Introduction to Systematic Reviews

Madhukar Pai, MD, PhD
Assistant Professor
Department of Epidemiology & Biostatistics
McGill University, Montreal, Canada
Email: madhukar.pai@mcgill.ca
Case Study 1: “Egg on their faces: the story of human albumin solution”*

Human albumin solution, a blood product, has been used in the treatment of blood loss and burns since the attack on Pearl Harbour over half a century ago.

In the UK alone, an estimated 100,000 patients are treated with human albumin solution each year, at a cost to the NHS of close to 12 million.

In 1996, the global albumin market was worth £900,000.

But is human albumin administration beneficial?

“Egg on their faces: the story of human albumin solution”

To answer this question a systematic review of controlled trials comparing albumin with crystalloid was conducted by the Cochrane Injuries Group. 30 RCTs including 1419 randomised patients identified. A meta-analysis showed that the risk of death among those treated with albumin was higher than in the comparison groups. The pooled risk ratio was 1.68 (95% CI 1.26, 2.23) The data suggested that for every seventeen critically ill patients treated with albumin there is one extra death.

“Egg on their faces: the story of human albumin solution”

The results were widely reported in the media and stimulated an immediate response from the regulatory agencies, the industry and the medical profession.

The industry launched a “Albumin Support Programme” to resuscitate the ailing US $1.5 billion global albumin market.

- The objective was to disseminate evidence supporting albumin:
  1. the preparation of literature reviews supporting the use of albumin to be sent to leading regulatory authorities
  2. preparation and dissemination of a Cochrane critique dossier
  3. the establishment of a medical advisory panel to write articles supporting the use of albumin.

The industry set aside US $2.2 million for the program.

“Egg on their faces: the story of human albumin solution”

“Despite vigorous attempts by the plasma products industry to limit the impact of the systematic review on albumin sales, the use of albumin declined steeply.

Throughout the UK albumin sales fell by 40%.

The decline in albumin use occurred despite vigorous criticism of the review in the letters pages of the BMJ.

The decline in albumin sales is a clear indication that doctors took into account the evidence presented in the systematic review and that many doctors changed their practice in response.”

"Egg on their faces: the story of human albumin solution"

Figure 1: Albumin Sales in Scotland and Northern Ireland Before and After Publication of Systematic Review on Human Albumin Administration in Critically Ill Patients

Case study 2: “A patient with MI in 1981”

When discussing the discharge of a patient with uncomplicated acute myocardial infarction (MI), the resident asks whether beta-blockers should be administered to prevent further episodes and death.

Consultant asks her to find the “evidence” on this topic and report back during the next grand rounds.

Residents searches MEDLINE and finds 4 clinical trials that report conflicting results:

Case study 2: “A patient with MI in 1981”

Conclusions from 4 RCTs:

- “mortality and hospital readmission rates were not significantly different in the two groups” – Reynolds & Whitlock, 1972
- “until the results of further trials are reported long-term beta blockade is recommended after uncomplicated MI” – Multicenter International Study 1977
- “the trial was designed to detect a 50% reduction in mortality and this was not shown” – Baber, 1980
- “long-term treatment with timolol… reduces mortality and rate of reinfarction” – Norwegian Study Group, 1981

Case study 2: “A patient with MI in 1981”

The resident is confused and seeks Consultant’s help

Consultant advises her to find a review article

Resident finds a traditional review article (Mitchell, BMJ 1981)

“Thus, despite claims that they reduce arrhythmias, cardiac work and infarct size, we still have no clear evidence that beta-blockers improve long-term survival after infarction despite almost 20 years of trials.”

Resident presents these findings in the rounds and patient is discharged without a beta-blocker.

Case study 2: “A patient with MI in 1981”

What conclusions would our resident have reached if she had access to a meta-analysis?

More than 33 RCTs have been conducted on this research question and the evidence from meta-analysis shows:

- Pooled relative risk \(0.80\) (20% reduction in mortality among those who received beta-blockers after AMI)
- Beta-blockade is now the standard of care worldwide

Case study 3: “Same trials, different takes”

- Mammography for breast cancer is an established screening method
- Is screening with mammography justifiable?
- They identified 8 large RCTs on this topic, with over 182,000 women randomized

“Same trials, different takes”

The authors found that no trial data were of high quality
  - Two were of medium quality, and the rest were poor quality or flawed.

When the results of the two medium quality trials were combined, the risk ratio was 1.00 (95% CI 0.96, 1.05)

They concluded that “screening for breast cancer with mammography is unjustified” and “any hope or claim that screening mammography with more modern technologies than applied in these trials will reduce mortality without causing too much harm will have to be tested in large, well-conducted randomised trials...”

**“Same trials, different takes”**

<table>
<thead>
<tr>
<th>Study</th>
<th>Screened Number of deaths/ number of women</th>
<th>Not screened Number of deaths/ number of women</th>
<th>Relative risk* (95% CI)</th>
<th>Weight (%)</th>
<th>Relative risk* (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malmö 1976</td>
<td>2537/21088</td>
<td>2593/21135</td>
<td>0.98 (0.93-1.04)</td>
<td>70.08</td>
<td></td>
</tr>
<tr>
<td>Canada 1980a</td>
<td>418/25214</td>
<td>414/25216</td>
<td>1.01 (0.88-1.16)</td>
<td>11.22</td>
<td></td>
</tr>
<tr>
<td>Canada 1980b</td>
<td>734/19711</td>
<td>690/19694</td>
<td>1.06 (0.96-1.18)</td>
<td>18.70</td>
<td></td>
</tr>
<tr>
<td>Subtotal</td>
<td>3689/66013</td>
<td>3697/66105</td>
<td>1.00 (0.96-1.05)</td>
<td>100.00</td>
<td></td>
</tr>
</tbody>
</table>

Test for heterogeneity: $\chi^2=1.80, df=2$ (p=0.41)
Test for overall effect: $z=0.05$ (p=0.96)

**All-cause mortality in medium-quality screening trials after 13 years**

*Fixed-effects model.

The US Preventive Services Task Force used the same set of 8 trials:

- Recently, a 2001 Cochrane Collaboration review of the same trials concluded that six of the eight trials were "flawed" or of "poor quality" and that the pooled results from the remaining two better trials did not support a benefit from mammography.

- Although the USPSTF was concerned about many (but not all) of the flaws identified in this review, it did not consider the presence of flaws sufficient reason in itself for rejecting trial results.

- The meta-analysis performed for the USPSTF on the most current published data found that the pooled effect size of the combined trials was sizable and statistically significant: the summary relative risk (RR) of breast cancer death among women randomized to screening in seven trials that included women older than 50 was 0.77 (95 percent CI, 0.67-0.89).

- The USPSTF recommends screening mammography, with or without clinical breast examination, every 1-2 years for women aged 40 and older.”
Case study 4: “Passive smoking & lung cancer”

- A topic of great debate and controversy for many years
- First few epidemiologic studies were published in 1981
- Vigorously attacked by the tobacco industry
  - Too small an association
  - Potential bias
  - Potential confounding
  - Lack of biological proof
- Evidence accumulated over the next 2 decades
- It was not until about 10 years ago when several official bodies/agencies concluded that passive smoking is a cause of lung cancer
  - The tobacco industry continues to dispute this claim!!

Case study 4: “Passive smoking & lung cancer”

Hackshaw et al. conducted a very comprehensive systematic review in 1997:

- They identified 37 published studies that reported risk of lung cancer among lifelong non-smoking women according to the husband’s smoking status.
- Their meta-analysis revealed that the overall risk of lung cancer among lifelong non-smoking women was 1.24 times higher when their husbands smoked, as compared to those women whose husbands did not smoke.

Case study 4: “Passive smoking & lung cancer”

Case study 4: “Passive smoking & lung cancer”

What is evidence-based medicine?

The practice of EBM is the integration of

- **individual clinical expertise**
  
  with the

- **best available external clinical evidence**
  
  from systematic research

  and

- **patient’s values and expectations**

http://www.cebm.net/index.asp
What is evidence-based public health?

Evidence-based public health is “the development, implementation, and evaluation of effective programs and policies in public health through application of principles of scientific reasoning including systematic uses of data and program planning models.”

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.
- To keep up to date in Internal Medicine, need to read 17 articles a day, 365 days a year!
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>Grade of Recommendation</th>
<th>Level of Evidence</th>
<th>Therapy/Prevention, Aetiology/Harm</th>
<th>Prognosis</th>
<th>Diagnosis</th>
<th>Economic analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1a</td>
<td>SR (with homogeneity*) of RCTs</td>
<td>SR (with homogeneity*) of inception cohort studies; or a CPG validated on a test set.</td>
<td>SR (with homogeneity*) of Level 1 diagnostic studies; or a CPG validated on a test set.</td>
<td>SR (with homogeneity*) of Level 1 economic studies</td>
</tr>
<tr>
<td>A</td>
<td>1b</td>
<td>Individual RCT (with narrow Confidence Interval*)</td>
<td>Individual inception cohort study with ≥ 80% follow-up</td>
<td>Independent blind comparison of an appropriate spectrum of consecutive patients, all of whom have undergone both the diagnostic test and the reference standard.</td>
<td>Analysis comparing all (critically-validated) alternative outcomes against appropriate cost measurement, and including a sensitivity analysis incorporating clinically sensible variations in important variables.</td>
</tr>
<tr>
<td>1c</td>
<td>All or none⁹</td>
<td>All or none case-series</td>
<td>Absolute SpPins and SnNouts¹⁰</td>
<td>Clearly as good or better⁴⁴, but cheaper. Clearly as bad or worse but more expensive. Clearly better or worse at the same cost.</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>2a</td>
<td>SR (with homogeneity*) of cohort studies</td>
<td>SR (with homogeneity*) of either retrospective cohort studies or untreated control groups in RCTs.</td>
<td>SR (with homogeneity*) of Level ≥2 diagnostic studies</td>
<td>SR (with homogeneity*) of Level ≥2 economic studies</td>
</tr>
<tr>
<td>B</td>
<td>2b</td>
<td>Individual cohort study (including low quality RCT; e.g., &lt;80% follow-up)</td>
<td>Retrospective cohort study or follow-up of untreated control patients in an RCT; or CPG not validated in a test set.</td>
<td>Independent blind comparison but either in non-consecutive patients, or confined to a narrow spectrum of study individuals (or both); all of whom have undergone both the diagnostic test and the reference standard; or a diagnostic CPG not validated in a test set.</td>
<td>Analysis comparing a limited number of alternative outcomes against appropriate cost measurement, and including a sensitivity analysis incorporating clinically sensible variations in important variables.</td>
</tr>
<tr>
<td>2c</td>
<td>&quot;Outcomes&quot; Research</td>
<td>&quot;Outcomes&quot; Research</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3a</td>
<td>SR (with homogeneity*) of case-control studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3b</td>
<td>Individual Case-Control Study</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>4</td>
<td>Case-series (and poor quality cohort and case-control studies⁹⁹)</td>
<td>Case-series (and poor quality prognostic cohort studies )</td>
<td>Reference standard was not applied independently or not applied blindly</td>
<td>Analysis with no sensitivity analysis</td>
</tr>
<tr>
<td>D</td>
<td>5</td>
<td>Expert opinion without explicit critical appraisal, or based on physiology, bench research or &quot;first principles&quot;</td>
<td>Expert opinion without explicit critical appraisal, or based on physiology, bench research or &quot;first principles&quot;</td>
<td>Expert opinion without explicit critical appraisal, or based on physiology, bench research or &quot;first principles&quot;</td>
<td>Expert opinion without explicit critical appraisal, or based on economic theory</td>
</tr>
</tbody>
</table>
Karl Pearson is probably the first medical researcher to use formal techniques to combine data from different studies (1904):

- He synthesized data from several studies on efficacy of typhoid vaccination

His rationale for pooling data:

- “Many of the groups… are far too small to allow of any definite opinion being formed at all, having regard to the size of the probable error involved.”

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. His encouragement, and the endorsement of his views by others, led to the opening of the first Cochrane centre (in Oxford, UK) in 1992 and the founding of The Cochrane Collaboration in 1993.

Source: http://www.cochrane.org/cochrane/archieco.htm
Systematic reviews/meta-analyses indexed in PubMed – 10 years

Search: meta-analysis(MeSH) OR meta-analysis(tw) OR systematic review(tw)
ACP Journal Club
Linking Research to Practice in Internal Medicine

VOLUME 126 • NUMBER 2
March/April 1997

THERAPEUTICS
- Lowering LDL-cholesterol levels reduced fatal coronary events in patients with acute MI and average cholesterol levels... 29
- Amiodarone did not increase mortality or mortality rates in severe heart failure... 30
- Meta-analysis: Mortality is reduced when fibrinolytic therapy is started soon after the onset of MI symptoms... 31
- Hirudin was no more effective than unfractionated heparin for acute MI... 32
- Hirudin reduced death or MI more than heparin at 48 hours but not at 30 days... 33
- Meta-analysis: β-blockers improve function in dilated cardiomyopathy... 34
- Meta-analysis: Thrombolytic therapy increases the risk for early death and intracranial hemorrhage after acute ischemic stroke... 35
- Meta-analysis: Misoprostol reduces NSAID-induced gastrointestinal mucosal injury... 36
- Meta-analysis: Pentoxifylline improves walking in intermittent claudication... 37
- Meta-analysis: Respiratory rehabilitation relieves dyspnea in COPD... 38
- Review: Antiinflammatories are ineffective for acute bronchitis... 39
- Implantable insulin pump improved quality of life in NIDDM... 40
- Valdueño accelerated healing in recurrent herpes genitalis... 41
- Subcutaneous tunneling reduced catheter-related sepsis in critically ill patients... 42

DIAGNOSIS
- γ-dimer levels detected DVT in patients hospitalized for stroke rehabilitation... 43
- CSF proteins 150 and 131 were specific for diagnosing Creutzfeldt-Jakob disease... 44
- Brain protein 14-3-3 was a sensitive test for Creutzfeldt-Jakob disease... 45
- Meta-analysis: Glycosylated hemoglobin levels are useful for diagnosing diabetes... 46

PROGNOSIS
- Ischemic stroke with accompanying atrial fibrillation was associated with reduced survival and functional status... 47

ETIOLOGY
- LDL particle size was smaller in CAD, but other lipid parameters were stronger predictors of CAD... 48
- High glucose level but not LDL particle size was an independent risk factor for MI in men... 49
- Low cholesterol levels were associated with suicide in men... 50

ECONOMICS
- Low-molecular-weight heparin was cost-effective for perioperative prevention of DVT... 51
- Cost-effectiveness of pravastatin in the secondary prevention of CAD in men varied with the risk profile of the patient... 52
- Intensive therapy extended life and was cost-effective for IDEFI... 53

Other Articles Noted
Glossary... 54
Contents... 56
Purpose and Procedure... A-7
Editorial: Transferring evidence from research into practice: 3. Developing evidence-based clinical policy... A-14
Centre for Reviews and Dissemination databases

CRD was established in January 1994, and produces and promotes the use of research based knowledge in health and social care.

DARE – (Database of Abstracts of Reviews of Effects) contains over 4000 abstracts of quality assessed and critically appraised systematic reviews. The database focuses on the effects of interventions used in health and social care.

NHS Economic Evaluation Database (NHS EED) contains over 6000 abstracts of quality assessed economic evaluations. The database aims to assist decision-makers by systematically identifying and describing economic evaluations, appraising their quality and highlighting their relative strengths and weaknesses.

Both DARE and NHS EED include details of abstracts in the process of being written and these can be 'fast-tracked' on request.

The HTA database brings together details of completed and ongoing health technology assessments from around the world. The abstracts in the database are descriptive rather than analytical and do not form critical appraisals of the reports. The database is produced in collaboration with the INAHTA Secretariat, based at SBU, Sweden.
Are textbooks good sources of current evidence?

Not always!
- They are better for background questions than foreground questions
- They are not updated frequently and often lag behind current evidence by many years

Exceptions:
- ACP Medicine [Scientific American Medicine]
- UpToDate
- Clinical Evidence
- Harrison’s Online
- Emedicine (totally online text)
Evidence vs. textbook recommendations

A. Thrombolytic Therapy

<table>
<thead>
<tr>
<th>Year</th>
<th>Cumulative RCTs</th>
<th>Odds Ratio (Log Scale)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960</td>
<td>1 23</td>
<td></td>
</tr>
<tr>
<td>1965</td>
<td>3 149</td>
<td></td>
</tr>
<tr>
<td>1970</td>
<td>7 1793</td>
<td></td>
</tr>
<tr>
<td>1975</td>
<td>10 2544</td>
<td></td>
</tr>
<tr>
<td>1980</td>
<td>23 5767</td>
<td></td>
</tr>
<tr>
<td>1985</td>
<td>27 6125</td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td>70 48154</td>
<td></td>
</tr>
</tbody>
</table>

Textbook/Review Recommendations

<table>
<thead>
<tr>
<th>Routine</th>
<th>Specific</th>
<th>Rare/Never</th>
<th>Experimental</th>
<th>Not Mentioned</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>10</td>
<td></td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>6</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>10</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>4</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>8</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>8</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Antman et al. JAMA 1992
Are textbooks good sources of current evidence?
Systematic reviews are done in different domains

Tuberculosis among Health-Care Workers in Low- and Middle-Income Countries: A Systematic Review

Rajnish Joshi¹², Arthur L. Reingold¹, Dick Menzies³, Madhukar Pai³*

1 Division of Epidemiology, School of Public Health, University of California Berkeley, Berkeley, California, United States of America, 2 Department of Medicine, Mahatma Gandhi Institute of Medical Sciences, Sevagram, Maharashtra, India, 3 Montreal Chest Institute, McGill University, Montreal, Canada

Meta-analysis of “rates”
Systematic reviews are done in different domains


Dick Menzies, MD, MSc; Madhukar Pai, MD, PhD; and George Comstock, MD, DrPH

Meta-analysis of “diagnostic accuracy [diagnosis]”
Chloroquine or amodiaquine combined with sulfadoxine–pyrimethamine for uncomplicated malaria: a systematic review

Jimee Hwang¹, Edward Bitarakwate², Madhukar Pai³, Arthur Reingold³, Philip J. Rosenthal⁴ and Grant Dorsey⁴

1 Department of Internal Medicine, University of California San Francisco, San Francisco, CA, USA
2 Elizabeth Glaser Pediatric AIDS Foundation, Kampala, Uganda
3 Division of Epidemiology, School of Public Health, University of California Berkeley, Berkeley, CA, USA
4 Department of Infectious Diseases, University of California San Francisco, San Francisco, CA, USA
Systematic reviews are done in different domains

Risk of Tuberculosis From Exposure to Tobacco Smoke

A Systematic Review and Meta-analysis

Michael N. Bates, PhD; Asheena Khalakdina, PhD; Madhukar Pai, MD, PhD; Lisa Chang, MPH; Fernanda Lessa, MD, MPH; Kirk R. Smith, PhD

Arch Intern Med. 2007;167:335-342

Meta-analysis of “observational studies [etiology]”
Are these the same or different?

- Traditional, narrative review
- Systematic review
- Overview
- Meta-analysis
- Pooled analysis
Types of review articles

All reviews
(also called overviews)

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

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

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

<table>
<thead>
<tr>
<th>Components of a 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>Method section</td>
<td>Usually not present, or not well-described</td>
<td>Clearly described with pre-stated 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 usually 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, the effects of this 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 onto pre-tested data forms; 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

Road map for systematic reviews

A "road map" for diagnostic reviews

Define a focused diagnostic review question.1 (Patient/Disease, Index test, Reference standard, and Outcomes)

1. **Published databases:** Medline, EMBASE, BIOSIS, Web of Science, Cachexia CENTRAL, MEDITEX, and subset specific databases.
2. **Search strategy:** Use sensitive filters for diagnostic studies.3,5,7,11
3. **Identify appropriate databases and sources of diagnostic studies:**
4. **Search directly or via reference manager software:**
5. **Include a librarian:**
6. **Software suggestions:** EndNote, Reference Manager, ProCite
7. **Need clear inclusion and exclusion criteria

Reviewer 1 screens all titles/abstracts and makes selections for second screen

Software suggestions: EndNote, Reference Manager, ProCite

Reviewer 2 screens all titles/abstracts and makes selections for second screen

Excluded after second screen

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

Articles considered eligible after full-text review (by 2 reviewers) is the final set of studies for inclusion (N)

Studies included in the final analysis (N)

Each article gets a unique ID number

Reviewer 1 extracts data (including quality assessment) from the final selected articles

Reviewer 2 extracts data (including quality assessment) from the final selected articles

Collect outcomes as TP, FP, FN, and TN; or raw ROC data

Software suggestions: Access, Excel

Software suggestions: PRODIGY, PROMETHEUS

Software suggestions: QUOROM, MOOSE as general guides for report writing (acknowledging that they are not meant for diagnostic reviews)

Interpret, discuss results, and write the report

Discuss applicability of results, and limitations of the review

Make recommendations for practice or policy, and research

Se = sensitivity; Sp = specificity; LR = likelihood ratio; DOR = diagnostic odds ratio; ROC = receiver operating characteristic; SROC = summary receiver operating characteristic; TP = true positives; FP = false positives; TN = true negatives; FN = false negatives; TPR = true positive rate; FPR = false positive rate. Superscripts indicate reference numbers.

www.evidence-basedmedicines.com

Road map for diagnostic reviews

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.”

Are all reviews equal?

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

Are all reviews equal?

In 1999, Cynthia Mulrow’s survey was repeated.

This time 158 reviews published in 6 major general medical journals [Annals of Internal Med; JAMA; New Engl J Med; BMJ; Am J Med; J of Int Med]

Findings:
- 34% addressed a focused review question
- 28% described the method of locating evidence
- 14% used explicit criteria for selecting studies for inclusion
- 9% assessed the quality of the primary studies
- 21% performed a quantitative analysis

McAlister et al. The medical review article revisited: has the science improved? Annals Int Med 1999;131:947-51
Why aren’t traditional, narrative reviews good enough?

- Can be subjective, prone to bias and error
- Literature search may be patchy and inadequate
- Selective citation of literature
- No description of the methods used by the review
- Usually uses vote-counting: can be misleading
- Usually not quantitative: can’t pick up small effects
- Readers can’t judge the quality of the review
- Readers can’t replicate or verify the review
- Hard to separate research evidence from anecdotal experiences
- Narrative reviews may disagree with each other
How are systematic reviews better?

- You don’t have to be an “expert” to do one!
- More objective, less prone to bias and error
- Literature search is comprehensive, exhaustive and repeatable
- Clear description of the methods used
- Explicit criteria for choosing studies
- Includes assessment and discussion of quality of primary studies
- Quantitative synthesis avoid vote counting
- Can pick up small effects by pooling data
- Readers can replicate or verify the review
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 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

How to read a systematic review?

- A clearly defined, explicit question
- Comprehensive and systematic search for studies
- Explicit, reproducible strategy for screening and including studies (inclusion/exclusion criteria)
- Assessment of quality of primary studies
- Explicit, reproducible data extraction
- Appropriate analysis and reporting of results
  - Exploration of heterogeneity, publication bias, etc.
- Discussion should consider limitations and strength of evidence
- Interpretation supported by data
- Implications for patient care and future research