

From: Cochrane Handbook for Systematic Reviews of Interventions 2.4.5, May 2005

APPENDIX 6b. Including adverse effects

Edited by Yoon K Loke, Deirdre Price and Andrew Herxheimer on behalf of the Cochrane Adverse Effects Subgroup.

6b.1. Introduction

The policy of The Cochrane Collaboration has always been that reviews look at all relevant outcomes of a healthcare intervention. In practice, however, review authors have frequently avoided studying unintended effects, and have concentrated instead on the intended, beneficial outcomes. Methodological guidance on how to review adverse effects has also been lacking. This section provides guidance from the Adverse Effects Subgroup of the Non-randomised Studies Methods Group.

Every healthcare intervention comes with the risk, great or small, of harmful or adverse effects. A Cochrane review that considers only the favourable outcomes of the interventions that it examines, without also assessing the adverse effects, will lack balance and may make the intervention look more favourable. This source of bias, like others, should be minimised. All reviews should therefore include some evaluation of adverse effects.

An intervention may have many potential adverse effects. The systematic assessment of adverse effects can make substantial demands on time and resources. This needs to be considered in the early stages of the protocol design. Many adverse effects are too uncommon to be observed in randomized controlled trials, which are most appropriate for the assessment of common, known effects. Full evaluation of adverse effects, therefore, often requires other types of evidence.

The extent and nature of the adverse effects analysis should be formulated based on the principles laid out in Section 2.3.1 of the main Handbook. It is worth highlighting two aspects that are of special relevance here:

1. The selected adverse outcomes should be those that are important in guiding the decisions of healthcare providers, researchers, policymakers and consumers;
2. There is often a major trade-off between comprehensiveness and the quality of the adverse effects data included in a review. It may not help to include evidence that is likely to be biased, even if no better evidence exists.
Nevertheless, it is recognized that important information on rare, serious harms may only be available from sources that are susceptible to bias. In these instances, the limitations of the data should be rigorously appraised and critically discussed.

If the review will not evaluate adverse effects, this should be stated explicitly and a reason given.

6b.1.1 Definitions

Many terms are used to describe harmful effects of healthcare interventions; several of these are defined in Table 1. Published papers often use the terms ‘adverse effect’, ‘adverse drug reaction’, ‘side effect’, ‘toxic effect’, ‘adverse event’ and ‘complications’ loosely and interchangeably.

Table 1. Definitions of terms related to adverse outcomes

Adverse event	An unfavourable outcome that occurs during or after the use of a drug or other intervention but is not necessarily caused by it. It can be defined as “any abnormal sign, symptom, or laboratory test, or any syndromic combination of such abnormalities, any untoward or unplanned occurrence (for example, an accident or unplanned pregnancy), or any unexpected worsening or improvement in a concurrent illness” (Aronson 2005).
Adverse effect	An adverse event for which the causal relation between the drug/intervention and the event is at least a reasonable possibility. This term applies to all interventions.
Adverse drug reaction (ADR)	This term is used only with drugs. The terms ADR and adverse effect are used interchangeably with respect to drugs (Edwards 2000).
Complications	This term is widely used to describe adverse events following surgical and other invasive interventions. It can be considered to be synonymous with ‘adverse event’ or ‘adverse effect’.
Seriousness and intensity (or severity) of the adverse effect	Often confused with seriousness, severity is better termed-‘intensity’. WHO terminology differentiates between the terms ‘serious’ and ‘severe’ in this way: ‘serious’ refers to adverse effects that have significant medical consequences, e.g. lead to death, permanent disability or prolonged hospitalisation. A review should state whether ‘serious’ is defined in this way, or whether it also includes other effects that the patient considers serious. In contrast, ‘severe’ refers to the intensity of a particular adverse effect. For example, a non-serious adverse effect, such as headache, may be severe in intensity (as opposed to mild or moderate).
Side effect	This is any unintended effect of a pharmaceutical product that occurs at doses normally used for therapeutic purposes in humans and is related to the pharmacological properties of the drug. While some side effects may be harmful (and can thus be considered adverse effects), there are also side effects

	that are beneficial.
Safety	This word usually refers to (the relative lack of) serious adverse reactions, such as those that threaten life, require or prolong hospitalization, result in permanent disability, or cause birth defects. But, serious, indirect adverse effects, such as traffic accidents, violence, and damaging consequences of mood change, can also be categorized by this term. They may or may not be detected in trials (depending on participant numbers, intensity of monitoring, and length of follow up), and data on such adverse effects may be available only from non-randomised studies.
Tolerability	The term is usually used in referring to medically less important, that is, without serious or permanent sequelae, but unpleasant adverse effects of drugs. These include symptoms such as dry mouth, tiredness, etc, that can affect a person's quality of life and willingness to continue the treatment. As these adverse effects usually develop early and are relatively frequent, RCTs may yield reliable data on their incidence.

6b.2. Formulating the problem

6b.2.1 Scope of an assessment of adverse effects

It would be impractical for review authors to carry out exhaustive safety analyses for every intervention. Table 2 describes some specific therapeutic situations in which a detailed evaluation of adverse effects is warranted.

The scope of the adverse effects evaluation needs to be defined during protocol development as the subsequent direction of the review depend critically on the chosen approach, which may be:

1. *Assess intended and unintended (adverse) effects together, applying common inclusion criteria (in terms of types of studies, types of participants and types of interventions).*

Here a single search strategy would be used. The critical issue is how the review authors intend to deal with the three datasets that may potentially arise:

- (a) studies that report both the intended effects and adverse effects of interest
- (b) studies that report intended effects but not adverse effects
- (c) studies that report adverse effects, but not the beneficial outcomes of interest

A review based on the first dataset (a) is relatively easy to perform, and has the important advantage that benefits and harms can be compared directly since the data are derived from the same population and setting. Furthermore,

evidence on benefits and harms arise from studies with similar designs and quality. However, data on adverse effects may be very limited and biased towards short-term harms.

Evaluation of benefit and harm using some combination of the three datasets (rather than (a) alone) will increase the amount of information available. For instance, datasets (a) and (b) could be used to evaluate beneficial effects, while (a) and (c) could be used to assess adverse effects. However, as the studies addressing adverse effects are different from studies addressing beneficial effects, authors should note benefits and harms cannot be easily compared directly.

2. *Assess intended and unintended (adverse) effects together but use different inclusion criteria for selecting studies that address unintended (adverse) effects*

The application of different inclusion criteria is a method of specifically addressing the problem that most experimental studies (such as RCTs) are insufficient to evaluate rare, long-term or previously unrecognized adverse effects. The approach allows a more rigorous evaluation of adverse effects, but is more costly in time and resources, tends to increase the quantity of data with higher risk of bias, and means that benefits and harms can often not be compared directly.

3. *Undertake a separate review only of adverse effects*

A separate review might be considered for an intervention that is given for a variety of diseases or conditions, yet whose adverse effect profile might be expected to be similar in different populations and settings. For example, aspirin is used in a wide variety of patients, such as those with stroke, or peripheral vascular disease, and also in those with coronary artery disease. The main effects of aspirin would typically be addressed in separate Cochrane reviews, but adverse effects (such as intracerebral or gastrointestinal bleeding) are probably similar within the different disease groups and might be addressed together in an independent review. Indeed, unless trials exist on combined populations, such a question would be difficult to address in any other way. This approach might reduce the workload.

Table 2. Contexts and examples warranting detailed examination of adverse effects

When there is a narrow margin between benefit and harm	
Treatment is of modest or uncertain benefit, with some possibility of harm.	<ul style="list-style-type: none"> • Aspirin for prevention of cardiovascular events in a healthy patient; increase in haemorrhage. • Antibiotics for sore throat and respiratory tract infections; risk of rash and diarrhoea. • Finasteride for the treatment of male pattern baldness; causes erectile dysfunction. • Urgent direct current cardioversion in

	patients with new atrial fibrillation who are cardiovascularly stable; risk of stroke from cardioversion
Treatment potentially highly beneficial, but there are major safety concerns	<ul style="list-style-type: none"> • Aspirin for a patient with a stroke, but who has a past history of gastrointestinal haemorrhage. • Carotid endarterectomy in elderly patients with ischaemic heart disease who present with stroke
Treatment potentially beneficial in long-term, or to community, but no immediate direct benefit to individual.	<ul style="list-style-type: none"> • Improving uptake of a vaccine to promote herd immunity, while trying to assuage fears about early serious neurological adverse effects.
When there are a number of efficacious treatments with differing safety profiles	
Treatments are of equivalent efficacy, but they have different safety profiles	<ul style="list-style-type: none"> • Antiepileptic drugs for women with epilepsy who plan on becoming pregnant • A new insulin injection device is thought to cause less pain than the existing device
The balance of benefits and harms differ substantially e.g. the most efficacious intervention may have serious adverse effects, while the less effective intervention is potentially safer.	<ul style="list-style-type: none"> • Warfarin or aspirin in a healthy middle aged man with lone atrial fibrillation. • Disease-modifying drug in erosive rheumatoid arthritis e.g. using hydroxychloroquine (relatively safe) or methotrexate (potentially more effective, but less safe). • Radical mastectomy for breast cancer as opposed to limited, breast-conserving surgery
When adverse effects deter a patient from continuing on an efficacious treatment	
Treatment is of considerable benefit but adverse effects threaten patient's adherence.	<ul style="list-style-type: none"> • Patient with severe heart failure has responded well to an ACE inhibitor, but now complains of cough. Which is the best option - stopping the medication altogether, trying a lower dose, or changing to an angiotensin receptor blocker?

6b.2.2 What types of outcomes?

Selection of adverse outcomes can be difficult. Specific adverse effects associated with an intervention may be known in advance of the review, others will not. It may

not be possible to identify beforehand exactly which effects will be most relevant to the review. The following general strategies may be used depending on the study question and the therapeutic or preventive context.

Narrow focused:

A detailed analysis of one or two known or a few of the most serious adverse effects that are of special concern to patients and health professionals;

Pros: Easiest approach, especially with regard to data extraction. Can focus on important adverse effects and reach a meaningful conclusion on issues that have a major impact on the treatment decision (McIntosh 2004).

Cons: Scope may be too narrow. Method is only really suitable for adverse events that are known in advance.

Broad sweep:

To detect a variety of adverse effects, whether known or previously unrecognized, in the included studies.

Pros: Wider coverage, and can evaluate new adverse effects that we may not have previously been aware of.

Cons: Potentially large volume of work with particular difficulties in the data extraction process. Some researchers have found broad, non-specific evaluations to be very resource-intensive, with little useful information to show for the effort expended (McIntosh 2004). These researchers also point out that detection of previously unrecognized adverse effects may be best addressed through primary surveillance (see Section 3.3), rather than in a systematic review.

In order to address adverse effects in a more organized manner, review authors may choose to narrow down the broad sweep into some of the following areas:

- the five to ten most frequent adverse effects
- all adverse effects that either the patient or the clinician considers to be serious
- the most common adverse effects that lead the patient to stop using the intervention (caution – see also section 5.4 in this chapter);
- By category, for example:
 - diagnosed by clinician (e.g. gastrointestinal haemorrhage)
 - diagnosed by lab results (e.g. hypokalaemia)
 - patient-reported symptoms (e.g. pain).
 - biomarkers that may be early indicators of possible adverse effects (for example, abnormal liver enzymes); offering a means of collecting relevant information even from short-term studies.

This is not a comprehensive list, but the use of any of the above strategies should help authors approach the adverse effects analysis in a systematic, manageable and clinically useful fashion.

6b.2.3 What types of studies?

The decisions on what types of studies to include will be based primarily on the research question, balancing the elements of comprehensiveness, type of adverse effect(s) of interest, as well as the time and resources available.

Although most Cochrane systematic reviews focus on RCTs, which provide the most reliable estimates of effect, rare adverse events are unlikely to be observed in clinical trials, and a thorough investigation may require the inclusion of cohort studies, case-control studies and even case series.

6b.3. Locating and selecting studies

6b.3.1 Choice of search method

The scope of the review (see Section 2.1 Scope of an assessment of adverse effects) determines the nature of the search strategy. The general approaches for searching and selection are:

(i) Apply a standard search strategy as recommended by the author's Collaborative Review Group (CRG). Check all retrieved studies to identify those that report the intended effects and/or unintended effects of interest. This strategy is relatively simple and less resource intensive. However, it is likely to lead to different lists of potentially relevant studies for intended and unintended effects, with some overlap between them. Further evaluation of these lists depends on the proposed scope of the review, as described in Section 2.1.

(ii) Conduct a separate, additional adverse effects search to supplement the standard strategy. This is far more comprehensive but is likely to be time-consuming and resource intensive.

6b.3.2 Additional adverse effects search

The optimal search strategy for specifically identifying reports of adverse effects has yet to be established, although work on this area is ongoing (Golder 2004a, Golder 2004b). Two main approaches can be used, both of which have their own limitations and so a combination of these approaches is advisable to maximise sensitivity (the likelihood of not missing studies that might be relevant):

Searching electronic databases using index terms (also called controlled vocabulary or thesaurus terms)

Index terms such as MeSH or Medical Subject Headings in MEDLINE and Emtree in EMBASE are assigned to records in electronic databases in order to describe the studies. Subheadings can also be added to index terms to describe specific aspects for example, side effects of drugs, or complications of surgery. There are differences in index terms used to denote data on adverse effects in the major databases (MEDLINE and EMBASE), for example:

Aspirin/adverse effects (MEDLINE)

Acetylsalicylic-acid/ adverse-drug-reaction (EMBASE)

In the above example, Aspirin is the MeSH term and adverse effects is the subheading; Acetylsalicylic-acid is the Emtree term and adverse-drug-reaction is the subheading.

Within a database, studies may be (i) indexed under the name of the intervention together with a subheading to denote that adverse effects occurred, for example, Aspirin/adverse effects or Mastectomy/complications; or (ii) the adverse event itself

may be indexed, together with the nature of the intervention, for example, Gastrointestinal Hemorrhage/ and Aspirin/ or Lymphedema/ and surgery/; or (iii) occasionally, an article may be indexed only under the adverse event, for example, Hemorrhage/chemically-induced.

Thus, no single index or subheading search term can be relied on to identify all data on adverse effects, although a combination of index terms and subheadings is useful in detecting reports of major adverse effects which are likely to be considered of significance by the indexers (Derry 2001).

Subheadings which may prove useful in MEDLINE are:

/adverse effects

/poisoning

/toxicity

/chemically induced

/contraindications

/complications

Subheadings which may prove useful in EMBASE are:

/side effect

/adverse drug reaction

/drug toxicity

/complication

Searching electronic databases using free text terms (also called text words)

Free text terms are used by authors in the title and abstract of their studies when published as journal articles and these terms are then searchable in the title and abstract of electronic records in databases. There are two important problems that severely limit the usefulness of free text searching:

- there is a wide range of terms used by authors to describe adverse effects, both in a general sense (toxicity, side-effect, adverse effects) and more specifically (for example, lethargy, tiredness, malaise may be used synonymously). Therefore, as many relevant synonyms as possible should be included in the search.
- adverse effects that are not mentioned in the title or abstract of the study and are, therefore, not included in the electronic record (even though they are described in the full report), will not be detected using the free text search (Derry 2001).

A highly sensitive free text search should incorporate this potentially wide variety of synonymous terms used to denote data on adverse effects in studies while also taking into account different conventions in spelling and variations in the endings of terms to include, for example, singular and plural terms, for example, adverse or side or hemorrhage or haemorrhage or bleed or bleeding or blood loss. These terms used to describe adverse effects should then be combined with free text terms used to describe

the intervention of interest, for example (aspirin or acetylsalicylic acid) and (adverse or side or hemorrhage or haemorrhage or bleed or bleeding or blood loss).

It is clear that no single approach can be relied on to yield all the studies that have data on adverse effects of an intervention. The search, therefore, needs to combine index terms and free text terms and is likely to take several iterations. For instance, it may be necessary to repeat the electronic search incorporating additional index terms, subheadings and free text terms derived from the terms used to index and describe the studies initially identified as relevant. In deciding which combination of terms to use, authors will need to balance comprehensiveness (sensitivity) against precision. For example, an electronic search that retrieves 20,000 studies is likely to contain the majority of all relevant studies but if only 300 are relevant (1.5%), then it is very imprecise and will have a cost implication in terms of time and resources. (See Section 4).

6b.3.3 Additional sources of information

Review authors who are planning an exhaustive search may wish to consider checking the following sources:

- Standard reference books on adverse effects such as Meyler's Side Effects of Drugs and its annual update, the Side Effects of Drugs Annuals, and screening the papers they summarise.
- Regulatory agencies, for example:
 - in Australia the Australian Adverse Drug Reactions Bulletin (<http://www.tga.gov.au/adr/aadrb.htm>)
 - and the European Public Assessment Reports from the European Medicines Evaluation Agency (<http://www.emea.eu.int/#>).
 - in the UK Current Problems in Pharmacovigilance (<http://medicines.mhra.gov.uk/ourwork/monitorsafequalmed/currentproblems/cpprevious.htm>)
 - in the US, MedWatch, the Food and Drug Administration Safety information and Adverse Events Reporting Program (<http://www.fda.gov/medwatch/elist.htm>)

Authors can also apply to the WHO Uppsala Monitoring Centre (UMC; <http://www.who-umc.org>) for special searches of their spontaneous reporting database; this was for example done for melatonin (Herxheimer 2002). However, frequencies of adverse effects calculated from UMC data may differ from the figures derived from a meta-analysis of double-blind, randomized controlled trials (Loke 2004).

Information on the safety of medical devices and surgical interventions is also available from a variety of regulatory authorities. Some examples include:

- UK National Joint Registry, which records details of hip and knee replacement operations in England and Wales (<http://www.njrcentre.org.uk>)
- The Medical Devices section of the UK Medicines and Healthcare Products Regulatory Agency (<http://devices.mhra.gov.uk/>)

- the US Food and Drug Administration, MedWatch for devices (<http://www.fda.gov/medwatch/index.html>)

6b.4. Assessment of study quality

The usual tools for assessment of methodological quality (see Section 6. Assessment of study quality) should identify more rigorous studies with results closer to the ‘truth’ – presumably for both therapeutic and adverse effects. However, we lack empirical evidence for the relevance of quality tools to adverse effect analysis. The author should use the standard quality assessment tools cautiously as the study quality assessed may apply only to the primary focus of the study, which would usually be the intended effects of the intervention. For example, the primary outcome measure of an intervention may have been studied in a placebo controlled, triple-blind, adequately concealed randomized trial, with standard laboratory measurements. In contrast, the adverse effects of the same treatment may be collected retrospectively, when treatment allocation is known to one or more of the parties (patients, clinician, analyst) via a self-assessment questionnaire. Although a high quality grade may be given to the primary portion of the study, the design to monitor the harmful affects of the treatments falls far short of this standard.

However, there is evidence that the methods used in monitoring or detecting adverse effects have a major influence on adverse effect frequencies. For example, in a group of hypertensive patients, passive monitoring based on spontaneous reports yielded rates of 16%, while active surveillance using specific questioning found a rate of 62% (Olsen 1999). Studies in which adverse effects are carefully sought will report a higher frequency than studies in which they are sought less carefully. Different methods of monitoring adverse effects will yield different results, which may make comparisons between studies, or a formal meta-analysis, impossible (Edwards 1999).

Selective reporting of results is also a particular problem with adverse effects (Ioannidis 2001, Loke 2001). For example:

- Certain categories only may be reported (for example, the study states that they looked for events defined by: several body systems, methods of collection, time periods (3, 6, 12 months), dose (20mg, 40mg, 80mg), but report only laboratory results for neurological disorders after 6 months with the 40mg dose).
- Adverse event categories may not be clearly defined (for example, ‘system = cardiovascular’ but, without indicating seriousness, intensity, duration, diagnostic method, or final outcome).
- Treatment groups may be combined (for example, “x participants withdrew from the study because of adverse effects”).
- Generic statements (for example, “no unexpected adverse effects were seen”/“there was no difference between the groups in adverse effects reported”/“the drugs were well tolerated”).

In many instances (particularly with the generic statements above), authors may have to take greater account of what was left unsaid rather than what was actually reported. Authors will have to choose either to exclude the study from the adverse effect analysis, or to include the study on the assumption that there were indeed no adverse effects (this should be the exception rather than the rule).

Thus authors should take into account two important aspects in assessing the quality of adverse effects:

- How rigorous were the methods used in detecting adverse effects?
- How good is the quality of reporting?

Examples of potentially useful questions in each area are:

On conduct:

Are definitions of reported adverse effects given?

How were adverse effects data collected: prospective/routine monitoring, spontaneous reporting, patient checklist/ questionnaire/diary; systematic survey of patients?

On reporting:

Were any patients excluded from the adverse effects analysis?

Were the methods used for monitoring adverse effects reported?

Did the report provide numerical data by intervention group?

Which categories of adverse effects were reported by the investigators?

Did the investigators report on all important or serious adverse effects?

Finally, non-randomized studies are prone to biases, which can be hard to identify and deal with and authors planning to include such data should seek guidance from the Cochrane Non-randomised Studies Methods Group.

6b.5. Collecting data

6b.5.1 Terms

We suggest that information falling under any of these terms 'adverse effect', 'adverse drug reaction', 'side effect', 'toxic effect', and 'adverse event' be considered as being potentially suitable for data extraction when evaluating the harmful effects of a treatment. For further details see Glossary and Table 1, Section 1 above.

6b.5.2 Exclusions

Remember that no mention of adverse effects does not necessarily mean that no adverse effects occurred. It is usually safest to assume that they were not ascertained or not recorded: authors have to choose between excluding the study from the adverse effect analysis, and including it on the assumption that the incidence was zero (that should be the exception).

6b.5.3 Data collection forms

Authors may find it useful to design and use a separate data collection form for safety outcomes. Some reviews may include additional studies beyond those included in the therapeutic portion of a review.

6b.5.4 Outcome characteristics

The definition of a particular adverse effect may vary between studies, as can definitions of intensity. For example, in a review of aspirin and gastrointestinal haemorrhage, some trials simply reported “gastrointestinal bleeds”; others reported specific categories of bleeding, such as haematemesis, melaena, and proctorrhagia, (Derry 2000). The definition and reporting of severity of the haemorrhages (for example, major, severe, requiring hospital admission) also varied considerably among the trials. (Zanchetti 1999).

Moreover, a particular adverse effect may be described and/or measured in different ways among the trials – take for example, tiredness, fatigue or lethargy, all of which might be terms used in adverse effects reports. Authors may also use different thresholds for ‘abnormal’ results (for example, hypokalaemia diagnosed at a serum potassium concentration of 3.0 mmol/l or 3.5 mmol/l).

Are the adverse effects terms comparable across studies? Authors will need to decide which categories are similar enough to collect data on and justify lumping together in the analysis. For example, gastrointestinal bleed, haematemesis, and melaena were included in the aspirin analysis, but proctorrhagia was excluded.

There are a number of initiatives aimed at harmonizing adverse effects terms (Bankowski 1999), and the National Cancer Institute set of toxicity criteria is an example of a standardized scheme for judging severity of adverse effects across trials of cancer therapy. (<http://ctep.cancer.gov/reporting/CTC-3.html>). The WHO uses the system-organ class categories (<http://www.who-umc.org/pdfs/ardguide.pdf>) which allows authors to collate adverse effects data into one of several system-organ classes such as ‘gastrointestinal system disorders’ or ‘vision disorders’ (MacLehose 2003). However, some researchers have found that the standard ‘preferred terms’ used by regulators and industry can distort descriptions in the original reports of adverse events and blur distinctions between them (Medawar 2003).

Withdrawal or drop-outs as outcome measure

These outcome measures are often seen in trial reports. We urge authors to be cautious in interpreting such data as surrogate markers for safety or tolerability because of the potential for bias:

- The attribution of reason(s) for discontinuation is complex and may be due to mild but irritating side effects, toxicity, lack of efficacy, non-medical reasons, or a combination of causes (Ioannidis 2004).
- The pressures on patients and investigators under trial conditions to keep the number of withdrawals and drop-outs low can result in rates that do not reflect the experience of adverse events within the study population.
- Unblinding of treatment assignment often takes place prior to the decision to withdraw. This can lead to an over-estimate of the intervention’s effect on patient withdrawal. For example, symptoms of patients in the placebo arm are less likely to lead to discontinuation. Conversely, patients in the active intervention group who complained of symptoms suggestive of adverse effects would have been more readily withdrawn.

Quality of Life Indicators

These are usually general measures that do not look specifically at particular adverse effects of the intervention. While quality of life scales can be used to gauge the overall well-being, they should not be regarded as substitutes for a detailed evaluation of safety and tolerability.

6b.6. Analysing and presenting results

In addition to the advice given in Section 8, there are number of issues especially relevant to the analysis of adverse effects.

If different types of studies are being used to evaluate beneficial and harmful effects, then an author must consider how to analyse potentially disparate datasets where studies reporting intended effects are different from those that report adverse effects. Special techniques might be used to synthesise data from a diverse range of sources (Wald 2003, Jefferson 2003).

The analysis of zero events in either arm (for example, “the drug was safe”, and “no serious adverse effects were seen”) needs careful consideration. Data of this type need to be evaluated in the following contexts:

- How thorough were the methods used to detect adverse effects?
- How many patients were studied and for how long?

It is not possible (based on zero events detected) to conclude that a drug does not cause a suspected adverse effect. However, we can use the rule of 3, which states that the 95% confidence intervals of zero are 0-3 events in the observed sample, to estimate an upper limit for the frequency of the adverse effect (Eyspach 1995). For example, if no adverse effects occur in 300 participants, then any adverse effects associated with the intervention might be as frequent as 1 in 100, but are unlikely to be more frequent. Note that studies with no events in either arm can be included in a meta-analysis of risk differences, although they cannot be included in a meta-analysis of odds ratios or risk ratios.

It is important to remember that a systematic review is not synonymous with a meta-analysis. There may be occasions when adverse effect information is best summarised in a qualitative or descriptive manner. For instance, data derived from divergent sources (for example, different study design, different populations, different data collection methods) cannot be combined. It may not be possible to compare benefits and harms directly. In practice this means that adverse effects from RCTs, case reports, case series, cohorts, and case controls cannot all be pooled together using standard meta-analysis principles. Moreover, the data from non-randomised studies are more prone to bias, and are often heterogeneous; combining them to produce a summary statistic may not be appropriate.

6b.7. Interpreting results

6b.7.1 Applicability

Many RCTs are restricted to carefully selected subgroups of the population, and it is generally inappropriate to extrapolate adverse effects data from such studies to the wider population, which includes more vulnerable people, for example, with co-

morbidities, co-medications. In interpreting adverse effects data, authors must take into account the inclusion and exclusion criteria used during recruitment of participants.

6b.7.2 Trade-offs

Including studies beyond those included in the analysis of intended effects means that the analysis of harm is carried out in studies whose participants may differ from those included in the studies used in the analysis of benefit. This creates potential difficulties in assessing the trade-off between benefits and harms. Review authors will need to consider how much, if at all, the participants in the additional studies can differ from those in the benefit studies, and remain comparable.

For example, in a study of the benefits and harms of aspirin used as an antiplatelet drug to reduce cardiovascular events, a review author might want to include in the adverse effect analysis a study in which aspirin was used as an antiplatelet drug to reduce scarring after mastectomy. Predefined inclusion criteria, other than indication for treatment (for example, dose, duration of treatment, reporting of adverse effects), would need to be met. The decision to include the study or not should depend on whether there is evidence that these women differ systematically in their risk of gastrointestinal haemorrhage from people who take the drug to prevent cardiovascular problems.

Extending the review to observational studies and anecdotal case reports can create additional difficulties in evaluating the benefit: harm trade-off. Authors will need to consider how efficacy data from high-quality trials can be weighed up against adverse effects from low quality studies.

6b.8. Contributions

Contributing authors: Jeff Aronson, Anne-Marie Bagnall, Andrea Clarke, Sheena Derry, Andrew Herxheimer, Yoon Loke, Heather McIntosh, Harriet MacLehose, Deirdre Price, Nerys Woolacott

Comments on drafts: Phil Alderson, Jon Deeks, Anne Eisinga, Su Golder, Sally Green, Julian Higgins, Tom Jefferson, Carol Lefebvre, Philippa Middleton

Editors: Yoon Loke, Deirdre Price and Andrew Herxheimer on behalf of the Cochrane Adverse Effects Subgroup.

6b.9. References

- Aronson 2005.** Aronson JK, Ferner RE. Clarification of terminology in drug safety. *Drug Safety*. 2005; in press
- Derry 2001.** Derry S, Kong Loke Y, Aronson JK. Incomplete evidence: the inadequacy of databases in tracing published adverse drug reactions in clinical trials. *BMC Medical Research Methodology* 2001; 1: 7
- Derry 2000.** Derry S, Loke YK. Risk of gastrointestinal haemorrhage with long term use of aspirin: meta-analysis. *BMJ* 2000; 321: 1183-7
- Edwards 2000.** Edwards IR, Aronson JK. Adverse drug reactions: definitions, diagnosis, and management. *Lancet* 2000; 356: 1255-1259
- Edwards 1999.** Edwards JE, McQuay HJ, Moore RA, Collins SL. Reporting of adverse effects in clinical trials should be improved: lessons from acute postoperative pain. *Journal of Pain and Symptom Management* 1999; 18: 427-37
- Eypasch 1995.** Eypasch E, Lefering R, Kum CK, Troidl H. Probability of adverse events that have not yet occurred: a statistical reminder. *BMJ* 1995; 311: 619-20
- Golder 2004a.** Golder S, Duffy S, Glanville J, McIntosh H, Miles J. Developing efficient search strategies to identify papers on adverse events. A: testing and precision sensitivity [abstract]. In: 12th Cochrane Colloquium 2004 Oct 2-6; Ottawa, Ontario, Canada:75-6
- Golder 2004b.** Golder S, Duffy S, Glanville J, McIntosh H. Developing efficient search strategies to identify papers on adverse events. B: using statistical analysis [abstract]. In: 12th Cochrane Colloquium 2004 Oct 2-6; Ottawa, Ontario, Canada:75.
- Herxheimer 2002.** Herxheimer A, Petrie KJ. Melatonin for the prevention and treatment of jet lag (Cochrane review). *The Cochrane Database of Systematic Reviews* 2002, Issue 2 Art. No.: CD001520. DOI: 10.1002/14651858.CD001520
- Ioannidis 2004.** Ioannidis JP, Evans SJ, Gotzsche PC, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Annals of Internal Medicine* 2004; 141: 781-8
- Ioannidis 2001.** Ioannidis JP, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. *Journal of the American Medical Association* 2001; 285: 437-43
- Jefferson 2003.** Jefferson T, Demicheli V. Balancing benefits and harms in health care: observational data on harm are already included in systematic reviews. *BMJ* 2003; 327: 750
- Loke 2004.** Loke YK, Derry S, Aronson JK. A comparison of three different sources of data in assessing the frequencies of adverse reactions to amiodarone. *British Journal of Clinical Pharmacology* 2004; 57: 616-21
- Loke 2001.** Loke YK, Derry S. Reporting of adverse drug reactions in randomised controlled trials - a systematic survey. *BMC Clinical Pharmacology* 2001; 1: 3
- McIntosh 2004.** McIntosh HM, Woolacott NF, Bagnall AM. Assessing harmful effects in systematic reviews. *BMC Medical Research Methodology* 2004; 4: 19
- MacLehose 2003.** MacLehose H, Klaes D, Garner P. Amodiaquine: a systematic review of adverse events. Version 1, March 2003. WHO. Available at: <http://www.who.int/medicines/oranization/par/edl/expcom13/expcom03add.shtml>.
- Medawar 2003.** Medawar C, Herxheimer A. A comparison of adverse drug reaction reports from professionals and users, relating to risk of dependence and suicidal behaviour with paroxetine. *International Journal of Risk and Safety in Medicine* 2003; 16: 5-19
- Olsen 1999.** Olsen H, Klemetsrud T, Stokke HP, Tretli S, Westheim A. Adverse drug reactions in current antihypertensive therapy: A general practice survey of 2586 patients in Norway. *Blood Pressure* 1999; 8: 94-101
- Wald 2003.** Wald NJ, Morris JK. Teleanalysis: Combining data from different types of study. *BMJ* 2003; 327: 616-8
- Zanchetti 1999.** Zanchetti A, Hansson L. Risk of major gastrointestinal bleeding with aspirin. *Lancet* 1999; 353: 148-50