Summarizing Evidence I: Systematic Reviews

EBCP Module #15
Outline of Module

- This module will review or introduce the following topics:
  - Types of reviews
  - Locating systematic reviews
  - Selection of studies for systematic review
  - Appraisal of studies in systematic reviews
Objectives of Module

Students who complete this module should be able to:

- Distinguish between systematic and other types of reviews
- Distinguish between a systematic review and a meta-analysis
- Locate systematic reviews
- Identify the steps required to perform systematic reviews
- Explain how these steps can reduce sources of bias in summarizing evidence
- Identify factors that impact the credibility of the review methodology
- Identify factors that impact the quality of the evidence base
- Recognize common standard systems for rating levels of evidence
Evidence Synthesis
In previous modules, you have learned to:
1. Read and make sense of individual clinical studies
2. Evaluate the internal validity of these studies, including identifying the impact of systematic error (bias) and random error (chance)
3. Consider how the results of these studies apply to real-life patients

We have suggested a method to order how you read individual studies:

1. Get your bearings
   - Identify question, design, etc.
2. Evaluate internal validity
   - Bias (information and selection)
   - Confounding
   - Chance
3. Evaluate external validity
Bodies of Evidence

- You may find in clinical practice that there are not single studies but “bodies” of evidence related to a given question.
  - Studies usually do not stand alone – when a body of evidence exists on a topic, relying on a single study is likely to be misleading.
  - Ideally, important research findings should be replicated (repeated) before they are incorporated into routine clinical practice.
  - At times, the existence of multiple studies on the same topic may lead to multiple—and sometimes contradictory—findings about the same question.

- The purpose of this module is to introduce one method used to synthesize evidence across multiple sources.
  - This method is called the systematic review.
Types of Review
There are many possible ways to summarize evidence. Some are better than others.

Consider, for example, the summaries of evidence found in the daily news—what might be called “journalistic” or “narrative” reviews.

These summaries may highlight recent studies and provide some background on previous research.

However, journalistic reviews are intended to draw attention to research, not to evaluate its validity.

They are not sufficient for making medical decisions.
Another common type of evidence summary is the expert review.

An expert review is evidence summarized by a topic expert and often includes his or her own perspectives on research.

- In medical journals, this may take the form of a “Commentary” or “Editorial.”
- You may also find some articles published as “Reviews” that are not systematic and are, effectively, expert review.
Summarizing Evidence: Systematic Review

- Both *journalistic reviews* and *expert reviews* are subject to the same limitations.
  - Consider how easy it is to “cherry pick” evidence to support one side of an argument without adequately representing the other.
  - Surprising, scary, or unexpected findings may be emphasized because they draw in an audience.

- The concept of *selection bias* has been previously used to describe bias in the selection of participants in an *individual* study.
  - However, it also applies to the selection of clinical studies that are evaluated in a review of research evidence.
  - “Cherry picking” evidence in journalistic and expert reviews leads to a type of systematic error analogous to selection bias.
  - The *systematic review* is designed to prevent such bias.
What counts as “systematic”?

- Systematic reviews are distinguished from other forms of review by:
  1. Explicitly stating reviewers’ objectives and methods
  2. The use of transparent methodologies to obtain and assess research evidence
  3. Laying out specific inclusion and exclusion criteria for the articles that will be used for the review

- A key point to keep in mind when reading systematic reviews:
  - Systematic reviews, like all science, should be reproducible.
Systematic Review Vs. Meta-Analysis

- **Systematic Reviews:** Articles that contain rigorous methodologies that are described within the documents. (They often provide a mainly ‘qualitative’ summary of the evidence base on a topic.)

- **Meta – Analysis:** A subset of systematic reviews in which a statistical technique is used to combine data from homogenous studies. The goal is to combine multiple smaller studies into one large study. (Usually provides a ‘quantitative’ summary of the evidence base on a topic, which is not necessarily better than a pure systematic review.)

- **NOTE:** A systematic review will often contain a meta-analysis, but not always. Conversely, a meta-analysis should always be part of a systematic review. You’ll learn more about meta-analyses in Module 16.
Where to Look for Systematic Reviews

- PubMed – Several approaches may work:
  1. Filter ‘Article Types’ (left column) for systematic reviews
     - Always double-check the results though – use the skills gained in this module to review abstracts and determine whether or not a study is truly a systematic review.
  2. Filter ‘Article Types’ (left column) for meta-analysis
     - Similar to the filter for RCTs. Remember, a meta-analysis should be part of a systematic review.
  3. Use the Clinical Queries search tool – Look in the Systematic Reviews column

- Cochrane Library
  - Cochrane Database of Systematic Reviews – Contains systematic reviews written by Cochrane.
  - Database of Abstracts of Reviews of Effects (DARE) – Contains systematic reviews that Cochrane has critically appraised.
Using a Systematic Review

Once you find a Systematic Review that addresses a clinical question of interest, there are two separate and equally important questions to answer:

1. Is the methodology of the review itself credible?
   - Credibility may be decreased by an overly wide or vague question, inappropriate eligibility criteria for articles, inadequate literate search, or failure to appropriately assess individual articles for validity.

2. How confident can we be about the quality of the evidence?
   - Confidence about the body of evidence may be decreased by serious flaws in study designs, high risk of bias, inconsistency between studies, and poor precision.
First Judgment:
Is the review methodology credible?
Evaluating Credibility of Systematic Review Methodology

- Consider the following issues – each addressed in more detail later in this module:
  - **The Question**: Did the review address a sensible and appropriately focused clinical question (or was it too wide, too vague, or so focused as to be of very limited use?)
  - **The Search**: Was the search for relevant data thorough, so that it is unlikely that important evidence was missed?
  - **Choosing Studies to Include**: Were the processes for selection of studies transparent and reproducible?
  - **Appraising the Studies**: Did the reviewers do a good job of evaluating the quality and risk of bias in the individual studies they chose to include in the systematic review?
Steps in Conducting a Systematic Review

A systematic review can be thought to have four basic parts or steps:

1. Defining a question (e.g., by PICO)
2. Conducting a literature search
3. Applying study inclusion criteria to choose studies to include in the review (i.e., screening)
4. Appraising studies

Systematic reviews are written to describe each of these steps in order.

Therefore, thoughtful readers (including you!) should be aware of some of the relevant questions that are raised at each of these steps.
As with all studies, the first step in reading a systematic review is to get your bearings.

Ask: What is the authors’ clinical question?
  - Use PICO or a similar method.

You must also consider whether the clinical question is appropriate.

Does the question meaningfully reflect clinical practice? Is it too broad? Too narrow?

Consider, for example, the following systematic review:

- **P:** Adults (>18 years old) with any form of cancer
- **I:** Chemotherapy
- **C:** No chemotherapy
- **O:** 5-year survival

Is it clinically meaningful to synthesize research on the effectiveness of all types of chemotherapy for all types of cancer on 5-year survival?
  - Probably not.
Part 1: Clinical Question

- It is important, also, to consider the type (or domain) of question being asked in a systematic review.
  - Remember that we generally consider four types of clinical questions: intervention, diagnosis, prognosis, and harm/etiology/risk.

- The type of question relates to the type of studies that should be included in the systematic review.
  - Criteria for type(s) of study design to include has a strong impact on the validity of a systematic reviews.
  - Systematic reviews are limited by the reliability of the studies they summarize.
  - For example, randomized, placebo-controlled clinical trials provide the best evidence for questions about interventions.
Part 2: The Literature Search

- Readers of systematic reviews should be familiar with methods to search available medical research databases using appropriate search terms.

- Search strategies used for systematic reviews should be comprehensive and should focus on the primary literature.
  - Multiple synonyms and search terms to describe each concept should be used by the review authors and reported
  - Searches are often improved by assistance from a medical librarian

- Common resources used for systematic review include:
  - PubMed (which includes the MEDLINE database)
  - EMBASE (sometimes referred to as the “European MEDLINE”)
  - CINAHL (database that includes nursing and allied health literature)
  - Cochrane Central Register of Clinical Trails (a registry of clinical trial reports)
Part 2: Literature Search – Grey Literature

- Additionally, some systematic reviews may evaluate grey literature.
  - Grey literature consists of unpublished studies and reports not available from traditional sources.
  - Grey literature includes unpublished data sets, doctoral theses, meeting abstracts, and data collected by the FDA, pharmaceutical companies, or other industry.

- Including grey literature in systematic reviews helps assure that authors made an exhaustive search for all relevant studies
  - It may help to broaden the scope of included data
  - It may decrease the risk of publication bias (addressed later)
  - Grey literature often has not been peer reviewed, so must be evaluated carefully for internal validity
Part 2: Literature Search Example

Below is an example description of a literature search from a systematic review on the toxicity of vancomycin in organ transplants:

"An experienced health sciences librarian ran extensive literature searches in MEDLINE via Ovid, Cochrane Database of Systematic Reviews via Wiley (CDSR), Database of Abstracts of Reviews of Effects via Wiley (DARE), Cochrane Central Register of Controlled Trials via Wiley (CENTRAL), Embase via Wiley, Web of Science, and ClinicalTrials.gov in June of 2015. Some terms included subject headings and keywords for toxicity, adverse effects, vancomycin, and organ transplants. No filters for language, year, etc. were applied to the searches. Full search strategies from all databases are listed in the appendix."

Part 2: Literature Search

- Systematic reviews should also specify if and how languages, publication dates, and other factors were specified in the search.
  - Ideally, as many languages as possible should be included, and any date limitations should have a logical justification.

- For example:
  - “Where possible, filters were set for studies pertaining to humans but articles written in all languages were included. The search was performed in September 2013. No time limits were set in an attempt to gather all articles published up until the end of September 2013. Once duplicate references were removed the titles and abstracts of the references were screened.”

Part 2: Literature Search

- In summary, readers should consider multiple aspects of literature searching when reading a systematic review:
  - The databases/grey literature accessed
  - Search terms
  - Search limits (including publication dates, languages, etc.)

- You should also take note of who performed the literature search.
  - The Institute of Medicine recommends working with a librarian or other information specialist to plan out your search strategy and to peer-review the final strategy used.
  - Multiple people may independently perform other aspects of systematic reviews in order to reduce error as you will learn in the following slides.

Part 3: Choosing Studies ("screening")

- After an appropriate clinical question is specified and a comprehensive search is performed, authors of systematic reviews must determine which search results will get further review.
  - This step is known as screening.

- Readers of systematic reviews should be attentive to who performs the screening process and the criteria that are used.
Part 3: Choosing Studies (“screening”)

- Study screening is performed based on explicitly stated inclusion and exclusion criteria.

- As a rule, study screening should be performed by more than one independent reviewer.
  - This reduces the risk of “selection bias” in choosing which studies are to be selected for further evaluation by the authors.

- “Blinding” of the screening process by removing access to journal and author names can also decrease bias at this stage.
Part 3: Study Screening Example

Here is an example of how study screening is described in the methods section of a systematic review. This paper reviewed the association of chronic pain and mortality:

“Identified studies were initially filtered with a title search by a single observer (DS) based on the following inclusion criteria:

Study type – observational studies
Participants – community dwelling adults
Exposure – chronic (>3 months) or widespread pain including fibromyalgia
Outcome – mortality
Papers published in English

A review of abstracts and keywords was then undertaken by two reviewers (DS) (RW) before the retrieval of full text articles for further screening. Disagreements were discussed during a consensus meeting with a third reviewer (JM) for final selection of studies to be included in the review.”

Part 3: Criteria for Study Choice / Screening

- Criteria for choice of studies to include in a review should be rational, reproducible, and relate directly to the clinical question.

- The domain or type of clinical question is the starting place for choosing appropriate study designs to include.

  - Examples:
    - Systematic reviews of Intervention questions with a large evidence base available might reasonably limit study type to RCT’s, while if the evidence base is smaller, prospective cohort studies might also be included.
    - Systematic reviews of Prognosis or Etiology type questions might reasonably limit study type to prospective cohort studies (or might include retrospective cohort, case-control, and even cross-sectional studies if the evidence base is limited).

- Our confidence in the overall conclusions of the Systematic Review will depend on the study design and quality (risk of bias) of the individual studies.
Part 4: Study Appraisal

After the clinical question, search strategy, and screening process, the last step of a systematic review is study appraisal.

In systematic reviews, study appraisal should:
- Avoid reviewer bias
- Describe study similarities and differences
- Assess risk of bias in each individual study and overall
- Be reproducible

Different study designs require the use of different tools or checklists for assessing risk of bias
Part 4: Study Appraisal

Aspects of studies that are assessed in systematic reviews are the same as those that you have previously learned to evaluate:

- **Internal validity** – can the results be trusted?
  - Bias
    - Information bias
    - Selection bias
  - Confounding
  - Chance

- **External validity** – can the results be generalized? In what settings and to which populations can they be applied?
Part 4: Study Appraisal

Some factors that might be assessed in qualitatively comparing multiple studies include:

- Factors affecting *internal validity*, such as:
  - Experimental vs. observational study design
  - Selection of study subjects
  - Exposure and outcome measurement
  - Differences in follow-up
  - Control of confounding factors (such as by randomization or statistical adjustment)

- Factors affecting *external validity*, such as:
  - Demographic characteristics of patients enrolled in the studies
Part 4: Study Appraisal

- To help minimize error in assessing studies, systematic reviews should:
  - Have multiple independent reviewers
  - Have a specified process to resolve conflicts between reviewers (such as a “tie-breaking” additional reviewer)
  - Use standard forms for data collection
  - Eliminate identifiers from articles being reviewed
    - This last feature is analogous to “blinding” in other types of studies, such as clinical trials.
    - This process reduces the possibility that particular authors or journals might be given preferential treatment by reviewers.
Below is a table from a systematic review on the relationship between chronic pain and mortality. It shows the characteristics of included studies that were reviewed:

<table>
<thead>
<tr>
<th>Study</th>
<th>n</th>
<th>Age</th>
<th>% female</th>
<th>Location</th>
<th>Follow-up</th>
<th>Pain phenotype</th>
<th>All-cause mortality</th>
<th>Cause specific mortality (adjusted results)</th>
<th>Adjusted for</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macfarlane, G.J. et al. (2001)</td>
<td>6569</td>
<td>18-85</td>
<td>58</td>
<td>NW England</td>
<td>8 yrs</td>
<td>Widespread pain ACR (1990) criteria</td>
<td>MRR (95% CI) 1.31 (1.05–1.65)</td>
<td>MRR (95% CI) Cancer 2.07 (1.37–3.13) Cardiovascular disease 1.12 (0.78–1.61) Respiratory disease 1.01 (0.57–1.79) Other diseases 0.91 (0.45–1.85) All external causes 5.21 (0.64–28.78)</td>
<td>age, sex, study location</td>
</tr>
<tr>
<td>Macfarlane, G.J. et al. (2007)</td>
<td>7182</td>
<td>30 and over</td>
<td>54</td>
<td>Finland</td>
<td>14–16 yrs</td>
<td>Widespread pain - In at least 4 sites (face validity with ACR (1990) criteria)</td>
<td>MRR (95% CI) 0.86 (0.74–1.00)</td>
<td>MRR (95% CI) Cardiovascular disease 0.83 (0.68–1.02) Cancer 0.64 (0.46–0.91) Respiratory diseases 0.89 (0.54–1.49) Other disease related 1.39 (0.88–2.19) Non disease related 1.39 (0.75–2.58)</td>
<td>age, gender, education, physical work stress, mental work stress, alcohol consumption, tobacco smoking, BMI</td>
</tr>
<tr>
<td>Andersson, H.I. (2009)</td>
<td>1609</td>
<td>25–74</td>
<td>50</td>
<td>Sweden</td>
<td>14 yrs</td>
<td>Widespread pain - In more than four pain locations including upper and lower body and axial pain (to get close to ACR criteria)</td>
<td>MRR (95% CI) 1.95 (3.26–3.03) Adjusted 1.09 (0.62–1.90)</td>
<td>MRR (95% CI) Cardiovascular disease 2.17 (1.12–4.21) Cancer 1.15 (0.52–2.53) Other 1.18 (0.47–2.99)</td>
<td>Age, sex, living alone, contact with friends, club membership, chronic disease, smoking, physical activity, perception of stress, BMI, insomnia (cause specific results adjusted for age and sex)</td>
</tr>
<tr>
<td>McBeth, J. et al. (2009)</td>
<td>4515</td>
<td>16 and over</td>
<td>51.6</td>
<td>NW England</td>
<td>8.2 yrs</td>
<td>Widespread pain ACR (1990) criteria. Number of pain sites</td>
<td>MRR (95% CI) 2.4 (1.9–2.9) Adjusted 1.3 (1.1–1.5)</td>
<td>MRR (95% CI) Cancer 1.6 (1.3–2.6) Cardiovascular disease 1.3 (0.99–1.60) Respiratory disease 1.0 (0.7–1.6) All external causes 0.6 (0.3–1.8) Other 0.8 (0.5–1.4)</td>
<td>age, sex, practice, ethnic group, Townsendi score of deprivation</td>
</tr>
<tr>
<td>Sjogren P. et al. (2010)</td>
<td>2242</td>
<td>16 and over</td>
<td>51.3</td>
<td>Denmark</td>
<td>8 years</td>
<td>Chronic pain (6 months or more)</td>
<td>MRR (95% CI) 1.21 (1.02–1.44)</td>
<td>MRR (95% CI) Cancer 1.6 (1.3–2.6) Cardiovascular disease 1.3 (0.99–1.60) Respiratory disease 1.0 (0.7–1.6) All external causes 0.6 (0.3–1.8) Other 0.8 (0.5–1.4)</td>
<td>age, sex, education, marital status, BMI, smoking, antidepressant use, anxiolytic use, self-reported circulatory diseases, infectious or parasitic diseases, diabetes and mental disorders</td>
</tr>
</tbody>
</table>

Second Judgment: How good is the evidence base?
How good is the evidence?

Once you have decided that the methodology of the review is credible, the next question is whether the body of evidence currently available on the topic is strong enough to use for clinical decision-making.

The authors of the review should address this issue, and often will use a standard grading system.

Multiple organizations have grading systems for “levels of evidence”:

- **GRADE (Grading of Recommendations Assessment, Development and Evaluation)**: high, moderate, low, very low
- **U.S. Preventive Services Task Force**: A, B, C, D, I, and levels of certainty about benefit of high, moderate, and low
- **Oxford Centre for Evidence-Based Medicine**: 1a, 1b, 1c, 2a, 2b, 2c, 3a, 3b, 4, 5
How good is the evidence?

- It is quite possible to have a credible review that concludes the quality of the evidence on a topic is poor.

- Or, a poor quality review on a topic for which the evidence is actually strong
  - but you should avoid using poor quality reviews, due to the risk of misleading results

- Conversely, just because the quality of the body of evidence on a topic is poor, it does not mean that the systematic review itself is not valid! These are separate issues and require separate judgements.
How good is the evidence?

- Your level of trust in the body of evidence should be based on:
  - *Type of study design* – remember the evidence pyramid! Experimental designs are stronger than observational designs; prospective designs are stronger than retrospective designs
  - *Risk of bias in individual studies* (review authors may use checklists or standardized rating tools)
  - *Precision of the results*, which is closely related to sample sizes, and which can be evaluated by looking at the confidence intervals around the estimate of effect size
  - *Consistency of results* across multiple studies, evaluated by measures of heterogeneity and forest plots (addressed further in Module 16)
  - Likelihood of *publication bias* (addressed further in Module 16)
  - Whether the populations, interventions, and outcomes measured by the studies are similar to your patient
PRISMA

- Detailed guidelines for the conduct and reporting of systematic reviews are available at: prisma-statement.org

- These guidelines describe many of the issues raised in this module, as well as other aspects of the systematic review process.

- Additional resources related to conducting and reviewing systematic reviews can be found here http://guides.lib.uiowa.edu/systematicreviews
Key Points for Module 15

- When searching for evidence on a clinical question, Systematic Reviews are a preferred source of information for clinicians because they reflect the whole body of evidence on the topic.
  - Remember, systematic reviews sit at the top of the “evidence pyramid”

- Systematic reviews are distinguished from other reviews by:
  - Explicitly stating reviewers’ objectives and methods
  - Identifying rational inclusion and exclusion criteria for selecting articles to include in the review
  - Using transparent methodologies to obtain and assess the quality of research evidence.

- Systematic reviews require four basic steps:
  1. Defining a question
  2. Conducting a literature search
  3. Applying study inclusion criteria to choose articles (i.e., screening)
  4. Appraising studies

- Appraisal of systematic reviews requires judgments about two separate issues:
  1. The rigor of the review methodology
  2. The quality of the evidence base itself – which in turn impacts our confidence about applying the results of the systematic review to patient care
Please complete Module 15 and 16 prior to taking online quiz.

Thank you!

Module Design Team: Matthew Rysavy, Theresa Hegmann, Amy Blevins
Review date: August 2017
Key References


