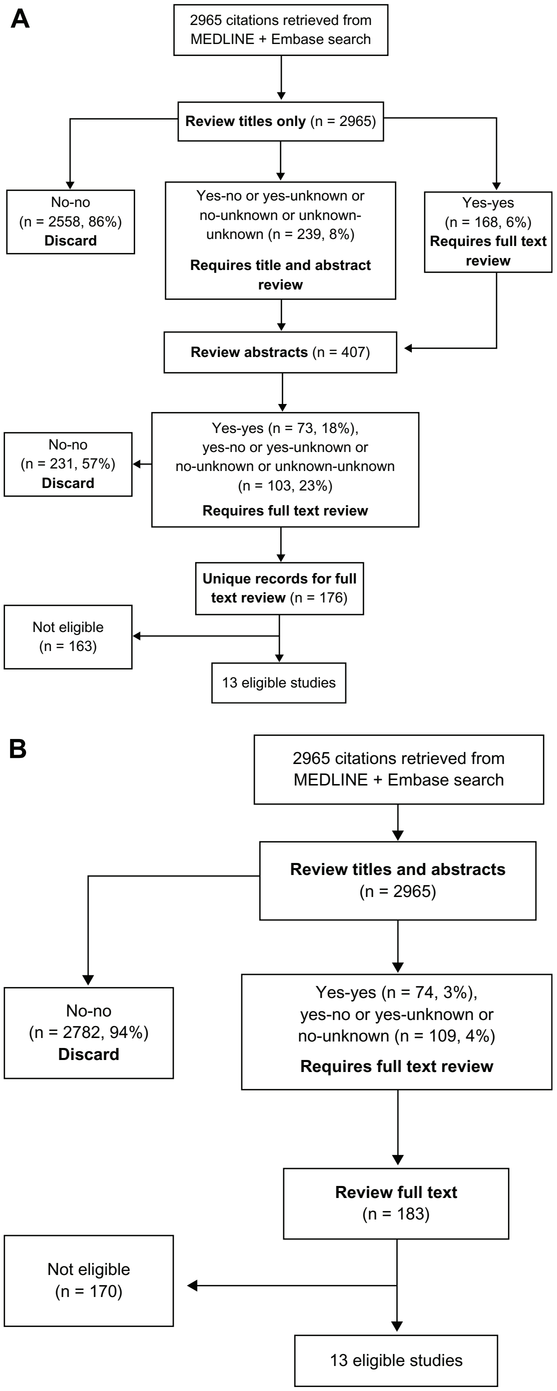
A
2965 citations retrieved from MEDLINE + Embase search
Review titles only (n = 2965)
No-no (n = 2558, 86%) Discard
Yes-no or yes-unknown or no-unknown or unknown-unknown (n = 239, 8%)
Requires title and abstract review
Yes-yes (n = 168, 6%) Requires full text review
Review abstracts (n = 407)
No-no (n = 231, 57%) Discard
Yes-yes (n = 73, 18%), yes-no or yes-unknown or no-unknown or unknown-unknown (n = 103, 23%)
Requires full text review
Unique records for full text review (n = 176)
Not eligible (n = 163)
13 eligible studies
B
2965 citations retrieved from MEDLINE + Embase search
Review titles and abstracts (n = 2965)
No-no (n = 2782, 94%) Discard
Yes-yes (n = 74, 3%), yes-no or yes-unknown or no-unknown (n = 109, 4%)
Requires full text review
Review full text (n = 183)
Not eligible (n = 170)
13 eligible studies

Figure 1 Flow diagram of article selection and review using a titles first approach (**A**) and titles abstract screening process (**B**).

strategies for performing systematic reviews (Table 1, 2).[6,7] Precision is defined here as:[6]

$$\text{Precision} = \frac{\text{Number of full-text documents correctly classified}}{\text{Total number of citations classified as “yes”}}. \tag{1}$$

Recall is defined as:

$$\text{Recall} = \frac{\text{Number of full-text documents correctly classified}}{\text{Total number of full-text documents in final collection}}. \tag{2}$$

Precision and recall are combined in a single metric, the F-measure, which is a weighted mean of precision and recall, given as:

$$\text{F-measure} = \frac{2\ \times \text{Precision} \times\ \text{Recal}}{(\text{Precision} + \text{Recall})} \tag{3}$$

Yield is defined as the fraction of citations that are included in the final systematic review using a given screening approach.[7] Burden is a measure of the total number of citations that a person has to review given a screening approach and should be minimized as much as possible.[7] The 2 × 2 table is generated as shown in Tables 1 and 2.

$$\text{Yield} = \frac{\text{tp}^{\text{TA}} + \text{tp}^{\text{T}}}{\text{tp}^{\text{TA}} + \text{tp}^{\text{T}} + \text{fn}^{\text{T}}} \tag{4}$$

and

$$\text{Burden} = \frac{\text{tp}^{\text{TA}} + \text{tn}^{\text{TA}} + \text{fp}^{\text{TA}} + \text{tp}^{\text{T}} + \text{fp}^{\text{T}}}{\text{N}}, \tag{5}$$

where “N” is the total number of citations. All counts and calculations were performed using Microsoft Excel.

Results

The systematic review search resulted in 2965 citations after removal of duplicates. Each of the five reviewer pairs reviewed approximately 20% of the 2965 citations (average: 593, range: 584–614 articles per pair). Each reviewer belonged to two separate pairs and reviewed approximately 1186 citations (593 × 2 pairs). The titles first screening strategy resulted in an immediate rejection of 2558 (86%) of the retrieved records after reading the title alone, with the need to review 239 abstracts, and subsequently 176 full text papers (Figure 1A). In

Table 1 Reviewer provides the labels based on titles and abstracts review

	Reviewer reported “yes” for inclusion	
	Yes	**No**
Ultimately included in systematic review?		
Yes	tp^{TA}	0
No	fp^{TA}	tn^{TA}

Abbreviations: TA, titles and abstracts together; tp, true positive; fp, false positive; tn, true negative.

Table 2 Reviewer provides the labels based on titles-only review

	Reviewer reported "yes" for inclusion	
	Yes	**No**
Ultimately included in systematic review?		
Yes	tp^T	fn^T
No	fp^T	tn^T

contrast, the simultaneous titles and abstracts review resulted in rejection of 2782 citations (94%) after reading both title and abstract, and the need to review 183 full text articles (Figure 1B). Both methods led to the same 13 articles being selected for the final systematic review and meta-analysis.

Using the titles and abstracts as the gold standard, the sensitivity of the titles-first search strategy was 96.2% (95% confidence interval [CI]: 92.3–98.4) and the specificity was 91.7% (95% CI: 90.6–92.7). Interreviewer agreement between reviewer pairs with the titles-first screening strategy was 91.9% (range = 87.8% to 93.8%, κ = 0.54 [95% CI: 0.49–0.59]) (Table 3). Interreviewer agreement with a titles and abstracts screening strategy was 96.3% between reviewer pairs (range = 95.4%– 97.4%, κ = 0.56 [95% CI: 0.48–0.63]). The medical specialty training of the reviewers in relationship to the interreviewer agreement results are also provided in Table 3.

Precision, recall, and F-measure were based on Tables 4–6. The yield of the titles-first method compared to the titles and abstracts method was 100%. The burden was 114%.

Discussion

This pilot study demonstrated the feasibility of a titles-first then abstracts screening strategy. The two methods identified the same set of articles for the final systematic review. The titles first approach had the advantage of immediately discarding 86% of all citations, reducing the time required to read abstracts for citations irrelevant to the study question. The important difference between the two screening methods was reading fewer abstracts in the titles-first method. The titles-first step had a lower agreement rate between reviewer pairs than the simultaneous title and abstract approach and a lower precision, suggesting that time saved in titles-only review may be expended during the resolution of interreviewer disagreements. Previously described metrics support the commonly held concept that titles and abstracts screening is more precise (7.1% versus 3.2%), but we show here that recall is the same. This essentially means that the final list of citations for the systematic review did not differ between the methods. The F-measure for the titles and abstracts method is higher due solely to higher precision that occurs when including abstracts. Similarly, the yield of the titles-first strategy was 100%.

Table 3 Interreviewer agreement on the titles-first screening step versus the title and abstract screening step, by reviewer pair

Reviewer pair	**1**	**2**	**3**	**4**	**5**	**Overall**
	Neurologist A and pulmonologist	**Pulmonologist and neurologist B**	**Neurologist B and gynecologist/obstetrician**	**Gynecologist/obstetrician and general surgeon**	**General surgeon and neurologist A**	
Citations reviewed (n)	586	584	614	588	593	2965
% agreement on titles-alone classification (95% CI)	90.6 (88.2–93.0)	87.8 (85.2–90.5)	93.8 (91.9–95.7)	93.5 (91.5–95.5)	93.7 (91.8–95.7)	91.9 (91.0–92.9)
Kappa statistic on titles-alone classification (95% CI)	0.51 (0.40–0.62)	0.40 (0.30–0.51)	0.63 (0.52–0.74)	0.67 (0.57–0.77)	0.53 (0.39–0.66)	0.54 (0.49–0.59)
% agreement on titles and abstracts classification (95% CI)	96.2 (94.7–97.8)	95.3 (93.7–97.1)	95.2 (94.4–97.5)	97.4 (96.2–98.7)	96.6 (95.2–98.1)	96.3 (95.6–97.0)
Kappa statistic on titles and abstracts classification (95% CI)	0.56 (0.39–0.72)	0.43 (0.25–0.60)	0.56 (0.40–0.71)	0.72 (0.59–0.86)	0.49 (0.29–0.68)	0.56 (0.49–0.63)

Note: Percent agreement = (number of citations with a consensus of a "yes" "yes" or "no" "no" response in a reviewer pair)/total number citations screened.
Abbreviation: CI, confidence interval.

Table 4 Titles-first classified as relevant

	n	n	n
Yes	13	0	13
No	394	2558	2952
Total	407	2558	2965

Notes: Precision = 3.2%; recall = 100%; F-measure = 0.0619; n = number in sample.

It is unclear whether the fairly low interreviewer agreement is particular to the present study or a general feature common to systematic review teams. This group of reviewers had several similarities including enrollment in the same graduate course on systematic reviews. None had formal systematic review experience prior to the class. All reviewers had simultaneous instruction on systematic review methodology via the same in person lectures from experts in the field of systematic review methodology. Notably, the review group was comprised of five practicing specialist physicians of four different specialty backgrounds. Most specialties (gynecology, general surgery, sleep pulmonology) were directly relevant to the disease outcome of interest (breast cancer) or exposure of interest (light at night). The inconsistent interrater agreements, all relatively low, may be an area worthy of further study since systematic reviews are performed by teams with varying levels of experience. It is possible that reviewers employ their own methods or modifications of citation classification even in the setting of formal instruction.

Our small study can be improved by future researchers. We performed the titles-first screening before the simultaneous titles and abstracts screening on the same set of 2965 articles. The same reviewer pairs were assigned to the same articles. It is possible that this ordering, and seeing the titles and abstracts a second time, led to a selection bias and unequal comparison. The simultaneous titles and abstracts method may have been more accurate simply because it was the second method. However, given that each individual reviewed nearly 1200 citations, it is unlikely that reviewers were able to recall past decisions. Timing of the process of review was not performed in either method. Recording time expenditure by the reviewers during each step would strengthen our assertion that titles-first review is more efficient.

Table 5 Titles and abstracts classified as relevant

	n	n	n
Yes	13	0	13
No	170	2782	2952
Total	183	2782	2965

Notes: Precision = 7.1%; recall = 100%; F-measure = 0.1327; n = number in sample.

Table 6 Overall data for both titles-only and titles and abstracts classified as relevant

	n	n	n
Yes	26	0	26
No	564	5340	5904
Total	590	5340	5930

Note: n = number in sample.

Conclusion

If future studies confirm our findings, it may be reasonable to use and even recommend a titles-first screening strategy in lieu of the standard titles and abstracts strategy. Since little to no guidance is currently available on strategies for reviewing citation lists, the benefits of an accurate, less time-consuming process that does not compromise the quality of the final review are notable. Abstracts themselves are imperfect and may reflect bias, spin, and nondisclosure of negative findings on primary study endpoints.[8]

Although our study question evaluated observational epidemiology research, it is likely that systematic reviews and meta-analyses for clinical trials could also employ a titles-first method. This is likely superior to proposed automated techniques that are highly time-efficient but remove human participation and topical expertise from the initial screening altogether.[9] Whether a titles-first methodology is as successful for different types of study exposures and outcomes, including treatment benefit, adverse events, and other endpoints of interest, is yet to be determined. Meanwhile, the burgeoning number of clinical trials[10] makes it especially important to synthesize medical information in a timely and accurate manner.

Acknowledgments

The authors thank Dr Kay Dickersin, MA, PhD, at the Center for Clinical Trials, Department of Epidemiology, and Dr Satyanarayana Vedula MBBS, PhD, from the Department of Epidemiology at the Bloomberg School of Public Health, the Johns Hopkins University for their insightful comments, which were integral to the development of an earlier version of this manuscript. Publication of this article was funded in part by the Open Access Promotion Fund of the Johns Hopkins University Libraries. This work is supported by a fellowship grant of the Canadian Institute of Health Research to Dr Farrah Mateen and a Ruth L. Kirschstein NRSA award from the National Institutes of Health to Dr Biren Kamdar (grant # F32 HL104901).

Disclosure

The authors report no conflicts of interest in this work.

References

1. Eden J, Levit L, Berg A, Morton S, editors. *Finding What Works in Health Care: Standards For Systematic Reviews*. Washington, DC: The National Academies Press; 2011.
2. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet*. 2009;374(9683):86–89.
3. Barker FH, Veal DC, Wyatt BK. Comparative efficiency of searching titles, abstracts, and index terms in a free-text data base. *J Doc*. 1972;28(1):22–36.
4. *Cochrane Style Guide*. *4.1 Edition*. The Cochrane Collaboration; 2010. Available from: http://www.cochrane.org/sites/default/files/uploads/Cochrane-Style-Guide_4-1-edition.pdf. Accessed Feb 23, 2013.
5. Kamdar BB, Tergas AI, Mateen FJ, Bhayani NH, Oh J. Night-shift work and risk of breast cancer: a systematic review and meta-analysis. *Breast Cancer Re Treat*. 2013 Feb 12 [Epub ahead of print].
6. Cohen AM, Hersh WR, Peterson K, Yen P. Reducing workload in systematic review preparation using automated citation classification. *J Am Med Inform Assoc*. 2006;13(2):206–219.
7. Wallace BC, Trikalinos TA, Lau J, Broadley C, Schmid CH. Semi-automated screening of biomedical citations for systematic reviews. *BMC Bioinformatics*. 2010;11:55.
8. Vera-Badillo FE, Shapiro R, Ocana A, Amir E, Tannock IF. Bias in reporting of end points of efficacy and toxicity in randomized, clinical trials for women with breast cancer. *Ann Oncol*. Epub January 9, 2013.
9. Sampson M, Barrowman NJ, Moher D, et al. Can electronic search engines optimize screening of search results in systematic reviews: an empirical study. *BMC Med Res Method*. 2006;6:7.
10. Dickersin K, Rennie D. The evolution of trial registries and their use to assess the clinical trial enterprise. *JAMA*. 2012;307(17):1861–1864.

Appendix

Search in Medline (PubMed)

"Nurses"[MeSH Terms] OR "Employment"[MeSH Terms] OR airline* crew*[tw] OR cabin*[tw] OR attendant*[tw] OR crew*[tw] OR flight*[tw] OR personnel*[tw] OR night*[tw] OR work[tw] OR nightshift*[tw] OR shift[tw] OR stewardess*[tw]) AND ("Risk Factors"[MeSH Terms] OR "Occupational Diseases"[MeSH Terms] OR "Life Style"[MeSH Terms] OR "Occupational Exposure"[MeSH Terms] OR "Work Schedule Tolerance"[MeSH Terms] OR "Circadian Rhythm"[MeSH Terms] OR "Time Factors"[MeSH Terms] OR "Melatonin"[MeSH Terms] OR "Carcinogens"[MeSH Terms] OR "Light"[MeSH Terms] OR "Lighting"[MeSH Terms] OR "Personnel Staffing and Scheduling"[MeSH Terms] OR "Sleep"[MeSH Terms] OR "Workplace"[MeSH Terms]) AND("Breast Neoplasms"[MeSH Terms] OR "breast cancer"[tw]).

Search in Embase

'Nurse'/exp OR 'employment'/exp OR airline* AND crew* OR cabin* OR attendant* OR 'airplane crew'/exp OR crew* OR flight* OR personnel* OR night* OR 'shift worker'/exp OR nightshift* OR 'shift' OR 'work'/exp OR stewardess* AND ('cancer incidence'/exp OR 'cancer epidemiology'/exp OR 'cancer risk'/exp OR 'risk factor'/exp OR 'occupational disease'/exp OR 'life style'/exp OR 'occupational exposure'/exp OR 'work schedule'/exp OR 'circadian rhythm'/exp OR 'time'/exp OR 'melatonin'/exp OR 'carcinogen'/exp OR 'light'/exp OR 'illumination'/exp OR 'personnel management'/exp OR 'sleep'/exp OR 'workplace'/exp) AND ('breast tumor'/exp OR 'breast cancer'/exp).