# ILCOR Task Force Systematic Reviews (TFSR) of interventions: step by step guide

## Background

The International Liaison Committee on Resuscitation (ILCOR) has created a number of processes to assist in the evaluation of the published science for resuscitation and related first aid. The Task Force Based Systematic Review is designed to be a rigorous process following a strict methodology, which can result in the construction of a Consensus on Science Statement and Treatment Recommendation (CoSTR) as well as a publishable manuscript. The methodology is based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA: http://www.prisma-statement.org) and the approach proposed by the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) working group (<https://gdt.gradepro.org/app/handbook/handbook.html>). The overall process is coordinated by the Scientific Advisory Committee (SAC) of ILCOR, but guided by the task force SAC representatives.

The individual task forces have prioritised the topics that they wish to review. These topics have originated or evolved from (i) previously created Population Intervention Comparison Outcome (PICO) questions (eg. from the 2010 and 2015 processes), (ii) topics submitted for consideration by stakeholders (eg. national/international Councils or Guideline writing bodies), or (iii) topics generated by the task force.

### Access to previous worksheets

The 2010 and 2015 digital worksheets can be accessed using the instruction documents “Access to the 2010 Digital Worksheets” and “Access to the 2015 Digital Worksheets” at <https://www.ilcor.org/ilcor-documents/continuous-evidence-evaluation/#Docs>.

### Next steps

Each topic is allocated to a team of international experts to review. The following steps are listed to help these Evidence Reviewers systematically work their way through this process.

## Confirmation of Task Force Systematic Review team

The task force chair with the assistance of the task force SAC representative confirms the membership of the TFSR team. The team consists of a TF SR Team lead who is expected to be a task force member, a member of the SAC, and a number of other content experts. The other content experts can be members of the task force (or other task forces), Domain Leads, other SAC members, and other invited experts outside of the task force. All members will be required to complete an AHA/ILCOR general conflict of interest statement, as well as make specific declaration of conflicts (both financial and intellectual) that relate to the topic being reviewed.

The TF Chairs must check COI disclosures for the team members and resolve any potential conflicts according to the ILCOR COI policy, usually by replacing those members with potential conflicts. More difficult COI questions can be sent to the COI Co-Chairs.

## Complete the SAC TF review PICOST template

The TFSR team leader, under the guidance of the TFSR SAC representative, will be responsible for the completion of the PICOST template for TF reviews (“TFSRScR\_PICOST\_template” on <https://www.ilcor.org/ilcor-documents/continuous-evidence-evaluation/#Docs>), and its submission to SAC for documentation.

The topics to be evaluated need to be converted into the Patient/population, Intervention, Comparison and Outcome (PICO) format. These 4 individual components need to be explicit. Outcomes

As part of the PICO development, the GRADE process requires designation of key outcomes (up to 7) for each PICO question (which are then categorised as critical, important, of limited importance). The outcomes should generally be considered from the perspective of the patient and should be allocated a score on a scale from 1 to 9. Critical outcomes are scored 7-9, important outcomes 4-6, and limited importance 1-3.

Outcomes that need to be considered for most recommendations will include significant adverse effects/morbidity and, of course, survival. Examples of such outcomes are: Neurologically intact survival (critical 9), Discharge from hospital alive (critical 8), and Return of spontaneous circulation (important 6). The listed (key) outcomes should be those previously prioritised by the task force, or in the case of a new topic, agreed by the task force and consistent with previously prioritised outcomes.

The expectation is that all of these key outcomes will be included in the Evidence Profile tables and the CoSTR. If evidence is lacking for a key outcome, this should be acknowledged rather than excluding that outcome.

It is expected that the initial searches are not restricted by these TF prioritized outcomes. Further refinement of the inclusion and exclusion criteria or the prioritised outcomes can occur after the initial search, but this should be clearly documented for inclusion in the task force summary.

### Key studies

Existing key studies or systematic reviews that have been published should be recorded on the PICOST template, as these publications will assist in the development of initial search strategies and updating of the search strategies for the future process.

## Develop/Confirm Search Strategies

The complete ILCOR systematic review process (allowing peer reviewed publication of the TFSR) requires searching of Medline, Embase and Cochrane databases. An initial search could be performed using a Medline search alone (eg as a rapid review: *Tsertsvadze et al. How to conduct systematic reviews more expeditiously? Systematic Reviews (2015) 4:160*), but this will be insufficient to update or re-affirm a CoSTR. if it is considered that the literature would support the development/update/affirmation of a CoSTR, this will require a search of Embase and the Cochrane database as well.

### 2015 PICO topics

The PICO topics evaluated in 2015 had search strategies created in conjunction with an Information Specialist team, and these strategies can be used to search the three databases. These searches should be run to compliment the 2015 searches and need to include dates since the last run search (see individual worksheets: accessed as described above). If dates of last search are not clear, searches should be run from 1st January 2014.

### 2010 PICO topics

The PICO topics evaluated in 2010 had search strategies developed to a variable standard. These search strategies are certainly sufficient for an initial search of Medline/Embase/Cochrane, and may be sufficient for a complete review (and creation of an updated CoSTR). The SAC representative will be able to facilitate the development of new or modified strategies if necessary (using either contracted Information Specialists or individual librarians).

These searches should be run to compliment the 2010 searches and need to include dates since the last run search (see individual worksheets: accessed as described above). If dates of last search are not clear, searches should be run from 1st January 2009.

### New topics

New topics will require a new search strategy. The SAC representative will be able to facilitate the development of these strategies using either contracted Information Specialists (IS) or individual librarians. Please include information regarding recent Systematic Reviews and suggested specific search terms in the submitted TFSRScR PICOST template.

## Inclusion and exclusion criteria

The inclusion and exclusion criteria that are planned to be used to screen the search results should be documented. They may need to be altered after initial review of the studies identified by the search, but the updated (final) inclusion and exclusion criteria need to be recorded.

The criteria would normally refer to the PICO components (eg. “animal studies and manikin studies will be excluded”) but need to also refer to study characteristics (eg. “Unpublished studies (eg. conference abstracts, trial protocols), and case series will be excluded”).

An example of inclusion/exclusion criteria is:

*Included all studies with concurrent controls which addressed the PICOST question. Excluded review articles, studies with historical controls, animal studies, and studies that did not specifically answer the question. Excluded unpublished studies, and studies only published in abstract form, unless accepted for publication.*

Further refinement of the inclusion and exclusion criteria can occur after the initial search, but this should be clearly documented for inclusion in the summary if necessary.

### Task force review

The list of inclusion and exclusion criteria should be presented to the task force for feedback.

## PROSPERO registration

If the TFSR team would like to publish their review, then the PROSPERO registration process is required to be completed before data abstraction (<https://www.crd.york.ac.uk/prospero/>) and approved by the SAC representative. Guidance for completing this registration is available on <https://www.ilcor.org/ilcor-documents/continuous-evidence-evaluation/#Docs> (Prospero Registration template).

## Initial search results

The list of articles from the initial searches of Medline, Embase and the Cochrane Central Register of ControlLed Trials (Central) should be collated and stored. Exclusion of duplicate publications can be performed manually (eg. using word or Excel files) or using programs such as Endnote or Covidence. If the searches are performed by an IS or librarian, the results should be collated (ideally in Word and Endnote files), and the list of articles need to be forwarded to the TFSR team as an EndNote file and a word file. We would expect that the number of studies included for initial review would be less than 1000 but will depend on the breadth of the search and the years being searched.

## Review of titles and abstracts

The overwhelming majority of the studies will be able to be excluded according to the specific inclusion and exclusion criteria by completing the review of title/abstract. This initial screening should be performed by two content experts, these roles will be allocated by the TFSR team lead.

## Review of full text of studies

The next step is to review the full text of the studies identified for further review by the initial “Review of titles and abstracts”.

The articles identified for further review need to be retrieved for critical appraisal. If the content experts are unable to access any articles, they should contact the SAC representative who will assist with this process or escalate to SAC.

This full text screening should be performed by the same two content experts who were allocated to review the titles and abstracts. Disagreements about decisions should be resolved by the TFSR team lead (or SAC representative if the TFSR team lead is one of the two allocated to review the studies).

The literature relevant to the PICO question is then evaluated, and a re-evaluation of the inclusion/exclusion criteria and the outcomes listed as key can be done at this stage.

Studies that did not include any of the final key outcome measures would normally be excluded from further analysis. Studies that identified significant harm may be exceptions to this.

### Task force review

At the point that the list of included studies is complete, the list should be provided to the task force to ensure there are no obvious omissions.

## Evaluation of included studies

The next step involves extracting the data from the studies and performing a bias assessment for each study (based on each key outcome).

### Extracting data

A member of the TFSR team needs to be delegated to extract the study data. Extracted data from each study needs to be entered into a separate row of a standardised spreadsheet (eg. Excel file) or table (eg. Word file). The data elements extracted should include as a minimum the following elements (but should be comprehensive enough to allow meaningful comparisons):

* **Reference:** eg. Aufderheide 2005, 734
* **Methods:** eg. Randomised Controlled Trial
* **Participants:** eg. 230 OOHCA (presumed ≥ 21 years), presumed cardiac, ventilatable with facemask, then intubateable with ETT. Milwaukee, WI, USA
* **Interventions:** eg. Impedance threshold device (facemask then ETT) plus standard CPR
* **Comparisons:** eg. Sham ITD (facemask then ETT) plus standard CPR
* **Outcomes:** eg. Primary: survival to ICU admission I 29/114 (25.4%) vs C 20/116 (17.2%) NS. Secondary: 24 hr survival I 19/114 (16.7%) vs C 14/116 (12.1%) NS; 1 year survival I 4/114 vs C 2/116. No hospital discharge data.
* **Notes:** eg. Enrolled over 8 months. 15:2 (2000 guidelines). Similar adverse effects. Discontinued early. Concern about ventilation rates. Subgroup analysis: better ICU admission with PEA

The end result may look like this (in either Word or Excel):

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Reference | Methods | Participants | Interventions | Comparisons | Outcomes | Notes |
| Aufderheide 2005, 734 | Randomised Controlled Trial | 230 OOHCA (presumed ≥ 21 years), presumed cardiac, ventilatable with facemask, then intubateable with ETT. Milwaukee, WI, USA | Impedance threshold device (facemask then ETT) plus standard CPR | Sham ITD (facemask then ETT) plus standard CPR | Primary: survival to ICU admission I 29/114 (25.4%) vs C 20/116 (17.2%) NS  Secondary: 24 hr survival I 19/114 (16.7%) vs C 14/116 (12.1%) NS; 1 year survival I 4/114 vs C 2/116. No hospital discharge data. | Enrolled over 8 months. 15:2 (2000 guidelines). Similar adverse effects. Discontinued early. Concern about ventilation rates. Subgroup analysis: better ICU admission with PEA |

### Further data analysis

At the point of data extraction, this is the time that a priori sub-group analyses are also created (eg. in separate Tables and Figures). It is expected that the vast majority of TFSRs will not be eligible for or require meta-analyses: please discuss with the SAC representative. Meta-analyses will require further data verification and cleaning and preparation of draft meta- analyses for review by the TFSR team.

### Bias assessment

Each included study needs to have an assessment of its potential bias. A designated member of the TFSR team needs to complete this for the individual key outcomes, as this information will be included in the rows of Evidence Profile tables. A second member of the TFSR team should be allocated to confirm these assessments.

It may be that the estimate for each category of bias is the same for each of the key outcomes, or some may be different (eg. incomplete reporting of long-term survival outcomes). Either way, it should be made clear from the table what the results are for each key outcome. The tool to be used depends on the study type, and on whether the review is planned to be published in the peer reviewed literature. The tools currently recommended by GRADE for the common study types that are encountered for intervention studies are as follows:

#### Randomised Controlled Trials:

As a minimum, the GRADE Handbook lists the following criteria: “Table 5.4: Study limitations in randomized controlled trials” (<https://gdt.gradepro.org/app/handbook/handbook.html#h.m9385o5z3li7> )

* Lack of allocation concealment
* Lack of blinding
* Incomplete accounting of patients and outcome events
* Selective outcome reporting
* Other limitations

For each of these categories, the risk of bias should be allocated as “Low”, “Unclear” or “High”. A table (in Word or Excel) would look something like this:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Study** | **Lack of Allocation concealment\*** | **Lack of Blinding\*** | **Incomplete accounting of patients and outcome events\*** | **Selective outcome reporting\*** | **Other limitations** | **Outcomes to which these assessments apply** | **Overall risk of bias for outcome(s) for study\*\*** |
| Aufderheide 2005, 734  ITD+SCPR vs ShamITD+SCPR | Low | Low | Low | Low | Discontinued early. Indirectness: 2000 guidelines. | All | Low |
| Aufderheide 2011, 798  ITD+SCPR vs ShamITD+SCPR | Low | Low | Low | Low | Indirectness: 2005 guidelines | All | Low |
| Aufderheide 2011, 301  ITD+ACD vs SCPR  (subset of Frascone 2013, 1214) | High (randomised according to weekly block randomization schedule) | High  (only outcome assessors) | Low | Unclear, some exclusions based on difficultly with airway, border on deviation from IT analysis. | High: Significant differences in real time feedback about CPR quality. Increase enrollment numbers then stop early.  Significant involvement of sponsor. | All | High |

The more comprehensive tool which is currently recommended by Cochrane for RCTs is the “Revised Cochrane risk-of-bias tool for randomized trials (RoB 2)”, details of which can be accessed at <https://sites.google.com/site/riskofbiastool/welcome/rob-2-0-tool/current-version-of-rob-2> . If a publishable systematic review is the goal, then this tool should be used. Please discuss with your SAC representative.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Manuscript | Randomization process | Deviation from intended intervention | Missing outcome data | Measurement of outcome | Selection of reported result | Outcomes to which this assessment applies | Overall |
| Perkins et al, 2018 | Low | Low | Some concernsa | Low | Low | All outcomes | Some concernsa |
| Jacobs et al, 2011 | Low | Low | Low | Low | Low | All outcomes | Low |
| Olasveengen et al, 2009 | Low | Highb | Low | Some concernsc | Low | All outcomes | High |

The end result of this process would look something like this:

#### Non-RCTs:

As a minimum, the GRADE Handbook lists the following criteria in “Table 5.5: Study limitations in observational studies”

(<https://gdt.gradepro.org/app/handbook/handbook.html#h.m9385o5z3li7> )

* Failure to develop and apply appropriate eligibility criteria (inclusion of control population)
* Flawed measurement of both exposure and outcome
* Failure to adequately control confounding
* Incomplete or inadequately short follow-up

For each of these categories, the risk of bias should be allocated as “Low”, “Unclear” or “High”. An example of how this might look is pasted below:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Study** | **Year** | **Industry Funding** | **Eligibility Criteria** | **Exposure/Outcome** | **Confounding** | **Follow up** |
| **Survival Hosp Disch** |  |  |  |  |  |  |
| **Vadeboncoeur** | 2014 | No | Low | Low | Low | Low |
| **Stiell** | 2012 | No | Low | Low | Low | Low |
| **Stiell** | 2014 | No | Low | Low | Low | Low |

The more comprehensive tool which is currently recommended by Cochrane for Non-Randomised studies of Interventions is the “Risk Of Bias In Non-randomized Studies - of Interventions (ROBINS-I)” tool, details of which can be accessed at <https://sites.google.com/site/riskofbiastool/welcome/home?authuser=0> . (see also Schünemann HJ, Cuello C, Akl EA, et al. GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence. J Clin Epidemiol. 2019;111:105–114).

If a publishable systematic review is the goal, then this tool should be used. Please discuss with your SAC representative. An example of the use of this tool would look something like this:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Study**  **(author, year)** | **Confounding** | **Selection** | **Classification of interventions** | **Deviation from intended intervention** | **Missing data** | **Outcomes** | **Selective reporting** | **Outcomes to which this assessment applies** | **Overall** |
| Agostinucci, 20111 | Critical | Low | Low | Low | Low | Low | Moderate | All outcomes | Critical |
| Blumenstein, 20152 | Critical | Low | Low | Low | Moderate | Low | Moderate | All outcomes | Critical |
| Cesana, 20173 | Critical | Critical | Low | Low | Low | Low | Moderate | All outcomes | Critical |
| Chen, 20084 | Critical | Low | Low | Low | Low | Low | Moderate | All outcomes | Critical |

## Creating "Evidence Profile" tables

The next step in the GRADE process is fill out an “Evidence Profile” table to allow documentation of assessment of the evidence to support the key outcomes, across the studies. The Evidence Profile tables are created using the GRADEPro website (<https://gradepro.org> ). Your SAC representative can assist with this process.

There should be one table for each PICO question, and one row for each of the key outcome variables. The tables have a number of criteria that need assessment. They are:

• Study design

• Risk of bias

• Inconsistency

• Indirectness

• Imprecision

• Other (including publication bias)

• Quality of evidence for outcome

The end result is a table which can be exported, for example in the following format:

| **Certainty assessment** | | | | | | | **№ of patients** | | **Effect** | | **Certainty** | **Importance** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **№ of studies** | **Study design** | **Risk of bias** | **Inconsistency** | **Indirectness** | **Imprecision** | **Other considerations** | **ITD** | **no ITD** | **Relative (95% CI)** | **Absolute (95% CI)** |
| **Neurologically intact survival at hospital discharge (Aufderheide 2011, 798) (assessed with: Modified Rankin ≤ 3)** | | | | | | | | | | | | |
| 1 | randomised trials | not serious | not serious | not serious | serious a | none | 254/4373 (5.8%) | 260/4345 (6.0%) | **RR 0.97** (0.82 to 1.15) | **2 fewer per 1,000** (from 11 fewer to 9 more) | ⨁⨁⨁◯ MODERATE | CRITICAL |

When updating a 2015 PICO, there will be existing Evidence Profile (EP) tables. If the new search has identified studies which should be incorporated into these EP tables, the information can be used to update the content and evaluation of certainty within the EP tables (eg. Imprecision, Inconsistency etc.). If any modification of the existing data is required (eg. Relative or Absolute Effect, meta-analyses etc.), this should be coordinated with the SAC representative and may need to be escalated to SAC for assistance.

Further guidance for completing these columns is listed below:

### Study design

The study design(s) for the studies included in assessment of each outcome should be described here. In general studies should be grouped together. If there is evidence from RCTs, then lower levels of evidence may not need to be included: please discuss with your SAC representative. Please list the study according to standard criteria: Randomised Controlled Trial, Observational studies, Case reports, animal studies etc.

### Risk of bias

The overall risk of bias for each study relevant to each key outcome was allocated in the “Risk of bias in individual studies” table. In the Evidence Profile table, a summary assessment is required for each outcome across the studies. The possible outcomes are:

• No serious limitations: Most information is from studies at low risk of bias.

• Serious limitations: Most information is from studies at moderate risk of bias.

• Very serious limitations: Most information is from studies at high risk of bias.

### Inconsistency

Variability in the magnitude of effect may be due to differences in population, interventions, comparisons, outcomes or differences in study design (and inherent risk of bias). Inconsistency is a concept that also considers the extent to which the studies which look at the same outcomes agree with each other yielding consistent findings. Again reviewers are asked to assess the studies that report that outcome as having:

• No serious inconsistency

• Serious inconsistency

• Very serious inconsistency

Reviewers should document limitations when (1) point estimates vary widely across studies, (2) confidence intervals show minimal or no overlap (ie. studies appear to have different effects), or (3) statistical tests of heterogeneity are suggestive (these would normally be performed by the SAC representative if needed).

### Indirectness of evidence

The GRADE process describes direct evidence as “research that directly compares the interventions in which we are interested, delivered to the populations in which we are interested, and measures the outcomes important to patients”. Concerns about directness therefore arise when there are differences in the population (eg. patients in cardiac arrest vs not in cardiac arrest), intervention (eg. standard CPR using 2015 guidelines vs standard CPR using 2005 guidelines), or outcomes (eg. ROSC vs termination of VF for 5 seconds), or where there are no head to head comparisons between interventions.

Again reviewers are asked to assess the studies that report that outcome as:

• No serious indirectness

• Serious indirectness

• Very serious indirectness

In general, allocating limitations as serious or very serious should only be considered where there is a compelling reason to think that the biology in the population of interest is so different that the magnitude of effect will differ substantially (eg. cardiac arrest victim vs stroke victim). Evidence from animal studies, manikins or other models would generally be rated as having very serious limitations (but this would be dependent on the key outcomes listed).

Interventions may be delivered/implemented differently in different settings (eg. therapeutic hypothermia). Limitations should therefore be considered if the intervention differs in some way from what was defined in the PICOST.

Important differences in outcome measures include time frame (eg. hospital discharge vs 6 month survival) or other surrogate outcomes (eg. hospital admission vs neurologically intact survival). Usually data that relies on surrogate outcomes would result in an allocation of serious or very serious limitations.

Limitations in more than one type of directness may also suggest a need to rate the studies as having very serious limitations.

### Imprecision

The assessment of precision and imprecision is a complex matter. Again reviewers are asked to assess the studies that report that outcome as:

• No serious imprecision

• Serious imprecision

• Very serious imprecision

The Confidence Intervals around a result allow us to assess the range in which the true effect lies. If the Confidence Intervals were not sufficiently narrow (such as overlap with a clinical decision threshold, eg. 1% absolute difference) the quality would be rated as serious limitations (or as very serious limitations if the CI is very wide). Another way of describing this is where the recommendation would be altered if upper boundary of the CI or the lower boundary of the CI represented the true effect. Factors that may further influence this decision include the importance of the outcome, the adverse effects, the burden to the patient, resource use, and the difficulty of introducing a measure into practice. Trials stopped early, and early publications (particularly if small) both overestimate the actual treatment effect. The total number of patients included in the review should exceed the number of patients generated by a conventional sample size calculation for an adequately powered clinical trial (assuming alpha of 0.05, beta of 0.2, and Relative Risk Reduction or Relative Risk increase of 25% or more [for binary outcomes], or a Minimally Important Difference, and Standard Deviation from a relevant study [for continuous outcomes]). This “Optimal Information Size” can be estimated from tables in Appendix A. If however the baseline risk is small (<5%) and sample sizes are large (such as >2000 patients per group) it can be assumed that no serious limitations are present. If the total number of patients across studies experiencing the outcome of interest is less than 400 for both dichotomous and continuous outcomes, reviewers should consider rating the limitations as serious.

### Publication bias

Unidentified studies may yield systematically different estimates of beneficial effects of an intervention. Studies with positive results are much more likely to be published (OR 3.9; 95% CI 2.68-5.68). Biased conclusions can result from early review (missing studies with delayed publication [more likely with negative studies]), restricting the search to English language journals, or not including “gray” literature (eg. clinical trial registers, abstracts, theses). Discrepancies between meta-analyses of small studies and subsequent large RCTs occur in approximately 20% of cases, in part due to publication bias.

GRADE suggests that the rating for publication bias across studies should be allocated:

• undetected, or

• strongly suspected

Reviewers should allocate strongly suspected when the evidence consists of a number of small studies, especially if these are industry sponsored or if the investigators share another conflict of interest. The risk of publication bias in observational studies is probably larger than RCTs (particularly small studies, data collected automatically or data collected for a previous study). The use of graphical or statistical testing for publication bias may be useful but has limitations and is not routinely recommended. Additional information about unpublished trials can be found in databases such as [www.clinicaltrials.gov](http://www.clinicaltrials.gov) or [www.who.int/ictrp/en/](http://www.who.int/ictrp/en/) .

### Other considerations for evidence based on observational studies

Observational studies that are methodologically rigorous may have their quality rated up where there is a large magnitude of effect, where there is a dose-response gradient, or when all plausible confounders or biases would reduce the demonstrated effect. To be considered methodologically rigorous, observational studies should:

• comprehensively and accurately measure prognostic factors associated with the outcome of interest;

• minimize loss to follow-up;

• accurately measure outcome; and

• conduct an adjusted analysis that accounts for differences in the distribution of prognostic factors between intervention and control groups.

#### Magnitude of effect

A large magnitude effect would be considered as justification to increase the rating by one level (from low to moderate) if:

• relative risk [RR] = 2-5 or RR 0.5-0.2 with no plausible confounders); or

• RR very large >5 or <0.2 and no serious problems with risk of bias or precision (sufficiently narrow confidence intervals);

The reviewer would be more likely to rate up if the above size of effects occurred rapidly and out of keeping with prior gradient of change; in these situations, they would usually be supported by indirect or lower levels of evidence.

If above criteria are all met, and the RR is very large (eg. >5-10) or very low (RR<0.2), rating up by two levels (from low to high) could be considered, but obviously not if other concerns were present (such as risk of bias, imprecision, inconsistency, indirectness and publication bias).

#### Dose-response effect

A dose-response gradient (such as increased effect with increased dose, or decreased time to intervention, or increased intensity or duration of an educational intervention) increases the confidence in the findings of observational studies. In this setting, the rating up by one level could be considered.

#### Issues around confounding

If all plausible prognostic factors are accurately measured in observational studies, and if all the observed residual confounders and biases would diminish the observed effect, then the effect estimate would be strengthened. In this setting, the rating up by one level could be considered.

## Preparation of the summary of the review

The summary outcome of the TFSR could be:

* a new Consensus on Science statement with Treatment Recommendations (CoSTR),
* an update or modification of a previously existing CoSTR, or
* a “Task force Insight”

In all cases, a draft summary is provided by the TFSR team, but the final summary product is a result of task force discussions, based on the information from the Evidence Profile tables, and considered in the light of the categories of the GRADE Evidence To Decision table on the GRADEPro website (<https://gradepro.org> ). These categories are:

* Is there a problem priority?
* What is the overall certainty of this evidence?
* Is there important uncertainty about how much people value the main outcomes?
* Are the desirable anticipated effects large?
* Are the undesirable anticipated effects small?
* Are the desirable effects large relative to undesirable effects?
* Are the resources required small?
* Is the incremental cost small relative to the net benefits?
* What would be the impact on health inequities?
* Is the option acceptable to key stakeholders?
* Is the option feasible to implement?

### Consensus on Science and Treatment Recommendations

The format of the Consensus on Science statement with Treatment Recommendations (CoSTR) is outlined in the CoSTR Instructions document on the ILCOR website (<https://www.ilcor.org/ilcor-documents/continuous-evidence-evaluation/> ). The relevant components of this template should be completed (some may well be not applicable [N/A]), and the completion can be checked by the SAC representative using the relevant (some may well be N/A) components of the “Checklist for CoSTR” document (also on the ILCOR website: <https://www.ilcor.org/ilcor-documents/continuous-evidence-evaluation/> ).

### No recommendation/Task force Insights

A Task force Insights approach may be preferred to a formal CoSTR when there is insufficient data to support a recommendation.

The GRADE process supports not making a recommendation but the reason for the decision should be specified (<https://gdt.gradepro.org/app/handbook/handbook.html#h.gf7oy3kwftmn> ):

1. The confidence in effect estimates is so low that the panels feel a recommendation is too speculative (see the US Preventative Services Task Force discussion on the topic [Petitti 2009; PMID: 19189910].

2. Irrespective of the confidence in effect estimates, the trade-offs are so closely balanced, and the values and preferences and resource implications not known or too variable, that the panel has great difficulty deciding on the direction of a recommendation.

3. Two management options have very different undesirable consequences, and individual patients’ reactions to these consequences are likely to be so different that it makes little sense to think about typical values and preferences.”

The format for the “Task force Insights” includes the opportunity to include a Consensus on Science statement if relevant (sufficiently rigorous evaluation of the studies was performed) but follows with a discussion of the key components of the deliberations of the task force.

Suggested format includes:

#### Statement about why this topic was reviewed.

Examples of these statements are:

* *“This topic was prioritized by the ALS task force because of ongoing controversy in the published literature.”*
* *“This topic was re-evaluated by the BLS task force because it had not been reviewed by ILCOR since 2010.”*

#### Narrative summary of evidence identified

Examples of these statements are:

* *“Three observational studies were identified that were published since 2010. They compare the use of “intervention X” with “comparator Y” in “population Z” in “1234 patients”.”*
* *“The results are inconsistent and do not support making any treatment recommendations.”*

#### Narrative Reporting of the Evidence to Decision table and Task force discussions

The task force should document the key issues that were considered in their deliberations, to provide more transparency about the complexity of the discussions.

Examples of these statements are:

* *“We considered the reported increase in the important but short-term outcome of ROSC but weighed this against the potential costs of implementation.”*
* *“We considered the reported increase in the important but short-term outcome of ROSC and the absence of data suggestion long term harm.”*
* *“There were large differences between the actual interventions between studies, the diversity of populations studied, and the inconsistency of the reported results.”*

### Consider re-running the search

If the time since last running of the search was more than 1 year, or if the task force is aware of new publications, the search should be rerun, and the summary documents (including Bias tables and EP tables) updated if necessary. The SAC representative can advise.

### Prepare manuscript for peer reviewed publication

If the TFSR team consider that the work completed should be submitted for peer reviewed publication, the finalised document should be completed according to the SR checklist, and may be submitted with a SR cover letter. Both documents are posted on <https://www.ilcor.org/ilcor-documents/continuous-evidence-evaluation/#Docs> .

## Posting of the “draft” task force summary for public comment

The summary documents, as outlined in the previous section, are then posted (eg. on ILCOR.org) for public comment.

## Final document prepared for publication

The task force will incorporate the information obtained from public comments into their final summary documentation for the PICO question. This information will then be able to be incorporated into the relevant ILCOR publication.

Good luck!

Peter Morley

Chair, ILCOR Scientific Advisory Committee

23rd September 2019

## References

Alonso-Coello P, Schünemann HJ, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, Treweek S, Mustafa RA, Rada G, Rosenbaum S, Morelli A. **GRADE Evidence to Decision (EtD) frameworks**: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. bmj. 2016 Jun 28;353:i2016.

Alonso-Coello P, Oxman AD, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, Treweek S, Mustafa RA, Vandvik PO, Meerpohl J, Guyatt GH. **GRADE Evidence to Decision (EtD) frameworks**: a systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines. bmj. 2016 Jun 30;353:i2089.

Corbett MS, Eldridge SM, Emberson JR. **RoB 2**: a revised tool for assessing risk of bias in randomised trials. bmj. 2019 Aug 28;366:l4898.

Guyatt GH, Thorlund K, Oxman AD, Walter SD, Patrick D, Furukawa TA, Johnston BC, Karanicolas P, Akl EA, Vist G, Kunz R. GRADE guidelines: 13. **Preparing summary of findings tables and evidence profiles—continuous outcomes**. Journal of clinical epidemiology. 2013 Feb 1;66(2):173-83.

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The **PRISMA statement for reporting systematic reviews** and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS medicine. 2009 Jul 21;6(7):e1000100.

Schünemann HJ, Cuello C, Akl EA, et al. GRADE guidelines: 18. How **ROBINS-I** and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence. J Clin Epidemiol. 2019;111:105–114

Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, Cates CJ, Cheng HY,

Guyatt GH, Oxman AD, Santesso N, Helfand M, Vist G, Kunz R, Brozek J, Norris S, Meerpohl J, Djulbegovic B, Alonso-Coello P. GRADE guidelines: 12. **Preparing summary of findings tables—binary outcomes.** Journal of clinical epidemiology. 2013 Feb 1;66(2):158-72.

Tsertsvadze et al. How to conduct **systematic reviews more expeditiously**? Systematic Reviews (2015) 4:160

## Key websites

<https://gdt.gradepro.org/app/handbook/handbook.html>

<https://gradepro.org> (requires log in or sign up)

<http://www.gradeworkinggroup.org>

<https://www.ilcor.org/ilcor-documents/continuous-evidence-evaluation/#Docs>

## Appendix A: Estimating the Optimal Information Size



