Systematic Reviews: step-by-step overview

Madhukar Pai, MD, PhD
Assistant Professor
Department of Epidemiology & Biostatistics
McGill University, Montreal, Canada
Email: madhukar.pai@mcgill.ca

Disclosure

• I co-chair the Stop TB Partnership’s New Diagnostics Working Group
• I consult for the Bill & Melinda Gates Foundation
• I serve on editorial boards of:
  □ Cochrane Infectious Diseases Review Group
  □ PLoS One
  □ PLoS Medicine
  □ IJTLID
Information explosion - a key barrier to use of knowledge

The importance of research synthesis

- We need evidence for both clinical practice and for public health decision making
- Where does evidence come from?
  - An good review is a state-of-the-art synthesis of current evidence on a given research question
  - Given the explosion of medical literature, and the fact that time is always scarce, review articles play a big role in decision-making
The importance of research synthesis

- Given that most clinicians and public health professionals do not have the time to track down all the original articles, critically read them, and obtain the evidence they need for their questions,
  - Systematic reviews and clinical practice guidelines may be their best source of evidence
    - Several “pre-digested” sources of evidence are currently available
    - The EBM movement is heavily dependent on these pre-appraised evidence sources

Hierarchy of evidence

<table>
<thead>
<tr>
<th>Level of Evidence</th>
<th>Methodology</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>SR (with homogeneity of RCTs)</td>
</tr>
<tr>
<td>1b</td>
<td>Individual RCT (with remote confidence interval)</td>
</tr>
<tr>
<td>2a</td>
<td>SR (with homogeneity of cohort studies)</td>
</tr>
<tr>
<td>2b</td>
<td>Individual cohort study (including RCTs)</td>
</tr>
<tr>
<td>3a</td>
<td>SR (with homogeneity of case-control studies)</td>
</tr>
<tr>
<td>3b</td>
<td>Individual case-control study</td>
</tr>
<tr>
<td>4</td>
<td>Case-series (not used as evidence)</td>
</tr>
<tr>
<td>5</td>
<td>Expert opinion without explicit critical appraisal, or based on expert consensus or “best evidence”</td>
</tr>
</tbody>
</table>

Guidelines and recommendations: GRADE

Systematic reviews are a key component of the GRADE process

But evidence from SR is NOT sufficient for policy!

Doing New Research? Don’t Forget the Old

Nobody should do a trial without reviewing what is known

Mike Clarke

On May 2, 1998, George Orwell addressed the Association of American Medical Colleges in Philadelphia to present a vision of the future of health information. ‘I look forward to the day when every medical student in any part of the civilized world shall in an hour be able to gain a knowledge pertaining to a subject of the experience of every other man in the world’ [1]. Has his vision been realized? Good quality, but some of it is not. How anyone wishing to use the health literature to make well-informed decisions must both identify the relevant research from among this vast amount of information and then appraise it. This is an impossible task for many. Even though making access to the literature easier and cheaper will increase the ability of people to find research, it will also reveal just how much information there is out there and how daunting is the task of making sense of it.

Box 1. Practical Suggestions for Researchers

- Conduct a systematic review of your research question before embarking on a new study, or identify a relevant review done by someone else.
- Design your study to take account of the relevant successes and failures of the prior studies, and of the evidence within them.
- Discuss the findings of your study in the context of an updated systematic review of relevant research.
- Publish the systematic review within, alongside, or shortly after the report of your study.
- Provide information from your study to others doing systematic reviews of similar topics.

Prof Archibald Cochrane, CBE (1909 - 1988)

- The Cochrane Collaboration is named in honour of Archie Cochrane, a British researcher.
- In 1979 he wrote, "It is surely a great criticism of our profession that we have not organised a critical summary, by specialty or subspecialty, adapted periodically, of all relevant randomized controlled trials”

Source: http://www.cochrane.org/cochrane/archieco.htm

The Cochrane Collaboration

- Archie Cochrane’s challenge led to the establishment during the 1980s of an international collaboration to develop the Oxford Database of Perinatal Trials.

Source: http://www.cochrane.org/cochrane/archieco.htm
Steady rise in number of SRs published

Bastian et al. PLoS Med 2010
Since textbooks are not good sources of current evidence, emergence of online, updated resources:

Note: not all chapters are well supported by SRs
Systematic reviews in TB

- A large number in the past 5 – 10 years
- We are aware of over 250 SRs focused on TB
- On diagnostics alone, there are 40+
- Increasingly being used for policy and guideline development in TB (with or without GRADE)
  - WHO
  - ATS
  - ERS
  - CDC
  - IUATLD
  - IDSA
  - PHAC

Examples: treatment
Examples: diagnostics

**Annals of Internal Medicine**

**Systematic Review: T-Cell-based Assays for the Diagnosis of Latent Tuberculosis Infection: An Update**

Muthitakul Re, MD, PhD, Kinza Lewand, MD, and Dick Inaarsen, MD, PhD

**Annals of Internal Medicine**

**Article**


Dick Inama, MD, MPH, Muthitakul Re, MD, PhD, and George Comstock, MD, MPH

Updated Guidelines for Using Interferon Gamma Release Assays to Detect Mycobacterium Tuberculosis Infection—United States, 2010

Examples: etiology/epidemiology

**Review Article**

**Risk of Tuberculosis From Exposure to Tobacco Smoke**

Michael N. Berto, PhD, Andreas Khulahri, PhD, Muthitakul Re, MD, PhD, Lisa Chang, MPH, Fumio Oka, MD, MPH, Kon R. Smith, PhD

**Tobacco Smoke, Indoor Air Pollution and Tuberculosis: A Systematic Review and Meta-Analysis**

Huang Ke Liu, Haidi Gao, Yiren Wang

**Tobacco and tuberculosis: a qualitative systematic review and meta-analysis**

K. Sharma, C. S. Zhang, S. S. Davor, K. N. Dhawan, M. S. Tran, M. Gupta, C. Ray

1. The International Union Against Tuberculosis and Lung Disease, Pole Mérive, The University of Michigan
2. University of North Carolina, Winston-Salem, North Carolina, USA
4. Health Institute, New Delhi, India
Examples: etiology/epidemiology

Tuberculosis among Health-Care Workers in Low- and Middle-Income Countries: A Systematic Review
Jaeghe Joo,1,2,1 Arthur L. Reingold,3 Dick Mantel,4,5 and Marlous Pou1

Tuberculosis and latent tuberculosis infection in close contacts of people with pulmonary tuberculosis in low-income and middle-income countries: a systematic review and meta-analysis

Upcoming WHO policy on serologic tests

Commercial Serological Antibody Detection Tests for the Diagnosis of Pulmonary Tuberculosis: A Systematic Review
Karen R. Steinberg,1,2 Maggie Hooper,3 Susan Leal,4,5 Philip C. Hapewill,3 Andrew Ramsay,1 Dick Mantel,4,5 Jane Cunningham,6 Kevin Wedder,7 and Zhikai Pan1

A systematic review of commercial serological antibody detection tests for the diagnosis of extrapulmonary tuberculosis
Karen R. Steinberg, Maggie Hooper, Susan Leal,3 Philip C. Hapewill,3 Andrew Ramsay,1 Dick Mantel,4,5 Jane Cunningham,6 Kevin Wedder,7 and Zhikai Pan1

Performance of Purified Antigens for Serodiagnosis of Pulmonary Tuberculosis: a Meta-Analysis
Karen R. Steinberg,1,2 Nicole Chandwani,1 Megan Green,2 Iye Sakiya,1 Fazum Nabi,4 Philip C. Hapewill,3 Andrew Ramsay,1 Zhikai Pan1,2, and Susan Leal3,5

11
Are these the same or different?

- Traditional, narrative review
- Systematic review
- Overview
- Meta-analysis
- Pooled analysis

Types of review articles

- Meta-analyses
- Systematic reviews
- Individual patient data (IPD) meta-analyses
- Reviews that are not systematic (traditional, narrative reviews)

In practice, not all meta-analyses are conducted as part of systematic reviews

Meta-analyses

Individual patient data (IPD) meta-analyses

Reviews that are not systematic (traditional, narrative reviews)

Systematic reviews

All reviews (also called overviews)

Some definitions

- **Traditional, narrative reviews**, usually written by experts in the field, are qualitative, narrative summaries of evidence on a given topic. Typically, they involve informal and subjective methods to collect and interpret information.
- “A **systematic review** is a review in which there is a comprehensive search for relevant studies on a specific topic, and those identified are then appraised and synthesized according to a predetermined and explicit method.”

Some definitions

• "A meta-analysis is the statistical combination of at least 2 studies to produce a single estimate of the effect of the healthcare intervention under consideration."

• Individual patient data meta-analyses (pooled analyses) involves obtaining raw data on all patients from each of the trials directly and then re-analyzing them.


Narrative vs. Systematic Reviews

<table>
<thead>
<tr>
<th>Component of review</th>
<th>Traditional, narrative reviews</th>
<th>Systematic reviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formulation of the question</td>
<td>Usually address broad questions</td>
<td>Usually address focused questions</td>
</tr>
<tr>
<td>Methodology</td>
<td>Usually not present, or not well-described</td>
<td>Usually described with pre-defined criteria about participants, interventions and outcomes</td>
</tr>
<tr>
<td>Search strategy to identify studies</td>
<td>Usually not described; mostly limited by reviewers’ abilities to retrieve relevant studies; usually not reproducible and prone to selective citation</td>
<td>Clearly described and normally exhaustive; transparent, reproducible and less prone to selective citation</td>
</tr>
<tr>
<td>Quality assessment of identified studies</td>
<td>Usually all identified studies are included without explicit quality assessment</td>
<td>Only high-quality studies are included using pre-stated criteria; if lower-quality studies included, their effects are tested in subgroup analyses</td>
</tr>
<tr>
<td>Data extraction</td>
<td>Methods usually not described</td>
<td>Usually undertaken by more than one reviewer, output pre-tested criteria; attempts often made to obtain missing data from authors of primary studies</td>
</tr>
<tr>
<td>Data synthesis</td>
<td>Qualitative description employing the ‘vote-counting’ approach, where each included study is given equal weight, irrespective of study size and quality</td>
<td>Meta-analysis assigns higher weights to effect measures from more precise studies; pooled, weighted effect measures with confidence limits provide power and precision to results</td>
</tr>
<tr>
<td>Heterogeneity</td>
<td>Usually dealt with in a narrative fashion</td>
<td>Heterogeneity dealt with by graphical and statistical methods; attempts are often made to identify sources of heterogeneity</td>
</tr>
<tr>
<td>Interpreting results</td>
<td>Prone to cumulative systematic biases and personal opinion</td>
<td>Less prone to systematic biases and personal opinion</td>
</tr>
</tbody>
</table>

Elements of a Systematic Review

- Formulate the review question & write a protocol
- Search for and include primary studies
- Assess study quality
- Extract data
- Analyze data
- Interpret results & write a report

All systematic reviews are not meta-analyses!

• “...it is always appropriate and desirable to systematically review a body of data, but it may sometimes be inappropriate, or even misleading, to statistically pool results from separate studies. Indeed, it is our impression that reviewers often find it hard to resist the temptation of combining studies even when such meta-analysis is questionable or clearly inappropriate.”

All reviews are not systematic!

- In 1987, Cynthia Mulrow published an interesting article entitled “The Medical Review Article: State of the Science.”
- She examined 50 review articles published in 4 major general medical journals [Annals of Internal Med; Archives of Internal Med; JAMA; New Engl J Med]
- Findings:
  - 80% addressed a focused review question
  - 2% described the method of locating evidence
  - 2% used explicit criteria for selecting studies for inclusion
  - 2% assessed the quality of the primary studies
  - 6% performed a quantitative analysis


All systematic reviews are not systematic!

Many Reviews Are Systematic but Some Are More Transparent and Completely Reported than Others

The PLoS Medicine Editors

Epidemiology and Reporting Characteristics of Systematic Reviews

All systematic reviews are not systematic!

- 300 SRs were identified (one month)
- Majority (272 [90.7%]) reported in specialty journals
- Most reviews (213 [71.0%]) were categorized as therapeutic, and included a median of 16 studies
- Reviews typically searched a median of three electronic databases and two other sources
- Most (197/295 [66.8%]) reviews reported information about quality assessment, while few (68/294 [23.1%]) reported assessing for publication bias.
- A little over half (161/300 [53.7%]) reported combining their results statistically, of which most (147/161 [91.3%]) assessed for consistency across studies.
- There were large differences between Cochrane reviews and non-Cochrane reviews in the quality of reporting


When can meta-analyses mislead?

- When a meta-analysis is done outside of a systematic review
- When poor quality studies are included or when quality issues are ignored
- When small and inconclusive studies are included
- When inadequate attention is given to heterogeneity
  - Indiscriminate data aggregation can lead to inaccurate conclusions
- When reporting biases are a problem
  - Publication bias
  - Time lag bias
  - Duplicate publication bias
  - Language bias
  - Outcome reporting bias

While almost all trials with “positive” results on antidepressants had been published, trials with “negative” results submitted to the US Food and Drug Administration, with few exceptions, remained either unpublished or were published with the results presented so that they would appear “positive.”

Optimism bias, non-replicated studies, and selective reporting

About 50% of studies on TB diagnostics have some sort of industry involvement — these studies tend to report conclusions that are favorable to the diagnostic (Fontela P et al. PLoS One 2009)
If exposure and disease are not associated

False positive study

Publication Bias

100 studies will be designed

5 studies show false positive results

5 studies will be published

Likely to be meta-analyzed

THE FALSE POSITIVE RESEARCH CYCLE
(Choi, 1998)

![Diagram showing the false positive research cycle with nodes for if exposure and disease are not associated, false positive study, publication bias, 100 studies will be designed, 5 studies show false positive results, 5 studies will be published, and likely to be meta-analyzed.]

Essay

Why Most Published Research Findings Are False
John M. A. Pacheco

Summary
There is increasing concern that most published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and reporting the ratio of true to false relationships among the relationships studied. In each scientific field, a research finding is less likely to be true when the studies conducted in a field are few, when effect sizes are small, when there is a greater number of known, and when there is less evidence of a true relationship, where there is greater flexibility in design, definitions, outcomes, and analytic models, and when there is greater financial and other interest in a particular procedure where more teams are involved in a scientific field or field's statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be true than false. Moreover, for many current scientific fields, claimed research findings are often less accurate estimates of the true relationship because of publication bias. Furthermore, if the study findings are claimed to be significant, the implications of these results for the conduct of interpretation of research findings are unclear due to the uncertainty in their reliability and some coordinates of the research findings. Modeling the framework for false positive findings is a necessary step to improve research findings. Several methodologies have been used (0.11) that the high rate of nonsignificance (lack of confirmation) of research discoveries is a consequence of the conditions, i.e., the rate of true research findings that are confirmed after a single study are assessed as being statistically significant, typically for a p-value less than 0.05. Research is often appropriate and summarized by positive, but unfortunately, there is a widespread notion that medical research articles can be proven that most claimed research findings are false.

It can be proven that most claimed research findings are false.

Research findings should be interpreted based only on positive. Research findings are defined here as any relationship reaching formal statistical significance, e.g., effective intervention, intermediate protection, risk factors, or associations. "Negative" research is also more prevalent in research fields and can vary depending on whether the field is large, the field is large, the field is small, and the field is small. The probability that an observed relationship is true is small, the probability that a false relationship is true is small, and the probability that a true relationship is true is small. The probability that a false relationship is true is small, and the probability that a true relationship is true is small.

PPM (p-value or probability) has been used to quantify the false positive rate (FPR). The PPM is also the complementary probability of this. A PPM of 0.1 would indicate a 10% chance of a false positive result. If a research finding has been claimed to be significant based on achieving formal statistical significance, the power is another probability that is true in the positive predictive value (PPV).

PLoS Med 2005
Empirical Evidence for Selective Reporting of Outcomes in Randomized Trials: Comparison of Protocols to Published Articles

Selection in Reported Epidemiological Risks: An Empirical Assessment

Concontradicted and Initially Stronger Effects in Highly Cited Clinical Research

Why Most Discovered True Associations Are Inflated

Evidence About Inflated Early Effect Sizes
So, you still want to take this course?

“meta-analysis has made and continues to make major contributions to medical research, clinical decision making, and standards of research reportage. However, it is no panacea. Readers need to examine any meta-analyses critically to see whether researchers have overlooked important sources of clinical heterogeneity among the included trials. They should demand evidence that the authors undertook a comprehensive search, avoiding covert duplicate data and unearthing unpublished trials and data. Lastly, readers and researchers alike need to appreciate that not every systematic review should lead to an actual meta-analysis...”

David Naylor. BMJ 1997;315:617-619
Road Map for Systematic Reviews

Define a focused 4-part review question (Patient, Intervention, Comparison and Outcome)

Search electronically or via reference manager. Avoid language restrictions at this stage. Involves a librarian.

Software suggestions: EndNote, Reference Manager, ProCite.

Search on all relevant databases and sources.

Save all citations (titles/abstracts) in a reference manager. Document search strategies that were employed. These citations are ready for first screen (N1).

Reviewers meet and resolve disagreements on citations they do not agree on. The final number (N) selected after this process is ready for second screen (review of full-text articles).

Excluded after second screen.

Get full texts of all articles identified for second screen (N).

Articles considered eligible after full-text review (by two reviewers) is the final set of studies for inclusion (n).

Studies included in the final analysis (n).

Excluded from the final analysis (n).

Paper data extraction forms (one per study).

Keep a log of excluded studies with reasons for exclusion.
Formulating the Review Question & Writing a Protocol
How are these questions different?

- Is polymerase chain reaction (PCR) useful in TB diagnosis?

- In patients suspected to have pulmonary tuberculosis, is PCR more sensitive and specific than culture?

Architecture of a focused question: a 4-part review question

**P** - Who is the patient or what problem is being addressed?

**I** - What is the intervention or exposure?

**C** - What is the comparison group?

**O** - What is the outcome or endpoint?

± study design

Richardson et al. The well-built clinical question: a key to evidence-based decisions. ACP Journal Club 1995;A-12
Architecture of a focused question:
a 4-part review question

- **PICO + study design**

  - **Study designs (domains):**
    - Etiology [cohort, case-control]
    - Therapy [RCT]
    - Prognosis [cohort]
    - Harm [cohort, case-control]
    - Diagnosis [cross-sectional, case-control]
    - Economic [cost-effectiveness analysis, etc.]
Formulation of a therapy question

Intervention Outcome
---
Is IPT effective in treating LTBI?

Patient/problem Intervention
---
In adults with HIV infection, is 9 months of INH effective in reducing the risk of active TB, as compared to placebo?

Outcome + RCTs Comparison

Formulation of an etiology question

Exposure Outcome
---
Is smoking a risk factor for tuberculosis?

Patient Exposure
---
Are people who smoke tobacco regularly at a greater risk of developing pulmonary tuberculosis as compared to those who do not smoke?

Outcome + cohort & case-control studies Comparison
Formulation of a diagnosis question

Is urine LAM antigen detection accurate for TB?

Is urine LAM a more sensitive and specific test in diagnosing active pulmonary TB disease in HIV-infected persons as compared to culture?

+ diagnostic studies [cross-sectional]

How a focused question helps in searching for studies

PICO + STUDY DESIGN FILTER

Studies most likely to address the question
Once a review question is defined

- Search the literature and see if a review has been done already!
  - This is called a “scoping search”
  - Use sources like the Cochrane Library, DARE database
  - Use Clinical Query in PubMed to identify systematic reviews
- If a review has been done, see if there some way you can improve on it
- If a high-quality systematic review already exists, consider an alternative question!

Once you decide to do a review

- Once you decide to do a review, write a short, draft protocol
- Could be 3 – 4 pages long (background, 4-part question (PICO), study designs to be included, and methods)
  - Why?
    - Gets you started!
    - Forces you to read and understand the context
    - Makes you formulate a focused question
    - Makes you plan the search strategy
    - Makes you describe inclusion/exclusion criteria clearly
    - Makes you think about the data you want to collect and the methods you will use to analyze them
Outline of a full protocol

• Cochrane protocol format*:
  ▫ Background
  ▫ Objectives
  ▫ Criteria for considering studies for this review (PICO)
    • Types of studies (study designs)
    • Types of participants
    • Types of interventions
    • Types of outcome measures
  ▫ Search strategy for identification of studies
  ▫ Methods of the review
    • Eligibility
    • Data collection
    • Assessment of methodological quality
    • Data analysis
  ▫ References

*Cochrane Reviewers' Handbook http://www.cochrane.org/index.htm