

International Child Health Review Collaboration: The Reviewer's Toolkit

INTRODUCTION: AIMS OF A SYSTEMATIC REVIEW

The ICHRC is an international collaborative project to determine the evidence basis for paediatric guidelines developed by the World Health Organisation (WHO) for use by health care workers in resource limited settings.

The systematic reviews aim to give a concise, reliable and quality appraisal of evidence on a given topic. This will enable decisions to be made about local implementation of WHO policies and guidelines and will demonstrate the evidence basis behind such guidelines for doctors, clinical officers, nurses and other health care workers in resource limited settings.

As a reviewer you will be responsible for accessing all of the available evidence on a given clinical question, appraising the evidence, judging it's quality and importance and answering the given question based on the available evidence.

The starting point is the WHO clinical guidelines, as outlined in the Pocket Book of Hospital Care for Children. You should determine the given guidelines governing your subject and from here work through the stages of:

- defining a search question,
- accessing relevant articles,
- extracting data,
- evaluating the evidence and
- preparing and writing the review.

After you have submitted a draft of your review, we then ask content experts to review and provide you with critical feedback before publication.

This project shares several characteristics with Cochrane systematic reviews, but has clearly defined scope in terms of topics and the target setting. It uses a widely available search strategy ('Clinical Queries' in Pubmed), and the reviews are short summaries of the evidence. These reviews will be updated regularly, and will form the resource needed by WHO to modify treatment recommendations as new evidence is published.

This methodology is based upon the WHO Handbook for Guideline Development, currently in draft form.

1. DEFINING THE SEARCH QUESTION

Before starting, it is important to establish a search question. The PICO check list can help formulate a question:

- Patient population or problem
- Intervention or Exposure
- Comparison
- Outcome

By using this checklist, an appropriate query can be drawn up. For example, 'Should zinc be used in the treatment of acute gastroenteritis in children in developing countries?'. The population is children in developing countries with acute gastroenteritis, the intervention is zinc, the comparison is with those not treated with zinc and the outcome would be clinical improvement. It is instructive to underline the essential words and phrases of the query. In this example, zinc, treatment, acute gastroenteritis, children and developing countries.

You will find that from the outset, it is important to properly understand your question.

To this end it is a good idea to learn from reading a few recent review articles or textbooks about your subject to be able to understand the issues and controversies involved, especially if it concerns an area that is new to you. Try to find out the wider implications of the subject, eg. why it is important, epidemiology, public health burden, diagnosis, treatment etc. This will help you to understand when you read the relevant studies, you may discover potential synonyms for the literature search and it will hopefully give you motivation for carrying out the project

The starting point for literature searching is to use Pubmed's clinical queries, which is freely available from going to www.pubmed.com and clicking on "clinical queries".

Keywords, MeSH terms, synonyms and spelling

Keywords and synonyms

Keywords are often those established by underlining key components of your search query. They are essentially free-text searches that search that exact term and thus can retrieve articles in which the term is incidental. PubMed attempts to assign a MeSH (Medical Subject Heading) term to your keyword, though will not search for other more-specific terms that may be more appropriate. Consequently many useful papers may be missed if MeSH terms are not considered as well as keywords.

It is useful to keyword search when the term you are searching for does not appear in the MeSH database, or the term is new or highly specific. New literature may not yet have been assigned a MeSH term and hence would only be found by keyword search. With keywords, there is a requirement to think of all synonyms or use a broad enough term to cover the topic. Brainstorm (or use a thesaurus) for alternative terms. Synonyms can be included in the search query using the Boolean operator 'OR' (explained later).

Truncations

If the keyword should have variant endings – singular, plural, adjective – truncation can side-step the need to combine them all with 'OR'.

With truncation, use the common 'stem' of the word – for example, child - followed by the truncation symbol (in Pubmed this is *). Thus child* could retrieve child, children, childs, childrens etc. There is a risk that truncation may yield false 'hits'.

MeSH terms

Every article entered into PubMed / Medline is assigned a MeSH term by a third party – essentially allocating the article a topic heading. MeSH contains almost 17,000 terms. Each of these terms represents a single concept appearing in the medical literature. As important new concepts appear, a new MeSH keyword is created. When a new reference is added to MEDLINE, indexers review the article and then choose/add the appropriate MeSH keywords (usually 10 to 20) to represent the contents of the article. Using MeSH terms in the search means that articles can be retrieved by topic and bypasses the problem of whether the right term has been chosen with keywords. A useful technique is to search the database of MeSH terms available at:

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=mesh>

Establish the correct MeSH term by looking at the definition, and then look at the MeSH tree of Subheadings to see where this term fits in. By looking up and down the Subheadings tree, it is possible to make your search term more or less specific. Using the example of gastroenteritis, gastrointestinal diseases offer a far broader search, whereas dysentery or gastritis would restrict the yield of papers.

Using the MeSH terms means the most recognised form of the term is used and enables greater control over your search. However, as mentioned above, newer articles or new terms may not yet be allocated a MeSH term.

For completeness, both MeSH terms and keywords should be used in the search strategy. The importance of this is illustrated in the example below.

Remember:

- Keywords and MeSH terms
- Related terms
- Alternative spellings (particularly between UK and US English)
- Synonyms
- Truncations

Some advocate the use of a mind-map / spider diagram or columns to expand on these major topics. This is a matter for personal preference.

MeSH vs. keywords: an example

There will be vastly different results achieved depending on whether or not keywords or MeSH terms are used. Below is an example where MeSH terms do not entrain crucial articles to answer a question.

Take the question "What is the evidence behind dexamethasone therapy in bacterial meningitis in developing countries?"

Most authors will use keywords but these often do not match up with the MeSH terms. Using only one set will not produce the same results. The red example below uses the MeSH terms in the clinical filter and yet even with broader terms does not entrain a number of important articles. The keyword search is in blue and one can see that once put through the filter it entrains 6 further articles, (including the seminal work by Dr. Liz Molyneux.)

This of course occurs if the author when using the keyword search thinks of all synonyms or uses a broad enough term to cover the topic. They also have to take into account all truncations (hence "countr*"). The results in bold are based upon PubMed clinical queries searching although other databases/information repositories need to be searched as well.

#1 Search "Meningitis, Bacterial"[MeSH] AND "Developing Countries"[MeSH] 105 articles

» This is then put through the Clinical Queries filter/other database search engine:

#2 Search (#1) AND (randomized controlled trial[Publication Type] OR (randomized[Title/Abstract] AND controlled[Title/Abstract] AND trial[Title/Abstract])) 4 articles

#3 Search bacterial meningitis AND developing countr* 216 articles

» This is then put through the Clinical Queries filter:

#4 Search (#3) AND (randomized controlled trial[Publication Type] OR (randomized[Title/Abstract] AND controlled[Title/Abstract] AND trial[Title/Abstract])) 10 articles

So why not use keywords alone? The answer lies that in the example above it is relatively easy to work out the terms that will get the correct relevant articles. This is obviously not always the case. Running a search for skin grafting and burns the MeSH terms turned up a number of extra articles compared with keywords...61 Vs. 24. This is because the MeSH term is "skin transplantation" whilst with keywords the reviewer has to include a number of terms other than skin graft, such as ;

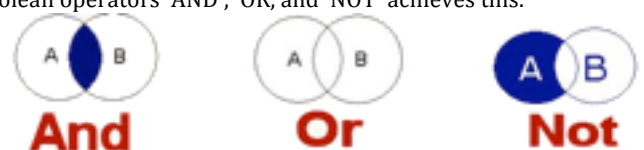
- Dermatoplasty
- Dermatoplasties
- Grafting, Skin
- Graftings, Skin
- Skin Grafting
- Skin Graftings
- Transplantation, Skin
- Skin Transplantations
- Transplantations, Skin

"Transplantations. skin [MeSH]" picks them all up.

Another reason not to use keywords alone is that for broader topics then a number of irrelevant articles are captured which defeats the purpose of the clinical filter somewhat.

Combining the search terms

The keywords identified in the last section should be combined to search the databases. This refines the search. Use of the Boolean operators 'AND', 'OR', and 'NOT' achieves this.



AND retrieves records containing both / all of the combined terms. In the figure below, the shaded area represents the papers found when A (e.g. gastroenteritis) AND B (e.g. zinc) are combined.

OR retrieves records containing either term. This is useful when synonymous terms exist. In this example, A represents

gastroenteritis and B acute diarrhoea. For this search the terms are interchangeable so can be combined with OR.

NOT is a term of exclusion, retrieving records that use one term but not another. In the figure below, B could represent gastritis but NOT A, appendicitis. This risks excluding records that could be useful, shown in the overlap.

In this example then, a final search strategy could be:
Zinc AND (acute diarrhoea OR gastroenteritis OR infectious diarrhoea OR acute gastroenteritis OR diarrhea).
Further examples can be found in the 'examples of search queries' in the reviewer's toolkit.

As the search strategy is for the WHO review, it should focus on children in developing countries. At this stage however, it is preferable not to restrict the findings.

Inclusion and Exclusion Criteria - Setting Parameters

Before searching, set out on paper what you would and would not like to include. Obviously, the key terms should be included. Decide what sort of papers you need - systematic reviews and randomised controlled trials are the order of the day for the WHO review, but would you consider cohort studies or retrospective reviews? Humans only? Paper age? Paper language? Certain journals? Treatment versus prophylaxis? How will you select your papers?

Setting this out at the start avoids ambiguity later.

The ICHRC search strategy is based on PubMed (www.pubmed.com). By using the PubMed 'Clinical Queries' or 'Special Queries' options, you can set up specific searches to identify systematic reviews or different types of studies. This includes searches of the Cochrane Database of Systematic Reviews. Systematic reviews of policy interventions (such as pricing of pharmaceuticals) can be difficult to find and other search strategies will be needed.

It is important to try to include studies from the developing world as well as the more standard literature. Developing country journals are not well represented in PubMed and commercial databases such as EMBASE, CAB Abstracts. Regional databases (AIM, IMEMR, HELLIS; LILACS, WPRIM) grouped under the general heading of the Global Health Index <http://www.who.int/ghl/medicus/en/> contain unique citations and in many cases full text articles.

Regional offices of WHO have supported the development of these indexes to highlight the health research literature of developed world. The majority of journals indexed by Regional databases are not indexed in PubMed and other databases. An information specialist (eg at the WHO Library) can also suggest other databases to search depending on the topic areas.

Retrieval of 'grey literature' such as Ministry of Health reports, case studies and unpublished studies, is best done based on the results from journal article searching from PubMed, Regional databases and other sources. Terms such as the key authors and institutions can be used with Google or other search engines to identify grey literature cited on the internet. It is also important to scan key web sites individually as general search engines are not capable of retrieving all the relevant information on a web site. Obtaining training in using search engines is important to limit research results to pertinent information. The combination of efficient use of search engines and targeted website/authors is much more effective for identifying unique information than large unfocused searches. Personal contact with key informants

will help to identify sources of information not found in the published journals or cited on website.

Limiting or Expanding Search Results

Your strategy will almost invariably retrieve too many, too few or irrelevant hits at first try.

If too few hits:

- Use truncation
- Use a thesaurus - ensure all relevant synonyms are included
- Spelling - check, or try an alternative
- Check terminology
- Combine keywords with the Boolean operative 'OR'
- Select 'All Subheadings' if presented with the option
- Free-text searches - for terms too new or not widely used enough to be Subject Headings
- Search other databases
- Use the "broad, sensitive" search checkbox with PubMed Clinical Queries

If too many hits:

- Use Boolean operators - particularly 'AND' or 'NOT'.
- Use specific Subheadings
- Use filters - this will be done by using the Clinical Queries tool on PubMed to restrict to randomised controlled trials
- Focus - available on many databases to focus on Subject Heading terms, retrieving articles with the Subject Heading as a main subject.
- Use thesaurus terms
- Use the "narrow, specific" search checkbox with PubMed Clinical Queries

Note: the goal of evidence based medicine is to elucidate the evidence behind interventions. Arbitrarily excluding papers due to a large volume of research risks excluding key findings

If too many irrelevant hits:

- Use Subject Headings rather than free-text alternatives
- Use the subject tree (see MeSH) to find more precise Subject Headings
- Boolean operator 'NOT' - use with caution!

For thoroughness:

- Avoid limits
- Search reference lists
- Specify inclusion and exclusion criteria
- Include ALL relevant data
- Should be reproducible

What if no relevant papers can be found?

Having searched the Cochrane Library, PubMed Clinical Queries and other databases/grey literature mentioned above, for systematic reviews and randomised controlled trials/ cohort studies/case control studies etc, the final stage would be to include a good quality review. This only really refers to a small subset of questions not easily answered by the other types of studies.

It may simply be that the current recommendations are based on institutional traditions, hence the importance of evidence-based medicine and its documentation.

Having formulated a search strategy, the next stage is to retrieve the articles.

2. ACCESSING RELEVANT ARTICLES

Abstracts off all studies from the literature searching must be read. This minimises the risk of missing a relevant article.

For any articles that may be relevant to the clinical question, the full text version must be sourced and reviewed.

Some papers can be accessed directly from the database, such as PubMed, by a link to the appropriate journal. Most other articles can be retrieved by searching for the journal online and retrieving the electronic article from the archive or through an online partner database. Otherwise, a paper copy should be sought in your library or by inter-library loan where such facilities exist.

Reviewers at institutions with access to library facilities should use these. For those in developing countries that may not have such access, HINARI may help (Health InterNetwork Access to Research Initiative - <http://www.who.int/hinari/en/>). This has been set up by the WHO and major publishers to enable developing countries to gain access to over 3100 journal titles. Currently this resource is available to 113 countries.

3. EXTRACTING DATA FROM ARTICLES

Read the collected articles to understand the background and key points, and highlight the relevant findings that answer your query.

Establish:

- Paper type – systematic review, randomised controlled trial etc
- Author, setting, country, urban or rural, community or hospital, tertiary centre or peripheral centre.
- Year of publication
- Does the paper meet your inclusion / exclusion strategy? – for example, dated information, adults not children
- What is the clinical question the paper addresses?
- Patient numbers
- Key results and outcomes – actual figures are useful here
- P values and confidence intervals – what statistical tests have been performed?
- Assess the methodology (see below)
- GRADE the level of evidence

Assessing the Methodology

The methodology of the paper must be considered before papers are included. The strategies differ slightly between systematic reviews and randomised controlled trials.

Quality criteria for systematic studies

- Were the questions and methods clearly stated?
- Was the search method comprehensive and the methodology described?
- Were explicit methods used to determine which studies were included in the review?
- Was the methodological quality of primary studies assessed?
- Was the selection and assessment of primary studies reproducible and free from bias?
- Were differences in individual study results adequately explained?
- Were the results of primary studies combined appropriately?
- Were the reviewers' conclusions supported by data cited?

Quality criteria for randomised controlled trials

- Were the setting and study patients clearly described?
- Was assignment randomised and similarity between groups documented?

Was allocation to study groups adequately concealed from patients and investigators, including blind assessment of outcome?

Were all clinically relevant outcomes reported?

Were > 80% of patients who entered the study accounted for at its conclusion?

Were they analysed in the groups to which they were randomised (intention to treat)?

Were both statistical and clinical significance considered?

Quality for cohort studies / retrospective studies

Were the recruitment setting, diagnostic criteria, disease severity, co-morbidity and demographic details documented?

Was the referral pattern described?

Referral or diagnostics access bias avoided?

Was an adequate follow up rate achieved?

Were > 80% patients entered accounted for in results and clinical status known?

Were objective outcome criteria developed and used?

Was outcome assessment blind?

Was adjustment for extraneous prognostic factors carried out?

4. EVALUATING THE EVIDENCE

Each paper eligible for inclusion should be analysed using evidence summaries that should be created applying the GRADE approach. This approach allows a structured and transparent judgment of the quality of evidence for each outcome rated as important by the guideline panel. These evidence profiles are based on systematic review(s).

For each question, data should be extracted from the systematic review for all of the outcomes (benefits and harms) that were rated to be important. If there is more than one systematic review, start with the best one, and supplement it as needed with additional data from other good quality systematic reviews. If the published systematic reviews are not recently updated (within the last 2 years), a search should be conducted to retrieve more recent studies or studies in press that could be relevant to the specific issue.

USE OF GRADE IN THE INTERNATIONAL CHILD HEALTH REVIEW COLLABORATION

GRADE stands for Grading of Recommendations Assessment, Development and Evaluation. It is a systematic method of assessing the quality of studies included in a systematic review and developing recommendations or guidelines based upon the evidence. The basis of the evaluation of evidence is the systematic review(s) of available studies identified from the process of evidence retrieval. It is easiest to do this in a tabular format, using a GRADE table (See example, figure 1, below).

This methodology has been adopted by a number of international organisations, including the World Health Organisation, as a standardised approach to quality assessment in systematic reviewing and guideline development. More information about GRADE can be found at www.gradeworkinggroup.org

The International Child Health Review Collaboration (www.ichrc.org) has decided to use the GRADE method to assess the quality of studies in its systematic reviews and as a means of developing recommendations based upon the evidence.

Figure 1: Example GRADE tables from the World Health Organisation Paediatric ART guideline (2008)

Comparison: EARLY vs DEFERRED ANTIRETROVIRAL TREATMENT IN INFANTS (≤ 1 YEAR)							
Outcome: EARLY MORTALITY ($<1^{st}$) YEAR							
Population group: HIV INFECTED INFANTS (≤ 1 YEAR)							
No of studies	Design	Limitations	Consistency	Directness	Imprecise or sparse data	Other factors	QUALITY RANK
2 [*] (440 infants)	RCT	Minor Limitations ¹	No Serious inconsistency	Serious indirectness ²	No Serious imprecision ³		3 MODERATE QUALITY
Outcome: LATE MORTALITY ($<5^{th}$) YEAR							
5 ^{***} (616 infants)	Observational	Serious Limitations ⁴	No serious inconsistency	Serious indirectness ⁵	Good sample sizes		1 VERY LOW QUALITY
Outcome: DISEASE PROGRESSION							
2 (440 infants)	RCT	Minor Limitations ¹	No Serious inconsistency	Serious indirectness ²	No Serious imprecision ³		3 MODERATE QUALITY
Outcome: SEVERE/LIFE THREATENING EVENTS							
2 [*] (440 infants)	RCT	Minor Limitations ¹	No serious inconsistency ⁶	No serious indirectness ⁷	Serious imprecision ⁸		3 MODERATE QUALITY

1. Prendergast study was not blinded (minor limitation, not downgraded).

2. Indirect population: 80% in CHER and 60% in Prendergast study were not breastfed. Indirect comparison: symptomatic children have been included in the early treatment group.

Figure 2: Example GRADE table from WHO Malaria Treatment Guidelines 2010**GRADE Table A8.1.1**

Is artesunate superior to quinine for treating severe malaria in endemic areas?

Quality assessment			Summary of findings								Quality	Importance
No. of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	No. of patients		Effect				
						AS	QN	Relative risk (95% CI)	Absolute			
EFFICACY: death												
6	Randomized trial	No serious limitations ¹	No serious inconsistency	No serious indirectness ²	No serious imprecision	133/975 (13.6%)	214/963 (22.2%)	RR 0.62 (0.51–0.75)	84 fewer per 1000 (from 56 to 109 fewer)	HIGH	CRITICAL	
EFFICACY: neurological sequelae at discharge												
2	Randomized trial	No serious limitations ³	No serious inconsistency	Serious ⁴	Very serious ⁵	8/656 (1.2%)	3/597 (0.5%)	RR 2.21 (0.64–7.63)	6 more per 1000 (from 2 fewer to 33 more)	VERY LOW	CRITICAL	
EFFICACY: time to hospital discharge (days)												
1	Randomized trial	Very serious ⁶	Not applicable ⁷	No serious indirectness	Very serious ⁸	59	54	–	MD 0.10 (–1.34 to 1.54)	VERY LOW	IMPORTANT	
HARMS: hypoglycaemia (routine monitoring)												
2	Randomized trial	Serious ⁸	No serious inconsistency	Serious ⁹	No serious imprecision	12/96 (12.5%)	24/89 (27%)	RR 0.46 (0.25–0.87)	146 fewer per 1000 (from 35 fewer to 203 fewer)	LOW	CRITICAL	

Panel comment: There is very little evidence of artesunate versus quinine in children.

Panel conclusion: Intravenous artesunate is more effective at reducing deaths in severe malaria (high quality evidence). No difference was shown in the rate of neurological sequelae (very low quality evidence). Artesunate may result in less hypoglycaemia (low quality evidence).

- No serious limitations: two out of six trials had inadequate allocation concealment; however, the panel chose not to downgrade for this as a sub-group analysis excluding these trials did not affect the significance of the result or the absolute magnitude of the effect.
- No serious indirectness: four trials enrolled adults only, one trial enrolled only children and one large trial enrolled children and adults; in a total of 1938 participants, 1664 were adults and 274 were children; a sub-group analysis of trials which only enrolled adults did not affect the significance of this result and it was, therefore, decided not to downgrade this outcome on the basis of indirectness.
- No serious limitations: this large RCT was open label although allocation concealment was classified as adequate; of the 10 patients discharged from hospital with residual neurological sequelae, five had psychiatric sequelae, four had persisting problems with balance (one of whom had psychiatric sequelae and tremor) and two had hemiparesis; it was felt that these adverse events were minimally subjective and it was, therefore, decided not to downgrade due to absence of blinding.
- Serious indirectness: one large trial enrolled both adults and children and one trial enrolled only children; 1259 out of 1533 participants were adults but it was not reported how many adults had data for this outcome; adults are less likely to develop neurological sequelae than children.
- Very serious imprecision: the 95% CI of the pooled estimate includes appreciable benefit with both artesunate and quinine.
- Very serious limitations: this trial used inadequate allocation concealment and was open label.
- Not applicable: only one trial.
- Serious limitations: one of the two trials used inadequate allocation concealment and was open label.

Using the GRADE method:

The GRADE system generates two different outcomes: quality of a study and strength of recommendation.

Study quality is defined as “the extent to which one can be confident that an estimate of effect or association is correct.”

Assessing the quality of a study is important because our confidence in the estimate of effect (or results) generated by a study determines how useful the estimate is.

Making a recommendation on the basis of the evidence is important because it enables consideration of other important outcomes for patients including harms or undesirable effects, patient preferences or values and the wise use of resources.

THE GRADE METHODOLOGY

1. Assessing the quality of the evidence

GRADE tables which summarise the evidence can be created using the GRADEpro software available freely on the internet from:

<http://www.gradeworkinggroup.org/members/balance.htm>

Tables may also be created in an alternative software package such as Microsoft Word.

Creating GRADE tables:

- Choose the first and most important outcome.
- Identify the studies that report on that particular outcome.
- Fill in the number of studies that report the outcome and reference the studies.
- Assess the quality of the evidence by looking at:
 1. Study design
 2. Limitations of the studies, in terms of their conduct and analysis
 3. Consistency of the results across the available studies
 4. Directness (or applicability or external validity) of the evidence with respect to the populations, interventions and settings where the proposed intervention may be used
 5. Precision of the summary estimate of effect
 6. Other considerations

Study design:

There are two main types:

- RCTs
- Observational, cohort, case-control, and other types including case-series and case reports

The study design forms the baseline starting point for the quality level.

Underlying Methodology Quality Rating

RCT	High
Downgraded RCTs or upgraded observational studies	Moderate
Well-done observational studies with control groups	Low
Others (e.g., case reports or case series)	Very low

From this baseline level of quality, evidence grade can be increased or decreased according to a number of criteria as follows:

Decrease grade if:
<ul style="list-style-type: none"> • Serious (-1) or very serious (-2) limitation to study quality • Important inconsistency (-1) • Some (-1) or major (-2) uncertainty about directness • Imprecise or sparse data (-1) • High probability of publication bias (-1)
Increase grade if:
<ul style="list-style-type: none"> • Strong evidence of association- significant relative risk of >2 or <0.5 based on consistent evidence from two or more observational studies with no plausible confounders (+1) • Very strong evidence of association – significant relative risk of >5 or <0.2 based on direct evidence with no major threats to validity (+2) • All plausible confounders would have reduced the effect (+1) • Dose-response gradient (+1)

FACTORS THAT DECREASE QUALITY GRADE

Limitations:

The following factors reduce the study quality for RCTs:

- Lack of allocation concealment
- Lack of blinding (especially if subjective outcome)
- Large loss to follow up
- Failure to use intention to treat analysis
- Stopping trial early for benefit
- Selective reporting of outcomes

An overall judgement should be made about the quality of the studies and if a decision is made to downgrade the quality a clear explanation should be given:

- No limitations: majority of studies meet the minimum quality criteria for the study design
- Minor limitations: Make note of these, but do not downgrade quality
- Serious limitations: One of the minimum criteria for quality is not met by the majority of studies in the review. Downgrade by one grade
- Very serious limitations: At least two of the minimum criteria for quality

For other study types, different criteria can be used to assess quality. The Newcastle-Ottawa checklist may be used for observational studies. The manual and checklist are available at: http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm

Inconsistent results

Widely differing estimates of the treatment effect (heterogeneity or variability in results) across studies suggest true differences in the underlying treatment effect.

Sources of variability between studies include:

- Difference in populations eg. Larger effects seen in sicker patients

- Difference in interventions eg. Larger effects seen with higher drug doses
- Difference in outcomes eg. When a treatment is less effective over time and different durations of outcome effect are measured

If such heterogeneity exists but investigators are unable to identify a plausible explanation the quality of evidence decreases. For example in two different trials involving mildly and severely unwell patients you might expect a difference in outcome and the quality would be unaffected by significantly different results where there is a good explanation for it.

Assessing the (in)directness of evidence (or applicability or external validity)

Two types of indirectness are considered. The first type occurs when considering the use of two or more different interventions, for example when comparing two different drugs. Trials evaluating the two drugs don't compare them head-to-head but only with a third drug or with placebo. In this situation, only indirect comparison is possible and the quality of evidence is lower.

The second type of indirectness occurs when studies have different study populations, interventions, comparisons against which the intervention is compared and outcomes of interest. This is only relevant where such differences are likely to lead to differences in the size of effect and would not necessarily be applied where an intervention would be expected to have the same effect across most groups of patients.

Imprecision

Imprecision occurs when studies are underpowered and the confidence intervals are wide and consistent with both the potential for harm and benefit, and as a result, the outcome is uncertain. The quality is reduced.

For **dichotomous outcomes**, downgrade the quality of evidence for any of the following reasons:

1. Total (cumulative) sample size is lower than the calculated optimal information size
2. Total number of events is less than 300
3. 95% confidence interval (or alternative estimate of precision) around the pooled or best estimate of effect includes no effect and:
 - a. If recommending in favour of an intervention – the upper confidence limit includes an effect that, if it were real, would represent a benefit that, given the downsides, would still be worth it
 - b. If recommending against an intervention – the lower confidence limit includes an effect that, if it were real, would represent a harm that, given the benefits, would still be unacceptable.
4. 95% confidence interval (or alternative estimate of precision) around the pooled or best estimate of effect excludes no effect but:
 - a. If recommending in favour of an intervention – the lower confidence limit crosses a threshold below which, given the downsides of the intervention, one would not recommend the intervention

- b. If recommending against an intervention – the upper confidence limit crosses a threshold above which, given the benefits of an intervention, one would recommend the intervention.

When even rates are very low, 95% confidence intervals around relative effects can be very wide, but 95% confidence intervals around absolute effects may be narrow. Under such circumstances one may not downgrade the quality of evidence for imprecision.

For **continuous outcomes**, downgrade the quality of evidence when:

1. 95% confidence interval includes no effect and the upper or lower confidence limit crosses the minimal important difference, either for benefit or harm
2. If the minimal important difference is not known or use of different outcomes measures required calculation of an effect size, downgrade if the upper or lower confidence limit causes an effect size of 0.5 in either direction.

Publication bias

Publication bias is where the decision to publish a study is based upon the direction or statistical significance of results and in most cases occurs when only studies with statistically significant are published. It is difficult to assess publication bias and is based upon judgement of clinicians. In areas where only small trials funded by industry are available, the risk of publication bias is more likely.

Reporting bias

Selective reporting of studies, typically by neglecting to report those that show no effect. Be suspicious when the published evidence is limited to a small number of trials, all of which were funded by a for-profit organization

FACTORS THAT INCREASE QUALITY GRADE

Large magnitude of effect

When observational studies yield large or very large and consistent estimates about the magnitude of an effect we can be confident about the result. The observational study design is likely to overestimate the true effect size. But where there is a large effect size, it cannot be explained purely from the study design and is likely to represent a true effect. The larger the magnitude of effect, the stronger the evidence.

Magnitude of effect	Effect measure	Quality of evidence
Large	RR>2 or <0.5	Upgrade 1 level
Very Large	RR>5 or <0.2	Upgrade 2 levels

Plausible confounding

In observational studies, plausible biases may cause an underestimation of the true treatment effect.

For example, if only sicker patients receive an experimental intervention yet the outcome is better than less sick patients, it is likely that the actual effect size is larger than described in the study. This will increase the quality, and you should upgrade by one level.

Dose response gradient

In an observational study where a dose-response gradient is seen, this may increase confidence in the findings and increase the quality.

The reasons for upgrading or downgrading the quality of evidence should be stated and given as a footnote below the table.

The “quality of the evidence” is then categorised as ‘high’, ‘moderate’, ‘low’ or ‘very low’ as per the definitions below:

Quality	Definition
High	Further research is very unlikely to change our confidence in the estimate
Moderate	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
Low	Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
Very Low	Any estimate of effect is very uncertain

This can then be entered under the column, “Quality” in the “Summary of findings” half of the GRADE table

The second aspect of a GRADE table is the summary of findings. The results will need to be extracted from the review or meta-analysis and represented as following:

Dichotomous outcomes:

- Total number of patients in each group
- Total number with event
- An estimate of control group list (control event rate)
- Effect size: Relative risk or odds ratios, absolute differences and 95% confidence intervals

Continuous outcomes:

- Total number of patients in each group
- Summary estimate of effect (weighted mean difference or standardised mean difference and 95% confidence interval)

It is advisable that one reviewer extracts data from the systematic reviews and/or from single studies and prepares drafts of the GRADE evidence profiles with detailed footnotes explaining the judgements that were made. Each judgement should be made explicit and available to the reader in order to increase the transparency of the whole process. These should be checked by at least one other member of the team.

2. Recommendation

The recommendation is a trade-off between benefits and harms.

It starts by considering the quality of evidence available, the results of that evidence, and the various benefits and harms that apply to the use of the intervention in question. Next, it involves the consideration of the values and the relative importance placed on each adverse and beneficial outcome considered, by the patients who will be affected by an intervention.

Finally, resource implications of an intervention should be considered. This includes consideration of cost-effectiveness and availability of specific resources in an area.

For the ICHRC this will often involve the consideration of a resource-poor setting and the necessity to balance the use of an intervention against other cheaper (or more expensive) alternatives.

The recommendation should be as follows:

Strong recommendation:

- For patients—most people in your situation would want the recommended course of action and only a small proportion would not; request discussion if the intervention is not offered
- For clinicians—most patients should receive the recommended course of action
- For policy makers—the recommendation can be adopted as a policy in most situations.

Weak recommendation:

- For patients—most people in your situation would want the recommended course of action, but many would not
- For clinicians—you should recognise that different choices will be appropriate for different patients and that you must help each patient to arrive at a management decision consistent with her or his values and preferences
- For policy makers—policy making will require substantial debate and involvement of many stakeholders.

This enables clinicians to individualise treatment options recognising that patients may have different values even when faced with overwhelming evidence for benefit or harm.

More information on GRADE

This guideline for use of GRADE in the ICHRC is based upon papers by the GRADE working group: www.gradeworkinggroup.org

GRADE guideline papers are available under the publication section of the website.

5. PREPARING AND WRITING THE REVIEW

Having formulated a search strategy, retrieved papers, evaluated the evidence, produced GRADE tables and tabulated key results, the final stage is to bring this together into the WHO report.

Though this process has much in common with Cochrane Reviews, there is a very clearly defined scope in terms of topic and target settings.

The basic template should be as follows:

TITLE

INTRODUCTION

This should be direct and pertinent. The people using the reviews will understand the background to the topic, though a brief explanation of any controversy is welcome. It should conclude with: 'This review intends to answer the question: e.g. Should zinc be used in the treatment of acute gastroenteritis?'

METHODS

There needs to be a clearly documented search strategy that specifies details of the databases (including web sites) to be searched, and the search strategy to be applied to each database. This will be very similar between reviews. Mention the databases searched, how articles were selected and graded. Also state the number of papers identified.

RESULTS

It is helpful to consider the results in terms of outcomes and include the relevant data. P values, confidence intervals and other statistical tests are vital. Systematic reviews and randomised controlled trials should really be the ideal source of your data where possible. Discuss any methodological points, and dosage information if relevant. The GRADE tables should be referenced in this section with the table presented in appendices at the end.

Ideally, only good data should be included. Aim to transmit the information in as clear a form as possible.

DISCUSSION

This should briefly put into context the findings. The main conclusions should be stated. Disclaimers on the quality of the relevant literature may be mentioned, and options for future work discussed.

SUMMARY

A concise statement of the evidence.

REFERENCES

Use the Pubmed format (which can be obtained from selecting the reference and choosing display settings (at the top of the page) then summary (text).)

The final report should be sent to one of:

Yoko Askawa: yoko.asakawa@rch.org.au

Lilian Downie: lilian.downie@rch.org.au

Julian Kelly: julian.kelly@rch.org.au .

From here they will be sent to secondary reviewers, and returned to you in order to make any requested changes.

Once the secondary review stage has been completed and any amendments made, the review will be published on www.ichrc.org. Selected reviews are also being serialised in the Journal of Tropical Paediatrics.

Reviews will be updated every few years to reflect the changing evidence base.

REFERENCES

The GRADE working group. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328:1490-1494 printed, abridged version.

<http://bmj.bmjournals.com/cgi/content/full/328/7454/1490>

Sample GRADE tables:

Example GRADE tables can be viewed in the annex or appendices of the following publications:

Report of the WHO Technical Reference Group, Paediatric HIV/ART Care Guideline Group Meeting WHO Headquarters, Geneva, Switzerland 10-11 April 2008. Available at:

www.who.int/hiv/pub/paediatric/WHO_Paediatric_ART_guideline_rev_mreport_2008.pdf

WHO Guidelines for the treatment of malaria- 2nd edition. 2010

Available at: http://whqlibdoc.who.int/publications/2010/9789241547925_eng.pdf

WHO guidelines for Management of Postpartum Haemorrhage and retained placenta. 2009.

http://www.who.int/reproductivehealth/publications/maternal_perinatal_health/9789241598514/en/

ACKNOWLEDGEMENTS

With thanks to Marshall Dozier and the University of Edinburgh Library Staff for their advice on literature searching.

Appendix: Summary of GRADE methodology

THE GRADE APPROACH TO ASSESSMENT OF EVIDENCE**Table 1: Ranking the Quality of Evidence**

Quality of evidence (summary score)	Study design	Lower if *	Higher if *
High (4)	Randomized trial or valid accuracy study for diagnostic tests	Study quality: -1 Serious limitations -2 Very serious limitations	Strong association: +1 Strong, no plausible confounders, consistent and direct evidence
Moderate (3)			
Low (2)	Observational study or indirect accuracy studies for diagnostic tests	-1 Important inconsistency Directness: -1 Some uncertainty -2 Major uncertainty	+2 Very strong, no major threats to validity and direct evidence +1 Evidence of a dose response gradient
Very low (1)		-1 Sparse or imprecise data -1 High probability of reporting bias	

* 1 = move up or down one grade (for example from high to intermediate)

2 = move up or down two grades (for example from high to low)

- ⊕⊕⊕⊕ High = Further research is very unlikely to change our confidence in the estimate of effect.
- ⊕⊕⊕○ Moderate = Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
- ⊕⊕○○ Low = Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
- ⊕○○○ Very low = Any estimate of effect is very uncertain.

Limitations = include problems in study design, such as for RCTs, lack of blinding or allocation concealment, incomplete reporting, selective outcome reporting, or use of unvalidated outcomes measures.

Inconsistency = Differences exist in the direction and size of the effect across the studies.

Uncertainty = Indirect comparisons or indirect populations have been considered across the studies, and there may be compelling reasons to expect important differences in the size of the effect.

Validity = Patients participating in RCTs are assessed to have same risk and/or mortality as non enrolled patients in whom the intervention is expected to be required.