

Module 7: Collecting data from relevant studies

This module covers ways in which we collect the data we need for the review from the included studies. You will work through it in a slightly different way from other modules. At points in the module, you are asked to move to other modules so that you understand the type of information you want as well as the mechanics of collecting it. You will find links to modules 8, 9 and 10.

Learning objectives

- Be able to decide what data to collect from studies
- Devise a method and appropriate form for collecting data from included studies
- Be aware of ways to minimise errors and bias in collecting data
- What to do about obtaining and reporting missing data

Relevant sections of the *Cochrane Handbook for Systematic Reviews of Intervention*

- Chapter 7: Selecting studies and collecting data

Where does this go in a Cochrane review?

- You will need to include some description of your methods for collecting data in the Methods section of your protocol and review.

Collecting data you want, in a format you want it

Now that we have identified studies we are going to consider for inclusion in the review, we need to work out how we are going to use them and what information we need to extract from the studies to include in the review. Like all other components of the review we try to do this in a systematic way. At first it may seem odd to have a whole module devoted to the issue of collecting data from the studies you've found. But you've somehow got to manage the process of getting information about a study from a paper publication (or more than one paper, or a letter or a phone call...), organise it and then present it in your review. If you don't have a system for doing this, it's easy to forget where information came from, just collect what is easiest to get out of papers, or make mistakes.

You need a system for collecting the information you want

In addition to collecting information from the studies, you will need a system for collecting information about the process you follow, for example the number of studies you find, how many of those you include in your review and so on.

What information do we need to collect?

There are four main types of information or data we need for the review

There are four main types or categories of information we need to collect from studies considered for inclusion in our review:

- How the study measures up against our eligibility criteria and whether it should be included in the review (and if not, why not)
- Descriptive information about the study to complete the table of 'Characteristics of included studies' or the table of 'Characteristics of excluded studies' in RevMan
- Information about the quality of the study
- Information about the results of the study in the form of data to be used in your review

Why bother with a specific form?



Read [Section 7.5](#) of the Handbook

You may be thinking that taking the time and effort to design and test a structured form to extract data from studies is not worth it when you could just sit down with your included studies and enter the information you need straight into RevMan. Experience suggests that it is worth taking time over this step and Section 7.5 of the *Cochrane Handbook for Systematic Reviews of Interventions* gives you some good reasons why we bother with data collection forms. You should read that now.

In summary we need data collection forms as they:

- Specifically mirror your review question and so allow you to reassess all important components of your question and ensure you extract the relevant data relevant
- Document the history of the process of taking the information from the study into the review and so allow you to backtrack and identify any errors or assumptions you have made in your review
- Record the information you need to generate the results of your review, without any additional, unnecessary information. This will allow more efficient data entry into RevMan and highlight any gaps in the data
- Allow more than one reviewer to extract data from a study and compare the results of their work, increasing the reliability of the data collected for the review. This is an important step in reducing bias.

Format of data collection forms

There are several different approaches to the format of your data collection form and no single style is necessarily the best. Some reviewers prefer an open ended form with a few headings for structure and lots of space to record information in flexible formats. Others prefer to list every item they wish to extract in a highly structured, checklist type approach.

The two examples at the end of this module will give you some idea of the variation in data extraction forms. You should design your form to suit your preferred style of working.

Paper versus electronic forms



Read [Section 7.5.2](#) of the Handbook

Section 7.5.2 of the *Cochrane Handbook for Systematic Reviews of Interventions* devotes a little time to discussion of the benefits of paper versus electronic data collection forms. On the surface it may appear that an electronic form where data can be transferred directly from the form into RevMan would save time, however, at this stage that can only occur for transferring numerical data in spreadsheet form.

If however you prefer working with electronic media you may prefer to extract data into an electronic form and enter into RevMan from there. To date, most reviewers have used paper data collection forms.

Some review groups may have a template for a data collection form you can modify to your particular review but in many cases individual reviews are so specific that a form needs to be designed to match that review.



Read [Section 7.5.3](#) of the Handbook

As discussed already there are some key components to data collection forms and these are outlined in Section 7.5.3 of the *Cochrane Handbook for Systematic Reviews of Interventions*, which you should read before proceeding. Once you have read this return to this point and progress through each section.

Part One of a Data Collection Form: Review, Reviewer and Study Information

Even if you and your co-reviewers are working on one review at the moment, it is possible in the future you will do more, or your review may be passed on to someone else in your review group for comment or updating. Because of this it is important that it is clear which review this form relates to, as well as which reviewer and which study. The following list of suggestions for data fields may help you in designing Part 1 of your form.

- Title of review
- Review ID (as given by your review group when you register your review)
- Reviewer
- Version or date of designing form
- Date of completing form
- Study ID (to match how you have coded it in revman)
- Year of publication of study



Activity: design Part One of your form

Now have a go at designing Part 1 of your data collection form. There may be more information specific to your review to include and you may want to exclude some of the above list.

Part Two of a Data Collection Form: Study Eligibility

Complete Module 8 before you progress further.

Before we design this part of the data collection form we need to understand the process of selecting studies for you review. Skip to Module 8 and complete it, then return to this point to design Part 2 of your form.

To make decisions about including or excluding studies, as we discussed in Module 8, most people use some kind of eligibility form. These forms help you structure the way you go about judging whether a study is eligible for your review. They also help you to find out at a later date why you reached certain judgements.

Example of a study eligibility form

You will find an example of an eligibility form at the end of this module.

You'll see that it collects some information about the study ID (useful in case the form later gets detached from the study report), and then asks a series of questions about the study design, participants, interventions and outcomes. These will come directly from the inclusion criteria you set up earlier and will be a way of seeing if each study matches your criteria.

Try to design an eligibility form for your review. Use the table you generated in Module 5 as your guide.

One final tip: write on your eligibility form whereabouts in the report the relevant bit of information came from. You could also write on the report of the study which part of the form a particular section helps with. It's much easier to come back to this and find out why you came to a decision.



Read [Section 7.3](#) of the *Cochrane Handbook for Systematic Reviews of Interventions*



Find a 'Characteristics of included studies' table on the *Cochrane Library* and familiarise yourself with it

Part Three of a Data Collection Form: Qualitative Information about the Study - The table of 'Characteristics of included studies'

Section 7.3 of the *Cochrane Handbook for Systematic Reviews of Interventions* outlines this part of the data collection form, but in summary, the information we need to take from the studies to complete this part of the review is the description of the study, in particular its participants, interventions and outcomes. This information will be used to complete the table of 'Characteristics of included studies', which has a section also for methods, but it may be best to leave this until after the quality assessment is completed (we will cover that in Part Four). The table of 'Characteristics of included studies' is an important part of a systematic review as it is here that the reader can find detailed information about the included study and interpret the results of the review. It may also provide information about why the results of some included studies differ from others.

It may be worthwhile familiarising yourself with a table of 'Characteristics of included studies' by finding a review on *The Cochrane Library* you are interested in and going to the "Characteristics of included studies" section, then follow the link to the Table.

The types of information needed to complete this part of the review are well described in the *Cochrane Handbook for Systematic Reviews of Interventions*, but some hints to make things easier when data are entered later into RevMan are

- Be consistent in the order and style you use to describe the information (for example when extracting information about participants, start with sample size, then diagnostic description, then demographic information etc). This will make it easier to complete the Table of Included Studies, prevent you from overlooking information and make reading of the review easier.
- Highlight any missing information as unclear or not described to make it clear to the reader of your review that the information was not included in the description of the study, not that you forgot to extract it
- Record any additional, study specific information (for example that the report of the study was translated from a language other than English, or that it was a duplicate publication) in the notes section.

Skip to Module 9: "Assessing the Quality of Studies" before advancing further.

Part Four of a Data Collection Form: Study Quality

Before we discuss the design and use of a data extraction form for recording information about the quality of the included studies, we first need to understand some principles and issues surrounding the assessment of study quality and how it relates to Systematic Reviews. Before progressing further, complete Module 9: Assessing the Quality of Studies, then return to here.



Think about which method of quality assessment you want to use.

Now that you have a grasp of the process of assessing the quality or validity of your included trials, we need to think about how you are going to do this in your review.

Think about the method you wish to use, describing set criteria such as allocation concealment, blinding and withdrawals. Look at your review group's module on *The Cochrane Library* and see if they have a preferred method.



Read [Section 8.5](#) of the Cochrane Handbook for Systematic Reviews of Interventions

Once you know how you are going to do your validity assessment, you can design this component of your data collection form. If you are assessing whether a study met certain key criteria you need to list them, using either a checklist or categories of the 'Risk of bias' table as outlined in Section 8.5 of the Cochrane Handbook for Systematic Reviews of Interventions. Some examples of criteria you may wish to include are:

- Selection Bias (Was the study randomised? Was allocation concealment adequate?)
- Performance Bias (Were participants and care providers blind to the intervention? Were there any co-interventions?)
- Attrition Bias (Were all participants randomised accounted for in the analysis? Were any withdrawers described?)
- Detection Bias (Was there a blinded assessment of outcome)
- Analysis (Was the analysis appropriate?)



Design your validity assessment form

Design your form now, bearing in mind all the aspects of validity assessment you wish to use. You will need to test out how your form works, but we will discuss that later.

Part Five of a Data Collection Form: Data for Results

The final section of your data collection form is where you record the results of the included studies in a format to allow later entry into RevMan. It is this part of the form that may be transferred electronically into RevMan if it is set up as a spreadsheet, although many reviewers prefer a paper based system with later data entry.

*Skip to Module 10:
Planning the
Analysis*

Before we go on to discuss the design and use of this part of your data collection form, you need an understanding of the type of data you will need to extract in your review. Go to Module 10: Planning the Analysis and complete it, along with the Additional Module A1 on continuous data if any of the outcomes in your review are measured on a continuous scale (for example blood pressure, pain on a visual analogue scale), then return to here and complete this module.

To design the rest of your data extraction form you will need to formulate a table to allow recording of data available from the study for each outcome under each comparison. Your table may look something like this:

COMPARISON 1		
	Intervention A	Intervention B
Outcome 1		
Outcome 2		
COMPARISON 2		
	Intervention C	Intervention B
Outcome 1		
Outcome 2		

If you are performing subgroup or sensitivity analyses, you may also want to include room in the table to collect information about which group the study belongs to.

You should by now have an idea about the type of data you are likely to be collecting from the results sections of your included studies. This will relate to your outcomes. Think about whether your results are likely to be in the form of dichotomous (for example dead/alive, smoking/not smoking) or continuous (for example blood pressure, pain on a visual analogue scale) outcomes. For dichotomous outcomes you will need to extract the number of participants experiencing the outcome and the total number in the group. For continuous outcomes you will need to extract the number of participants, the mean and the standard deviation for each group.



Try designing the results part of your data collection form

A useful hint is to allow sufficient space on your data collection form to record any conversion calculations (for example calculating a standard deviation from a standard error) or notes.

Now have a try at designing this section of the form for your review.

Testing the form

Now you have designed your form, you need to test it to see if it works. The important features of a data collection form are that it is easy to use, comprehensive (it is very frustrating to get further along your review and find you need a piece of data you didn't record) and consistent across reviewers. Questions or items need to be phrased in an unambiguous way and you need to check they mean the same thing to all your reviewers.



Read [Section 7.6.3](#) of the Handbook

Section 7.6.3 of the *Cochrane Handbook for Systematic Reviews of Interventions* discusses the testing of data collection forms and you should read it now.

In summary the main components to testing your data collection form are

- Pilot testing the form with a sample of studies to ensure it is understandable, easy to complete and comprehensive
- Reliability testing the form in a more formal way by comparing a sample of completed forms by two or more reviewers to ensure they are in agreement.

Using the form

Two issues you may need to consider when starting to use your form to extract and record data are whether or not you need to blind reviewers and what to do about missing data. These issues are covered in Section 7.6.3 and 7.7 of the *Cochrane Handbook for Systematic Reviews of Interventions*. As with the decision about trial inclusion and quality assessment, bias and error are likely to be reduced if more than one reviewer extracts the data, independent of each other. Whether or not bias is further reduced by blinding the reviewer to the source and authors of the study is yet to be determined, but some reviewers have gone to considerable lengths to mask studies of all identifying information.

By the completion of this module you should have designed and tested your data extraction form and will have also completed Modules 8, 9 and 10 as you worked your way through designing your data extraction form. Progress now to Module 11 for some more detailed information about the analysis components of your review.

Example from a review in schizophrenia

Trial ID
Action
Methods Allocation: Blindness: Duration:
Participants Diagnosis: N= Age: Sex: History: Excluded:
Interventions 1. 2. 3.
Outcomes - able to use Service outcomes Global impression Mental state Social functioning Adverse effects Economic data Leaving the study early Outcomes - unable to use
Notes

Example from the Effective Practice and Organization of Care CRG

Study design:

- RCT
- CCT
- CBA
- ITS

Interventions:

- **Type of intervention** (for each comparison group the components/categories of the intervention received are recorded, as listed under TOPICS)

Professional

Financial

Organisational

Regulatory

- Controls

- Characteristics of the intervention

Evidence base

Purpose of recommendations (eg appropriate management, cost containment)

Nature of desired change (initiation of new management, stopping introduction of new management, reduction, increase, cessation or modification of established management)

Format

Source

Based upon implementation of clinical guidelines

Guidelines developed through formal consensus process

Recipient

Deliverer

Timing

Setting

Source of funding

Ethical approval

Type of targeted behaviour

Participants:

- Characteristics of participating providers

Profession

Level of training

Clinical specialty

Age

Time since graduation/years in practice

- Characteristics of participating subjects

Clinical problem

Number of subjects included in the study

Episodes of care

Patients

Providers

Practices

Hospitals

Communities or regions

Other characteristics (eg age, gender, ethnicity)

Setting:

- Reimbursement system

- Location of care
- Academic status
- Country
- Proportion of eligible providers from the sampling frame

Methods:

- Unit of allocation
- Unit of analysis
- Study power
- Methodological quality

RCT/CCT

CBA

ITS

- Consumer involvement

Prospective identification of barriers to change

Outcomes:

- Description of the main outcome measures (health professional outcomes/process measures, patient outcomes, costs)
- Length of time outcomes measured after initiation of the intervention
- Length of post-intervention follow-up period
- Possible ceiling effect eg little room for improvement in provider performance because it was adequate without the intervention (based on baseline measurements or control group performance)

Results:

- RCTs and CCTs

The results for the main outcomes in natural units

The baseline performance and post-intervention differences between study and control groups (including statistical significance if reported; and indicating if the units of randomisation and analysis were different)

- CBAs

The results for the main outcomes in natural units

The baseline difference in the pre-post intervention change between groups

For each available comparison, the difference across study groups of the pre-post intervention change (including statistical significance if reported; and in all cases reporting a more favourable provider/patient outcome in the intervention group as a positive finding ie where differences in the groups are in the intended direction)

- ITSs

The results for the main outcomes in natural units (in all cases reporting a more favourable provider/patient outcome attributable to the intervention as a positive finding ie where changes in the outcome are in the intended direction)

Amiodarone for the prevention of sudden death in people with heart disease

Study eligibility form

Study ID:

Type of study

Q1. Is the study described as randomised?

Yes	Unclear	No
↓	↓	↓
Go to		
next question		Exclude

Participants in the study

Q2. Did the participants in the study have heart disease?

NB Answer 'no' if participants had hypertrophic cardiomyopathy

Yes	Unclear	No
↓	↓	↓
Go to		
next question		Exclude

Q3. Were the participants reported to have, or be at risk of, ventricular ectopics (VE or VEA), ventricular tachycardia (VT) or ventricular fibrillation (VF)?

NB Answer 'no' if the study reports that participants had, or were at risk of, atrial arrhythmia or had no rhythm problems

Yes	Unclear	No
↓	↓	↓
Go to		
next question		Exclude

Interventions in the study

Q4. Was one group given amiodarone by mouth/orally (*not intravenously*) for at least 3 months?

Yes	Unclear	No
↓	↓	↓
Go to		
next question		Exclude

Q5. Did another group receive the same care, with or without placebo, but without amiodarone? *NB Answer 'no' if this group could receive an implantable defibrillator*

Yes	Unclear	No
↓	↓	↓
Go to		
next question		Exclude

Outcomes in the study

Q6. Did the study report deaths and follow participants up for at least 3 months?

Yes	Unclear	No
↓	↓	↓
Include , subject to clarification of 'unclear' points		Exclude

Final decision

Include	Unclear	Exclude
---------	---------	---------