CHAPTER 2:

INTRODUCTION TO SYSTEMATIC REVIEWS

Kelly Rushton and Owen Price

CHAPTER OVERVIEW

Health professionals must make sure that the treatments they provide are safe and have been proven to work. This means they must be able to access and understand all the available research relating to treatments that they use. There can be hundreds of research studies conducted on each treatment, so it is often not practical for busy health professionals to gather all of the evidence themselves.

Deciding on a treatment on the basis of a single study or a select number of studies is not recommended. Individual studies may have been poorly conducted and have misleading results, or the findings of the study may conflict with other studies that were not accessed by the health professional.

This is why it is important that researchers gather all of the available evidence relating to each treatment, analyse and combine that evidence, and make it available to healthcare professionals in a comprehensible way. This allows professionals, and their patients, to make accurate judgements about a treatment's risks and benefits.

The method used by researchers to ensure that this process of identifying, analysing and combining the findings of multiple studies is sufficiently rigorous, is called a systematic review. This chapter will detail the steps involved in a well-conducted systematic review.

LEARNING OBJECTIVES

By the end of this chapter you should be able to:

1. Understand what systematic reviews are
2. Understand the stages involved in undertaking a systematic review
3. Understand why systematic reviews are important

INTRODUCTION

Traditionally, systematic reviews have combined evidence from treatment trials and have answered 'yes/no' questions such as 'Does this treatment work?', 'Is it safe?' This involves combining 'quantitative' data (data related to numbers) across a large number of participants and studies and calculating what the average benefit (or risk) of a treatment might be. More recently, systematic reviews have also combined findings from 'qualitative' studies. Qualitative studies explore people's views, experiences and beliefs in an in-depth way, using, for example, individual or group interviews. This type of evidence is important. For example, understanding the way service users feel about a service or treatment can provide important indications of how likely they are to continue to use it. Systematic reviews based on qualitative data can answer important questions about how, why and when a service or treatment might work, rather than 'does it work.' The information this provides can be used to change treatments and services to better meet the needs of service users. Again, a more complete understanding of views and experiences can be generated from combining the findings of many qualitative studies than from relying on the findings of one study alone.

Conducting a review

Irrespective of whether you are conducting a systematic review involving qualitative or quantitative evidence, there are a series of steps that must be followed to ensure that your review is properly conducted. These are as follows:

Defining the question

(what is not already known and what will the review provide an answer to?).

Data synthesis

(the process used to combine and summarise all of the available data).

1**2****3****4****5**

Searching the literature

(how and where studies will be identified).

Data extraction

(how results will be transferred from the original studies to a review spreadsheet, so that the results of all the included studies can be viewed together).

Quality assessment

(to judge how well the original studies were conducted, and to identify any particular strengths or weaknesses that might have influenced their results).

The following sections will describe these steps in more detail.

Importantly, how each of these stages is conducted should always be agreed and written up in a **review protocol** before starting the systematic review. The review protocol should provide detail on all the stages of the proposed review. This helps to prevent researchers making on-the-spot decisions. It reduces error or 'bias' in the findings and helps others to judge how well the review was completed i.e. did it address its original aims and objectives and follow the right methods?

Defining the review question

Developing a good review question is perhaps the most important step in conducting a systematic review. The question will inform the search methods, guide researchers in deciding which studies should be included and excluded and determine whether or not the review produces findings that are meaningful and useful to health services and their users.

It is really important that the review question is well-constructed. The review question must have potential to generate new knowledge and understanding. Put simply, is the review worth doing? Is the question worth answering?

The review question must be precise enough to ensure that the review can be completed. Systematic reviews are normally undertaken by a team of people, but can sometimes still take one or two years to complete. A review question that would need hundreds of thousands of studies to answer it would not be feasible to complete.

There are some easy ways to ensure that a proper review question is developed. 'Does it work' (or effectiveness) questions use a technique called **PICO** (Population, Intervention, Comparison, Outcomes). This ensures the following categories are specified in the review question:

- **Population:** the patient or service user group that the review is interested in e.g. children, working-age adults, people living with psychosis
- **Intervention:** the treatment or therapy that the review will evaluate e.g. counselling, anti-psychotic medicines, care-planning
- **Comparison:** the treatment or therapy that the intervention will be compared with e.g. education, usual care or 'no treatment'
- **Outcomes:** the outcomes that will be used to evaluate the intervention's effect e.g. mental health symptoms, quality of life, patient satisfaction

Figure 5 provides an example of how these four categories can be used to develop a focused research question.

Figure 5 Developing a review question using PICO

However, not all review questions will be addressing 'does it work' or 'effectiveness' questions. Reviews that are interested in understanding people's experiences will be searching for qualitative evidence. These types of review use a different 'PICO' technique to structure their questions: PICO (Population, Interest, Context). This ensures that the following categories are specified in the review question:

- **Population:** the patient or service user group you are interested in e.g. adults with bi-polar disorder
- **Interest:** the activity or issue you are interested in e.g. care-planning
- **Context:** the setting of interest e.g. community mental health teams

Figure 6 provides an example of how these three categories can be used to develop a focused research question for a review of qualitative evidence.

Figure 6 Developing a review question using PICO (review of qualitative evidence)

Searching the literature

Once you have your question it is time to see what studies are already out there. To search the literature in a rigorous way you must:

- Know which electronic databases you are going to search (i.e. identify your **data sources**)
- Have a systematic method for searching the databases to make sure no relevant studies are missed (i.e. develop a **search strategy**)

Searching databases will generate many studies that are potentially relevant to your review question. You will need to sift through these studies to identify those that are and discard those that aren't. To ensure that this is done objectively, without bias, each study is assessed against pre-agreed criteria. These are called the review's **inclusion and exclusion criteria**.

Remember, the arrangements for all of these processes should be pre-agreed and written up in your review protocol before your systematic review begins. Changing the protocol after the review begins is possible but is not desirable.

Selecting the data sources – where to search

Systematic reviews are normally conducted by searching online databases. These databases store nearly all of the research conducted in healthcare and their records stretch back many decades. Figure 7 shows some of the databases that are typically used for systematic reviews. You will notice that some of these are quite general and keep records related to many different areas of health research (e.g. MEDline). Others are more subject-specific. PsycInfo, for example, only includes records of mental health and psychological research. Which databases you choose will depend partly on your review question. As a general rule, it is desirable to include all the databases that you think might hold relevant records. Although this will make your review more time-consuming, it will reduce the likelihood of you missing relevant studies and biasing your review.

All of the databases in Figure 7 can be accessed via a simple Google search but they require expensive subscriptions. Most people access them using a University or hospital subscription, so you may need to ask these organisations for a username and password. Although using these databases is not difficult once you've learnt how to use them, they can seem fiddly and frustrating at first. Each has its own quirks in terms of how they work, so it is advisable to enlist the help of someone who is experienced in using them. Hospital and/or University librarians are often happy to book you in for an individual session on request.

Figure 7 Examples of electronic databases

Developing a search strategy

Developing an effective search strategy is vital, because it is this which ensures all relevant studies are included in the review. The term 'search strategy' refers to the terms that are entered into the electronic databases to retrieve studies. It also refers to how these terms are combined to ensure that only studies relevant to the review question are generated by the databases you use. This really is important - if your search generates 20,000 results, you might not have the time check them for relevance. The aim is to capture all relevant results without generating lots of irrelevant results.

A useful technique is to structure your search terms around the PICO categories of your research question. Using the PICO question in Figure 5 (In working-age adults with anxiety, is Cognitive Behavioural Therapy more effective than Citalopram in reducing anxiety symptoms?), you can see that for each PICO component (Population, Intervention, Comparison, Outcome) there are many possible ways of referring to these. For example, Cognitive Behavioural Therapy is often referred to by the acronym 'CBT' and Citalopram can be referred to by its trade name 'Cipramil,' or a more general term such as 'medication.' Of course, different research studies will opt for different terms, so it is important that your search strategy can capture these differences. Figure 8 shows how search terms can be organised around the PICO categories (Population, Intervention, Comparison, Outcomes) using all possible term variants for each category.

Of course, inputting all your search terms into the databases would simply generate a long list of the studies relevant to each term. Really, you want the database to only generate a list of studies that include all of the terms from all your PICO categories (e.g. Population, Intervention, Comparison, Outcome). This will minimise the number of results your search generates, without missing relevant studies. To do this, you have to let the database know which terms are variants within each PICO category, and how the different categories should be combined to generate the best set of results. This is achieved using 'boolean operators.' Boolean operators are simply the words 'AND' or 'OR.' You can enter these after each search word to let the database know whether the word you have used is a variant of other terms in the same category (by using 'OR') or whether you are combining different categories of search terms (by using 'AND.'). Look again at Figure 8 to see how AND and OR have been used to distinguish search term variants and PICO categories.

Figure 8 Search terms generated using PICO technique

Inclusion and exclusion criteria

Once all your databases have been searched, you will have a long list of studies that are potentially relevant. These need to be sifted to check their relevance and to decide whether or not they will be included in the review. To ensure that this process of sifting and checking is unbiased, it is important that pre-agreed criteria are followed. These criteria are called your inclusion and exclusion criteria. To minimise bias, it is important that two researchers independently check each study against these criteria and compare whether or not they think it should be kept in or out of the review.

The easiest way to develop inclusion and exclusion criteria for a review is to use the PICO framework. Table 1 provides an example of inclusion and exclusion criteria developed for the example review question in Figure 5.

Table 1 Inclusion and exclusion criteria generated using PICO technique

PICO component	Inclusion criteria	Exclusion criteria
Population	Participants are working age adults (18-65 years)	Participants are older adults (65+) Participants are children and/or adolescents (<18)
Intervention	Intervention is Cognitive Behavioural Therapy-based	Intervention is a pharmacological treatment Intervention is a non-Cognitive Behaviour Therapy-based psychological therapy
Comparison	Comparison treatment is Citalopram	Comparison treatment is an anti-depressant but not Citalopram Comparison treatment is a psychological therapy
Outcome	Anxiety symptoms are measured as a criterion for intervention success	Outcomes measured relate to depression, no anxiety measure is included

Data extraction

The term 'data extraction' refers to the process of transferring relevant findings from the included studies to a working spreadsheet. This enables the data from all the included studies to be viewed together. Your data extraction spreadsheet might include columns for findings that answer your review question (e.g. study results) as well as columns for summarising important information on study context or quality (for example, the country where the study was conducted, study design, or whether or not the study was conducted in a 'real world' setting). The spreadsheet will have individual rows for each included study.

Quality assessment

Assessing the quality of each study included in a systematic review is important. No research study is perfect. Individual studies may be prone to error or have limitations that influence the findings. If your review does not account for these biases, the results of your review may also be biased.

Broadly speaking, quality assessment draws the reader's attention to important strengths or weaknesses in the research evidence. Sometimes it is used to apportion more or less weight to individual studies in a review, depending upon whether they are judged to be of a higher or lesser quality.

The way in which quality is assessed in a systematic review often depends upon the subject matter and the type of studies being reviewed. Quality checklists can provide a useful way of assessing all of the evidence in a methodical and standardised way. There is not one single accepted tool for quality assessment, but a variety of published quality assessment guidelines and tools are available that might help you in this process.

As a general rule, quality assessment will usually consider:

- How each study was designed and conducted
- How appropriate this design was to fulfil the study's objectives
- How well the intervention was delivered
- How data were measured, collected, reported and analysed
- How well the findings were interpreted
- How relevant or generalizable these results might be to other people outside of the study

Data synthesis method

The term 'data synthesis' refers to the method that is used to combine findings across all of the studies in your review.

The techniques that are used to combine qualitative and quantitative research evidence are very different. For quantitative systematic reviews (those which aim to answer 'does it work' questions), a statistical technique called 'meta-analysis' is often used. Meta-analysis usually combines data from randomised controlled trials that have been carried out to evaluate treatment effectiveness. Numerical data from many different trials are combined, to explore the average benefit of a particular treatment. An example of the type of data used in a meta-analysis might be service users' scores on a depression scale before and after anti-depressant treatment. Meta-analysis is very important because it enables researchers to assess a treatment's benefit in a much larger group of people than could be achieved in a single trial. Broadly speaking, the larger the number of participants, the more confidence you can have in your estimation of a result. Of course, the relevance of your result depends very much on the type of data you combine. Trials conducted in very different groups of people, or with very different interventions, may not be able to be grouped together.

Qualitative systematic reviews group together data, or text, that summarises people's views and experiences. There are many different techniques that can be used to combine this type of data (Barnett-Page and Thomas, 2009), although most share two key stages.

The first stage seeks to summarise the findings of the included studies, by organising them into common themes (often referred to as descriptive themes). The second stage looks more closely at these themes, to try to understand how they relate to each other. This second stage generates 'analytical themes.' Analytical themes provide a higher level of analysis. They extend our understanding beyond the findings of the original studies, and can be a useful way of examining the combined evidence on a topic.

PPI stories from EQUIP

Next Lauren discusses her experiences of being involved in a systematic review.

LAUREN'S STORY

I joined the EQUIP programme at its inception as a participant on a research methods training course for service users and carers. One of the first things we had to do was to decide how we were going to measure service user 'involvement' in care planning. We conducted a systematic review to identify all the research that had already been published on this topic, so that we could understand all the different ways that involvement had been measured in the past. Reviewing was one of the skills that I had learnt on my research methods course.

Taking part in a systematic review can be daunting, but there are many ways that PPI representatives can contribute. With the right training, knowledge and education work we can work as equals and even lead on such work.

We met as a group to plan the review together and to decide exactly what we needed to find out. My first job was to help with the literature search. I helped to develop search terms for the library database searches.

Then I agreed with another researcher which studies should be included in our review and which should be left out. This was a long process but more manageable than I expected. Information, guidance and attention to detail made the task possible. I enjoyed this piece of work and was proud of my emerging skills in research. It felt good to have my capabilities recognised.

Next, I had another important role to play - I helped to shape the way that our data was extracted and synthesised. In our review we looked at all the different ways that involvement could be measured. We scored each one according to its quality i.e. how well it would measure involvement from a researcher's perspective, but we also scored each one according to how easy or hard a service user might find them to complete.

I worked with our service user and carer advisory group (SUCAG) to come up with a 'wish list' for a good involvement measure. Our literature review had found a number of ways of measuring service users' involvement in care decisions, but none of them were acceptable to people with lived experience. PPI involvement had shown us that we needed to develop a new measure!

Like me, you might be asked to shape a review to make sure its results are meaningful and relevant to others. It was great to be able to make sure service users' views were included in our review, and to help write up our findings.

REFLECTIVE EXERCISE

- **Why might you carry out a systematic review?**
- **What are the main stages involved in conducting a systematic review?**
- **How can the quality of studies in a systematic review be assessed?**

ALLIED PAPERS

- 1) Bee, P., Price, O., Baker, J., and Lovell, K. (2015) Systematic synthesis of barriers and facilitators to service user-led care planning. *The British Journal of Psychiatry* 207:2, 104-114; DOI: 10.1192/bjp.bp.114.152447
- 2) Price, O., Baker, J., Bee, P., and Lovell, K. (2015) Learning and performance outcomes of mental health staff training in de-escalation techniques for the management of violence and aggression. *The British Journal of Psychiatry* 206:6, 447-455; DOI: 10.1192/bjp.bp.114.144576

REFERENCES AND FURTHER READING

Barnett-Page, E. and Thomas, J. (2009) *Methods for Research Synthesis* (Internet). Available from: <http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=188>