# Information in practice

# Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources

Trisha Greenhalgh, Richard Peacock

## Abstract

**Objective** To describe where papers come from in a systematic review of complex evidence.

**Method** Audit of how the 495 primary sources for the review were originally identified.

**Results** Only 30% of sources were obtained from the protocol defined at the outset of the study (that is, from the database and hand searches). Fifty one per cent were identified by "snowballing" (such as pursuing references of references), and 24% by personal knowledge or personal contacts.

**Conclusion** Systematic reviews of complex evidence cannot rely solely on protocol-driven search strategies.

## Introduction

In Cochrane reviews of therapeutic interventions, most high quality primary studies could be identified by searching four standard databases—the Cochrane Controlled Trials Register (which contains 79% of studies listed in Cochrane systematic reviews), Medline (69%), Embase (65%), and Science and Social Sciences Citation Indexes (61%).[1] Searching 26 further databases identified only an extra 2.4% of trials. No comparable figures have been published for systematic reviews of complex evidence, which address broad policy questions and synthesise qualitative and quantitative evidence, usually from multiple and disparate sources.[2]

The aim of our study was to audit the origin of primary sources in a wide ranging systematic review of complex evidence.

## Method

We reviewed the diffusion of service-level innovations in healthcare organisations. The methods and full report have been published elsewhere.[3] [4] Briefly, six researchers mapped 13 different research traditions, compared their conceptual and theoretical approaches, and synthesised the empirical evidence. We report here on the search phase.

After extensive "browsing" in libraries and bookshops to get a feel for the overall research field, we used the following methods:

- Protocol driven (search strategy defined at the outset of the study)
  Hand search of 32 journals across 13 disciplinary fields
  Electronic search of 15 databases by index terms, free text, and named author
- "Snowballing" (emerging as the study unfolded)
  Reference tracking: we scanned the reference lists of all full text papers and used judgment to decide whether to pursue these further

Citation tracking: using special citation tracking databases (Science Citation Index, Social Science Citation Index, and Arts and Humanities Citation Index), we forward tracked selected key papers published more than three years previously, thereby identifying articles in mainstream journals that had subsequently cited those papers

- Personal knowledge (what we knew and who we knew)
  Our existing knowledge and resources
  Our personal contacts and academic networks
  Serendipitous discovery (such as finding a relevant paper when looking for something else).

We initially scanned over 6000 electronic abstracts and book titles, and selected 1004 full-text books or papers for further analysis. After appraising all these for quality and relevance, we cited 495 (of which 213 were empirical studies and 21 were systematic or quasi-systematic reviews) in the final report. We classified each according to its origin, using the taxonomy above.

## Results

The table shows the origins of our 495 sources. Twenty three per cent of the sources were known to us or were recommended by colleagues when we approached them by email, which took little time. Electronic searching, including developing and refining search strategies and adapting these to different databases, took about two weeks of specialist librarian time and yielded only about a quarter of the sources—an average of one useful paper every 40 minutes of searching. It took a month to hand search a total of 271 journal-years, from which we extracted only 24 papers that made the final report—an average of one paper per nine hours of searching.

Overall, the greatest yield was from pursuing selected references of references. It was impossible to isolate the time spent on reference tracking since this was done in parallel with the critical appraisal of each paper. Electronic citation tracking of selected papers took around a day in total and uncovered many important, recent sources, including five systematic reviews (three of which were not identified by any other method) and 12% of all empirical studies—around one useful paper for every 15 minutes of searching. Five papers came our way by chance, including a major systematic review that was passed to one of us in a coffee queue by someone who did not know about this research.

## Discussion

Systematic review of complex evidence cannot rely solely on predefined, protocol driven search strategies, no matter how many databases are searched. Strategies that might seem less

Methods of identifying 495 primary data sources in a systematic review of complex evidence (see text for definitions of terms). Values are numbers (percentages) of data sources

| Search method | Empirical research studies (n=213) | Theoretical or "overview" sources (n=261) | Systematic or quasi-systematic reviews (n=21) | Total (n=495) |
|---|---|---|---|---|
| Protocol driven: | 87 (41) | 53 (20) | 10 (48) | 150 (30) |
|   Electronic database search* | 75 (35) | 41 (16) | 10 (48) | 126 (25) |
|   Hand search (32 journals) | 12 (6) | 12 (5) | 0 | 24 (5) |
| "Snowballing": | 113 (53) | 130 (50) | 9 (43) | 252 (51) |
|   Reference tracking | 87 (41) | 127 (49) | 4 (19) | 218 (44) |
|   Citation tracking | 26 (12) | 3 (1) | 5 (24) | 34 (7) |
| Personal knowledge: | 21 (10) | 91 (35) | 5 (24) | 119 (24) |
|   Sources known to research team | 15 (7) | 66 (25) | 2 (10) | 85 (17) |
|   Social networks of research team | 4 (2) | 23 (9) | 2 (10) | 29 (6) |
|   Serendipitous | 2 (1) | 2 (1) | 1 (5) | 5 (1) |
| Raw total including double counting† | 220 (104) | 276 (106) | 24 (114) | 521 (105) |
| Total in final report | 213 | 261 | 21 | 495 |

*15 databases including Medline, EPOC, ASSCI, and ERIC.
†Numbers add up to more than 100% because some sources were located by more than one method. The proportion of sources double counted is probably a substantial underestimate since (for example) we did not flag a paper identified in a reference track if we already had it on file.

efficient (such as browsing library shelves, asking colleagues, pursuing references that look interesting, and simply being alert to serendipitous discovery) may have a better yield per hour spent and are likely to identify important sources that would otherwise be missed. Citation tracking is an important search method for identifying systematic reviews published in obscure journals.

This study builds on extensive secondary research by all the authors on the systematic review (see reference list), whose input we gratefully acknowledge. We also thank Jeanette Buckingham for comments on a previous draft of this paper.

Competing interests: None declared.

1  Royle P, Milne R. Literature searching for randomized controlled trials used in Cochrane reviews: rapid versus exhaustive searches. *Int J Technol Assess Health Care* 2003;19:591-603.
2  Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A. Synthesising qualitative and quantitative evidence: a review of possible methods. *J Health Serv Res Policy* 2005;10:45-53.
3  Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O, Peacock R. Storylines of research in diffusion of innovation: a meta-narrative approach to systematic review. *Soc Sci Med* 2005;61:417-30.
4  Greenhalgh T, Robert G, Bate P, Kyriakidou O, Macfarlane F. *Diffusion of innovations in health service organisations: a systematic literature review.* Oxford: Blackwells, 2005.

*(Accepted 14 September 2005)*

doi 10.1136/bmj.38636.593461.68

Department of Primary Care and Population Sciences, University College London Medical School, Holborn Union Building, London N19 5LW
Trisha Greenhalgh *professor of primary health care*

Archway Healthcare Library, Holborn Union Building, London
Richard Peacock *clinical librarian*

Correspondence to: T Greenhalgh p.greenhalgh@pcps.ucl.ac.uk

## What is already known on this topic

It is generally assumed that the more explicit and meticulous the search strategy, the more likely a systematic review is to pick up all the important papers

In systematic reviews of clinical treatments, most high quality primary studies can be identified by searching four standard electronic databases

## What this study adds

In systematic reviews of complex and heterogeneous evidence (such as those undertaken for management and policymaking questions) formal protocol-driven search strategies may fail to identify important evidence

Informal approaches such as browsing, "asking around," and being alert to serendipitous discovery can substantially increase the yield and efficiency of search efforts

"Snowball" methods such as pursuing references of references and electronic citation tracking are especially powerful for identifying high quality sources in obscure locations