



NIH PUBLIC ACCESS

Author Manuscript

Nurs Outlook. Author manuscript; available in PMC 2015 March 01.

Published in final edited form as:

Nurs Outlook. 2014 ; 62(2): 112–118. doi:10.1016/j.outlook.2013.11.002.

Managing Large-Volume Literature Searches in Research Synthesis Studies

Nancy L. Havill, CNM, PhD [Research Associate],
University of North Carolina at Chapel Hill School of Nursing

Jennifer Leeman, DrPH, MDiv [Assistant Professor],
University of North Carolina at Chapel Hill School of Nursing

Julia Shaw-Kokot, MSLS [Assistant Department Head, User Services, Health Sciences Library],
University of North Carolina at Chapel Hill

Kathleen Knafel, PhD, FAAN [Frances Hill Fox Distinguished Professor & Associate Dean for Research],
University of North Carolina at Chapel Hill School of Nursing

Jamie Crandell, PhD [Research Assistant Professor], and
University of North Carolina at Chapel Hill School of Nursing

Margarete Sandelowski, PhD, RN, FAAN [Cary C. Boshamer Distinguished Professor]
University of North Carolina at Chapel Hill School of Nursing

Abstract

Background—Systematic reviews typically require searching for, retrieving, and screening a large volume of literature, yet little guidance is available on how to manage this volume.

Purpose—We detail methods used to search for and manage the yield of relevant citations for a mixed-methods mixed research synthesis study focused on the intersection between family life and childhood chronic physical conditions.

Method—We designed inclusive search strings and searched nine bibliographic databases to identify relevant research regardless of methodological origin. We customized searches to individual databases, developed workarounds for transferring large volumes of citations and eliminating duplicate citations using reference management software, and used this software as a portal to select citations for inclusion or exclusion. We identified 67,555 citations, retrieved and screened 3,617 reports, and selected 802 reports for inclusion.

Discussion/Conclusions—Systematic reviews require search procedures to allow consistent and comprehensive approaches and the ability to work around technical obstacles.

© 2013 Mosby, Inc. All rights reserved.

Correspondence: 2810 Carrington Hall, CB 7460, Chapel Hill, NC 27599-7460. Tel: 919-966-8360. havill@email.unc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

The escalating interest in systematic reviews and specifically research synthesis studies has generated a burgeoning literature focused on searching for and retrieving relevant research reports. Among the diverse topics addressed are search strategies (e.g., pearl-growing, citation searching; Papaioannou, Sutton, Carroll, Booth, & Wong, 2009; Schlosser, Wendt, Bhavnani, & Nail-Chiwetalu, 2006); techniques for locating reports of quantitative, qualitative and mixed-methods studies (e.g., Cooke, Smith, & Booth, 2012; Walters, Wilczynski, & Haynes for the Hedges Team, 2006); comparisons of bibliographic databases to identify those yielding the best returns (e.g., McDonald, Taylor, & Adams, 1999; Stevinson & Lawlor, 2004); and recommendations for reporting search strategies and findings (e.g., Sampson, McGowan, Cogo, Grimshaw, Moher, & Lefebvre, 2009).

What has yet to be fully addressed, however, is the management of the large volume of literature likely to be found in even the most delimited review, the technical issues and workarounds necessary to search within diverse bibliographic databases across the social and behavioral science and practice disciplines, and the use of reference management software effectively and efficiently to track search activities and outcomes. Regardless of the scope of their reviews, reviewers will likely retrieve and therefore have to manage a much larger number of reports than they will ultimately include. The number of articles retrieved may be even greater when conducting mixed research synthesis studies, or reviews that include reports of qualitative, quantitative, and/or mixed-methods studies. Careful tracking of the references retrieved and of the decisions made throughout the search process is critical. Moreover, publication of systematic reviews of any kind now requires that the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; <http://www.prisma-statement.org/statement.htm>) guidelines be followed whereby reviewers detail the information sources, delimitations set for the search process, search strategies, and references identified, retrieved, and ultimately included in the review.

Accordingly, our purpose in this paper is to describe how we managed a literature search that initially yielded 67,555 documents in our ongoing National Institute of Nursing Research-funded research synthesis study—“Mixed-Methods Synthesis of Research on Childhood Chronic Conditions and Family”—hereafter referred to as the Family Synthesis study. We address how the search was designed, how reports retrieved were tracked, stored, organized, and evaluated for relevance, and how technical problems associated with managing this large volume of references were addressed.

The Family Synthesis Study

The purpose of the Family Synthesis study is to explore the intersection between family life and childhood chronic physical conditions. This is a mixed-methods mixed research synthesis study encompassing reports of empirical qualitative, quantitative, and mixed-methods studies, and qualitative and quantitative approaches for integrating the findings from these reports (Sandelowski, Voils, Crandell, & Leeman, 2013). Thus, the literature search was designed to be broadly inclusive, with the goal of identifying the full breadth of research findings related to the topic regardless of methodology. Team members include researchers with expertise in family research and synthesis methods, and an information specialist with expertise in developing search strategies effective for a range of health and behavioral and social science databases.

What follows is a detailed description of how we moved from an initial search yield of 67,555 documents to the 802 reports we accepted into our study. We detail the key phases in this recursive process and the strategies employed to address the challenges we encountered in each phase. We also draw from what we learned from an initial scoping study (Arksey &

O'Malley, 2005) we conducted to pilot test and refine elements of the search process we describe here.

Conducting and Managing the Search

Defining Key Concepts

As with all reviews of the literature, we began with an initial definition of the key concepts in our study: family, child, and chronic physical condition (Cooper, 2010). *Family* was defined broadly as constituting a group of intimates living together or in close geographic proximity with strong emotional bonds and with a history and a future (Fisher et al. 1998). *Child* was defined as an individual no older than 18 years. *Chronic physical condition* was defined as a medical condition lasting or expected to last at least 1 year and producing or expected to produce one or more of the following sequelae for the child: limitation in function or activity; dependence on medication, special diet, medical technology, assistive devices or persons; and/or the need for health services beyond what is usual for a child of the same age (Stein, Bauman, Westbrook, Coupey, & Ireys, 1993).

Identifying Bibliographic Databases

In consultation with the team's information specialist, and based on the results of our initial scoping study, we identified the databases most likely to include reports of research addressing the intersection between family life and childhood chronic physical conditions. During the scoping study, we had assessed the contribution of a range of databases, including Academic Search Premier, CINAHL, Cochrane Database of Systematic Reviews, EMBASE, ERIC, Family & Society Worldwide, PsychInfo, PubMed, Social Work Abstracts, Sociological Abstracts, and Web of Science. After comparing search yields, we retained all of the databases except the Cochrane and Web of Science databases, which yielded no relevant articles not already identified in searches of the other databases.

Selecting Limits and Search Terms

Bibliographic databases provide a range of options for limiting the overall scope of the search for literature. We limited the search only to English language publications and, to ensure the inclusion of relatively current research (Barroso, Sandelowski, & Voils, 2006), to the years 2000 to the present (or 2011). Consistent with the imperatives of a mixed research synthesis study, no limits were placed on particular types of research designs or methodologies.

The initial search was constructed as three separate topic-specific text-word search strings (i.e., lists of search terms), each of which addressed one of the three central concepts in our study, namely, family, child, and chronic physical condition. Each of these three topic search strings was pilot tested separately before being combined into a final strategy to ensure that the selected terms produced the desired results. By piloting each string separately, we were better able to troubleshoot when a group of terms yielded a much larger or smaller number of citations than anticipated. The family string included the terms *family*, *caregiver*, *mother*, *father*, *sibling*, *brother*, *sister*, *grandparent*, and *parent*. The child string included terms representing children from birth through adolescence, that is, the terms *child*, *infant*, *newborn*, *adolescent*, and *teenager*. To create the search string for chronic physical conditions we included both the general term *chronic illness* and terms for specific conditions because the results of the scoping study had demonstrated that the general term *chronic illness* identified many but not all relevant reports. The following disease-specific terms were included: *anemia*, *arthritis*, *asthma*, *cancer*, *cystic fibrosis*, *diabetes*, *end-stage renal disease*, *heart problems*, *muscular dystrophy*, and *seizure disorders*. These specific medical conditions were drawn from the physical diseases and conditions identified in the

National Survey of Children with Special Health Care Needs (Davidoff, 2004). To this list, we added cancer and end-stage renal disease because they were identified in our scoping study as conditions frequently addressed in studies of life in families with children with chronic physical conditions.

Customizing Searches to Individual Databases

Because databases have different rules regarding syntax and types of search terms, truncation rules, and limiters, the topic search strings were customized for use in each of the selected databases (Freund & Willett, 1982). With the exception of PubMed and CINAHL, we searched all databases using text-word searches with appropriate truncation. Truncation involves placing an asterisk after the base of a word with multiple alternate endings (e.g., child*) thereby cueing the database to identify all instances of words that begin with *child*, such as children, child, and childhood. Each of the words in the search string were thus entered and truncated as shown in the following illustration of a text-word search string for arthritis:

(child* or teen* or adolesc* or infant* or newborn*) AND (famil* or parent* or mother* or father* or caregiver* or "care giver*" or grandparent* or grandmother* or grandfather* or sister* or brother* or sibling*) AND arthriti*

Text-word searching with truncation was not used in PubMed because this database automatically stops searching after a maximum number of variations of the term have been identified and, therefore, all eligible articles may not be identified. Instead of text words, we used Medical Subject Heading (MeSH) terms to search PubMed. MeSH terms are a controlled vocabulary used to index articles within the bibliographic database. The vocabulary is hierarchically structured, with more specific terms located below broader terms. For example, the narrow terms *parents* and *siblings* are located below the broader term *family*. To select appropriate MeSH search terms, we assessed the MeSH database's definition for candidate broad terms (e.g., *family*) and the associated narrower terms to ensure that all desired narrower terms were captured. For example, the MeSH database's definition for the term *family* is "a social group consisting of parents or parent substitutes and children" and includes all of the narrower terms identified in our family search string. Therefore, we used the MeSH term *family*, which we "exploded" to include all the narrower terms included below it in the hierarchy (DeLuca et al. 2008). A final search string utilizing MeSH terms for PubMed for arthritis is shown below:

("infant"[Mesh] OR "child"[Mesh] OR "adolescent"[Mesh]) AND ("Family"[Mesh]) or "Caregivers"[Mesh] or grandparent* or grandmother* or grandfather* or aunt* or uncle*) AND ("Arthritis"[Mesh]) OR "Joint Diseases"[Mesh])

Because research reports are entered into the PubMed database prior to being indexed with MeSH terms, the use of MeSH terms has the disadvantage of failing to capture reports that have yet to be indexed. We addressed this limitation largely by searching in multiple other databases in addition to PubMed, knowing that relevant reports were likely to be included in more than one database. CINAHL indexing is completed before entry into the database, and all other databases were searched using appropriate text words, allowing retrieval of reports potentially missed by the MeSH searches. In addition, we plan to update our searches in the fourth year of the study and thereby capture reports not available at the time of our initial search.

Care was taken to ensure the same terms and truncation combinations were used in each of the databases. This can be challenging in some databases, like EMBASE, due to the complex choices and searching conventions. For example, Elsevier's EMBASE does not allow truncation within a phrase. With EMBASE we needed to turn off the MEDLINE

search feature because we had already searched PubMed. Knowing the idiosyncrasies of each database is critical for retrieving the desired citations.

Managing the Reports Retrieved

The broad search criteria and search terms applied resulted in 67,555 potentially relevant citations. We therefore had to develop a data management strategy that would transfer a large number of citations, eliminate duplicate citations, and preserve the results of all searches so that we could systematically review them for relevance or repeat any of the searches. We used RefWorks (<http://www.refworks.com/>), an online reference management software tool, to store, sort, and track the references identified through our searches. Most available reference management software tools can be electronically linked to a university library's bibliographic databases in a way that allows for the direct transfer of data. The data we transferred to RefWorks included the full citation and abstract of each article as well as electronic links to the complete texts to which the university library had electronic access. Transferring retrieved citations to a reference manager has the advantages of not only preserving the search precisely as it occurred, but also of providing a relatively straightforward platform for the research team to review titles, abstracts, and full-text articles.

Yet, there are limits to the amount of data that can be included in a single transfer that are imposed both by the bibliographic databases for proprietary reasons and by the reference manager's capacity to accept incoming files. For example, our full PubMed search yielded 15,239 references, which was too big for Refworks to accept as a single transfer. We then attempted to transfer the results of the search in segments, but we were limited to transferring a maximum of 500 citations at a time, which would have required over 30 separate transfers. Because transferring a large search in a series of smaller segments would have been extremely time consuming, we organized the search process as a series of condition-specific searches that each resulted in datasets that were small enough to be moved intact. This was accomplished by creating a separate RefWorks file for each bibliographic database, running each disease/condition as a separate search in each database, and then moving these smaller datasets into separate condition-specific folders within each RefWorks file. We also created text-files of all search results as a backup in case the reference manager databases became corrupted. Using this process, 9 RefWorks storage databases were created, 1 for each of the bibliographic search databases (e.g., PubMed, ERIC); each of these databases had 11 separate folders for the results of the individual condition-specific searches and the search on the general term *chronic illness* (Figure 1).

Duplicate Deletion

Conducting multiple searches individually resulted in numerous duplicate references both within and across the RefWorks databases. RefWorks allows users to identify duplicates across files within a database but not across databases. Accordingly, we developed a systematic approach whereby duplicate citations were identified and eliminated first within each of the nine RefWorks databases containing the files downloaded from the bibliographic databases. We then combined the condition-specific files from the original nine RefWorks databases into 11 new RefWorks databases, one for each condition. Within these condition-specific databases we were able to identify and then eliminate duplicate records occurring across the bibliographic databases (Figure 1).

RefWorks offers two options for viewing duplicates—exact and close duplicates—and both were applied. We used close duplicates because it identified duplicates that were missed by the exact duplicate function, such as when databases used different conventions to identify authors (e.g., full names versus initials) resulting in duplicate files that were not identified

by the exact duplicate function. We did not automatically delete duplicates but rather examined and manually deleted each identified duplicate. This was necessary because sometimes the reference manager mis-identified references as duplicates, such as identifying “Part II” of an article as a duplicate of “Part I.” Through this process, 24,584 references were identified and eliminated as duplicates, leaving an initial dataset of 43,114 references for review.

Report Review and Selection

The study team then reviewed titles, abstracts, and/or the full texts of these 43,114 references to identify those that met criteria for inclusion. The inclusion and exclusion criteria evolved over the course of the review as the team gained familiarity with the available literature and refined the study’s scope feasibly to accommodate the resources available to conduct the review (Levac, Colquhoun, & O’Brien, 2010). These criteria evolved primarily from a refined conceptualization of family that delimited the reports of studies to be included to those containing findings about: (a) family structure, defined as the ordered roles and relationships within the family, including routines and rituals of everyday family life, and division of labor; (b) family functioning, defined as characteristics of the family system (e.g., resilience, cohesiveness, environment, climate, values, family system stress) and interactions among family members (e.g., decision making, problem solving, information sharing, communication); (c) family relationships, defined as the nature and quality of relations among family members (e.g., marital adjustment, conflict and conflict resolution, withdrawal, attachment, relationship satisfaction); and/or (d) family resources, defined as factors external to the family that influence the quality of family life, including all types and sources of social support, including support from extended family and healthcare providers.

The team was able to use RefWorks as a portal to view titles and abstracts and, when needed, to link to the full text of most articles. In the rare cases that the university library did not own a subscription for a journal in which an article appeared, the article was requested through inter-library loan. Two members of the team reviewed each citation and maintained a hard-copy Excel spreadsheet that listed all citations in the RefWorks files, which they used to track references selected for inclusion and to document reasons for exclusion. This process was completed independently for each of the 11 condition-specific databases. The two reviewers then met to compare their decisions. In those cases where they differed, reviewers discussed the disputed reports to reach consensus. For all disputed reports, the full text of the article was retrieved and reviewed. Final decisions were recorded on a consensus spreadsheet, which documented the articles selected for inclusion and exclusion and the reasons for exclusion.

Completing the PRISMA Diagram

Throughout the process of search, retrieval, and selection of research reports, spreadsheets were maintained tracking the numbers of reports identified in each step and the reasons for exclusion. Therefore, completing the PRISMA diagram was a rather straightforward exercise in locating numbers and other information from existing spreadsheets (Figure 2).

Conclusion

We employed a broadly inclusive approach to search for all research reports of studies related to the intersection of family life and childhood chronic physical conditions regardless of methodology. Mixed-methods mixed research synthesis studies are increasingly being called for as a means of capturing more of the evidence available to guide practice. Such studies are especially well suited to contribute evidence on contextual factors that influence

intervention implementation and effectiveness, and on patients' and providers' perspectives on health problems and interventions to address those problems (Leeman, Voils, & Sandelowski, in press). Capturing findings generated by diverse research methods requires broad search and retrieval processes similar to the one described here whereby all relevant databases were searched deliberately to achieve topical, conceptual, and methodological scope with few a priori restrictions.

Our approach is in many ways similar to that used in scoping studies, which involve mapping the available literature to assess its breadth and depth and to identify potential gaps (Arksey & O'Malley, 2005; Daudt, van Mossel, & Scott, 2013; Davis, Drey, & Gould, 2009; Levac, Colquhoun, & O'Brien, 2010). Although our ultimate goal extends well beyond merely scoping the literature to include the actual synthesis of findings across studies, we used scoping largely as a search strategy because it allowed us to map the landscape of literature addressing families with children with chronic conditions. This map then served as the backdrop for choosing the literature that would be included in our research synthesis and for clearly defining what was not included. Knowing where we elected not to go on the map will serve to locate, contextualize, and clarify the boundaries of the research syntheses we will produce. In our case, using scoping as a search strategy also allowed us to develop a more refined conceptualization of family, a topic we will address in more detail in a future paper.

Conducting an inclusive research synthesis study mandates an inclusive search strategy that will likely yield thousands of references to review. This process requires not only procedures to organize search yields and allow systematic, consistent, and comprehensive approaches for review, but also the ability and creativity to work around technical obstacles.

Acknowledgments

The preparation of this article was supported by the National Institute of Nursing Research, National Institutes of Health under award number R01NR012445: "Mixed-Methods Synthesis of Research on Childhood Chronic Conditions and Family."

Funding

"Mixed-methods synthesis of research on childhood chronic conditions and family" (K. Knafl & M. Sandelowski, PIs; J. Leeman, J. Crandell, & J. Shaw-Kokot, co-Is). National Institute of Nursing Research, National Institutes of Health, R01NR012445, September 1, 2011-June 30, 2016.

References

- Arksey H, O'Malley L. Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*. 2005; 8:19–32.10.1080/1364557032000119616
- Barroso J, Sandelowski M, Voils C. Research results have expiration dates: Ensuring timely systematic reviews. *Journal of Evaluation in Clinical Practice*. 2006; 12:454–462.10.1111/j.1365-2753.2006.00729.x [PubMed: 16907690]
- Cooke A, Smith D, Booth A. Beyond PICO: The SPIDER tool for qualitative evidence synthesis. *Qualitative Health Research*. 2012; 22:1435–1443.10.1177/1049732312452938 [PubMed: 22829486]
- Cooper, H. *Research synthesis and meta-analysis: A step-by-step approach*. 4. Los Angeles, CA: Sage; 2010.
- Daudt HM, van Mossel C, Scott SJ. Enhancing the scoping study methodology: A large, inter-professional team's experience with Arksey and O'Malley's framework. *BMC Medical Research Methodology*. 2013; 13(1):1–9. Open access at www.biomedcentral.com/1471-2288/13/48. 10.1186/1471-2288-13-48 [PubMed: 23297754]

- Davidoff AJ. Identifying children with special health care needs in the National Health Interview Survey: A new resource for policy analysis. *Health Services Research*. 2004; 39:53–72.10.1111/j.1475-6773.2004.00215.x [PubMed: 14965077]
- Davis K, Drey N, Gould D. What are scoping studies? A review of the nursing literature. *International Journal of Nursing Studies*. 2009; 46:1386–1400.10.1016/j.ijnurstu.2009.02.010 [PubMed: 19328488]
- DeLuca JB, Mullins MM, Lyles CM, Crepaz N, Kay L, Thadiparthi S. Developing a comprehensive search strategy for evidence based systematic reviews. *Evidence Based Library and Information Practice*. 2008; 3:3–32.
- Fisher LW, Chesla CA, Bartz RJ, Gilliss C, Skaff MA, Sabogal F, Kanter RA, Lutz CP. The family and type 2 diabetes: A framework for intervention. *The Diabetes Educator*. 1998; 24:599–607.10.1177/014572179802400504 [PubMed: 9830956]
- Freund GE, Willett P. Online identification of word variants and arbitrary truncation searching using a string similarity measure. *Information Technology: Research & Development*. 1982; 1:177–187.
- Leeman, JL.; Voils, CI.; Sandelowski, M. Conducting mixed methods literature reviews: Synthesizing the evidence needed to develop and implement complex social and health interventions. In: Hesse-Biber, S.; Johnson, B., editors. *Oxford handbook of mixed and multimethod research*. New York: Oxford University Press; (in press)
- Levac D, Colquhoun H, O'Brien KK. Scoping studies: Advancing the methodology. *Implementation Science*. 2010; 5:69–77. Open access at www.implementationscience.com/content/5/1/69. 10.1186/1748-5908-5-69 [PubMed: 20854677]
- McDonald S, Taylor L, Adams C. Searching the right database: A comparison of four databases for psychiatry journals. *Health Libraries Review*. 1999; 16:151–156.10.1046/j.1365-2532.1999.00222.x [PubMed: 10620849]
- Papaioannou D, Sutton A, Carroll C, Booth A, Wong R. Literature searching for social science systematic reviews: Consideration of a range of search techniques. *Health Information and Libraries Journal*. 2009; 27:114–122.10.1111/j.1471-1842.2009.00863.x [PubMed: 20565552]
- Sampson M, McGowan J, Cogo E, Grimshaw J, Moher D, Lefebvre C. An evidence-based practice guideline for the peer review of electronic search strategies. *Journal of Clinical Epidemiology*. 2009; 62:944–952.10.1016/j.jclinepi.2008.10.012 [PubMed: 19230612]
- Sandelowski, M.; Voils, CI.; Crandell, J.; Leeman, J. Synthesizing qualitative and quantitative research findings. In: Beck, CT., editor. *Routledge international handbook of qualitative nursing research*. New York, NY: Routledge; 2013. p. 347-356.
- Schlosser RW, Wendt O, Bhavnani S, Nail-Chiwetalu B. Use of information-seeking strategies for developing systematic reviews and engaging in evidence-based practice: The application of traditional and comprehensive pearl growing. A review. *International Journal of Language & Communication Disorders*. 2006; 41:567–582.10.1080/13682820600742190 [PubMed: 17050471]
- Stein R, Bauman LJ, Westbrook LE, Coupey SM, Ireys HT. Framework for identifying children who have chronic conditions: A case for a new definition. *Journal of Pediatrics*. 1993; 122:342–347. [PubMed: 8441085]
- Stevinson C, Lawlor DA. Searching multiple databases for systematic reviews: Added value or diminishing returns? *Complementary Therapies in Medicine*. 2004; 12:228–232.10.1016/j.ctim.2004.09.003 [PubMed: 15649836]
- Walters LA, Wilczynski NL, Haynes RB. for the Hedges Team. Developing optimal search strategies for retrieving clinically relevant qualitative studies in EMBASE. *Qualitative Health Research*. 2006; 16:162–168.10.1177/1049732305284027 [PubMed: 16317183]

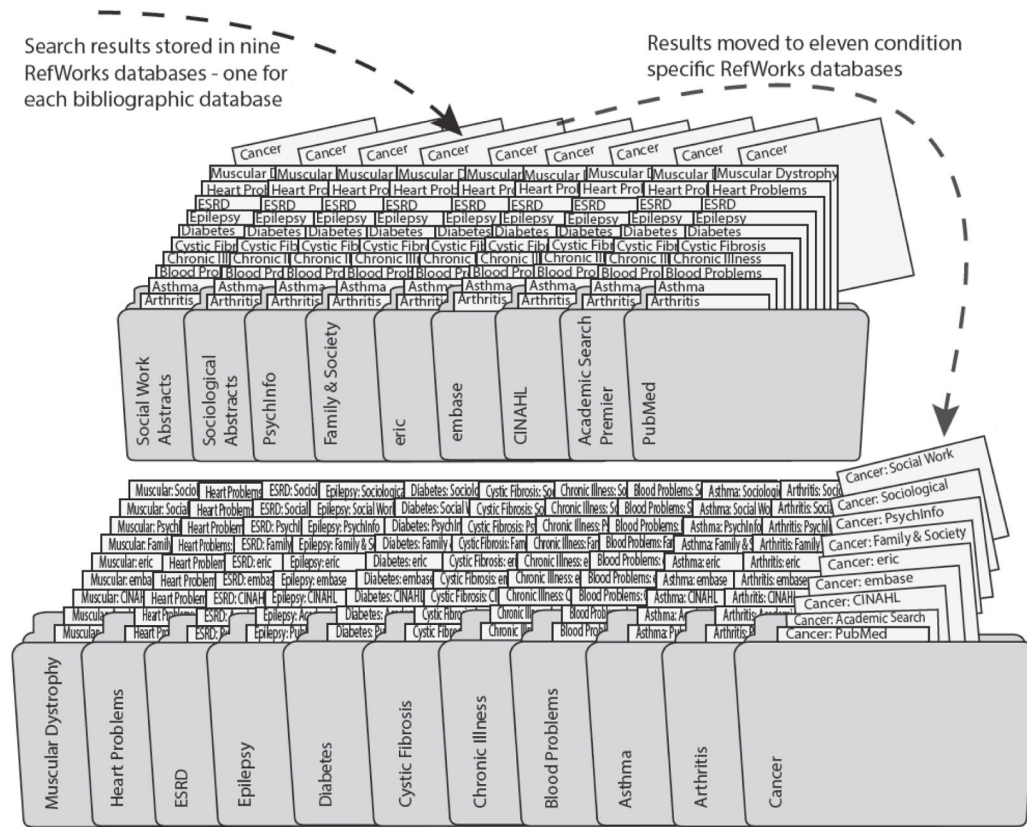


Figure 1.
Using RefWorks® to Manage Search Yields

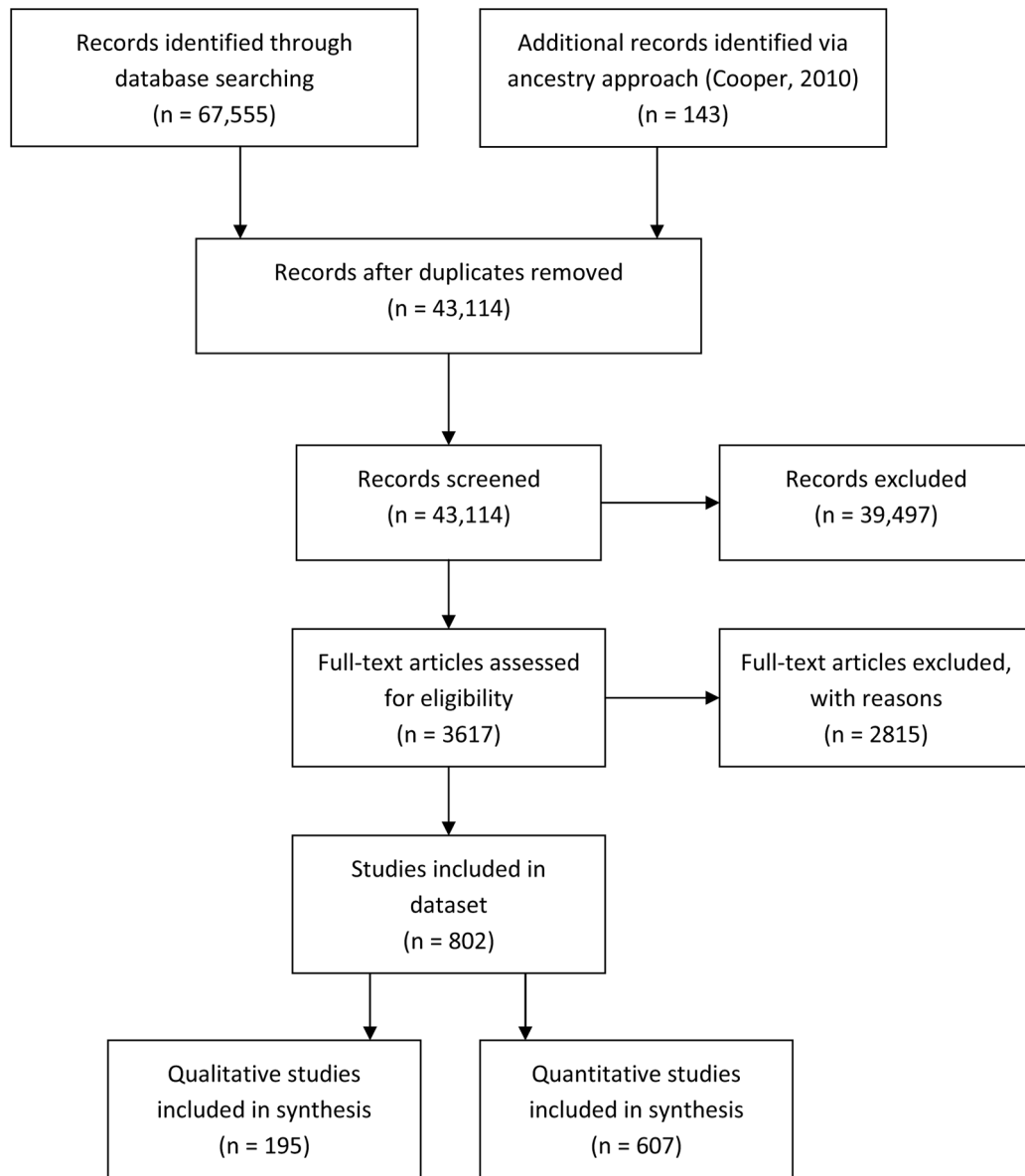


Figure 2.
PRISMA Flow Diagram