
Article:
Mandrik, O. orcid.org/0000-0003-3755-3031, Severens, J.L., Bardach, A. et al. (6 more authors) (2021) Critical appraisal of systematic reviews with costs and cost-effectiveness outcomes : an ISPOR Good Practices Task Force Report. Value in Health, 24 (4). pp. 463-472. ISSN 1098-3015

https://doi.org/10.1016/j.jval.2021.01.002

Critical Appraisal of Systematic Reviews with Costs and Cost-effectiveness Outcomes: an ISPOR Good Practices Task Force Report

Introduction

A systematic review (SR) can provide rigorous and complete evidence to support decision makers who consider both the effectiveness and cost-effectiveness of health interventions. A dramatic increase in published health economic (HE) studies, more specifically cost* and cost-effectiveness† studies, has resulted in the consequent proliferation of systematic reviews with cost and cost-effectiveness outcomes (SR-CCEO) (1, 2). First, such reviews help to identify strengths and weaknesses in HE studies, modelling methodologies, and data for modelling inputs. Secondly, SR-CCEOs may be informative for decision makers in resource allocation decisions for health interventions, especially in countries with limited capacity for health technology assessment (HTA).

However, it is challenging to appropriately interpret SR-CCEOs due to their heterogeneity in applied methods and reporting, and furthermore, due to variability in clinical and health settings in the original studies that they include. Methodological guidance and checklists that improve the quality of SRs on clinical evidence and/or decrease risk of bias in their interpretation or synthesis (3-6) have limited applicability for SR-CCEOs. There is little specific methodological guidance for SR-CCEOs (7-11). Although Chapter 20 of the Cochrane Handbook for Systematic Reviews of Interventions of the Cochrane Collaboration (Cochrane) (12) and three papers related to informing clinical practice guidelines (7-9) provide guidance, their recommendations do not focus on evaluating the quality of conduct or the risk of bias in SR-CCEOs. A critical analysis of guidelines on conducting and reporting SR-CCEOs identified multiple disagreements in these recommendations, suggesting that a standardised approach to conducting SR-CCEOs is needed (13).

Making universal recommendations for SR-CCEOs is difficult because they differ in several important aspects, in particular, with regard to their search and inclusion criteria, such as the

* For the purpose of this paper, cost studies are defined as studies analysing the costs of healthcare interventions including cost descriptions and cost-of-illness (economic burden of disease) studies. Sometimes cost studies might be based on an explicit comparison of alternatives.
† By cost-effectiveness studies we mean full economic evaluations, including cost-minimization, cost-effectiveness analysis, cost-utility analysis, cost-benefit analysis, and cost-consequence analysis.

types of studies included (trial or model-based, cost or cost-effectiveness), or in reporting solely economic characteristics or economic data alongside clinical outcomes. They also have different objectives, eg, to assess variability in outcomes and synthesize the findings, to identify the evidence gaps, or to assess the methods used.

Overall, SR-CCEO reliability and usefulness will improve with good practice guidance for SR-CCEOs with different objectives. Thus, ISPOR—The Professional Society for Health Economics and Outcomes Research established a global, multi-stakeholder, multi-disciplinary expert task force, to address this need (Appendix A).

While general recommendations on conducting SR-CCEOs are provided, the main goal is guidance on critical appraisal of SR-CCEOs regarding their quality and risk of bias. This report, which includes the ISPOR CrIteria for Cost (-Effectiveness) Review Outcomes (CiCERO) Checklist, will assist researchers, producers of health technologies and evidence users (decision makers / commissioners).

The task force categorized the recommendations according to the six stages of conducting an SR-CCEO. (Table 1).

<<INSERT TABLE 1 HERE>>

Stage 1. Planning and development

Each SR should be based on a comprehensive predefined protocol. It is a preferred practice to make the protocol of SR publicly available to prevent duplication of ongoing reviews, increase reproducibility of the research, and to avoid selective reporting. This can be achieved by registering the protocol with either immediate or delayed open access, (PROSPERO, the Centre for Open Science, or another independent online database), or by publishing it. Any deviations from this protocol should be included in the final report or publication. Independent of protocol availability, each review should have clearly stated objectives consistent with its reported results and conclusions, such as to synthesize the outcomes or to assess the methods.

It is routine practice to develop eligibility criteria around the PICO (population, intervention, comparator, and outcome) mnemonic in clinical reviews (14) or reviews of full economic evaluations (8). However, PICO or its derivatives are not fully applicable for methodological (eg, reviews appraising the design of economic models) or cost reviews (eg, cost of illness) in which the “comparator” or “intervention” component may be absent.

Depending on the objectives of the SR-CCEO, its design can be focused on:

- Model-based studies: eg, reviews assessing quality of models and reviews of studies using a life-time time horizon;
- Empirical health economic[‡] studies: eg, reviews assessing treatment costs and reviews of cost-effectiveness studies using a short time horizon;
- Or both, eg, reviews with broad perspectives and multiple time horizons.

Because SR-CCEOs are often used to inform decision makers, additional framing definitions are essential: time horizon and study perspective. These elements define which methods should be used for the literature search and synthesis.

Stage 2. Search for evidence

A review cannot be considered systematic if it is based on evidence identified through a non-targeted, unsound, incomplete, or non-reproducible search (15). The quality of the search depends on the experience of the person or group who developed the search (16, 17). Approaches to improving the quality of the search include involving information specialists or library scientists in search strategy development and using the peer-review electronic search strategies (PRESS) guideline (17, 18).

If a SR-CCEO is performed to update existing reviews, reusing the same search strategies may be appropriate. However, the quality of the initial search strategy should be re-evaluated. If a review uses search strategies from existing reviews to answer amended research questions, reviewers need to ensure that the adaptations in the objectives are reflected in the search strategy.

Conducting a SR is time-consuming. For clinical reviews it takes an average of 17 months from the registered project start to the publication date (19, 20). We expect that SR-CCEOs will have similar timelines: adding search words related to costs to the search line used in a clinical SR

[‡] The task force uses the term ‘empirical studies’ for single study-based economic evaluations, such as randomised and non-randomised trial-based economic evaluations, but also observational studies (single arm, multiple arm, real world data) that are used as a basis for cost-effectiveness analyses, often called piggy-back studies. Empirical studies are contrasted with modelling studies, explicitly synthesizing data using various sources.

will result in less hits, but a more complicated complementary search for grey literature will often be needed.

Cochrane requires the search date to be within 12 months of the publication date (12). This requirement is appropriate for SR-CCEOs summarizing outcomes. Therefore, a SR-CCEO should be conducted in the shortest time possible that does not compromise quality and comprehensiveness or should be updated prior to publication. Approaches that can decrease the review's time requirement include narrowing the SR-CCEO's objective or setting search restrictions if it is feasible and defensible. However, the task force believes that time duration may be less crucial for methodological than other reviews, given their objectives.

Selection of literature databases

Which sources to include in the systematic search should be justified primarily by the review's objectives, and it is unlikely that searching a single database will identify all relevant literature (22). There are different viewpoints on the best databases to search (7, 21, 23). However, an empirical study concluded that a search in Embase, HTA-journal database, MEDLINE/PubMed, and Scopus enabled identification of almost all the references in a SR-CCEO (23).

To minimize the risk of missing relevant studies, we recommend starting with the most commonly used international databases for cost and cost-effectiveness studies. A review of cost-effectiveness reviews, ie, an umbrella review, showed that the most commonly used resources (in order) were: MEDLINE, NHS EED (updated up to 2015), checking reference lists, Embase, and health technology assessment (HTA) report databases (21). See Appendix B for databases reflecting specific health topics and for SR-CCEOs with a regional focus.

Including multiple databases will likely identify more relevant studies, but it comes at the cost of additional records that need screening (24, 25). While we recommend searching at least three databases, if the reviewers chose not to, their decision should be well-justified and confirmed with evidence.

Developing and reporting a search strategy

The search strategy should be comprehensive enough to identify all relevant literature and reproducible, therefore, described in detail. Existing search filters can be used to identify cost and cost-effectiveness studies (26, 21) In addition, recommendations on search term and filter

selection (including Boolean operators), as well as considerations on sensitivity to specificity trade-offs and SR-CCEO objectives, are useful (7, 12).

Review authors should consider whether applying restrictions in the search (date of publication, study design, publication format, language, age of the subjects) might limit identification of all relevant literature. For example, if the review searched both clinical and cost-effectiveness studies and limits the search to RCTs, it misses possibly relevant model-based research.

Reviewers should consider that empirical studies measuring both clinical and cost outcomes are likely to report clinical and cost/cost-effectiveness results in separate publications. Therefore, for reviews with both clinical and economic studies, separate searching for articles reporting on either outcome may be preferable to increase the search results' comprehensiveness.

Supplemental searches

Even comprehensive search strategies may miss relevant studies, as approximately 4% of included studies were missed by database searches (23). In addition to database searches, other strategies to identify published literature include "snowballing" techniques (searching the bibliographies of all included studies), personal knowledge of existing studies, citation tracking or by contacting experts in the field (27). This means that the process of identifying relevant literature should include supplemental searches (28) using at least one-step back citation tracking of included studies.

Searching for grey literature

Searching grey literature[§] is challenging because the results are dependent on when the search is conducted, and therefore, potentially non-reproducible. However, grey literature may be particularly important to SR-CCEOs as one way to address publication bias. Thus, if a search of grey literature is not performed, it should be clearly justified.

We recommend including grey literature and to follow recommendations on grey literature searches (29). A supplementary search on HTA is especially important for SR-CCEOs because relevant reports may be not be in HTA databases. (See sources in Appendix B, sections 2 and 3). Furthermore, the authors may want to explore platforms that collect and aggregate grey

[§] Grey literature refers to research that is either unpublished or has been published outside of the traditional commercial or academic publishing and distribution channels. Examples of grey literature include: government reports, policy statements, and issues papers.

literature regarding specific topics, such as Program for Monitoring Emerging Diseases (ProMED) of the International Society for Infectious Diseases (https://promedmail.org/about-promed/).

As a general rule, we do not recommend that abstracts of conference proceedings be included in a search, even if technically possible. Scientific conference abstracts in SR-CCEOs could increase the risk of bias because it has been shown that more than half of such abstracts ultimately fail to publish their results after peer review in full (30), while other abstracts, eg, the Society of Medical Decision Making (SMDM), Health Technology Assessment international (HTAi) conference abstracts, are not indexed in international databases. Nevertheless, reviewers may include them if they make a solid argument for inclusion, for instance, to identify such abstracts for further follow-up for full text publications.

Social networks (a social media website or other application sharing information) may become additional sources of both clinical and economic data for SR-CCEOs. Although unknown, the risk of bias from these sources seems obvious. Reviewers should not apply information derived from such networks without first evaluating the risk of bias.

Stage 3. Study selection and eligibility

The study selection process includes screening of titles, abstracts, and full-text publications. Methods for study selection should promote transparency and minimize bias. The transparency in a SR-CCEO can be achieved by following SR reporting guidelines, such as the PRISMA statement (6). There is unlikely to be a “one size fits all” approach, so when evaluating a SR-CCEO, it is important to evaluate how the methodological approach may contribute to risk of bias.

For a SR-CCEO using the methods and/or outcomes from previously published reviews, the risk of bias increases when the previous reviews’ data analysis steps are applied. For example, the risk of bias would be higher if not only the search results are applied, but also full-text inclusion, due to the uncertainty in reliability of each of the literature selection and analysis steps.

Process of study selection

There are a number of tools and methodological recommendations on study selection in clinical SRs that are relevant for a SR-CCEO. For example, AMSTAR-2 (A MeaSurement Tool to

Assess systematic Reviews) appraises the quality of conduct around study selection (3), and Robson et al. (2018) summarizes the key conclusions of a SR related to study selection methods (31). The common recommendation to minimize the risk of excluding a relevant study or including an irrelevant study, is to perform each step of the study selection process - ideally independently - in duplicate, with conflicts resolved through discussion and/or by a third party while a combination of both is to be preferred.

One approach to address the risk of bias in literature selection if resources are limited, is to be more liberal in reviewing titles and/or abstracts for inclusion by a single reviewer and then at the full-text review stage, ensure that there is duplicate reviewing and stringent criteria application. This should mitigate any issues with a single reviewer and balance the risk of overinclusion (which comes with more research costs) with the risk of excluding relevant citations (30, 31).

Another strategy is using tools with machine learning capabilities, eg, Abstrackr, DistillerSR, SWIFT-Active Screener, and RobotAnalyst. In particular, these tools can be used to duplicate the manual selection. While machine-learning tools decrease screening time, the risk of bias in using such tools is currently uncertain. The available evidence is limited, and their performance is highly varied (32-34). If non-validated artificial intelligence tools are used, their literature screening accuracy should be tested on a sample and their use should be clearly reported.

Restrictions in eligibility criteria

It is difficult to characterize how the use of greater restrictions in study selection relates to the relevance and bias of a review's outcomes because such restrictions can increase or decrease these measures. For example, in clinical reviews, restricting the inclusion criteria to RCTs may increase the risk of bias with respect to adverse event rates (underestimation), but decrease the risk of bias in estimates of effectiveness.

For SR-CCEOs, there are a variety of relevant restrictions that might be considered beyond study design. The combination of these restrictions represents trade-offs between internal validity and broader generalizability (Box 1). Furthermore, restrictions on study perspective and cost methodologies (how and which costs are included in the analyses) may increase or decrease bias relative to the review's intended purpose.

<<INSERT BOX 1 HERE>>

Our experience suggests that applying restriction criteria during the search or when screening titles and abstracts is efficient. However, sometimes, full text reading is unavoidable. If

evidence quality is used as an exclusion criterion, another approach to assess the risk of bias would be to apply a scenario analysis where excluded sources are included to see if that changes the conclusion.

Stage 4. Critical appraisal of included studies

HTA bodies demand transparency and sound methods in original cost and cost-effectiveness studies to apply them in appraisals. Logically, to reduce flaws in synthesizing the evidence, a SR-CCEO should include a methodological quality assessment of included studies.

While assessing the quality of included studies, reviewers should provide a qualitative description and a critique of the evidence base. Reviewers should be explicit about: 1) the existence of and the type of biases that may exist in each study, eg, quality, quality of reporting, and sponsorship in the study, and 2) whether and how estimates were adjusted for transferability and with what assumptions. To increase the consistency in assessment of the methodological quality of each included study, one of the standard checklists (see below) is justified and should be used over self-designed evaluation approaches.

Appropriate methodological quality assessment for various kinds of cost and cost-effectiveness publications depends on the type of research conducted, eg, a trial-based study may need to focus more on consideration of population generalizability. Thus, assessment of quality in an empirical cost or cost-effectiveness study should not be handled in the same way as the assessment of a model.

There are a number of checklists developed to assess methodological quality and/or quality of reporting in included cost and cost-effectiveness studies (9). The most commonly used are:

- British Medical Journal checklist (35);
- Phillips checklist for model-based studies (36);
- Quality of Cost-Effectiveness Studies checklist for model-based evaluations (37)
- Consensus on Health Economic Criteria (CHEC) for trial-based studies (38);
- Consolidated Health Economic Evaluation Reporting Standards (CHEERS) (39);
- Bias in Economic Evaluation (ECOBIAS) Checklist for trial- and model-based studies (40);
- Second Panel on Cost-Effectiveness checklist (41).
- TRansparent Uncertainty ASsessmenT (TRUST) Tool for systematically identifying, assessing, and reporting uncertainties in decision models (42)
- Questionnaire to assess the relevance and credibility of a modeling studies (43).

Most of these tools are comparable in their coverage of key design characteristics. However, they differ in the extent to which they are suitable for empirical or model-based studies or whether their specific focus is on the quality of methods or on reporting. TRUST deviates in this respect; it is focused on identifying, assessing, and reporting uncertainty (42). In addition, the reviews of modelling studies will benefit from assessment of data source quality in the models (44).

The selection of the right methodological quality instrument will be a trade-off between the research question and objectives of the SR-CCEO, the available research capacity, the thoroughness of the evaluation of quality, and the requirements of the project funder or the target journal (if any). A comparative assessment of the checklists is reported by Wijnen et al. (2016) (9). No single checklist can be recommended, but a clear motivation must be given for use in the SR-CCEO.

To minimize systematic and non-systematic errors, at least, two reviewers should assess the quality of studies included in a SR-CCEO independently.

Stage 5. Data extraction and synthesis

Performing data extraction

The same data extraction standards and expectations that apply to SRs of clinical effectiveness should be applied to SR-CCEOs. Data extraction by a single reviewer results in more errors on average than does duplicated data extraction with the observed relative difference in accuracy of 21.7% (45). While duplicated extraction is preferred from the accuracy viewpoint, there is a trade-off between the accuracy and efforts required (30), especially since a SR-CCEO generally involves extracting a broad range of target outcomes (ie, clinical, cost, and cost-effectiveness outcomes), as well as data related to methodology. If an independent duplicated extraction is not possible, reviewers may consider performing a verification of study characteristics and extracting outcome data independently in duplicate (46).

Performing data synthesis

Considerations for synthesizing data depend on the purpose of the review, eg, synthesizing the outcomes or reporting methodological issues (47). There is no consensus on the best way to synthesize economic evidence. Possible approaches include structured narrative synthesis

(using descriptive methods instead of statistical approaches) (12), graphical synthesis (eg, cost-effectiveness diagram, permutation matrix) (48, 49), hierarchical matrix (50), or quantitative synthesis/meta-analysis (see "Meta-analysis in SR-CCEO") (51, 52). The stated order reflects the most applicable synthesis approach, ie, the approach that can be used under any circumstances to the least used synthesis based on lack of applicability.

One of the main challenges in choosing the "best" synthesis method for a particular SR-CCEO is matching the approach to synthesis to the review's scope and the observed variability among the studies it identifies. This variability can be methodological, clinical, or health setting (administrative or jurisdiction-related). It is especially challenging to make a single recommendation on a synthesis approach because SR-CCEOs themselves have broadly different scopes. Some reviews comment on the implication of the cost and cost-effectiveness studies for a broad range of jurisdictions, while others comment on the implication for a much narrower range, eg, HTA for a single government.

A premise to enable assessment of the synthesis's adequacy is a clearly defined objective that includes the intended audience (jurisdiction or health setting). Guiding questions should be used to assess clinical, health setting, and methodological compatibility (diversity or variability that cannot be measured statistically). These questions should be informed by tools for assessing transferability and applicability (53, 54), for instance using a decision chart for assessing the transferability of cost and cost-effectiveness results between countries (55).

Generally, results from modelling studies and empirical studies should be synthesized separately. Cost and cost-effectiveness studies based on trials or observational study designs, as well as probabilistic and deterministic analyses, should be synthesized separately, too. In addition, incorporating the results of sensitivity analyses should be considered (56, 57).

When synthesising numeric values, papers will likely be excluded based on missing information necessary for judging eligibility, applicability, homogeneity, etc. For example, missing demographic characteristics of the population analysed may make it impossible to determine if the study applies to an age group that the SR-CCEO focuses on. This should be properly documented. Ideally, sensitivity analysis should be done with and without the questionable sources.

In a SR-CCEO that summarizes cost or cost-effectiveness outcomes, all cost data should be converted into the same currency. In addition, it should be expressed in the same year, ie, inflation-adjusted, using the standard inflator for the country on which the analysis is focused, *before* the results are synthesized either narratively or quantitatively.

In the assessment of costs heterogeneity, the methodological, clinical, and setting compatibility should be considered where, in particular, the latter two will have their impact on resource use. For instance, the choice of conversion approach for costs would depend on settings' comparability (53, 54), with purchasing power parity (PPP) used to compare costs in heterogeneous settings. While standardization of costs should be undertaken for the synthesis, the *original* costs reported in the study should also be presented in the SR-CCEO as with all relevant original data, since valuation methods may differ (58).

SR-CCEOs that assess the methodology of included cost and cost-effectiveness studies have an exceptionally wide set of methodological questions on which they may focus (59). Hence, for such reviews, it is likely that only the broad criteria on narrative synthesis are applicable, unless the review is based on a narrow objective of only including studies that are comparable.

Exploring heterogeneity in data

Figure 1 illustrates that the "right" approach for summarizing cost and cost-effectiveness outcomes depends on the degree of clinical and methodological compatibility in the studies included. When studies are not comparable, narrative synthesis/comparison will be more appropriate. While not all of the differences in reported values can be explained, we strongly encourage the reviewers to attempt to do so by analysing characteristics of the studies and their impact on outcomes. Some factors, such as quality of reporting and conflicts of interest in the studies, can be direct indicators of risks of bias and may contribute to heterogeneity in outcomes. It is more challenging, though, to assess how methodological differences in the studies contribute to heterogeneity in outcomes.

Only in the case where the SR-CCEO's objective is very narrowly focused, is it feasible to explore associations between modelling methods and costs or cost-effectiveness outcomes using meta-regression analysis (60). If methodological factors that can potentially explain differences between the studies' outcomes are identified, they should be reported.

Meta-analysis

Only studies considered compatible with regard to clinical and health settings (eg, PICO,) and study methodology (ie, time horizon and study perspective) may be considered for synthesis. If a SR-CCEO pools outcomes in one common metric compatibility of different health settings (or jurisdictions) should be carefully assessed (Figure 1). Usually a very high degree of incompatibility will imply that pooling such results is not appropriate.

<<INSERT FIGURE 1 HERE>>

Therefore, a single quantitative synthesis may only be used in narrowly focused reviews with approaches to synthesis based on a distribution of outcomes rather than a single "true" outcome (eg, random-effects models) (12, 61). A SR-CCEO with a broad scope should report the results for compatible subgroups that are consciously selected, ideally based on predefined criteria, eg, results for high-income Asian countries.

It is the task force's opinion that the costs reported in various cost and cost-effectiveness studies are typically (although not always) more heterogeneous than effects (by heterogeneity we mean statistically-measured variability). Therefore, Figure 1 suggests a hierarchical approach in exploring data compatibility/homogeneity and pooling the data. This means each next level is possible on the condition that ALL of the previous levels have been achieved. In this way, homogeneity can be assessed in a similar manner as in clinical reviews (12, 61)

Data that can be pooled:

- for cost-effectiveness studies, the average and incremental effectiveness when there is sufficient homogeneity, as well as clinical and methodological comparability (the common effectiveness outcomes in cost-effectiveness studies, eg, QALYS or life-years gained),
- for costing studies, the average costs when there is methodological and health setting comparability,
- for cost-effectiveness studies, the average and incremental costs when there is methodological and health setting comparability,
- for cost-effectiveness studies, the net benefit (either net monetary benefit or net health benefit) when homogeneity and comparability is achieved in all above levels and willingness to pay threshold homogeneity is observed (or when the disaggregated costs and benefits can be combined using a common willingness to pay threshold).

To address incomparability among studies, a sensitivity (sub-group) analysis can be used in a SR-CCEO, similar to the clinical reviews.

Publication bias

Publication bias exists if the outcomes of a cost or cost-effectiveness study influence the publication decision. Bias in cost-effectiveness studies exists when published incremental cost-effectiveness ratios (ICERs) cluster around a proposed threshold, and it is likely to relate to the

origin of the sponsorship (47). Publication bias in SR-CCEO can be related to multiple reasons including:

(a) Failure to submit (sponsored) cost-effectiveness studies that have non-favorable results (an indicator of publication bias of this type can be a relationship in study sponsorship and reported incremental cost-effectiveness of technologies);
(b) Priority setting by target journals publishing cost and cost-effectiveness studies, eg, preference to publish methodological research, innovative evaluations (typically conducted for high-income settings) and to avoid model adaptations.

Assessment of publication bias may not be straightforward in SR-CCEOs. Researchers are advised to follow the task force's recommendations in Box 2. However, none of the proposed assessment methods is perfect and we encourage the development of new approaches.

<<INSERT BOX 2 HERE>>

Stage 6. Presentation and reporting

To optimize usefulness, it is important that the review reports, in sufficient detail, study characteristics and specific outcomes (at a minimum). More standardized reporting of SR-CCEOs will improve comparability between reviews and may influence future reporting in primary studies of cost and cost-effectiveness analyses.

For SR-CCEOs, the outcomes of interest, eg, total costs, life years, QALYs, as well as methodological aspects, eg, study perspective, health state valuation, type of costs, costs valuation, should be reported for each included study. Both cost and health outcomes should be presented separately for *each* strategy, within *each* study. Whether it is relevant to report one "base case" result or a range of results will depend on each specific research question posed in each separate SR-CCEO (39).

Economic outcomes and information regarding included studies, eg, the characteristics of patient populations and the methodological choices adopted in each included study, should be reported in summary tables. Box 3 presents the common elements in existing checklists assessing methodological quality and/or quality of reporting in cost or cost-effectiveness studies (the minimum reporting requirements) (35, 36, 38-40). Other elements that researchers may choose to report will depend on the review's objectives, the analysed interventions, and can, for

instance, include ethical and/or equity considerations as might have been reported in the studies included, and heterogeneity (subpopulation analysis). The reviewers should acknowledge the process behind the outcomes of interest choices, eg, whether expert opinion was involved.

A SR-CCEO that focuses on decision analytic models should also report the:

(a) Model type and characteristics, eg, clinical pathways, health states, cycle length, transition possibilities, half-cycle correction applied;
(b) Model validation, eg, face validity, cross-validation against other models, internal and external validity;
(c) Components of uncertainty analysis extracted and reported separately for probabilistic and deterministic sensitivity analyses, and scenario/subgroup analyses.

In some cases, there will be more aspects that are relevant to include, eg, disease specific modelling choices (62).

<<INSERT BOX 3 HERE>>

If a SR-CCEO includes studies performed without modeling, the specific reporting should include study type, eg, RCT or cohort, method(s) of cost calculation, eg, regression or descriptive; questionnaires, expert opinion, and control (or stratification) variables.

A compromise should be found between both the reporting of outcomes in summary tables and their narrative description, especially for items of interest. While a word limit demanded by peer-reviewed journals can restrict reporting, all the relevant information that cannot be included in the main paper should be presented in online appendices, supplementary materials, and/or study protocols.

Criteria for Cost (-Effectiveness) Review Outcomes (CiCERO) Checklist

Based on the considerations discussed above, the task force developed the ISPOR CiCERO Checklist - a tool to assess the quality of reporting, conduct, and risk of bias in SR-CCEOs. Using CiCERO leads to an overview of the quality and risk of bias in an SR-CCEO (without resulting in a single score). The general conclusion is dependent on the SR-CCEO's objectives and the data extracted. Assessing the quality and risk of bias will identify the review's critical weaknesses and give the user a feeling of overall confidence in the results of the SR-CCEO.

CiCERO includes 13 signalling questions to consider when evaluating the quality of reporting, conduct, and risk of bias in SR-CCEOs (Appendices C, D, and E for the PDF version and Appendix F for the Excel version). There are three versions of the CiCERO checklist for: 1) reviews of cost and cost-effectiveness studies, 2) reviews that summarize methods of cost and costs-effectiveness studies, and 3) SR-CCEOs that use the AMSTAR-2 instrument to assess quality in included studies.

The process of developing and validating CiCERO is reported in the Box 4. CiCERO's development was based on current SR-CCEO knowledge and experience. Because this is a rapidly developing research area, it is expected that the task force will update CiCERO and the report's recommendations in 5 – 7 years.

<<INSERT BOX 4 HERE>>

Limitations of the task force recommendations and the ISPOR CiCERO Checklist

While these recommendations were developed to evaluate the quality of conduct, reporting and risk of bias of SR-CCEOs, they may be used for conducting a rapid review. A poorly conducted systematic review, may not perform as well as a properly conducted, transparently reported rapid review (63). So far, limited information is available on biases related to social networks as a data source and artificial intelligence in screening and evaluating the literature. Thus, based on more empirical evidence, these topics should be detailed in future discussions regarding quality and risk of bias of SR-CCEO.

Conclusions

As the number of SR-CCEOs continues to increase, standardizing the preparation, reporting, and interpretation of their findings is of crucial and growing importance. Such standardization is required to make effective use of this evidence base to support healthcare decision making. This report describes good practice recommendations, organised in six stages, for critically appraising quality and risk of bias in SR- CCEOs. As such, it provides guidance to reviewers on how to minimize the risk of bias, as well as improve the quality of methods and reporting for conducting a SR-CCEO. In this way, SR-CCEOs can provide valuable evidence to healthcare decision makers.

517

References

1. Luhnen M, Prediger B, Neugebauer EAM, et al. Systematic reviews of health economic evaluations: A structured analysis of characteristics and methods applied. Research synthesis methods. 2019; 10: 195-206.
2. Mandrik O, Ekwunife OI, Meheus F, et al. Systematic reviews as a "lens of evidence": Determinants of cost-effectiveness of breast cancer screening. Cancer Med. 2019; 8: 7846-58.
3. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. Bmj. 2017; 358: j4008.
4. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol. 1991; 44: 1271-8.
5. Whiting P, Savovic J, Higgins JP, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol. 2016; 69: 225-34.
6. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009; 6: e1000097.
7. Thielen FW, Van Mastrigt G, Burgers LT, et al. How to prepare a systematic review of economic evaluations for clinical practice guidelines: database selection and search strategy development (part 2/3). Expert Rev Pharmacoecon Outcomes Res. 2016; 16: 705-21.
8. van Mastrigt GA, Hiligsmann M, Arts JJ, et al. How to prepare a systematic review of economic evaluations for informing evidence-based healthcare decisions: a five-step approach (part 1/3). Expert Rev Pharmacoecon Outcomes Res. 2016; 16: 689-704.
9. Wijnen B, Van Mastrigt G, Redekop WK, et al. How to prepare a systematic review of economic evaluations for informing evidence-based healthcare decisions: data extraction, risk of bias, and transferability (part 3/3). Expert Rev Pharmacoecon Outcomes Res. 2016; 16: 723-32.
10. Systematic reviews. CRD's guidance for undertaking reviews in health care. Center for Reviews and Dissemination of York. 2009.
11. Mathes T, Walgenbach M, Antoine SL, et al. Methods for systematic reviews of health economic evaluations: a systematic review, comparison, and synthesis of method literature. Med Decis Making. 2014; 34: 826-40.
12. Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). In: Higgins J, Thomas J, Chandler J, et al., eds., 2019.
13. Jacobsen E, Boyers D, Avenell A. Challenges of Systematic Reviews of Economic Evaluations: A Review of Recent Reviews and an Obesity Case Study. Pharmacoeconomics. 2020; 38: 259-67.
14. Richardson WS, Wilson MC, Nishikawa J, et al. The well-built clinical question: a key to evidence-based decisions. ACP J Club. 1995; 123: A12-3.
15. Levay P, Craven J. Systematic searching. Practical ideas for improving results. London: Facet Publishing, 2019.
16. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. Bmj. 1994; 309: 1286-91.
17. Rethlefsen ML, Farrell AM, Osterhaus Trzasko LC, et al. Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. J Clin Epidemiol. 2015; 68: 617-26.
18. McGowan J, Sampson M, Salzwedel DM, et al. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. J Clin Epidemiol. 2016; 75: 40-6.
19. Ganann R, Ciliska D, Thomas H. Expediting systematic reviews: methods and implications of rapid reviews. Implement Sci. 2010; 5: 56.
20. Borah R, Brown AW, Capers PL, et al. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ Open. 2017; 7: e012545.
21. Wood H, Arber M, Glanville JM. Systematic reviews of economic evaluations: How extensive are their searches? Int J Technol Assess Health Care. 2017; 33: 25-31.

572 22. Betran AP, Say L, Gulmezoglu AM, et al. Effectiveness of different databases in
573 identifying studies for systematic reviews: experience from the WHO systematic review of
574 maternal morbidity and mortality. BMC Med Res Methodol. 2005; 5: 6.
575 23. Arber M, Glanville J, Isojarvi J, Baragula E, Edwards M, Shaw A, Wood H. Which
576 databases should be used to identify studies for systematic reviews of economic evaluations?
577 Int J Technol Assess Health Care. 2018;34(6):547-554.
578 24. Hopewell S, McDonald S, Clarke M, et al. Grey literature in meta-analyses of
579 randomized trials of health care interventions. Cochrane Database Syst Rev. 2007; 18:
580 Mr000010.
581 25. Halladay CW, Trikalinos TA, Schmid IT, et al. Using data sources beyond PubMed has
582 a modest impact on the results of systematic reviews of therapeutic interventions. J Clin
583 Epidemiol. 2015; 68: 1076-84.
584 26. Filters to Identify Economic Evaluations. ISSG Search Filter Resource. The InterTASC
585 Information Specialists SubGroup.
586 27. Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in systematic
587 reviews of complex evidence: audit of primary sources. Bmj. 2005; 331: 1064-5.
588 28. How to review the evidence: systematic identification and review of the scientific
589 literature. Handbook series on preparing clinical practice guidelines. Canberra: National Health
590 and Medical Research Council, 1999.
591 29. Grey Matters: a practical tool for searching health-related grey literature. the Canadian
592 Agency for Drugs and Technologies in Health.
593 30. Shemilt I, Khan N, Park S, et al. Use of cost-effectiveness analysis to compare the
594 efficiency of study identification methods in systematic reviews. Syst Rev. 2016; 5: 140.
595 31. Robson RC, Pham B, Hwee J, et al. Few studies exist examining methods for selecting
596 studies, abstracting data, and appraising quality in a systematic review. J Clin Epidemiol. 2019;
597 106: 121-35.
598 32. Gates A, Guitard S, Pillay J, et al. AHRQ Methods for Effective Health Care.
599 Performance and Usability of Machine Learning for Screening in Systematic Reviews: A
600 Comparative Evaluation of Three Tools. Rockville (MD): Agency for Healthcare Research and
601 Quality (US), 2019.
602 33. Gartlehner G, Wagner G, Lux L, et al. Assessing the accuracy of machine-assisted
603 abstract screening with DistillerAI: a user study. Syst Rev. 2019; 8: 277.
604 34. Howard BE, Phillips J, Tandon A, et al. SWIFT-Active Screener: Accelerated document
605 screening through active learning and integrated recall estimation. Environment International.
606 2020; 138: 105623.
607 35. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic
608 submissions to the BMJ. The BMJ Economic Evaluation Working Party. Bmj. 1996; 313: 275-
609 83.
610 36. Philips Z, Ginnelly L, Sculpher M, et al. Review of guidelines for good practice in
611 decision-analytic modelling in health technology assessment. Health Technol Assess. 2004; 8:
612 iii-iv, ix-xi, 1-158.
613 37. Chiou CF, Hay JW, Wallace JF, et al. Development and validation of a grading system
614 for the quality of cost-effectiveness studies. Med Care. 2003; 41: 32-44.
615 38. Evers S, Goossens M, de Vet H, et al. Criteria list for assessment of methodological
616 quality of economic evaluations: Consensus on Health Economic Criteria. Int J Technol Assess
617 Health Care. 2005; 21: 240-5.
618 39. Husereau D, Drummond M, Petrou S, et al. Consolidated Health Economic Evaluation
619 Reporting Standards (CHEERS)--explanation and elaboration: a report of the ISPOR Health
620 Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. Value
621 Health. 2013; 16: 231-50.
622 40. Adarkwah CC, van Gils PF, Hiligsmann M, et al. Risk of bias in model-based economic
623 evaluations: the ECOBIAS checklist. Expert Rev Pharmacoecon Outcomes Res. 2016; 16: 513-
624 23.

625 41. Sanders GD, Neumann PJ, Basu A, et al. Recommendations for Conduct,
626 Methodological Practices, and Reporting of Cost-effectiveness Analyses: Second Panel on
627 Cost-Effectiveness in Health and Medicine. Jama. 2016; 316: 1093-103.
628 42. Grimm SE, Pouwels X, Ramaekers BLT, et al. Development and Validation of the
629 TRansparent Uncertainty ASsessmenT (TRUST) Tool for Assessing Uncertainties in Health
630 Economic Decision Models. Pharmacoeconomics. 2020; 38: 205-16.
631 43. Jaime Caro J, Eddy DM, Kan H, et al. Questionnaire to assess relevance and credibility
632 of modeling studies for informing health care decision making: an ISPOR-AMCP-NPC Good
633 Practice Task Force report. Value Health. 2014; 17: 174-82.
634 44. Cooper N, Coyle D, Abrams K, et al. Use of evidence in decision models: an appraisal
635 of health technology assessments in the UK since 1997. J Health Serv Res Policy. 2005; 10:
636 245-50.
637 45. Buscemi N, Hartling L, Vandermeer B, et al. Single data extraction generated more
638 errors than double data extraction in systematic reviews. J Clin Epidemiol. 2006; 59: 697-703.
639 46. Higgins J, Lasserson T, Chandler J, et al. Methodological Expectations of Cochrane
640 Intervention Reviews (MECIR) Standards for the conduct and reporting of new Cochrane
641 Intervention Reviews, reporting of protocols and the planning, conduct and reporting of
642 updates. 2019.
643 47. Bell CM, Urbach DR, Ray JG, et al. Bias in published cost effectiveness studies:
644 systematic review. Bmj. 2006; 332: 699-703.
645 48. Clarkson P, Davies L, Jasper R, et al. A Systematic Review of the Economic Evidence
646 for Home Support Interventions in Dementia. Value in health : the journal of the International
647 Society for Pharmacoeconomics and Outcomes Research. 2017; 20: 1198-209.
648 49. Gilbody S, Bower P, Whitty P. Costs and consequences of enhanced primary care for
649 depression: systematic review of randomised economic evaluations. The British journal of
650 psychiatry : the journal of mental science. 2006; 189: 297-308.
651 50. Nixon J, Khan KS, Kleijnen J. Summarising economic evaluations in systematic reviews:
652 a new approach. BMJ. 2001; 322: 1596-98.
653 51. Rotter T, Kinsman L, James E, et al. The effects of clinical pathways on professional
654 practice, patient outcomes, length of stay, and hospital costs: Cochrane systematic review and
655 meta-analysis. Evaluation & the health professions. 2012; 35: 3-27.
656 52. de Morton NA, Keating JL, Jeffs K. Exercise for acutely hospitalised older medical
657 patients. The Cochrane database of systematic reviews. 2007: Cd005955.
658 53. Barbieri M, Drummond M, Rutten F, et al. What do international pharmacoeconomic
659 guidelines say about economic data transferability? Value in health : the journal of the
660 International Society for Pharmacoeconomics and Outcomes Research. 2010; 13: 1028-37.
661 54. Drummond M, Barbieri M, Cook J, et al. Transferability of economic evaluations across
662 jurisdictions: ISPOR Good Research Practices Task Force report. Value Health. 2009; 12: 409-
663 18.
664 55. Welte R, Feenstra T, Jager H, et al. A decision chart for assessing and improving the
665 transferability of economic evaluation results between countries. Pharmacoeconomics. 2004;
666 22: 857-76.
667 56. Ruiz-Negron N, Menon J, King JB, et al. Cost-Effectiveness of Treatment Options for
668 Neuropathic Pain: a Systematic Review. PharmacoEconomics. 2019; 37: 669-88.
669 57. Squires H, Poku E, Bermejo I, et al. A systematic review and economic evaluation of
670 adalimumab and dexamethasone for treating non-infectious intermediate uveitis, posterior
671 uveitis or panuveitis in adults. Health technology assessment (Winchester, England). 2017; 21:
672 1-170.
673 58. Mattingly TJ, 2nd, Levy JF, Slejko JF, et al. Estimating Drug Costs: How do
674 Manufacturer Net Prices Compare with Other Common US Price References?
675 Pharmacoeconomics. 2018; 36: 1093-99.
676 59. Padmasawitri TIA, Frederix GW, Alisjahbana B, et al. Disparities in model-based cost-
677 effectiveness analyses of tuberculosis diagnosis: A systematic review. PloS one. 2018; 13:
678 e0193293.

679 60. Burgers LT, van de Wetering FT, Severens JL, et al. Using meta-regression analyses in
680 addition to conventional systematic review methods to examine the variation in cost-
681 effectiveness results - a case study. BMC Health Serv Res. 2016; 16: 23.
682 61. DerSimonian R, Laird N. Meta-Analysis in Clinical Trials. Controlled Clinical Trials. 1986;
683 7: 177-88
684 62. Chhatwal J, He T, Lopez-Olivo MA. Systematic Review of Modelling Approaches for the
685 Cost Effectiveness of Hepatitis C Treatment with Direct-Acting Antivirals. Pharmacoeconomics.
686 2016; 34: 551-67.
687 63. Tricco AC, Antony J, Zarin W, et al. A scoping review of rapid review methods. BMC
688 Med. 2015; 13: 224.
689

Critical Appraisal of Systematic Reviews with Costs and Cost-effectiveness Outcomes: an ISPOR Good Practices Task Force Report

Journal:	*Value in Health*
Manuscript ID	Draft
Article Type:	ISPOR Report
Health Areas List:	Health Areas
Methods of Interest List:	CER/HTA: systematic review < Methods of Interest
Keywords Enter Your Own:	appraisal of systematic reviews, quality of systematic reviews, bias in systematic reviews, cost-effectiveness outcomes
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
Appendix F Checklist CiCERO.xlsm	

ISPOR Systematic Reviews Good Practices Task Force Report

TITLE PAGE

Manuscript title: Critical Appraisal of Systematic Reviews with Costs and Cost-effectiveness Outcomes: an ISPOR Good Practices Task Force Report

Running title: Systematic Reviews Task Force Report

Key Words: appraisal of systematic reviews, quality of systematic reviews, bias in systematic reviews, cost-effectiveness outcomes, cost outcomes

Financial disclosures: No funding

Precis 25/25

This ISPOR Task Force report provides recommendations on the critical appraisal of quality and risk of bias in systematic reviews with cost and cost-effectiveness outcomes.

Acknowledgements:

The co-authors are grateful for contributions from Cristina Masseria (Pfizer), Mark Clowes, and Ruth Wong (the University of Sheffield). In addition, many thanks to Elizabeth Molsen-David at ISPOR for her continuous support from start to finish and her excellent editing.

The co-authors gratefully acknowledge the especially useful comments provided by: Chris Bartlett, Carolina Castañeda, Matt Chenoweth, Michael Drummond, Joost Enzing, Gunhild Hagen, Louise Hartley, Lisa Hess, Manthan Janodia, Rositsa Koleva-Koralova, Kelly Lenahan, Ghislaine van Mastrigt, Joey Mattingly, Celine Pribil, Chao Song, Gian Luca Di Tanna, and Yee Vern Yong.

The task force is very grateful to the following ISPOR student and new professional members who contributed their time to pilot the CiCERO Checklist: Naveel Atif, Paul Carty, Ziyan Chen, Jinal Chheda, Trupti Dhumal, Jagadeswara Rao Earla, Diána Elmer, Leonie Heron, Fidelia Ida, Chibuzo Iloabuchi, Praneetha Jain, Ki Jin Jeun, Aishwarya Kulkarni, Kelly Lenahan, Ee Vien Low, Rosa Lucchetta, Bhargavi Mahadik, Yves Mbous, Sreeranjani Sundaresan Menon, Abhishek Nair, Haritha S Nath, Nga Nguyen, Mona Nili, Chinyere Okoh, Apurva Pande, Jose Paredes, Jayeshkumar Patel, Uzma Pathan, Hellen Pham, Sivaraj Raman, Ajaree Rayanakorn, Aditi Shah, Ashna Talwar, Yuzhi Tang, Nirma Vadlamudi, Gintare Valentelyte, Felix Wolfrum, and Yan Wu.

ISPOR Systematic Reviews Good Practices Task Force Report

The co-authors also thank the following ISPOR reviewers for their time and insight on the report: Mariana Adwan, Sinaa Alageel, Reem AL-Mutairi, Naveel Atif, Reshmi Bhageerathy, Zafer Caliskan, Nancy El Hoyek, Ben Forrest, Gunhilde Hagen, Vida Hamidi, April Hebert, Fidelia Ida, Min-Hua Jen, Kirandeep Kaur, Usah Kreaden, Long Le, Yanli Li, Ee Vien Low, Rosa Lucchetta, Abdulrahman Mansir, Mansi Mathur, Sabina Nduaguba, Nga Nguyen, Sivaraj Raman, Anna Ringborg, Saeed Taheri, Bernice Tsoi, Gintare Valentelyte, Qingqing Xu, and Ana Yankovsky.

Number of tables = 1

Number of figures = 1

Number of boxes = 4

Number of appendices = 6

Number of words = 5252

Table 1. Overview of major quality and risk of bias criteria for systematic reviews of cost and cost-effectiveness outcomes

N	Title of the stage	Topics covered
Stage 1.	Planning and development	Clear objective Predefined and availability protocol Protocol deviations
Stage 2.	Search for evidence	Update or novel systematic review Comprehensive or rapid review Choice for database(s) Number of databases Comprehensiveness and reproducibility Use of supplementary materials Use of grey literature
Stage 3.	Study selection and eligibility	Process of study selection Eligibility criteria used
Stage 4.	Critical appraisal of included studies	Tools to appraise the included studies Process of appraisal
Stage 5.	Data extraction and synthesis	Process of data-extraction Assessment of heterogeneity

	Methods of synthesis
	Assessment of publication bias
Stage 6. Presentation and reporting	Reporting of included studies
	Reporting of the synthesis

Box 1. Study selection restrictions in eligibility criteria that represent trade-offs between internal validity and generalizability

- **Restriction by publication date**: If only including the last X years, the reviewer may actually increase generalizability to current and future years due to changes in research methods, standard of care, or other parameters.
- **Restriction by country/region**: Restrictions by country/region are frequently motivated by healthcare system and/or cost comparability, increasing internal validity for making statements about those settings (conditional on equally high quality of studies). However, this limits generalizability to those countries/regions included and perhaps to very similar country/regions.
- **Restriction by language**: This restriction can increase validity, but limit generalizability, eg, restricting to English-language publications while searching for US studies, or bias the outcomes, eg, restricting to English-language publications in studies with a global perspective. The challenge of including studies published in many languages is that the reviewer needs to be able to read/translate/interpret the text in each language, which may not be feasible. While in some circumstances, Google Translate or other tools can help to automate translations (31), the accuracy of these translations should be verified to avoid biases in interpretation.

Figure 1. Flowchart illustrating the method to determine data-synthesis in systematic reviews aiming to summarise cost and cost-effectiveness outcomes

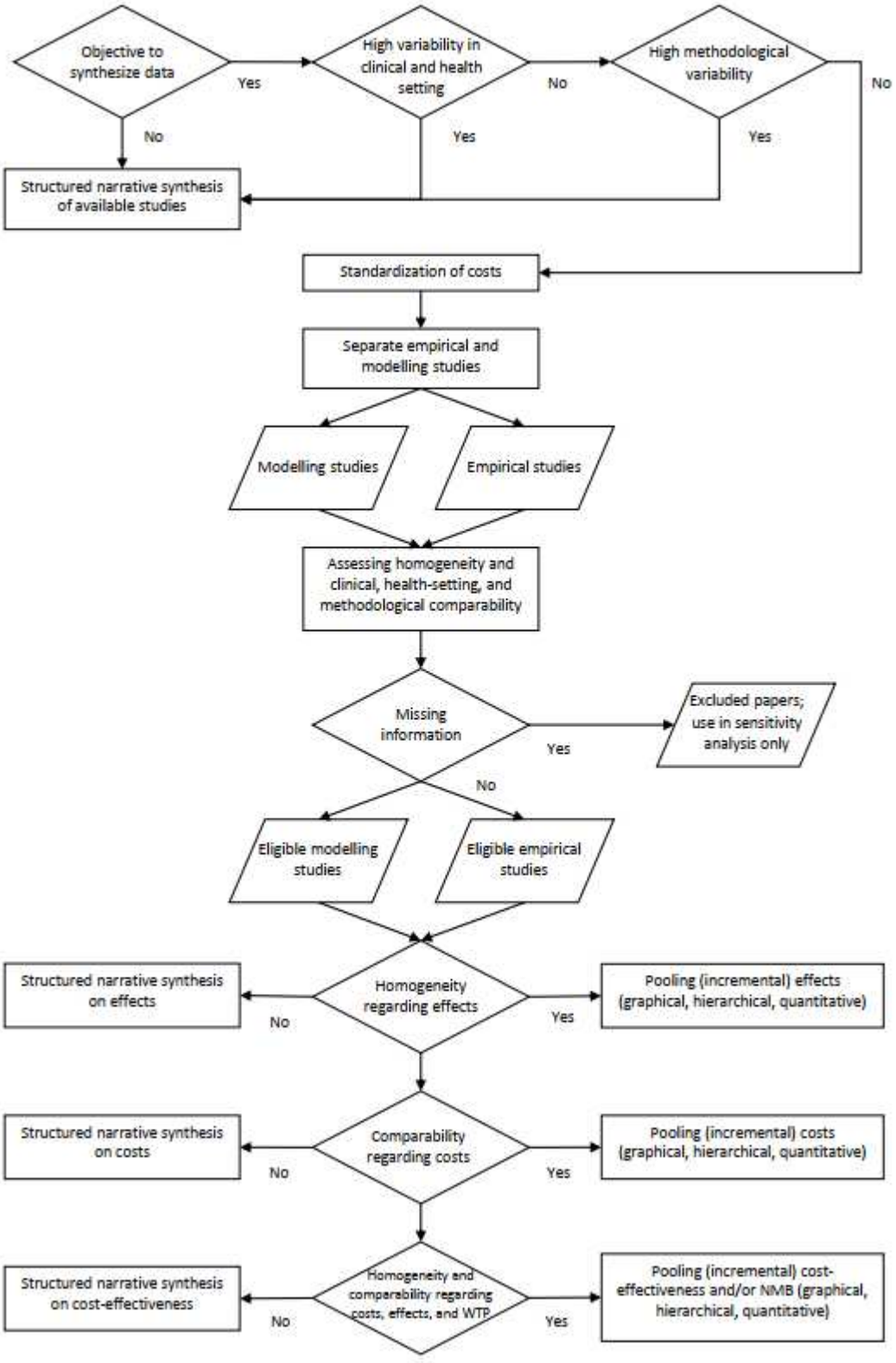

Figure legend: This flowchart can be used to determine what type of data synthesis is feasible in systematic reviews aiming to summarise cost and cost-effectiveness outcomes. If the clinical and health setting and methods are highly variable between included studies, a structured narrative synthesis is warranted. If these characteristics are considered compatible between included studies, reviewers could consider the pooling of effects, standardised costs, cost-effectiveness, or even net monetary benefit (NMB). While task force members agreed that pooling NMB outcomes is possible from a theoretical viewpoint, in practice, it is rarely the case due to the incompatibility of studies and the variation of willingness to pay (WTP) thresholds of the decision-making contexts.

Box 2. Task Force recommendations to assess publication bias in systematic reviews with cost and cost effectiveness outcomes (SR-CCEOs)

- Search relevant grey literature. (See the grey literature sub-section.)
- Search for conference proceedings with published abstracts that did not lead to peer-reviewed publications. (Note: Abstracts should not be searched for inclusion. However, they can be useful to assess possible publication bias.)
- Analyze conflicts of interest (sponsorship) reported in included studies.
- Analyze any differences in studies' outcomes by sponsorship and publication status, ie, differences between grey literature and published reports.
- Assess and explore the direction and magnitude of cost and effect differences in publications, for instance, placing the effectiveness results from cost-effectiveness analyses in the context of existing reviews of clinical effectiveness.
- Analyze the values and interpretations of reported sensitivity analyses (or their lack).
- Benchmark the approaches to exploring the publication bias applied in the clinical reviews, such as looking for the trials' protocols and exploring funnel plot asymmetry (if the SR-CCEO includes empirical studies).

Box 3. Common elements in the existing checklists assessing cost or cost-effectiveness studies

1.	Countries (setting of the study)
2.	Population of analysis (population characteristics)
3.	Audience and study perspective
4.	Time horizon and discounting
5.	Adjustment of inflation
6.	Interventions compared
7.	Method(s) for valuation of cost outcomes
8.	Method(s) for valuation of effectiveness and utility outcomes
9.	Compliance/adherence with intervention (eg, screening uptake)
10.	Decision analytic modelling or calculation approach
11.	Health outcomes (eg, gained life years, number of deaths avoided, QALYs)
12.	Uncertainty (eg, deterministic and probabilistic sensitivity analyses, scenario's, subgroup analyses)
13.	Conflicts of interest and sources of funding
14.	Software (including open source software)

Box 4. Developing the ISPOR CiCERO Checklist

The ISPOR **C**r**I**teria for **C**ost (-**E**ffectiveness) **R**eview **O**utcomes (CiCERO) Checklist is based on the ISPOR Critical Appraisal of Systematic Reviews with Cost and Cost-Effectiveness Outcomes Good Practices Task Force Report. CiCERO * has a series of questions to consider when evaluating the risk of bias in reviews reporting cost or cost-effectiveness outcomes or reviews reporting the methods of these studies.

CiCERO was based on combining aspects of existing instruments, such as the Cochrane Handbook for Systematic Reviews of Interventions, AMSTAR-2 (3), and ROBIS (5) plus the deliberation of international experts - task force members representing different stakeholder perspectives from academics to technology assessors and geographies around the world†.

To produce a final checklist, we used a two-stage validation approach to improve the readability and inter-rater agreement in use of the checklist:

(a) By the task force members (eight reviewers, eight reviews, two raters per publication).

(b) By members of the ISPOR student network group (minimum a relevant MSc-level) experienced in assessing cost and cost-effectiveness outcomes (CCEOs) publications and SR-CCEOs (36 reviewers, 27 reviews, 2-4 raters per publication).

The task force members piloted the initial instrument then it was adapted and used by the larger panel of students. Each reviewer assessed the risk of bias in the reviews independently. The reasons for disagreements were analysed resulting in amendments that provided details and clarifications of the checklist. We tested CiCERO on reviews with different objectives: 1) reviews of cost studies, 2) reviews of cost-effectiveness studies, and 3) reviews that summarize methods of cost and costs-effectiveness studies.

Selection of reviews for validation was based on manuscript diversity in terms of clinical areas, geographical focus, objectives (methodological vs synthesis), and outcomes (costs or cost-effectiveness). Comments received from the validation groups and the disagreement rates for each question were analysed to optimize understanding and interpretation of the final version of the checklist.

Finally, the task force report and checklist underwent two formal rounds of review to ensure that the good practice recommendations and checklist meet the high-quality consensus-developed standards of ISPOR Good Practices Task Force Reports.

*There is a shorter version of the CiCERO Checklist for reviews that summarize methods of cost and cost-effectiveness studies and a specific version for SR-CCEOs that are using AMSTAR-2.

† For more details on task force development, please see Appendix A or Criteria and Process for Initiating and Developing an ISPOR Good Practices Task Force.

Appendix A. Background on the ISPOR Good Practices Task Force

The proposal to initiate an ISPOR Good Practices Task Force on systematic reviews with economic outcomes was evaluated by the ISPOR Health Science Policy Council's Task Force Review Committee then recommended to the ISPOR Board of Directors for approval. The objective of the Task force was to provide recommendations on the critical appraisal of quality and risk of bias in systematic reviews with cost and cost-effectiveness outcomes.

The task force was comprised of international subject matter experts representing a diverse range of stakeholder perspectives (academia, research organizations, government, regulatory agencies and commercial entities). The task force met approximately every eigth weeks by teleconference and in person at ISPOR conferences. All task force members reviewed subsequent drafts of the report and provided frequent feedback in both oral and written comments.

To ensure that ISPOR Good Practices Task Force Reports are consensus reports, findings and recommendations are presented and discussed at ISPOR conferences. In addition, the first and final draft reports are circulated to the task force's review group. All reviewer comments are considered. Comments are addressed as appropriate in subsequent versions of the report. Most are constructive improving the report. All reviewers who submit substantive written comments are listed in the acknowledgements section.

For more information on ISPOR Good Practices Task Force, please see: Criteria and Process for Initiating and Developing an ISPOR Good Practices Task Force.

Appendix C. ISPOR CiCERO Checklist: CrIteria for Cost (-Effectiveness) Review Outcomes For systematic literature reviews that summarize cost and cost-effectiveness outcomes

The ISPOR CiCERO Checklist is a tool to assess the quality and risk of bias in systematic reviews of cost and cost-effectiveness outcomes[1].

Evaluation approach:
Y = "Yes" or "Probably Yes" N = "No", "Probably No", or "No Information", unless the question specifies otherwise NA = "Not Applicable"
General instructions:
• Answer each question ONLY after providing answers to ALL of the relevant sub-questions. • If **at least one** of the sub-questions is "No", then answer "No". • The questions answered as "NA" should be **excluded** from the grading.

Stage 1. Planning and development	**Possible answers**
Question 1. Is the review conducted according to the predefined protocol?	Y, N
1.1. Was evidence provided to document that the review methods were established *prior* to the conduct of the review?	Y, N
***Comment**:* • *Answer "Yes" if the full-text protocol is accessible. (The review provides a link or a reference to the protocol.)* • *Answer "No" in all other cases.*	
1.2. Did the review report whether there were any deviations from the protocol?	Y, N
Comment: • *Answer "Yes" if the review had deviations from the protocol and reported them or the review reported that there were no deviations from the protocol.* • *Answer "No" in the other cases.*	
	Y, N

[1] For the purpose of the CiCERO Checklist, cost studies are defined as studies analysing the costs of healthcare interventions including cost descriptions and cost-of-illness (economic burden of disease) studies. Sometimes cost studies might be based on an explicit comparison of alternatives. By cost-effectiveness studies we mean full economic evaluations, including cost-minimization, cost-effectiveness analysis, cost-utility analysis, cost-benefit analysis, and cost-consequence analysis.

Question 2. Does the review clearly report targeted population, outcomes, time horizon, study perspective, study design, and, when applicable, intervention(s) and comparator(s)?	

Comment:

- *Answer "Yes" if population, outcomes, study design, time horizon, and study perspective are reported for reviews not focused on comparison of interventions.*
- *Answer "Yes" if population, outcomes, study design, intervention and comparator, time horizon, and study perspective are reported for reviews on interventions (eg cost-effectiveness reviews).*
- *Answer "No" in all other cases.*

Stage 2. Search for evidence	Possible answers
Question 3. Did the review authors provide a detailed search strategy(-ies) for at least one database that includes the search month and year?	*Y, N*
Comment: ● *Answer "Yes" if the review authors provide the search strategy in either the main manuscript or an appendix AND report the search month and year.* ● *Answer "No" in all other cases.*	
Question 4. Is the search comprehensive and adequate?	*Y, N*
4.1 Did the search include an argued range of databases / electronic sources for published literature relevant to the aim of the review?	*Y, N*
Comment: ● *Answer "Yes" if a review has a global focus and includes more than two databases.* ● *Answer "Yes" if a review has a regional/local focus, AND it includes both global and region-specific sources.* ● *Answer "No" in all other cases.*	
4.2 Was supplemental searching conducted to identify relevant reports for cost - or cost-effectiveness outcomes that were not identified in the database search(es)?	*Y, N, NA*
Comment: ● *Answer "Yes" if at least one additional method was used (eg tracking citations, consulting experts or searching relevant websites or references.)). See recommendations on supplementary literature searching.* ● *Answer "NA" if review authors justify why supplementary search was not conducted.* ● *Answer "No" in all other cases.*	
4.3 Was a search for the relevant grey literature performed?	*Y, N, NA*
Comment: ● *Answer "Yes" if the review authors searched for grey literature relevant to the objective, (For example, did they search for HTA reports and/or scientific dissertations? See recommendations in subsection on grey literature search.)* ● *Answer "NA" if the review makes a strong argument on why grey literature was not searched.* ● *Answer "No" if the reviews did not search for the relevant grey literature or did not justify this decision.* ● *Answer "No" in all other cases.*	
4.4 Were the terms and structure of the search strategy sufficient to retrieve as many eligible studies as possible?	*Y, N*
Comment: ● *Answer "Yes" if the search terms were relevant to identify costs or cost-effectiveness studies. (See the recommendations in Stage 2.)*	
Question 5. Were the search dates for the review provided? If "Yes", was any justification for the search date provided?	Y, N
Comment: ● *Answer "Yes" if the review reports the date range, the search dates and the reasons for dates ranges searched.* ● *Answer "Yes" if the review provides the search dates while searching the evidence from commencement.* ● *Answer "No" in all other cases.*	*Y, N*

Stage 3. Study selection and eligibility	Possible answers
Question 6. Are the inclusion criteria relevant?	Y, N
6.1. Did the review authors clearly report their inclusion criteria?	*Y, N*
6.2. Are the inclusion criteria appropriate to answer the research question?	*Y, N*
Question 7. Is the study selection process appropriate?	*Y, N*
7.1 Did the review authors perform <u>each</u> step of the study selection independently in duplicate?	*Y, N*
Comment: • *If all of the steps of the selection process were performed in duplicate, say "Yes".* • *If review authors use the liberal accelerated approach in abstract screening and double reviewing in full-text screening, say "Yes".* • *If artificial intelligence is applied in the article search or screening, say "Yes" if the process was duplicated, and the review authors assess the possible biases by using this approach.* • *Answer "No" in all other cases.* *See the recommendations on the screening approaches.*	
7.2 If any restrictions to evidence inclusion were applied (ex. date, publication format or language), were they justified by the objectives of the review?	*Y, N, NA*
Comment: • *Answer "NA" if there were no restrictions mentioned.* • *Answer "Yes if a justification for restrictions was provided (eg, new technology, targeting the specific country or the region)* • *Answer "Yes" if broad timeline restrictions are applied (>10 years).* • *Answer "No" in all other cases.*	

Stage 4. Critical appraisal of included studies	Possible answers
Question 8. Was an assessment of the methodological quality of included studies performed?	*Y, N*
Comment: ● *Answer "Yes" if <u>any</u> peer-reviewed checklist (relevant to health economic studies) was used and reported to assess methodological quality in the original evidence. (See recommendations for the list of suggested instruments to use).* ● *Answer "Yes" if no checklist was used, but the reviewers considered all important criteria (See Drummond and Jefferson (1996)[2] for the minimum necessary criteria).* ● *Answer "Yes" if no studies were identified, but the methods section describes the methodological quality assessment approach in the manuscript or the protocol.* ● *Answer "No" in all other cases (including when review authors state that they used the checklist, but don't report the outcomes)*	

[2] Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *BMJ*. 1996;313:275-83.

Stage 5. Data extraction and synthesis	Possible answers
Question 9. Were the studies' risk of bias considered in the review's synthesis?	*Y, N, NA*
Comment: • *Answer "Yes" if the review authors identified and synthesized only the studies with a low risk of bias.* • *Answer "Yes" if the review authors excluded studies based on risk of bias but assessed the impact of such exclusion on the results.* • *Answer "NA" if no studies were identified or if the review's goal was to assess the methods, not synthesize the findings.* • *Answer "No" in all other cases.*	
Question 10. Were appropriate methods used to combine the results?	*Y, N*
10.1 Was the choice of the method(s) for data synthesis explained?	*Y, N*
***Comment**:* • *Answer "Yes" if the reviewers either explained their selection of the applied method(s) or argued why they did not select alternative method(s) of synthesis.* • *Answer "No" in other cases.*	
10.2 Were the cost data standardized?	*Y, N*
***Comment**:* *Answer "Yes" if <u>at least one</u> approach was applied:* • *all cost data was converted into the same currency and expressed in the same year or* • *costs were standardized to a percentage of GDP or healthcare expenditure or* • *another standardization approach was used.*	
10.3 Was the data synthesised in a de-aggregated manner, distinguishing individual components of effects, costs, and resource use from incremental results?	*Y, N*
10.4 Was the synthesis appropriate considering the target audience of the synthesis?	*Y, N, NA*
***Comment**:* • *Answer "Yes" if the review had a target audience specified and explained how synthesis was applicable to the target audience within a specified setting/ context, eg country-specific HTA.* • *Answer "No" if the target audience was specified, but not considered/explained in synthesis.* • *Answer "NA" if no target audience was specified by the review.* • *Answer "No" in other cases.*	
10.5 Was the synthesis appropriate, given the nature and similarity in the research questions (participants, interventions and comparators), study designs and outcomes across included studies?	*Y, N*
***Comment**:* • *Answer "Yes" if the review synthesized homogenous studies or applied qualitative synthesis with heterogeneous findings.*	

10.6. Was relevant between-study variation due to transferability (difference in jurisdiction/setting/context) described and addressed in the synthesis?	Y, N
10.7 If relevant, were the results from empirical cost or cost-effectiveness studies and modelling studies synthesized separately?	*Y, N, NA*
***Comment**:* • *Answer "NA" if the review did not include studies of different designs.*	
10.8 Were results from deterministic and probabilistic sensitivity analysis reported separately?	*Y, N, NA*
***Comment**:* • *Answer "NA" if the results report one type of synthesis only.*	
10.9 For meta-analysis: Was homogeneity of data properly assessed *prior* to pooling the data together? (For levels of homogeneity assessment, see Stage 5.) • **Was the weighting technique justified?**	*Y, N, NA*
***Comment**:* • *Answer "Yes" if homogeneity of data was properly assessed, and when applied, the weighting technique was justified.* • *Answer "No" if homogeneity of data was not properly assessed, or when applied, the weighting was not justified.* • *Answer "NA" if meta-analysis was not applied.*	
10.10 For narrative synthesis (including graphical synthesis): Was the data synthesized in a comprehensive, structured narrative way?	Y, N

Stage 6. Presentation and reporting	Possible answers
Question 11. Were the original studies included in the review described in adequate detail?	*Y, N, NA*
Comment: *Answer "NA" for each sub-question of question 11 if no studies were identified.*	
The reviews should report the following points for each of the included studies:	
11.1. Country of studied population	*Y, N, NA*
11.2. Description of the population of analysis	*Y, N, NA*
11.3. Time horizon, study perspective	*Y, N, NA*
11.4. Discount rate	*Y, N, NA*
Comment: *Answer NA if only short-term trials were involved, ie, one-year horizon or less.*	
11.5. Adjustment of inflation	*Y, N, NA*
11.6. Interventions compared	*Y, N, NA*
Comment: *Answer "NA" if comparing interventions was not an objective of the review (eg, cost-of-illness / burden of disease)*	
11.7. Method(s) for valuation of economic outcomes	*Y, N, NA*
(a) Cost(s) in the healthcare sector according to the horizon of interest (direct costs, capital costs)	*Y, N, NA*
(b) Indirect medical costs	*Y, N, NA*
(c) Costs outside the healthcare sector, such as productivity loss (indirect costs)	*Y, N, NA*
11.8. Method(s) for valuation of effectiveness outcomes, including source, type of source, estimates, duration (when relevant)	*Y, N, NA*
Comment: *Answer "NA" if assessing cost-effectiveness was not an objective of the review (eg cost-minimization, cost-of-illness / burden of disease or other costs analysis).*	
11.9. Compliance/adherence with treatment	*Y, N, NA*
Comment: *Answer "NA" if the review has a top-down macro-level approach or if the review analyzes an intervention performed without any follow up.*	
11.10. Decision analytic modelling or approach to calculation of economic outcomes	*Y, N, NA*
Comment: *Answer "NA" if the review includes only within-trial cost or cost-effectiveness studies.*	
11.11. Cost outcomes and/or health outcomes, eg, gained life years, number of deaths avoided, or QALY, and outcomes of economic value of an intervention, eg ICER or INHB.	*Y, N*

11.12. Uncertainty	*Y, N*
Comment: *Answer "Yes" if the review reported whether analyses are deterministic or probabilistic or based on other types of simulation.*	
11.13. Conflicts of interest and sources of funding	*Y, N*
11.14. Software used (R, STATA, SAS, Excel, SPSS etc)	*Y, N*
Question 12. Was any heterogeneity observed in the results of the review explored and discussed?	*Y, N, NA*
Comment: • *Answer "Yes" if the findings were homogenous or if the review authors explored heterogeneity in the results and discussed it.* • *Answer "NA" if no studies were identified or if the review aimed to assess the methods and not synthesize the findings.* • *Answer "No" in all other cases.*	
Question 13. Were the biases related to findings of the conducted review, including the conflicts of interest and funding of the reviewers, discussed?	*Y, N*

Appendix D. ISPOR CiCERO Checklist: Cost and cost-effectiveness considerations in reviews aiming to analyze the research methods

The CiCERO Checklist is a tool to assess the quality and risk of bias in systematic reviews of cost and cost-effectiveness outcomes[1].

Evaluation approach:
Y = "Yes" or "Probably Yes" N = "No", "Probably No", or "No Information", unless the question specifies otherwise NA = "Not Applicable"
General instructions:
• Answer each question ONLY after providing answers to ALL the relevant sub-questions. • If **at least one** of the sub-questions is "No", then answer "No". • The questions answered as "NA" should be **excluded** from the grading.

[1] For the purpose of the CiCERO Checklist, cost studies are defined as studies analysing the costs of healthcare interventions including cost descriptions and cost-of-illness (economic burden of disease) studies. Sometimes cost studies might be based on an explicit comparison of alternatives. By cost-effectiveness studies we mean full economic evaluations, including cost-minimization, cost-effectiveness analysis, cost-utility analysis, cost-benefit analysis, and cost-consequence analysis.

Stage 1. Planning and development	Possible answers
Question 1. Is the review conducted according to the predefined protocol?	Y, N
1.1. Was evidence provided to document that the review methods were established *prior* to the conduct of the review?	Y, N
Comment: ● *Answer "Yes" if the full-text protocol is accessible. (The review provides a link or a reference to the protocol.)* ● *Answer "No" in all other cases.*	
1.2. Did the review report whether there were any deviations from the protocol?	Y, N
Comment: ● *Answer "Yes" if the review had deviations from the protocol and reported them or the review reported that there were no deviations from the protocol* ● *Answer "No" in the other cases*	
Question 2. Does the review clearly report targeted population, outcomes, time horizon, study perspective, study design, and, when applicable, intervention(s) and comparator(s)?	Y, N
Comment: ● *Answer "Yes" for methodologic reviews, if at least the study design is reported.* ● *Answer "No" in all other cases.*	

Stage 2. Search for evidence	Possible answers
Question 3. Did the review authors provide a detailed search strategy(-ies) for at least one database that includes the search month and year?	*Y, N*
Comment: • *Answer "Yes" if the review authors provide the search strategy in either the main manuscript or an appendix AND report the search month and year.* • *Answer "No" in all other cases.*	
Question 4. Is the search comprehensive and adequate?	*Y, N*
4.1. Did the search include an argued range of databases / electronic sources for published literature relevant to the aim of the review?	*Y, N*
Comment: • *Answer "Yes" if a review has a global focus and includes more than two databases.* • *Answer "Yes" if a review has a regional/local focus, AND it includes both global and region-specific sources.* • *Answer "No" in all other cases.*	
4.2. Was supplemental searching conducted to identify relevant reports for cost - or cost-effectiveness outcomes that were not identified in the database search(es)?	*Y, N, NA*
Comment: • *Answer "Yes" if at least one additional method was used (eg tracking citations, consulting experts or searching relevant websites or references.)). See recommendations on supplementary literature searching.* • *Answer "NA" if review authors justify why supplementary search was not conducted.* • *Answer "No" in all other cases.*	
4.3. Was a search for the relevant grey literature performed?	*Y, N, NA*
Comment: • *Answer "Yes" if the review authors searched for grey literature relevant to the objective, (For example, did they search for HTA reports and/or scientific dissertations? See recommendations in subsection on grey literature search.)* • *Answer "NA" if the review makes a strong argument on why grey literature was not searched.* • *Answer "No" if the reviews did not search for the relevant grey literature or did not justify this decision.* • *Answer "No" in all other cases.*	
4.4. Were the terms and structure of the search strategy sufficient to retrieve as many eligible studies as possible?	*Y, N*
Comment: • *Answer "Yes" if the search terms were relevant to identify costs or cost-effectiveness studies. (See the recommendations in Stage 2.)*	
Question 5. Were the search dates for the review provided? If "Yes", was any justification for the search date provided?	Y, N
Comment: • *Answer "Yes" if the review reports the date range, the search dates and the reasons for dates ranges searched.* • *Answer "Yes" if the review provides the search dates while searching the evidence from commencement.* • *Answer "No" in all other cases.*	*Y, N*

Stage 3. Study selection and eligibility	Possible answers
Question 6. Are the inclusion criteria relevant?	Y, N
6.1. Did the review authors clearly report their inclusion criteria?	*Y, N*
6.2. Are the inclusion criteria appropriate to answer the research question?	*Y, N*
Question 7. Is the study selection process appropriate?	*Y, N*
7.1. Did the review authors perform <u>each</u> step of the study selection independently in duplicate?	*Y, N*
Comment: • *If not all of the steps of the selection process were performed in duplicate, say "Yes"* • *If review authors use the liberal accelerated approach in abstract screening and double reviewing in full-text screening, say "Yes"* • *If artificial intelligence is applied in the article search or screening, say "Yes" if the process was duplicated, and the review authors assess the possible biases by using this approach.* • *Answer "No" in all other cases.* *See the recommendations on the screening approaches.*	
7.2. If any restrictions to evidence inclusion were applied (ex. date, publication format or language), were they justified by the objectives of the review?	*Y, N, NA*
Comment: • *Answer "NA" if there were no restrictions mentioned.* • *Answer "Yes if a justification for restrictions was provided (eg, new technology, targeting the specific country or the region), or* • *Answer "Yes" if broad timeline restrictions are applied (>10 years).* • *Answer "No" in all other cases.*	

Stage 4. Critical appraisal of included studies	Possible answers
Question 8. Was an assessment of the methodological quality of included studies performed?	*Y, N*

Comment:

- *Answer "Yes" if <u>any</u> peer-reviewed checklist (relevant to health economic studies) was used and reported to assess methodological quality in the original evidence. (See recommendations for the list of suggested instruments to use).*
- *Answer "Yes" if no checklist was used, but the reviewers considered all important criteria (See Drummond and Jefferson (1996)[2] for the minimum necessary criteria).*
- *Answer "Yes" if no studies were identified, but the methods section describes the methodological quality assessment approach in the manuscript or the protocol.*

Answer "No" in all other cases (including when review authors state that they used the checklist, but don't report the outcomes)

[2] Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *BMJ*. 1996;313:275-83.

Stage 5. Data extraction and synthesis	Possible answers
9. Was the data synthesized in a comprehensive, structured narrative way?	Y, N

Stage 6. Presentation and reporting	Possible answers
Question 10. Were the original studies included in the review described in adequate detail?	*Y, N, NA*
Comment: *Answer "NA" for each sub-question of question 10 if no studies were identified.*	
The reviews should report the following points for each of the included studies:	
10.1. Country of studied population	*Y, N, NA*
10.2. Description of the population of analysis	*Y, N, NA*
10.3. Time horizon, study perspective	*Y, N, NA*
10.4. Interventions compared	*Y, N, NA*
Comment: *Answer "NA" if comparing interventions was not an objective of the review (eg, cost-of-illness/burden of disease)*	
10.5. Method(s) for valuation of economic outcomes	*Y, N, NA*
(a) Cost(s) in the health care sector according to the horizon of interest (direct costs, capital costs)	*Y, N, NA*
(b) Indirect medical costs	*Y, N, NA*
(c) Costs outside the healthcare sector such as productivity loss (indirect costs)	*Y, N, NA*
10.6. Method(s) for valuation of effectiveness outcomes, including source, type of source, estimates, duration (when relevant)	*Y, N, NA*
Comment: *Answer "NA" if assessing cost-effectiveness was not an objective of the review (eg cost-minimization, cost-of-illness/burden of disease or other costs analysis).*	
10.7. Decision analytic modelling or approach to calculation of economic outcomes	*Y, N, NA*
Comment: *Answer "NA" if the review includes only within-trial cost or cost-effectiveness studies.*	
10.8. Conflicts of interest and sources of funding	*Y, N*
10.9. Software used (R, STATA, SAS, Excel, SPSS etc)	*Y, N*
Question 11. Were the biases related to findings of the conducted review, including the conflicts of interest and funding of the reviewers, discussed?	*Y, N*

Appendix E. ISPOR CiCERO Checklist: Cost & Cost-Effectiveness Considerations for AMSTAR-2 Users

The CiCERO Checklist is a tool to assess the quality and risk of bias in systematic reviews of cost and cost-effectiveness outcomes[1].

Evaluation approach:
Y = "Yes" or "Probably Yes" N = "No", "Probably No", or "No Information", unless the question specifies otherwise NA = "Not Applicable"
General instructions:
• Answer each question ONLY after providing answers to ALL the relevant sub-questions. • If **at least one** of the sub-questions is "No", then answer "No". • The questions answered as "NA" should be **excluded** from the grading.

[1] For the purpose of the CiCERO Checklist, cost studies are defined as studies analysing the costs of healthcare interventions including cost descriptions and cost-of-illness (economic burden of disease) studies. Sometimes cost studies may be based on an explicit comparison of alternatives. By cost-effectiveness studies we mean full economic evaluations, including cost-minimization, cost-effectiveness analysis, cost-utility analysis, cost-benefit analysis, and cost-consequence analysis.

Stage 1. Planning and development	Possible answers
Question 1. Does the review clearly report targeted population, outcomes, time horizon, study perspective, study design, and, when applicable, intervention(s) and comparator(s)?	Y, N
Comment: • *Answer "Yes" if population, outcomes, study design, time horizon, and study perspective are reported for reviews not focused on comparison of interventions.* • *Answer "Yes" if population, outcomes, study design, intervention and comparator, time horizon, and study perspective are reported for reviews on interventions (eg cost-effectiveness reviews).* • *Answer "No" in all other cases.*	
Stage 2. Search for evidence	**Possible answers**
Question 2. Is the search comprehensive and adequate?	*Y, N*
2.1. Did the search include an argued range of databases / electronic sources for published literature relevant to the aim of the review?	*Y, N*
Comment*:* • *Answer "Yes" if a review has a global focus and includes more than two databases.* • *Answer "Yes" if a review has a regional/local focus, AND it includes both global and region-specific sources.* • *Answer "No" in all other cases.*	
2.2. Was supplemental searching conducted to identify relevant reports for cost - or cost-effectiveness outcomes that were not identified in the database search(es)?	*Y, N, NA*
Comment*:* • *Answer "Yes" if <u>at least one</u> additional method was used (eg tracking citations, consulting experts or searching relevant websites or references.)). See recommendations on supplementary literature searching.* • *Answer "NA" if review authors justify why supplementary search was not conducted.* • *Answer "No" in all other cases.*	
2.3. Was a search for the relevant grey literature performed?	*Y, N, NA*
Comment*:* • *Answer "Yes" if the review authors searched for grey literature relevant to the objective, (For example, did they search for HTA reports and/or scientific dissertations? See recommendations in subsection on grey literature search.)* • *Answer "NA" if the review makes a strong argument on why grey literature was not searched.* • *Answer "No" if the reviews did not search for the relevant grey literature or did not justify this decision.* • *Answer "No" in all other cases.*	
2.4. Were the terms and structure of the search strategy sufficient to retrieve as many eligible studies as possible?	*Y, N*
Comment: • *Answer "Yes" if the search terms were relevant to identify costs or cost-effectiveness studies. (See the recommendations in Stage 2.)*	

Stage 3. Study selection and eligibility	Possible answers
Question 3. If any restrictions to evidence inclusion were applied (ex. date, publication format or language), were they justified by the objectives of the review?	*Y, N, NA*
Comment: • *Answer "NA" if there were no restrictions mentioned.* • *Answer "Yes if a justification for restrictions was provided (eg, new technology, targeting the specific country or the region), or* • *Answer "Yes" if broad timeline restrictions are applied (>10 years).* • *Answer "No" in all other cases.*	
Stage 4. Critical appraisal of included studies	**Possible answers**
Question 4. Was an assessment of the methodological quality of included studies performed?	*Y, N*
Comment: • *Answer "Yes" if <u>any</u> peer-reviewed checklist (relevant to health economic studies) was used and reported to assess methodological quality in the original evidence. (See recommendations for the list of suggested instruments to use).* • *Answer "Yes" if no checklist was used, but the reviewers considered all important criteria (See Drummond and Jefferson (1996)[2] for the minimum necessary criteria).* • *Answer "Yes" if no studies were identified, but the methods section describes the methodological quality assessment approach in the manuscript or the protocol.* *Answer "No" in all other cases (including when review authors state that they used the checklist, but don't report the outcomes)*	
Stage 5. Data extraction and synthesis	**Possible answers**
Question 5. Were the studies' risk of bias considered in the review's synthesis?	*Y, N, NA*
Comment: • *Answer "Yes" if the review authors identified and synthesized only the studies with a low risk of bias.* • *Answer "Yes" if the review authors excluded studies based on risk of bias, but assessed the impact of such exclusion on the results.* • *Answer "NA" if no studies were identified or if the review's goal was to assess the methods, not synthesize the findings.* • *Answer "No" in all other cases.*	
Question 6. Were appropriate methods used to combine the results?	*Y, N*
6.1 Was the choice of the method(s) for data synthesis explained?	*Y, N*
***Comment**:* • *Answer "Yes" if the reviewer either explained their selection of the applied method(s) or argued why they did not select alternative method(s) of synthesis.* • *Answer "No" in other cases.*	

[2] Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *BMJ*. 1996;313:275-83.

6.2 Were the cost data standardized?	*Y, N*
***Comment**:* *Answer "Yes" if <u>at least one</u> approach was applied:* • *all cost data was converted into the same currency and expressed in the same year or* • *costs were standardized to a percentage of GDP or healthcare expenditure or* • *another standardization approach was used.*	
6.3 Was the data synthesised in a de-aggregated manner, distinguishing individual components of effects, costs, and resource use from incremental results?	*Y, N*
6.4 Was the synthesis appropriate considering the target audience of the synthesis?	*Y, N, NA*
***Comment**:* • *Answer "Yes" if the review had a target audience specified and explained how synthesis was applicable to target audience within a specified setting/ context, eg country specific HTA.* • *Answer "No" if the target audience was specified, but not considered/explained in synthesis.* • *Answer "NA" if no target audience was specified by the review.* • *Answer "No" in other cases.*	
6.5 Was the synthesis appropriate, given the nature and similarity in the research questions (participants, interventions and comparators), study designs and outcomes across included studies?	*Y, N*
***Comment**:* • *Answer "Yes" if the review synthesized homogenous studies or applied qualitative synthesis with heterogeneous findings.*	
6.6. Was relevant between-study variation due to transferability (difference in jurisdiction/setting/context) described and addressed in the synthesis?	Y, N
6.7 If relevant, were the results from empirical cost or cost-effectiveness studies and modelling studies synthesized separately?	*Y, N, NA*
***Comment**:* • *Answer "NA" if the review did not include studies of different designs.*	
6.8 Were results from deterministic and probabilistic sensitivity analysis reported separately?	*Y, N, NA*
***Comment**:* • *Answer "NA" if the results report one type of synthesis only.*	
6.9 For meta-analysis: Was homogeneity of data properly assessed *prior* to pooling the data together? (For levels of homogeneity assessment, see Stage 5.) • **Was the weighting technique justified?**	*Y, N, NA*
***Comment**:* • *Answer "Yes" if homogeneity of data was properly assessed, and when applied, the weighting technique was justified.* • *Answer "No" if homogeneity of data was not properly assessed, or when applied, the weighting was not justified.* • *Answer "NA" if meta-analysis was not applied.*	

6.10 For narrative synthesis (including graphical synthesis): Was the data synthesized in a comprehensive, structured narrative way?	Y, N
Stage 6. Presentation and reporting	**Possible answers**
Question 7. Were the original studies included in the review described in adequate detail?	*Y, N, NA*
Comment: *Answer "NA" for each sub-question of question 11 if no studies were identified.*	
The reviews should report the following points for each of the included studies:	
7.1. Country of studied population	*Y, N, NA*
7.2. Description of the population of analysis	*Y, N, NA*
7.3. Time horizon, study perspective	*Y, N, NA*
7.4. Discount rate	*Y, N, NA*
Comment: *Answer NA if:* *only short-term trials were involved, ie, one-year horizon or less.*	
7.5. Adjustment of inflation	*Y, N, NA*
7.6. Interventions compared	*Y, N, NA*
Comment: *Answer "NA" if comparing interventions was not an objective of the review (eg, cost-of-illness/burden of disease)*	
7.7. Method(s) for valuation of economic outcomes	*Y, N, NA*
(a) Cost(s) in the health care sector according to the horizon of interest (direct costs, capital costs)	*Y, N, NA*
(b) Indirect medical costs	*Y, N, NA*
(c) Costs outside the healthcare sector such as productivity loss (indirect costs)	*Y, N, NA*
7.8. Method(s) for valuation of effectiveness outcomes, including source, type of source, estimates, duration (when relevant)	*Y, N, NA*
Comment: *Answer "NA" if assessing cost-effectiveness was not an objective of the review (eg cost-minimization, cost-of-illness/burden of disease or other costs analysis).*	
7.9. Compliance/adherence with treatment	*Y, N, NA*
Comment: *Answer "NA" if the review has a top-down macro-level approach or if the review analyzes an intervention performed without any follow up.*	
7.10. Decision analytic modelling or approach to calculation of economic outcomes	*Y, N, NA*

Comment: *Answer "NA" if the review includes only empirical cost or cost-effectiveness studies.*	
7.11. Cost outcomes and/or health outcomes, eg, gained life years, number of deaths avoided, or QALY, and outcomes of economic value of an intervention, eg ICER or INHB.	*Y, N*
7.12. Uncertainty	*Y, N*
Comment: *Answer "Yes" if the review reported whether analyses are deterministic or probabilistic or based on other types of simulation.*	
7.13. Conflicts of interest and sources of funding	*Y, N*
7.14. Software used (R, STATA, SAS, Excel, SPSS etc)	*Y, N*