Health Evidence™
Helping public health use best evidence in practice

Quality Assessment Tool – Review Articles

Instructions for completion:

Please refer to the attached dictionary for definition of terms and instructions for completing each section. For each criteria, score by placing a check mark in the appropriate box.

First Author: _____
Year: _____
Journal: _____
Reviewer: _____

CRITERION	YES	NO
Q1. Did the authors have a clearly focused question [population, intervention (strategy), and outcome(s)]?		
Q2. Were appropriate inclusion criteria used to select primary studies?		
Q3. Did the authors describe a search strategy that was comprehensive? *Circle all strategies used*: ▪ health databases ▪ psychological databases ▪ social science databases ▪ educational databases ▪ other ▪ handsearching ▪ key informants ▪ reference lists ▪ unpublished		
Q4. Did search strategy cover an adequate number of years?		

For questions **5, 6**, and **8**, please choose the column relating to the appropriate methodology. Strike a line through the column that does not apply.

Q5. **Quantitative reviews:** Did the authors describe the level of evidence in the primary studies included in the review? ▪ Level I → RCTs only ▪ Level II → non-randomized, cohort, case-control ▪ Level III → uncontrolled studies	**Q5.** **Qualitative reviews:** Do the authors provide a clear description of the range of methods in each of the primary studies included in the review?		
Q6. **Quantitative reviews:** Did the review assess the methodological quality of the primary studies, including: *(Minimum requirement: **4/7** of the following)* ▪ Research design ▪ Study sample ▪ Participation rates ▪ Sources of bias (confounders, respondent bias) ▪ Data collection (measurement of independent/dependent variables) ▪ Follow-up/attrition rates ▪ Data analysis	**Q6.** **Qualitative reviews:** Did the review assess the methodological quality of the primary studies, including: *(Minimum requirement: **4/7** of the following)* ▪ Suitability of methodology/paradigm to the research question ▪ Sampling (selection of participants / settings / documentation) ▪ Clear description of context, data collection and data analysis ▪ Rigor: → Audit trail → Some coding by 2 or more coders, if appropriate → Deviant case analysis (negative cases) → Respondent validation (member checking) ▪ Triangulation ▪ Reflexivity (researcher and research process) ▪ Relevance (credibility, consistency, applicability, transferability)		
Q7. Are the results of the review transparent?			
Q8. **Quantitative reviews:** Was it appropriate to combine the findings of results across studies?	**Q8.** **Qualitative reviews:** Is there a description of how reviewers determined results were similar enough across studies to compare or combine them?		
Q9. Were appropriate methods used for combining or comparing results across studies?			
Q10. Do the data support the author's interpretation?			
		TOTAL SCORE:	

Quality Assessment Rating: (Circle one)

Strong (total score 8 – 10)

Moderate (total score 5 – 7)

Weak (total score 4 or less)

A systematic review is a research approach to accessing, acquiring, quality assessing, and synthesizing a body of research on a particular topic. All phases of systematic review development should be well described such that the process is transparent and replicable by others.

Q1 | Clearly focused research question

The review should have a clearly focused research question that contains the following components: **P**opulation, **I**ntervention, **C**omparisons, and **O**utcomes. ***NOTE:*** Remember **PICO**.

Population: How would you describe the population of interest?

→ Details on the population of interest should be clearly outlined to the level that it would be appropriate to determine whether the results apply directly to one's patient(s) / community / constituents.

Intervention: Which main intervention or exposure is being considered?

→ The intervention refers to a variety of actions that are undertaken with the expectation of promoting and achieving specific outcomes. This may include an intervention, a strategy, or a policy, including activities such as lobbying, coalitions, and legislation. The focus of the review is to evaluate the impact of these activities on specific outcomes for individuals, communities or the population. The activities being assessed should be similar enough that it is reasonable to assess their combined impact.

Comparison: What is the main alternative to compare with the intervention?

→ This might be a control group or another intervention. Often the comparison is not stated explicitly in the research question. Either a control group or another intervention can be used as the comparator. In some instances, due to the nature of public health, a comparison and/or control may not be feasible.

Outcome: What do the researchers hope to accomplish, measure, improve, or affect?

→ Outcomes relate to the measured impact of the activities and can be at the individual, community or population level. Outcomes may include health policies, health programs, coalition development, etc.

Any part of **PICO** that is not addressed in a review's main research question should be clearly stated in the inclusion criteria to receive a **Yes** for criterion #1. Outcomes can be general in *the research question* (e.g. to allow for a broader search strategy, especially if the topic at-hand has a limited body of literature available), and then be addressed more specifically in the evidence tables and/or highlighted through the process of data extraction. For example, a general question may read: "The aim of this study, therefore, was to systematically review evidence from controlled trials on the efficacy of motor development interventions in young children."

Overall Coding for Q1:

☐ If the answer to each of population, intervention and outcome is yes, then place a check mark in the **Yes** column. Otherwise, place a check mark in the **No** column.

Q2 | Provision of inclusion criteria

The review should clearly describe the criteria that were used to select primary studies. This includes decisions related to the target population, intervention, outcome(s), as well as the research design (i.e., RCT, cohort, participatory, etc). Using the descriptions "peer-reviewed" and/or "measurement of a quantitative outcome" in the inclusion criteria are NOT sufficient descriptions to count for study design. Mark a **No** for this criterion.

If authors mention in their *exclusion* criteria that they rejected reviews, letters, editorials and case reports, but do not specifically address what they chose to *include,* mark a **No** for this criterion.

Overall Coding for Q2:

☐ Place a check mark in the **Yes** column if selection criteria were clearly outlined.

Q3 | Comprehensive search strategy

A well-described comprehensive search strategy will include multiple database searches and a variety of other search strategies. Relevant databases, chosen based on the key concepts in the research question, will include those from health databases (Medline, CINAHL, BIOSIS, EMBASE, etc), psychological databases (PsycINFO), social science databases (sociological abstracts), and/or educational databases (ERIC). 'Other' databases may be used and should be described in the space provided. General web searches may be included in 'other'.

For reviews measuring *specifically health-related outcomes* (e.g. vaccine effectiveness), **at least 2 health databases** need to be employed to allow for only *ONE* type of database to be searched. (***NOTE:*** *The two do not have to include Medline*)

'Column 2' search strategies include:

- → **Handsearching** – journals of relevance to the review topic
- → **Reference lists** – reference lists of relevant reviews should be reviewed for potential titles
- → **Key informants** – should demonstrate consultation with experts in the field for relevant titles; this *can* include pharmaceutical representatives
- → **Unpublished (grey) literature** – efforts to locate unpublished literature should be described. This can include the use of the electronic database SIGLE (which is specific to grey literature), and the searching of conference proceedings or scientific meetings.

NOTE: Should the author(s) describe the manual searching of reference lists, it would be most appropriate to score as 'Reference Lists', NOT as both 'Handsearching' *and* 'Reference Lists'.

Overall Coding for Q3:

☐ To answer **Yes**, the author(s) should have used at least two strategies from each column (one database type may be appropriate, as described above). In other words, in addition to using at least two types of electronic databases, the author(s) must have utilized a minimum of two of the other strategies (i.e., handsearching; key informants; reference lists; and/or unpublished literature).

Q4 | Search strategy covers an adequate number of years

In order to ensure that the entire body of relevant research is included in the review, the search strategy should cover a sufficient time period. The number of years that are adequate to search for primary studies will vary depending on the topic and the amount of literature being developed in that field. Generally, at least 10 years should be used as a minimum length of time, however, this may be increased if there has been little published in that time frame, or may be shortened if there has been an extreme amount of literature published in the recent past. The duration may also be shortened if the review is an update, however the original search must have covered a sufficient number of years.

Overall Coding for Q4:

- ☐ Answer **Yes** if the search strategy covered enough years that it is unlikely that important studies were missed. If there is any doubt in the reviewer's mind, some additional consideration of the topic area with the librarian will be conducted to determine the final assessment.

Q5 | Rigour of studies included in review is described

The methodological quality of primary studies is powerful in helping to explain variations in results from study to study. Therefore, the methodological rigour of primary studies in the relevant topic area should be identified and clearly described. Should the author(s) describe the studies as 'observational', please consider these studies to be a Level III.

For reviews of reviews, select the level of evidence based on the types of primary studies that appeared in the systematic reviews/meta-analyses now under assessment.

Overall Coding for Q5:

- ☐ Place a check mark in the **Yes** column if the methodological rigour (ie. RCT, Cohort, qualitative, etc.) of the primary studies is clearly identified in the review and circle the appropriate level of evidence.

Q6 | Quality assessment of primary studies

Each primary study should be assessed for methodological quality using a standardized assessment tool/scale. These criteria apply to meta-analyses as well.

Review authors need to do more than just state quality-related data that was extracted. The implication of this data on a review's findings must be addressed. For example, just because review authors list sample sizes of the primary studies *does not* mean they have assessed study sample.

**Health-Evidence staff should not have to conduct the QA, based on study characteristics provided.*

A **minimum of four** of the following areas should be assessed and the results described (in narrative or table form for **each** included primary study) for **quantitative studies**:

- → *Research design* (most rigorous design given the research question)
- → *Study sample* (generalizability, baseline characteristics)
- → *Participation rate*
- → *Sources of bias* (confounders, respondent bias, blinding, allocation concealment)
- → *Data collection* (measurement of independent and dependent variables, assessment tools).
- → *Follow-up/attrition rates*
- → *Data analysis* (e.g., intention-to-treat)

For **Cochrane Reviews** authors are required to conduct a standardized 'Risk of Bias' assessment (*see http://www.cochrane-handbook.org/* Figure 8.6a). Their results are typically included in the *Characteristics of Included Studies* table. These characteristics translate to the Health Evidence QA tool as follows:

If Cochrane Authors assess…		**On the Health Evidence QA tool select…**
Sequence generation	→	Research design
Allocation concealment	→	Research design
Blinding	→	Source of bias
Free of selective reporting	→	Data collection
Incomplete long-term/short-term outcome data **Authors describe assessing intention-to-treat analysis & whether incomplete data was dealt with correctly.*	→	Data analysis

The ***JADAD*** and ***EPOC*** tools are well-reputed and typically code **Yes**. Systematic reviews from the Cochrane Library often employ criteria from the *Cochrane Reviewers' Handbook*, however it is important to clarify the areas of assessment as 4 out of the 7 are not always considered.

When review authors assess whether or not a primary study used a **"validated measure(s)"**, this counts toward a point for *Data Collection.*

Use of a **Funnel plot** *can* be used towards a point for *Sources of Bias*, as long as it appears in the body of the paper and is part of a larger QA.

In some instances, different quality assessment criteria may be used for different study designs included in the same review. For example the EPOC tool has different criteria for interrupted time series studies, compared to randomized controlled trials. In this case, as long as the *majority* of reviews are assessed with 4+ criteria then **Yes** is appropriate.

For reviews of **qualitative primary studies** the following should be assessed and described for **each** included primary study:

- → Suitability of methodology/paradigm to the research question
- → Sampling (selection of participants/settings/documentation)
- → Clear description of context, data collection and data analysis
- → Rigour:

i. Audit trail (authors have provided/kept record of sufficient evidence to show the process and data used to demonstrate that their interpretations were reasonable)
ii. Some coding by 2 or more coders, if appropriate
iii. Deviant case analysis (negative cases that do not fit the main findings/themes)
iv. Respondent validation (member checking)

→ Triangulation (use of 2 or more research methods or sources to answer the same research question, the idea being that using different methods that draw the same interpretations add to the strength/integrity of the study)

→ Reflexivity (regarding researcher and the research process – the researcher's reflections on their effect on the research and research process, and the effect of the research on them and how both of these may have affected the outcome/findings)

→ Relevance (credibility, consistency, applicability, transferability) – these refer to the research being consistent with the context of the population being studied. Consistency in the application of methods and fieldwork is important. Transferability refers to the need to give enough detail about the context to allow others to make decisions about the applicability of that body of research to other populations.

Overall Coding for Q6:

☐ For a review of quantitative studies, place a check mark in the **Yes** column if at least four of the seven criteria are assessed (study design, data collection methods and follow-up rate must be assessed). For a review of qualitative studies, place a check mark in the **Yes** column if at least four of the seven criteria are assessed (suitability; sampling; clear description of context, data collection and data analysis; and rigour must be assessed).

Q7 | Are quality assessments transparent?

For quality assessments to be transparent a minimum of two review authors should assess each primary study*, independently,* for methodological quality and the method of conflict resolution described. A numerical level of agreement may be identified (i.e., Kappa), but is not required. If only inter-rater agreement scores are reported, however, review authors must report a Kappa score of at least 0.80 in order to score a **Yes** for this criterion.

Overall Coding for Q7:

☐ Place a check mark in the **Yes** column if two (or more) independent reviewers assessed each primary study for methodological quality, with a method of conflict resolution identified.

Q8 | Did review authors assess appropriateness of combining study results (i.e., test of homogeneity, or assess similarity of results in some other way)?

It is important that primary study results be assessed for similarity prior to combining them (both statistically and/or non-statistically).

If a **meta-analysis** is conducted, a test for homogeneity or heterogeneity is the minimum requirement that should be assessed across studies prior to determining the overall effect size. If significant heterogeneity is detected, the author(s) should indicate use of a Random Effects Model, as opposed to a Fixed Effects Model.

On occasion, an author may indicate the presence of significant heterogeneity and still combine data using a Fixed Effects Model. This IS appropriate if analyses have been conducted with both the inclusion and exclusion of data sets that may notably skew results. The results of these separate analyses, however, MUST be reviewed for the reader's consideration. This process, often called 'sensitivity analysis', assesses the moderators that may have contributed to the heterogeneity.

If a **systematic review** or a **narrative review** is conducted for which statistical analysis is not appropriate, the results of each study should be depicted in graph/table format in order to assess similarity across the primary studies. Often the results will be in the form of a table, but in the case of a narrative review the results of each study will be described at length within the body of the review.

In some cases confidence intervals/effect sizes are *NOT* required. For a **review of reviews**, a narrative presentation is appropriate (e.g. "the intervention had a positive effect on 20% of participants); ideally, with a table listing main features of each of the systematic reviews under review, or thorough, CONSISTENT discussion of the main features in the body of the review. If the review of reviews doesn't consistently present the actual numerical (or other qualitative) results (e.g. effect sizes from the original reviews) in the text, then it should score a **No**.

In general, trust the review author(s)' judgment of what is significant heterogeneity. A declaration of the specific number that was calculated (e.g. Chi-square score) is not mandatory.

NOTE: Despite extensive search strategies, some Cochrane reviews are unable to retrieve any applicable studies. In this case, a priori methodologies are often described. Subheadings alone, however, are sufficient to score a **Yes**, as Cochrane requires that they are filled in adequately before publication. Without a **Yes** for these criteria, these types of reviews will be of only *Moderate* quality, which may result in them being missed by users who are looking only for *Strong* reviews.

Overall Coding for Q8:

- ☐ Place a check mark in the **Yes** column if a test of homo/heterogeneity has been conducted and the corresponding model applied, or if the individual study results have been disclosed graphically or narratively. Please note that if study results are listed narratively, the information must have been provided consistently for all studies within the review text.

Q9 | Weighting

Whether a meta-analysis or a systematic/narrative review, the overall measure of effect should be determined by assigning those studies of highest methodological quality greater

weight. In the case of *meta-analyses*, weighting may be based on sample size, which is also acceptable.

If review authors have named a specific statistical software package (e.g. RevMan) they have used to combine data, this is sufficient for weighting, as the vast majority of this software incorporates the weighting of studies by a number of participants. Review authors may describe using the DerSimonian and Laird approach to random-effects meta-analysis which *also* incorporates weighting. Higgins and Green (2009) explain that:

> "The random-effects method (DerSimonian 1986) incorporates an assumption that the different studies are estimating different, yet related, intervention effects [...] The method is based on the inverse-variance approach, making an adjustment to the study weights according to the extent of variation, or heterogeneity, among the varying intervention effects."
>
> *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.0.2., The Cochrane Collaboration, 2009. Available from http://www.cochrane-handbook.org

One may notice the inclusion of sensitivity analyses and/or funnel plot diagrams. These are useful for assessing the effect of study quality on results in the case of the former, and potential for publication bias in the case of the latter. While useful, these particular analyses are not mandatory for a review to acquire a **Yes** coding.

In a **narrative synthesis**, quality of EACH of the included studies must be discussed *consistently* throughout the conclusions/discussion section to receive a **Yes** for this criterion.

In some cases review authors disclose the QA scores of primary studies - in table format, for example - and discuss those scores, but do not actually 'weigh' them; essentially, allowing the readers to determine which ones have the most weight. This is *NOT* sufficient to score a **Yes** for this criterion, as the review authors should be doing all summative work. It *IS* appropriate, however, for review authors to state, for example: "only the studies with a quality score of 5 or above are included in the analysis."

Reviews that weight conclusions/discussion by primary study quality still receive a **Yes** even if < 3 quality parameters were assessed (as per QA criterion #6).

Overall Coding for Q9:

- ☐ Place a check mark in the **Yes** column if a weighting system has been used in determining the overall impact.

Q10 | Interpretation of results

Consider the reported data and assess whether the review author's interpretation of the results of the primary studies are supported by the data. If no numerical values or p values/confidence intervals are given, then the reviewer cannot determine whether any conclusions are supported by the data and should respond **No** to criteria #10. In addition, if review authors failed to adequately assess methodological quality of the primary studies (i.e. criteria #6 is **No**), and also failed to weight the studies by quality or sample size (for meta-analyses) in their synthesis of results (i.e. criteria #9 is **No**), then the response to #10 should also be **No**, since it is difficult to determine agreement with review authors' conclusion(s) if

no quality assessment has taken place, since it is possible that agreement with authors' overall conclusion(s) would differ if studies were of weak quality compared to very strong quality.

Overall Coding for Q10:

- ☐ Place a check mark in the **Yes** column if the data for the primary studies supports the interpretations outlined in the review.

Overall Coding for the Review

An overall assessment of the methodological quality of the review will be determined based on the results from each question. The total score is out of 10. Add all the check marks in the **Yes** column and add to the Total column under **Yes**. Do the same for the **No** column. Use the following decision rule to determine the overall assessment for the review based on the numbers in the Total columns.

- → Reviews with a score of **8 or higher** in the **Yes** column will be rated as **Strong**
- → Reviews with a score between **5-7** in the **Yes** column will be rated as **Moderate**
- → Reviews with a score of **4 or less** in the **Yes** column will be rated as **Weak**

In the case that a score does not necessarily reflect your impression of the actual quality of a review (i.e., Strong/Moderate/Weak), consider revisiting some of the criteria and **Yes** and/or **No** scores, or discuss with a second reviewer, so that the corresponding quality category is a reflection of the review's overall methods and the score will be an accurate reflection for use by public health decision-makers.

What to do if a criterion is Not Applicable:

If a response to a question is N/A the final denominator for determining the overall assessment will be reduced by one