Usefulness of Systematic Review Search Strategies in Finding Child Health Systematic Reviews in MEDLINE

Nicole Boluyt, MD, PhD; Lisa Tjosvold, BA, MLIS; Carol Lefebvre, MSc; Terry P. Klassen, MD, MSc, FRCP; Martin Offringa, MD, PhD

Objective: To determine the sensitivity and precision of existing search strategies for retrieving child health systematic reviews in MEDLINE using PubMed.

Design: Filter (diagnostic) accuracy study. We identified existing search strategies for systematic reviews, combined them with a filter that identifies articles relevant to child health, and applied the combination in MEDLINE to a reference set of child health systematic reviews.

Main Outcome Measures: Total number of records retrieved, sensitivity, and precision.

Results: We tested 9 search filters. Sensitivity of the systematic review filters combined with the child filter ranged from 68% to 96%; sensitivity of the child filter alone was 98%. The number of records retrieved with PubMed (limited to January 1990-January 2006) by the systematic review filters combined with the child filter ranged from 7861 to 618,053. Precision for the combined filters ranged from 2% to 52%. Because of poor reporting of specific systematic review criteria in both titles and abstracts, in 25% of the records screened we were unsure whether the article concerned a systematic review according to our definition.

Conclusions: The high numbers of records yielded by sensitive search strategies and the low precision threaten the use of systematic reviews for clinical decision making and guideline development. Reporting of specific systematic review criteria in titles and abstracts is poor, and reporting recommendations given by Quality of Reporting of Meta-analyses (QUOROM) should be used more strictly. To make identification using MEDLINE easier, there is an urgent need to set minimal criteria that any review should fulfill for it to be indexed as a systematic review.


Well-conducted systematic reviews provide the best evidence to guide clinical practice, are cornerstones for the recommendations of evidence-based practice guidelines, and should be an integral part of the planning of future research. In contrast to traditional narrative reviews, systematic reviews of the literature address a well-defined question, use an explicit search strategy to locate all relevant evidence, evaluate the retrieved studies using prospectively defined methodological criteria, and formally synthesize the results.

Clinicians and researchers should be able to reliably and quickly find systematic reviews. Two types of systematic reviews can be identified in current bibliographic databases: (1) the Cochrane systematic reviews, which can be found in the Cochrane Database of Systematic Reviews, the Cochrane Library and (2) systematic reviews published in journals (journal reviews). Journal reviews can also be found in the Cochrane Library Database of Abstracts of Reviews of Effects (DARE) and are available from the Centre for Reviews and Dissemination Web site. The DARE database provides quality assessments of systematic reviews published in journals after 1994, identified by regular searches of important bibliographic databases such as MEDLINE, EMBASE, and CINAHL and by hand searching some of the major general journals but no specific pediatric journals. Bibliographic databases such as MEDLINE index Cochrane reviews and can be used to identify other systematic reviews, but only those indexed in the respective database. Finding systematic reviews in MEDLINE poses 2 challenges. First, only a fraction of all citations in MEDLINE are for systematic reviews and, second, the MEDLINE indexing procedures do not include “systematic review” as a “publication type.” To limit the search results from a query in MEDLINE, therefore, it is recommended that a method-
EXISTING SEARCH STRATEGIES FOR SYSTEMATIC REVIEWS AND CHILD STUDIES

To identify articles reporting on the development and validation of systematic review search filters in MEDLINE, we searched MEDLINE from January 1995 to January 2006 with the following MeSH terms: MEDLINE, Information Storage and Retrieval/Methods, and Review, literature. In addition, reference lists of relevant articles were reviewed and content experts were contacted to find additional studies. To improve precision, we combined the systematic review filters with a sensitive child filter developed by the Cochrane Child Health Field\textsuperscript{10} to retrieve only studies in children.

REFERENCE STANDARD

To measure the sensitivity of the search strategies, a reference standard set of systematic reviews was established by searching for child health systematic reviews in DARE\textsuperscript{5} and by hand searching several pediatric journals for systematic reviews. All titles and abstracts in DARE (Cochrane Library, Issue 2, 2004) were screened to find child health systematic reviews also indexed in MEDLINE. We hand searched 7 pediatric journals indexed in MEDLINE with a range of impact factors and for which full-text electronic copies were available from our medical library (Table 1). All issues of each journal were searched for the following 5 years: 1994, 1997, 2000, 2002, and 2004. In addition, we were interested as to whether DARE contained all child health systematic reviews found by our hand search of pediatric journals. Inasmuch as DARE includes only systematic reviews on prevention, intervention, or diagnosis, we selected systematic reviews covering these domains. Next, we searched DARE in July 2006 using the Cochrane Library interface to determine how many child health systematic reviews on prevention, intervention, or diagnosis found by our hand search were also included in DARE.

SENSITIVITY

Because the child filter has not been validated, we measured the sensitivity of the child filter separately. The sensitivity of a search strategy was defined as the percentage of child health systematic reviews retrieved from our reference standard set of child health systematic reviews.

\[
\text{Sensitivity} = \frac{\text{Number of systematic reviews retrieved from reference standard}}{\text{Total number of systematic reviews in reference standard}} \times 100
\]

Sensitivity was measured for all systematic review filters separately and in combination with the child filter.

PRECISION

The precision of a combined search strategy was defined as the percentage of true child health systematic reviews retrieved in MEDLINE (limits 1990-2006) divided by the total number of records retrieved by the search.

\[
\text{Precision} = \frac{\text{Number of true systematic reviews retrieved}}{\text{All records in MEDLINE retrieved by the search}} \times 100
\]

We estimated precision rather than calculated precision because the yield of each combined search strategy was large, ranging from 7861 records (Montori 2 search filter) to 618,053 records (Montori 1 search filter). To estimate precision, we selected a random sample of 100 records from the yield of each combined search strategy. The random sample was generated using R software\textsuperscript{11} and stratified by year of publication to account for

### Table 1. Reference Standard: Number of Systematic Reviews Found by Hand Searching 7 Pediatric Journals With a Range of Impact Factors for 1994, 1997, 2000, 2002, and 2004 and Number of Child Health Systematic Reviews Included in DARE and Indexed in MEDLINE

<table>
<thead>
<tr>
<th>Journal</th>
<th>Impact Factor</th>
<th>No. of Systematic Reviews or Meta-analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pediatrics</td>
<td>3.4</td>
<td>36</td>
</tr>
<tr>
<td>Journal of Pediatrics</td>
<td>3.2</td>
<td>6</td>
</tr>
<tr>
<td>Pediatric Infectious Diseases Journal</td>
<td>2.4</td>
<td>10</td>
</tr>
<tr>
<td>Archives of Diseases in Childhood</td>
<td>2.1</td>
<td>34</td>
</tr>
<tr>
<td>Archives of Pediatric and Adolescent Medicine</td>
<td>2.1</td>
<td>20</td>
</tr>
<tr>
<td>European Journal of Pediatrics</td>
<td>1.2</td>
<td>5</td>
</tr>
<tr>
<td>Journal of Perinatal Medicine</td>
<td>0.9</td>
<td>4</td>
</tr>
<tr>
<td><strong>Total found by hand searching</strong></td>
<td></td>
<td><strong>115</strong></td>
</tr>
<tr>
<td>DARE\textsuperscript{a}</td>
<td></td>
<td><strong>298</strong></td>
</tr>
<tr>
<td>Overlap\textsuperscript{b}</td>
<td></td>
<td>26</td>
</tr>
<tr>
<td><strong>Total Reference Standard</strong></td>
<td></td>
<td><strong>387</strong></td>
</tr>
</tbody>
</table>

Abbreviation: DARE, Database of Abstracts of Reviews of Effects.\textsuperscript{a} Cochrane Library, Issue 2, 2004; reviews also indexed in MEDLINE.\textsuperscript{b} Records found in both DARE up to Issue 2, 2004, and 1 of the journals hand searched. In 2006, 52% of the journal reviews on prevention, intervention, or diagnosis were also indexed in DARE.

### METHODS

#### INCLUSION AND EXCLUSION CRITERIA

Herein, we use the term systematic review to refer to any literature review, meta-analysis, or other article that explicitly indicates the use of a strategy for locating evidence by mentioning at least the databases that were searched and reviewing the empirical evidence on children (aged <18 years). Excluded were systematic reviews that assessed the effects on the fetus of interventions in pregnant women.
changes in MeSH indexing with time. Each sample was screened independently by 2 investigators (N.B. and L.T.) for true systematic reviews. We considered an article a true systematic review only if the title, abstract, author-supplied key words, or publication type terms explicitly identified the article as a systematic review or meta-analysis or if the article abstract indicated a strategy for locating the literature reviewed and included children. We classified records as unsure if the title or abstract did not specifically state whether children were included, it was stated that a literature search was performed but without stating which databases were searched, or it was stated that the article was an evidence-based overview. In all of these unsure cases, the full text of the article was retrieved and a final decision was made as to whether it was a true systematic review. Next, we applied the search strategies to 7 priority topics in child health, developed by the Cochrane Child Health Field, that necessitate systematic reviews. The number of records retrieved for the most sensitive and the most precise systematic review search strategies is reported, as is the precision for the most precise filter. The search strings for each clinical topic were created by using the MeSH browser function. We reported the time it takes a pediatrician (N.B.) and a clinical librarian (L.T.) to screen 100 MEDLINE records.

## RESULTS

### EXISTING SEARCH STRATEGIES FOR IDENTIFYING SYSTEMATIC REVIEWS

We identified 4 studies reporting on the development and validation of several systematic review search strategies for MEDLINE. We evaluated 8 of the reported search strategies with varying sensitivity and precision and the PubMed Clinical Queries systematic review filter. Details of the systematic review search strategies (translated into PubMed format where necessary) and the child health search strategy are provided in the online supplement (eBox at http://www.archpediatrics.com).

### REFERENCE STANDARD

Our reference standard contained 387 child health systematic reviews indexed in MEDLINE (Table 1). A total of 431 child health systematic reviews were identified in DARE (Cochrane Library, Issue 2, 2004; N=4645), of which only 298 were indexed in MEDLINE. Hand searching MEDLINE-indexed pediatric journals yielded 115 child health systematic reviews, of which 26 records were also included in DARE (Cochrane Library, up to Issue 2, 2004). Of our hand-searched 115 reviews, 73 reported on prevention, intervention, or diagnosis. A 2006 search of DARE revealed that 38 of these 73 systematic reviews on prevention, intervention, or diagnosis (52%) were also indexed in DARE.

### SENSITIVITY

Sensitivity of the child filter when run separately in MEDLINE and tested against our reference standard was 98% (380/387; 95% confidence interval, 96%-99%). Sensitivity of the systematic review filters ranged from 68% to 96% (Table 2).

### PRECISION

Precision Not Limited to Clinical Topics

Figure 1 shows the number of records retrieved with the systematic review filters in PubMed (1990-2006) with and without the child filter added. The number of records retrieved with the systematic review filters ranged from more than 38,000 (Montori 2) to more than 3 million (Montori 1) and for the combined filters from more than 8000 (Montori 1) to more than 618,000 (Montori 2).

We screened 600 titles and abstracts for systematic reviews (a random sample of 100 records retrieved with 6 systematic review filters combined with the child filter) (Table 2). We were unsure in 150 records whether the article was a true systematic review and in 15 of those whether children were included in the review. For all of these unsure results, full text was retrieved, which eventually led to a yield of 137 of the 600 reviews being true systematic reviews in children. Precision ranged from 2% to 52% in our study with the combined systematic re-

---

Table 2. Reported Performance and Performance in Our Reference Standard of Combined Systematic Review Filter With Child Filter

<table>
<thead>
<tr>
<th>Search Filter</th>
<th>Reported Performance</th>
<th>Performance in Our Reference Standard</th>
<th>Performance in PubMed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity, %</td>
<td>Precision, %</td>
<td>Sensitivity, % (95% CI)</td>
</tr>
<tr>
<td>Shojania plus child</td>
<td>93</td>
<td>50&lt;sup&gt;c&lt;/sup&gt;</td>
<td>74 (69-78)</td>
</tr>
<tr>
<td>Boynton plus child</td>
<td>98</td>
<td>20</td>
<td>95 (92-97)</td>
</tr>
<tr>
<td>White 1 plus child</td>
<td>100</td>
<td>4</td>
<td>93 (91-96)</td>
</tr>
<tr>
<td>White 2 plus child</td>
<td>94</td>
<td>10</td>
<td>94 (91-96)</td>
</tr>
<tr>
<td>Montori 1 plus child</td>
<td>100</td>
<td>3</td>
<td>96 (93-97)</td>
</tr>
<tr>
<td>Montori 2 plus child</td>
<td>71</td>
<td>57</td>
<td>68 (64-73)</td>
</tr>
<tr>
<td>Montori 3 plus child</td>
<td>98</td>
<td>14</td>
<td>94 (91-96)</td>
</tr>
<tr>
<td>Montori 4 plus child</td>
<td>74</td>
<td>69</td>
<td>72 (67-76)</td>
</tr>
<tr>
<td>PubMed plus child</td>
<td>Unknown</td>
<td>Unknown</td>
<td>76 (72-80)</td>
</tr>
</tbody>
</table>

Abbreviation: CI, confidence interval.

<sup>a</sup>Reported performance without child filter. Sensitivity of child filter, 98% (95% confidence interval, 96-99). All numbers have been rounded.

<sup>b</sup>Precision not estimated for White 1, Montori 1, and Montori 4 filters because other filters with similar sensitivity have lower numbers of hits.

<sup>c</sup>Applied to the following 3 clinical topics: screening for colorectal cancer, thrombolytic therapy for venous thromboembolism, and treatment of dementia.
view and child filters, compared with 3% to 69% reported in the original studies (Table 2). On average, it took the pediatrician 34 minutes and the clinical librarian 47 minutes to screen the titles and abstracts of 100 retrieved records.

**Precision Limited to Clinical Topics**

**Figure 2** shows the number of records retrieved when applying a precise (Shojania) and a sensitive (Montori 3) systematic review filter, combined with the child filter, in PubMed (January 1990-January 2006) to clinical topics. ADHD indicates attention-deficit/hyperactivity disorder; UTI, urinary tract infection.

Although the DARE database contains quality assessments of systematic reviews published in peer-reviewed journals, DARE includes only 52% of our hand-searched child health journal reviews. However, DARE uses more stringent inclusion criteria for systematic reviews than we did. Although DARE was created in February 1995, all hand-searched child health journal reviews published in 1994 were indexed in DARE. The Cochrane Database of Systematic Reviews and DARE are already prefiltered in their focus on systematic reviews; thus, the only task is to find topic-relevant systematic reviews in these databases. In general medical databases such as MEDLINE, systematic reviews should ideally be indexed using an appropriate and specific publication type term. Until there is an agreed definition of a systematic review, indexers will be unable to introduce or apply such a term. Pending this, we will need
The usefulness of a systematic review filter is expressed by sensitivity and precision. The trade-off between these 2 features will determine the choice of filter. Researchers conducting a new systematic review or guideline developers would best be served by the most sensitive search. This search will have the highest probability of retrieving all relevant reviews, but it will have low precision, retrieving many irrelevant articles. Those with little time on their hands (for example, clinicians looking for answers to patient care questions) will likely be best served by a more precise search strategy. The sensitivity of the systematic review filters as reported in the original studies and compared with our reference standard were similar except for the Shojania filter (93% vs 74%, respectively). Shojania's reference standard set of papers mainly included high-quality systematic reviews that probably contained consistent descriptions of systematic review criteria in the abstracts’ “Methods” sections. A filter developed in such a sample will probably consist of another combination of text words and MeSH headings than would be optimal to retrieve systematic reviews in our reference standard, which includes both low- and high-quality systematic reviews.

Because the sensitivity of the child filter alone was excellent (98%), we combined the systematic review filters with the child filter to improve precision. Inasmuch as few systematic reviews were published before 1990, we limited our search to 1990 and after. The number of records retrieved with the various systematic review filters was in the hundreds of thousands. Adding the child filter reduced the number of records by a factor range of 5 to 7. Of the sensitive search filters, the Montori 3 has the best trade-off between sensitivity (94%), number of records retrieved (151 227), and precision (3%). Of the more precise filters, the Shojania filter has the best trade-off between sensitivity (74%), number of records retrieved (10 188), and precision (45%). Although we added a child filter to improve precision and used a broad definition of a systematic review, precision was lower than reported in the original studies for which we have comparative precision data. However, this is to be expected because precision changes with the prevalence of positive results (or richness of the database). The prevalence of child health systematic reviews in MEDLINE is lower than the prevalence of adult systematic reviews. In addition, most existing search strategies were developed from a small, selective subset of systematic reviews published in high-impact journals. Precision was also estimated from their subsets of journals. Clinicians searching for systematic reviews in MEDLINE using the filters search the universe of MEDLINE and not a specific set of journals. Therefore, we believe that clinicians will be interested in precision in the universe of MEDLINE. Shojania and Bero\(^8\) calculated precision in MEDLINE for 3 clinical topics (Table 2). As we have shown, precision varies considerably among different clinical topics and seemed to be higher than precision not limited to clinical topics. Therefore, the results will not be generalizable.

It was often difficult to determine from the title and abstract whether an article was a true systematic review. Abstracts often lacked descriptions of important criteria for systematic reviews. Instead, general terms were used, such as “A literature review was performed,” “We tried to collect all published literature,” “This is an evidence-based overview,” and “We performed a qualitative assessment of the literature.” Therefore, it is to be expected that the most sensitive systematic review filters also retrieve many irrelevant records. In 1999, the Quality of Reporting Meta-analyses (QUOROM) statement was published to improve the quality of reporting meta-analyses or systematic reviews of clinical, randomized, controlled trials.\(^14\) To be able to differentiate narrative reviews from systematic reviews, we advise that authors follow these recommendations when writing abstracts. In addition, when children are included, this should be explicitly reported in the title or abstract. To make identification using PubMed easier, there is an urgent need to set minimal criteria that any review should fulfill for it to be indexed as a systematic review. Ideally, a database of child health systematic reviews should be created. Pending this, validated child health filters are needed to be able to retrieve all child health–relevant articles. Since this study was conducted, a study has been published that determined the retrieval characteristics of age-specific terms in MEDLINE (Ovid interface; Ovid Technologies, Inc, New York, New York) for pediatric and neonatal medicine.\(^15\)

**LIMITATIONS**

We used a broad definition of a systematic review, and a number of these reviews will not fulfill some of the stringent criteria for systematic reviews used by others. Given the amount of time needed to perform a systematic review and the methodological skills required, our goal was not to exclude potentially relevant systematic reviews. The explicit mention of a strategy for locating evidence seems to be the most basic and least controversial.\(^7\)

Our reference standard contains a subset of the MEDLINE database because it is impossible to hand search all journals indexed in MEDLINE for child health systematic reviews, and sensitivity could have been overestimated or underestimated. We tried to compose a representative reference standard of systematic reviews by hand searching 7 pediatric journals with a range of impact factors. This should help avert selection bias in our search filter accuracy study because of higher standards of study description (eg, by the use of a structured abstract with an explicit “Methods” section) in high-impact journals. In an attempt to avert possible biases occurring in any one year, we sampled records from various years within a 10-year range.

All systematic review filters except the Shojania filter were developed and tested using the Ovid interface. Because PubMed is the only MEDLINE free interface available worldwide and is widely used, we converted the Ovid search filters into PubMed format.\(^6,7\) Translating a search filter from Ovid format to PubMed format is a factor that may influence the performance of the filter.
Figure 3. Recommended precise and sensitive PubMed systematic review and child search strategies.

**RECOMMENDATIONS FOR CLINICIANS AND RESEARCHERS**

We advise busy clinicians searching for systematic reviews on health care interventions to search the Cochrane Library (Cochrane Database of Systematic Reviews and DARE) for high-quality systematic reviews. Several studies have shown that Cochrane reviews are more rigorous than journal reviews.16,17 If no relevant systematic reviews are found in the Cochrane Library or the interest is other than prevention, treatment, or diagnosis, other databases such as MEDLINE should be searched next. For a specific search, the Shojania filter combined with the child filter can be used. For topics for which not many reviews are available (eg, constipation), it may be more efficient to search with the sensitive Montori 3 filter combined with the child filter. Researchers and guideline developers interested in finding nearly all systematic reviews can use the Montori 3 filter combined with the child filter. Readers can save the search filters (Figure 3) they wish to use in PubMed and use them in combination with a clinical topic.

Accepted for Publication: July 13, 2007.

**Correspondence:** Nicole Boluyt, MD, Center for Pediatric Clinical Epidemiology, Room H3-145, Emma Children’s Hospital, Academic Medical Centre, PO Box 22700, 1100 DD Amsterdam, the Netherlands (n.boluyt@amc.uva.nl).

**Author Contributions:** Dr Boluyt had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: Boluyt, Tjosvold, Lefebvre, Klassen, and Offringa. Acquisition of data: Boluyt and Tjosvold. Analysis and interpretation of data: Boluyt, Lefebvre, and Offringa. Drafting of the manuscript: Boluyt and Klassen. Critical revision of the manuscript for important intellectual content: Boluyt, Tjosvold, Lefebvre, and Offringa. Administrative, technical, and material support: Boluyt and Tjosvold. Study supervision: Boluyt, Klassen, and Offringa. Information science expertise: Lefebvre.

**Financial Disclosure:** None reported.

**Additional Information:** The eBox is available at http://www.archpediatrics.com.

**Additional Contributions:** Marjan Loep, MLIS (deceased), Marcel van der Paardt, PhD, and Arnold Leenders, MLIS, Medical Library, Academic Medical Center, Amsterdam, the Netherlands, searched DARE and hand searched the pediatric journals; Grace Liang, PhD, Alberta Research Centre for Child Health Evidence, University of Alberta, Edmonton, Alberta, Canada, developed a program to enable random sample selection; and Liza Bialy, BSc, Alberta Research Centre for Child Health Evidence, University of Alberta, Edmonton, Alberta, screened records for true systematic reviews. None received compensation for their work.
SYSTEMATIC REVIEW SEARCH STRATEGIES

Shojania


Boynton (Most Sensitive Strategy)


White 1 (Most Sensitive Strategy)


Montori 1 (Most Sensitive Strategy)


Montori 2 (Most Precise Strategy)

Medline [tw] OR (systematic [ti] AND review [ti])

Montori 3 (Minimizing the Difference Between Sensitivity and Specificity)


Montori 4 (Combining Most Precise Term With Most Sensitive Terms)


PubMed


Child Search Strategy

Infant [MeSH] OR Infant* OR infancy OR Newborn* OR Baby* OR Babies OR Neonat* OR Preterm* OR Prematur* OR Postmatur* OR Child* OR Child[MeSH] OR Child* OR Schoolchild* OR School age* OR Preschool* OR Kid or kids OR Toddler* OR Adolescent* OR Adole* OR Teen* OR Boy* OR Girl* OR Minors[MeSH] OR Minors* OR Puberty[MeSH] OR Pubert* OR Pubescen* OR Presubescen* OR Pediatrics[MeSH] OR Paediatric* OR Paediatric* OR Paediatric* OR Schools[MeSH] OR Nursery school* OR Kindergarten* OR Primary school* OR Secondary school* OR Elementary school* OR High school* OR Highschool*