

# Standards for CHERG reviews of intervention effects on child survival

Neff Walker,<sup>1\*</sup> Christa Fischer-Walker,<sup>1</sup> Jennifer Bryce,<sup>1</sup> Rajiv Bahl,<sup>2</sup> Simon Cousens<sup>3</sup> writing for the CHERG Review Groups on Intervention Effects

<sup>1</sup>Institute for International Programs, Bloomberg School of Public Health, The Johns Hopkins University, Baltimore, MD, USA,

<sup>2</sup>Department of Child and Adolescent Health, World Health Organization, Geneva, Switzerland and <sup>3</sup>Department of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK.

\*Corresponding author. Institute for International Programs, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, MD 21205, USA. E-mail: pneffwalker@yahoo.com

---

<b>Background</b>	The Lives Saved Tool (LiST) uses estimates of the effects of interventions on cause-specific child mortality as a basis for generating projections of child lives that could be saved by increasing coverage of effective interventions. Estimates of intervention effects are an essential element of LiST, and need to reflect the best available scientific evidence. This article describes the guidelines developed by the Child Health Epidemiology Reference Group (CHERG) that are applied by scientists conducting reviews of intervention effects for use in <i>LiST</i> .
<b>Methods</b>	The guidelines build on and extend those developed by the Cochrane Collaboration and the Working Group for Grading of Recommendations Assessment, Development and Evaluation (GRADE). They reflect the experience gained by the CHERG intervention review groups in conducting the reviews published in this volume, and will continue to be refined through future reviews.
<b>Presentation of the guidelines</b>	Expected products and guidelines are described for six steps in the CHERG intervention review process: (i) defining the scope of the review; (ii) conducting the literature search; (iii) extracting information from individual studies; (iv) assessing and summarizing the evidence; (v) translating the evidence into estimates of intervention effects and (vi) presenting the results.
<b>Conclusions</b>	The CHERG intervention reviews represent an ambitious effort to summarize existing evidence and use it as the basis for supporting sound public health decision making through <i>LiST</i> . These efforts will continue, and a similar process is now under way to assess intervention effects for reducing maternal mortality.
<b>Keywords</b>	Child survival, child mortality, interventions, modeling, projections, efficacy, effectiveness

---

## Introduction

An intervention is currently included in the Lives Saved Tool (LiST) only if there is evidence that it

reduces mortality among children <5 years of age, either directly or through effects on pregnant or recently delivered women. Future versions of LiST

will incorporate additional interventions and generate projections for maternal mortality and stillbirths. The Child Health Epidemiology Reference Group (CHERG) conducts the technical reviews used to make these inclusion decisions, as well as to determine the size of the effect and associated uncertainty. This article presents the standard guidelines for the intervention reviews presented in this volume and that will be applied in future CHERG intervention reviews. Our aim is to present the guidelines in sufficient detail to allow each review to be replicated and updated as new evidence becomes available, and to ensure that the results of the reviews are consistent, comparable and appropriate for use in LiST.

The guidelines developed for the CHERG reviews were based on the extensive work that has already been done on this topic elsewhere, such as by the Cochrane Collaboration<sup>1</sup> and the QUOROM<sup>2</sup> standards for meta-analysis of randomized controlled trials. However, as the end goal of the papers was to develop an estimate of the effect of an intervention in reducing cause-specific mortality, the CHERG had to develop methods to guide the systematic review, meta-analysis, and then rules for extrapolating from existing data to the specific outcome of interest. We have tried to incorporate existing standards for reviews of the scientific literature into the guidelines whenever possible. For example, the standards set by the Cochrane reviews<sup>1</sup> for specifying the search procedures and the need to be explicit about the inclusion and exclusion criteria used in selecting studies for consideration are clear and widely accepted, and we have therefore not re-invented them. In addition we have tried to abide by the recommendations for systematic reviews and meta-analyses laid out by several groups such as Cochrane<sup>1</sup>, QUOROM<sup>2</sup> in the reporting. CHERG has also adopted the system developed by the Working Group for Grading of Recommendations Assessment, Development and Evaluation (GRADE)<sup>3</sup> as the standard to be used in CHERG intervention reviews; we explain how GRADE was adapted to address the specific aims and content of the CHERG intervention reviews. For methodological issues not addressed in existing standards, CHERG has developed new standards that we describe here.

## Methods

Draft guidelines were prepared based on the previous experience of CHERG, and reviewed and revised in

two meetings of the leaders of the intervention review groups. The guidelines continue to evolve as new issues are identified in the process of conducting and reporting on the reviews and incorporating the results into LiST.

## Presentation of the guidelines

CHERG developed a conceptual framework for the intervention reviews that includes six steps. Figure 1 presents these steps as sequential, but in practice they are iterative with later steps often forcing a return to an earlier step for further work or clarification. We present the guidelines for each step below, beginning with a description of the expected product followed by an explanation of the process to be followed in generating the product.

### Step 1: Define the scope of the review

#### *Expected product*

The products of this step are clear descriptions of (i) the health outcomes addressed; (ii) the intervention(s) and (iii) the inclusion and exclusion criteria used in the search including geographic and temporal parameters.

#### *Process*

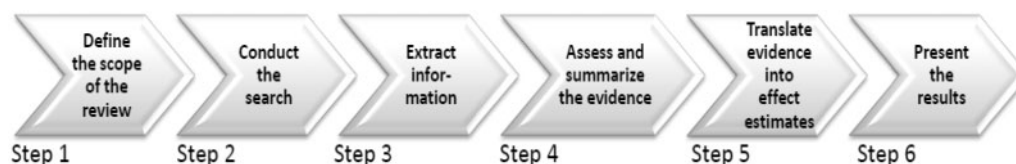
##### *Defining the health outcomes*

An intervention is considered for inclusion in LiST only if there is evidence suggesting an effect on mortality or severe morbidity among children <5 years of age, including effects realized through interventions delivered to pregnant and recently delivered women.

This effect must operate either directly on one of the causes of death considered by LiST or indirectly through diarrhoea incidence, stunting, severe acute malnutrition or intrauterine growth restriction (Box 1), although evidence of effects on all-cause mortality must also be considered. Where mortality data are scarce, review groups also review studies that provide evidence of effects on disease incidence, prevalence, duration or severity/hospitalizations. Evidence of potential harmful effects is also included as a part of the reviews.

##### *Defining the interventions*

The review group specified the characteristics of the intervention under review, including the range of intervention duration, dose or other factors. Whenever possible, reviews should focus on single



**Figure 1** Steps in the CHERG intervention reviews

**BOX 1**

Direct and indirect causes of death included in the LiST model	
<b>Direct causes</b>	
▪	Diarrhea
▪	Pneumonia
▪	Malaria
▪	Measles
▪	HIV
▪	Birth asphyxia
▪	Complications of prematurity
▪	Severe neonatal infection
▪	Neonatal tetanus
▪	Congenital abnormalities
▪	Other
<b>Indirect causes</b>	
▪	Diarrhea incidence
▪	Stunting
▪	Intrauterine growth retardation
▪	Severe acute malnutrition

preventive or treatment measures rather than intervention ‘packages’, because *LiST* models in general uses individual interventions and the effects are on individuals rather than populations. For example, *LiST* models the effect of treating an individual child suffering from acute watery diarrhoea with ORS, rather than the population-level effects of programs to distribute and promote ORS. For this reason the primary focus of the intervention reviews is on *proximal* interventions (e.g. treatment with ORS) rather than *distal* strategies (e.g. mass media promotion of ORS), although information on the levels of coverage of proximal interventions that can be achieved through distal strategies is important information for setting realistic target coverage rates for the proximal interventions in *LiST*.

*Time periods*

The reviews include relevant studies from all time periods for which data are available. Methods for taking into account how recent the data are in assessing the quality of the evidence are addressed in Step 4 below.

*Geographic scope*

Low-income countries account for over 99% of the burden of maternal and child mortality worldwide.<sup>4</sup> When sufficient studies from low-income countries are available, and in particular those with a high burden of the outcome addressed by the review, studies from high-income countries are excluded from the review. Data from all countries are considered

when few or no data are available from low-income countries. Similarly, reviews of interventions effective against causes that vary geographically (e.g. malaria or HIV) will focus primarily on geographic areas where these causes of death are prevalent.

*Other parameters*

Individual review groups are responsible for specifying the scope of the review with respect to other parameters of particular importance to the intervention under review or its effects. For example some interventions can have a pure placebo as the control arm of a trial while for other interventions this is not practical. The additional parameters to be included will be defined by the intervention review groups.

**Step 2: Conducting the search***Expected product*

The product of this step is a description of the databases and information sources that have been searched, in sufficient detail to allow the search to be reproduced.

*Process*

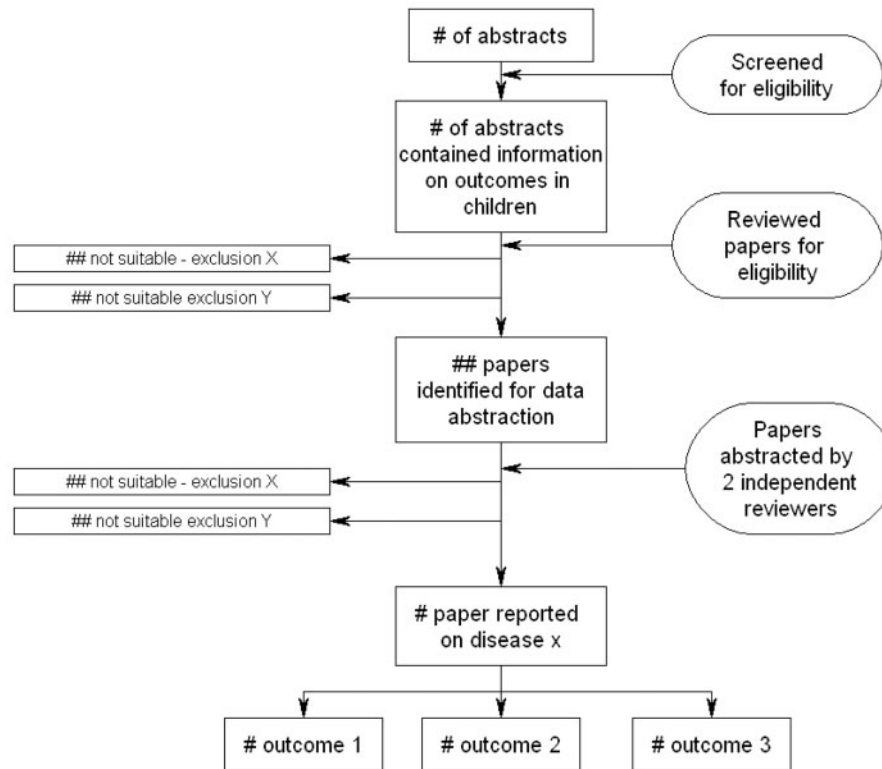
CHERG intervention reviews cast as wide a net as possible including published and unpublished evidence. Full documentation of all data sources is available in the Methods section of each report.

All searches include large publication databases such as PubMed, Cochrane Library and the WHO Regional Databases. When relevant, additional databases are also included such as CAB abstracts,<sup>5</sup> the System for Information on Grey Literature in Europe (SIGLE),<sup>6</sup> EMBASE,<sup>7</sup> Web of Science<sup>8</sup> and Popline.<sup>9</sup>

Because publication bias often excludes studies with negative findings, review groups make special efforts to include relevant studies available only in the grey literature. The strategies used include searching the System for Information on Grey Literature in Europe (SIGLE)<sup>10</sup> and databases that specialize in conference abstracts, as well as less formal strategies such as word of mouth and direct contact with the principal investigators. Results of unpublished research are included in a review when abstracts provide detailed methodological descriptions or full reports are available. Reviews do not include evidence substantiated only through personal communication.

Search terms relevant to the intervention are included in combination with ‘morbidity’, ‘incidence’, ‘prevalence’ and ‘mortality’. All combinations of search terms used are documented in the Methods section of the reviews.

The overall search strategy and findings are documented using a standard flow diagram, as shown in Figure 2. Previous reviews or meta-analyses are used as sources for the identification of possible studies to be included in the CHERG reviews, but each paper is re-reviewed based on the inclusion/exclusion criteria established by the CHERG review group and if



**Figure 2** Standard flow diagram for CHERG intervention reviews, showing the search strategy and outcomes

considered for inclusion is subject to the quality assessment procedures described in Step 4.

Reviews to date have been conducted using English search terms, but include papers published in other languages. CHERG is working to expand its scientific base to include searches conducted in other languages, and especially Chinese.

### Step 3: Extracting individual study information

#### *Expected product*

The product of this step is a table that presents for each study meeting the inclusion criteria and from which data are abstracted, characteristics of the study and all information needed to arrive at a judgement about the quality of the study methods and the relevance of its findings to the review.

#### *Process*

All research studies and reviews meeting the inclusion criteria are abstracted into a rectangular data base that can be accessed through Excel. Box 2 lists the core variables included in the database for each review; each review group determines the additional items that will be abstracted for their specific review. Each paper in this volume includes a web appendix that contains the full database as Supplementary

Table 1. These procedures are consistent with those recommended by GRADE.<sup>3,11</sup>

All CHERG reviews are based on abstraction of information from studies by two independent researchers, with discrepancies resolved by an outside supervisor. Exceptions to the double abstraction standard are explained, and details are provided about the steps taken to prevent, identify and correct abstraction errors.

### Step 4: Assessing and summarizing the evidence

#### *Expected product*

The product of this step is a table summarizing the overall quality and relevance of the available data that have the potential to contribute to an estimate of intervention effects, organized by health outcome.

#### *Process*

##### *Judging the quality of individual studies*

We use four categories of criteria for judging quality of study evidence, similar but not identical to those recommended by GRADE: (i) study design; (ii) study quality; (iii) relevance to the objectives of the review and (iv) consistency across studies. CHERG review groups applied the first three categories to individual studies in the database constructed in Step 3, to assign evidence

**BOX 2**

<b>Core variables for each study abstracted into the review data base*</b>
<b>Study identifiers and context</b>
<b>Study design and limitations</b>
<b>Intervention characteristics</b>
<b>Outcome effects</b>
<b>Summary of limitations</b>

\*This data base is available as supplementary table 1 for each of the reviews published in this volume.

quality scores to individual studies and then considered the consistency of findings across studies before assigning an overall evidence quality score.

We adapted the GRADE process to reflect the particular needs of the CHERG intervention reviews. Box 3 presents the specific criteria used for study design and study quality. Even the strongest studies, however, may not be directly relevant to the intervention review. The extent to which the study findings can be applied to the generation of effect estimates may be compromised to a greater or lesser degree because of variations in the health outcomes used as dependent variables; the duration, intensity or delivery strategy used in the intervention study; or the extent to which the study population is comparable to the women and children in low- and

middle-income countries where most deaths occur. The evidence quality score can also be downgraded for older studies, especially when changes in epidemiology or context suggest that the findings may have decreased relevance (e.g. antibiotic resistance).

Each review group makes a careful assessment of relevance for each study and records their decisions in Supplementary Table 1; how these judgements are applied in awarding a final study score are explained in detail in each paper.

The final step in judging the overall quality of evidence across studies is to assess the consistency of findings of an effect of the intervention on specific health outcomes. This is discussed in detail in Step 5, but affects Step 4 because any study that produces results that vary widely from those of several other studies of

**BOX 3**

Grading system for study design and study quality Used in the CHERG reviews of intervention effects
<p><b>Scoring system</b></p> <p>Each study is assigned a single score based on a four-point continuum: 'high', 'moderate', 'low' or 'very low'.</p>
<p><b>Score for study design</b></p> <p>Randomized trials receive an initial grade of 'high'; observational studies receive an initial grade of 'low'; evidence generated through other designs receives an initial grade of 'very low'.</p>
<p><b>Adjustment for study quality</b></p> <p>Review groups conduct a careful assessment of study methods and execution, and adjust the initial score assigned for study design if necessary based on the rules listed below. This two-step procedure allows for the promotion of well designed observational studies and the demotion of poorly executed RCTs, and is particularly appropriate given that many Cochrane reviews of RCTs conclude that the overall quality of the studies is low despite randomization.</p> <p><u>Rules for adjusting study design scores</u></p> <ul style="list-style-type: none"> <li>▪ A score should be downgraded one level if there are serious questions with regard to the overall quality of the study methods or sample size, evidence or suspicion of reporting bias or inconsistencies with other data sources that raise doubts about the validity of the results.</li> <li>▪ A score should be downgraded one level if there are very serious questions with regard to how the study was executed (e.g., methods used for surveyor training, quality of supervision, etc.)</li> <li>▪ A score should be upgraded one level if the researchers either controlled or accounted for all plausible confounders that would have reduced the effect of the intervention on the health outcome.</li> </ul>

acceptable quality is then re-examined to determine whether the reasons for this inconsistency can be explained. This sometimes includes contacting the study investigators to discuss the study methods in greater detail than is available in the written report.

The results of this part of the CHERG review paper are presented as Table 1 and in each paper in this volume. The table is organized by health outcome, including all-cause under-five mortality. A single paper may appear more than once in the table if the

**Table 1** Sample table for presenting results of the quality assessment for individual studies (Step 4 in the CHERG intervention review process)

Table 1. Quality assessment of trials of zinc for the treatment of diarrhea								
Quality Assessment						Summary of Findings		
No of studies (ref)	Design	Limitations	Consistency	Directness		No of events		Relative Risk (95% CI)
				Generalizability to population of interest	Generalizability to intervention of interest	Intervention	Control	
<i>Outcome (Outcome specific quality)</i>								

findings related to both all-cause and cause-specific mortality (or severe morbidity).

#### *Assessing the overall quality of evidence*

After all the individual studies have been reviewed, abstracted and graded, the review groups must follow a transparent process to arrive at a judgement about the overall quality of evidence supporting an effect of the intervention on specific health outcome(s). This judgement includes three separate components: (i) the volume and consistency of the evidence; (ii) the size of the effect, or risk ratio; and (iii) the strength of the statistical evidence for an association between the intervention and the health outcome as reflected in the *P*-value. At this point in the CHERG review process, each of these components has been quantified and the remaining challenge is combining the evidence generated from various sources into a single, replicable estimate of effects.

Data from a single study are unlikely to provide high-quality evidence on their own—one of the tenets of the scientific method is that results should be replicable (and replicated). Therefore, the production of a point estimate of an intervention's effect for use in *LiST* requires combining information from several studies into a single measure.

The basic approach is to take an average of the effect estimates from the individual studies through the use of meta-analysis methods. Standard meta-analyses involve calculating a weighted average of the study-specific estimates, giving more weight to studies that contain more information. The amount of information will depend on the sample size and the frequency of the event of interest in the study population among other things. Although many of the interventions addressed by CHERG and included in *LiST* have been reviewed systematically in the past (e.g. Cochrane database), as described in Step 3 above these, reviews will generally need to be updated to include new evidence identified through the CHERG reviews. For interventions for which no systematic review has ever been published or where systematic reviews have considered limited outcomes or used limited study designs, systematic reviews with meta-analysis are required as a part of the CHERG review process.

The type of meta-analysis performed depends on the heterogeneity of the findings across the available studies, which is assessed systematically as a part of the CHERG process. Unexplained heterogeneity in effects, or inconsistency in findings, is one reason to downgrade the overall quality of evidence score as explained in Box 3.

If there is not strong evidence of heterogeneity, review groups perform a *fixed effects* meta-analysis. This method assumes that the intervention effect is the same in all studies included in the analysis and calculates the point estimate for this fixed effect and its 95% confidence interval.

If there is evidence of heterogeneity across studies, a fixed effects meta-analysis is inappropriate. Review groups then work to identify and understand the causes of the heterogeneity. For example, perhaps the variability can be explained by the use of different study designs, such as consistent differences between results generated by cohort studies relative to those generated by case-control studies. In this instance, if there was no evidence of heterogeneity across cohort studies one might decide to perform a fixed effects meta-analysis limited to the available cohort studies.

If there remains unexplained heterogeneity in effect, but consistency in the direction of the effect, the review groups conduct a *random effects* meta-analysis, which assumes that the effect varies between studies rather than being constant. In practical terms, a random effects meta-analysis differs from a fixed effects analysis in the weighting system it uses; a random effects meta-analysis gives relatively more weight to small studies than a fixed effects meta-analysis. The interpretation of the point estimate also differs—it is now the average effect across studies—and the confidence interval around the point estimate will be wider.

If there is unexplained heterogeneity in the effect and the direction of the effect is positive in some studies and negative in others, further reviews of the evidence are conducted to clarify the reasons for these differences before inclusion of the intervention in *LiST*.

Consideration of heterogeneity should not be based solely on point estimates. It is entirely possible that one study among a set of 10 or more will produce a

negative point estimate, but perhaps with a wide confidence interval such that the data from the study are compatible with a positive effect of the intervention. For this reason, it is important to consider the uncertainty in the individual study estimates and whether the observed variation could have arisen by chance.

#### *Assigning a grade to the overall quality of evidence*

Each review group must assign an overall score reflecting their degree of confidence in the evidence supporting an association between an intervention and each of the relevant health outcomes, and their recommendation concerning inclusion of the intervention in the LiST model and the effectiveness value it should be assigned. The following guidelines are used to promote consistency across the CHERG review groups.

- A score of 'high' means that the reviewers are confident of both the inclusion of the intervention in the *LiST* model and the effectiveness estimate. Although additional evidence will continue to be incorporated, it is unlikely that new evidence will change the inclusion of intervention in the model or dramatically change the size of the effect estimate.
- A score of 'moderate' means that the reviewers are confident of the inclusion of the intervention in the model and given available information are presenting the best estimate of effectiveness. Additional research may alter the size of the effect but is not likely to change the inclusion in the model.
- A score of 'low' means that the reviewers recommend inclusion of the intervention in the model but that the size of the benefit should continue to be studied. Additional data are needed to increase confidence and will likely change the effect size.
- A score of 'very low' means that the reviewers do not believe there is sufficient evidence of benefit to recommend inclusion of the intervention in the LiST at this time unless there are compelling biological grounds for including the intervention. Additional data may change this classification in future versions of the tool.

### **Step 5: Translate the evidence into effect estimates**

#### *Expected product*

The product of this step is a 'best' estimate of the effect of the intervention on reducing under-five mortality due to a specific cause, with uncertainty expressed as 'lower' and 'higher' estimates.

#### *Process*

##### *Rules for generating best estimates of effect*

Each of the CHERG review groups produces a recommendation about whether the evidence of effect justifies inclusion of the intervention under study in

*LiST*, and if so, a point estimate of that effect and associated uncertainty. None of the CHERG intervention reviews conducted to date is based on evidence drawn solely from RCTs measuring effects on cause-specific mortality, and therefore judgements have to be made based on data of different levels of quality and relevance. CHERG therefore has developed a set of explicit rules to promote consistency and reproducibility. Box 4 presents the rules governing the estimation of effects that CHERG review groups have developed and applied to date; we expect this list to evolve as experience is gained through further review. All rules were developed to be conservative, leaning toward underestimation rather than overestimation of effects, because *LiST* will be applied in real-world settings where intervention quality and delivery are affected by many factors some of which may be beyond the control of program managers.<sup>12</sup>

#### *Translating statistical evidence into recommendations for LiST*

Meta-analyses generate summary risk/rate/odds ratios with associated confidence intervals, and all CHERG review groups report these uncertainty bounds as a standard part of their results. In addition, *P*-values are reported as a useful summary of the strength of the statistical evidence of an association between an intervention and a health outcome under the assumption of no bias due to confounding.

The product need for LiST is a point estimate with higher and lower bounds, and a summary judgement about inclusion of the intervention in the model. Table 2 presents the guidelines used by CHERG review groups to move from statistical evidence to recommendations about inclusion of an intervention in LiST, taking into account the number of studies contributing to the evidence, their study designs and the *P*-value emerging from the meta-analysis.

#### *Generating expert judgements of effects*

When the available evidence is insufficient to support the generation of an estimate, the review group uses systematic procedures based on the Delphi technique to generate an estimate based on expert opinion.<sup>13</sup> Standard methods for conducting a Delphi in ways that produce reliable consensus are available elsewhere;<sup>14</sup> for the CHERG intervention reviews, two particularly important aspects are the choice of Delphi participants (the group should represent the appropriate epidemiological experience in public health settings in low-income settings) and a full documentation of how the method is applied including a listing of all individuals who participate in the process as a subset of the intervention review group. The design of the Delphi process should aim to generate conservative estimates of effects.



**BOX 4****Rules for generating estimated intervention effects for use in *LiST*****Rule 0**

*If* there are less than 50 total events in an outcome category...

*Then* we consider there to be insufficient evidence at this level.

**Rule 1**

*If* there is no evidence of effect on cause-specific mortality,

*And* there is evidence of effect on all-cause mortality...

*Then* translate all-cause into cause-specific, and downgrade the quality score by one level.

**Rule 2**

*If* there is high- or moderate-quality evidence of effect on cause-specific mortality...

*Then* use the mortality effect.

**Rule 3**

*If* there is low- or very low-quality evidence of effect on cause-specific mortality,

*And* there is high- or moderate-quality evidence of effect on serious morbidity...

*Then* use the smaller of the two effects.

**Rule 4**

*If* there is low- or very low-quality evidence of effect on cause-specific mortality,

*And* there is low- or very low-quality evidence of effect on serious morbidity,

*And* there is high- or moderate-quality evidence of effect on mild/moderate morbidity...

*Then* take the smallest of the three effects.

**Rule 5**

*If* there is no evidence (see rule zero) relating to cause-specific mortality,

*And* there is high-quality evidence of effect on serious morbidity...

*Then* use this value as the estimate.

**Rule 6**

*If* there is moderate-, low- or very low-quality evidence of effect on serious morbidity,

*And* there is high- or moderate-quality evidence of effect on mild/moderate morbidity...

*Then* take the smaller of the two effects.

**Rule 7**

*If* there is no evidence relating to cause-specific mortality or serious morbidity,

*And* there is high-quality evidence of a reduction in mild/moderate morbidity...

*Then* use the estimate for the most severe outcome available as the estimate.

**Rule 8**




*If* there is no evidence that meets the conditions of rules 0-7 above,

*And* there is a clear and obvious clinical/preventive benefit of the intervention...

*Then* use expert judgment of effectiveness.

**Table 2** Guidelines for interpreting statistical test results from meta-analyses

Studies contributing to the estimate	P-value			
	P > 0.10	0.10 > P > 0.01	P < 0.01	P < 0.001
	Weak or no evidence of association (observational studies) or effect (trials).	Weak evidence of association (observational studies) or effect (trials).	Evidence of association (observational studies) or effect (trials).	Strong evidence of association (observational studies) or effect (trials).
Single RCT				
Multiple RCTs				
Single observational study				
Multiple observational studies				

-  Insufficient evidence to recommend inclusion unless additional evidence is available from other sources.
-  Will require careful consideration and judgment before recommendation for inclusion.
-  Sufficient statistical evidence to recommend inclusion if other GRADE criteria met.

**Step 6: Presenting the results**

*Expected product*

The product of this step is a written report that provides full documentation of the review methods and results and has been reviewed for completeness by appropriate technical experts at WHO and UNICEF, in a format that allows readers to make comparisons easily with other CHERG intervention review reports.

*Process*

*Standard report elements*

Although the presentation of results varies across interventions, all reports contain the following elements.

*Definition of the intervention(s):* For reviews covering multiple interventions, separate definitions are provided for each. This is especially important for sets

of interventions that are delivered together as ‘packages’; each biological or behavioural intervention is defined individually to avoid double counting of effects that may be included in more than one package or review.

*Explanation of rules applied and exceptions to guidelines:* The review should state clearly which of the rules of evidence (Box 4) were applied. The application of other parts of the guidelines is summarized with reference to this paper to minimize repetition. Any exception to these guidelines is justified and described in detail.

*Box summarizing conclusions, recommendations for LiST and LiST inputs if appropriate:* Each review group presents this information in a standard format at the close of the paper, preventing any misunderstanding of their final results and conclusions. Effect estimates are presented for the four age categories in the

postnatal period where possible; if insufficient data are available to support estimates by age categories, this is stated.

*Review limitations:* Major limitations in the review process are discussed in the Analysis and results sections of each review, but should be summarized succinctly in the discussion as a basis for the interpretation of the findings. Examples of important limitations include a lack of geographic diversity in study sites or the use of modelling to move from the measured impact on morbidity to the estimated impact on mortality.

*Recommendations for further research:* The Discussion section of each paper includes recommendations from the review group about priority gaps in evidence that should be addressed.

## Conclusions

The CHERG reviews of intervention effectiveness are the result of an ambitious and innovative process for bringing together scientists from multiple disciplines and with specific areas of expertise to generate state-of-the-art estimates of the extent to which individual biological and behavioural interventions can reduce deaths among children under five. This effort is motivated by a shared aim of supporting sound decision making about how best to use scarce public health resources to reduce deaths and achieve the Millennium Development Goals, and guided commitments to the scientific principles of transparency and replication. The guidelines presented in this article reflect thousands of hours of experience by members of the CHERG intervention review groups, and this is only the beginning. Reviews of interventions for reducing maternal mortality are under way now, and additional reviews will be commissioned as further evidence on existing interventions is produced and new interventions are developed.

Some important methodological issues remain to be addressed. They include the need for better estimates of uncertainty, and greater standardization in the use of expert judgements. CHERG will continue this work and disseminate it widely, leading over time to more complete use of existing evidence and clearer priorities for future research.

## Supplementary Data

Supplementary Data are available at *IJE* Online.

## Funding

The development of *LiST* is supported by the Bill & Melinda Gates Foundation through a grant to the United States Fund for UNICEF and implemented

by the Institute for International Programs of the Johns Hopkins Bloomberg School of Public Health. This work was supported in part by a grant to the US Fund for UNICEF from the Bill & Melinda Gates Foundation (grant 43386) to 'Promote evidence-based decision making in designing maternal, neonatal and child health interventions in low- and middle-income countries'.

## Acknowledgements

This work is the result of inputs from the many CHERG investigators identified as authors of the intervention review papers published in this volume.

**Conflict of interest:** None declared.

## References

- 1 Cochrane website. <http://www.cochrane.org> (13 June 2009, date last accessed).
- 2 Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF for the QUORM Group. Improving the quality of reports of meta-analyses of randomized controlled trial: the QUOROM statement. *Lancet* 1999;**354**:1896–900.
- 3 GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004;**328**:1490–98.
- 4 Bryce J, Boschi-Pinto C, Shibuya K, Black RE and the WHO Child Health Epidemiology Reference Group. WHO estimates of the causes of death in children. *Lancet* 2005;**365**:1147–52.
- 5 CAB abstracts. Available at <http://www.cabi.org/datapage.asp?iDocID=165> (13 June 2009, date last accessed).
- 6 System for Information on Grey Literature in Europe. Available at <http://opensigle.inist.fr/> (13 June 2009, date last accessed).
- 7 EMBASE. Available at <http://info.embase.com/about/what.shtml> (13 June 2009, date last accessed).
- 8 Web of Science. Available at <http://www.isiwebofknowledge.com/> (13 June 2009, date last accessed).
- 9 Popline. <http://db.jhuccp.org/ics-wpd/popweb/> (13 June 2009, date last accessed).
- 10 System for Information on Grey Literature in Europe. Available at <http://opensigle.inist.fr/> (13 June 2009, date last accessed).
- 11 Guyatt GH, Oxman AD, Vist GE *et al.* for the Grade Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;**336**:924–26.
- 12 Victora CG, Habicht JB, Bryce J. Evidence-based public health: moving beyond randomized trials. *Am J Public Health* 2004;**94**:400–5.
- 13 Linstone H, Turoff M (eds). *The Delphi Method: Techniques and Applications*. Reading, MA: Addison-Wesley, 1975.
- 14 Powell C. The Dephi technique: myths and realities. *J Adv Nurs* 2003;**31**:376–82.