ADOLOPMENT of a Published Systematic Review: Process

# 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. In some situations, a recently published systematic review (SR) may be identified, and duplication of effort may be considered wasteful of resources. The ADOLOPMENT process is designed to guide the adaptation of an existing published systematic review. It is a rigorous process following a strict methodology, which can result in the construction of a Consensus on Science Statement and Treatment Recommendation (CoSTR). The methodology is based on the GRADE-ADOLPMENT approach proposed by the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) working group). The overall process is coordinated by the Scientific Advisory Committee (SAC) of ILCOR but guided by the task force SAC representatives.

The expectation is that the SR is methodologically rigorous, this needs to be assessed using the AMSTAR 2 checklist. To assist us moving forward the SR needs to have the following qualities:

* Based on the TF approved PICOST
* Published search strategy/ies (including date of last search(es))
* Published inclusion and exclusion criteria
* Published data extraction tables (including results) from included studies
* Published Bias assessment tables (including criteria used)
* Published GRADE Evidence Profile (or Summary of Findings) tables or sufficient data to allow creation of GRADE Evidence Profile tables.
* Published meta-analyses (where appropriate)

**If the above criteria are not met**, a decision needs to be made about whether to proceed with the ADOLOPMENT process, to perform a systematic review de novo (TFSR or KSU/ESR), or not to proceed. The task force SAC representative can provide advice about this decision.

# **Confirmation of Task Force ADOLOPMENT team**

The task force chair with the assistance of the task force SAC representative confirms the membership of the TF ADOLOMENT (TFA) team. The team consists of a TFA 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), 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.

# **Review when Search Strategies last run**

The complete ILCOR systematic review requires searching of Medline, Embase and Cochrane databases. If the previous search meets this criterion, and searches were performed in the last 3 months, no further searching is required. If the searches were performed more than 3

months ago, it is recommended that the searches should be revisited (updated) and rerun. It is recommended that all search strategies be peer reviewed by peer information specialists. If published search strategies used in the original systematic review are not available for all databases, the authors should be contacted. The creation of a CoSTR or an update to a CoSTR requires a recent comprehensive search of at least these three databases.

# **Inclusion and exclusion criteria**

The inclusion and exclusion criteria used by the published SR to screen the search results should be documented and used by the TFA team for any new results. If these criteria are considered inadequate for the ADOLOPMENT process, a decision should be made about the next steps (as above, discuss with SAC representative).

# **PROSPERO registration**

Prospero registration of the original systematic review on which this adolopment is based should be cited.

# **Initial results from re-run of search**

If the searches need to be updated, peer reviewed and rerun, the list of articles identified 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 identified by the rerunning of the searches included for initial review would be small.

# **Review of titles and abstracts from re-run of search**

TFA team leads may use software to facilitate the conduct the update of this systematic review. The overwhelming majority of the studies will 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 TFA team lead.

# **Review of full text of studies from re-run of search**

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 additional 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.

Disagreements about decisions should be resolved by the TFA team lead (or SAC representative if the TFA team lead is one of the two allocated to review the studies). Studies that did not include any of the outcome measures identified in the original systematic review 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 TFA team needs to be delegated to extract the study data and add this in a way it can be considered in addition to the data from the previous SR. Extracted data from each study could be entered into a separate row of a standardised spreadsheet (eg. Excel file) or table (eg. Word file).

## Bias assessment

Each included study needs to have an assessment of its potential bias. A designated member of the TFA 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 TFA team should be allocated to confirm these assessments.

For consistency with the published SR, the same Bias assessment tool used in that SR could be used for any newly identified studies. The TFA team should consider using the newer risk of bias tools: eg. RoB 2 for RCTs, and ROBINS-I for non-RCTs for interventions.

# **Updating "Evidence Profile" tables**

If the new search has identified studies which should be incorporated into updated

“Evidence Profile” tables, the information can be used to update the content and evaluation of certainty within the EP tables (eg. Imprecision, Inconsistency etc.). The Evidence Profile tables are created using the GRADEPro website ([https://gradepro.org](https://gradepro.org/) ). Your SAC representative can assist with this process. There should be one table for each PICOST question, and one row for each of the key outcome variables.

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 in Appendix A.

# **Preparation of the summary of the review**

The summary outcome of the TFA 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 TFA 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](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?

Ideally, to support a new or updated CoSTR, an Evidence to Decision framework should be completed. This will obviously require a detailed review of the included studies.

## Consensus on Science and Treatment Recommendations

The format of the Consensus on Science statement with Treatment Recommendations (CoSTR) is outlined in the CoSTR template and Instructions document and a checklist that may be helpful. The relevant components of this template should be completed (some may well be not applicable [N/A]). The SAC representative should review the CoSTR using the relevant (some may well be N/A) components of the “Checklist for CoSTR” document.

Ideally, to support a new or updated CoSTR, a completed Evidence to Decision framework should be submitted as well.

## No recommendation

A Task force Insights approach may be preferred to a formal CoSTR when there is insufficient data to revise the consensus on science and the treatment recommendation.

The GRADE process supports not making a recommendation but the reason for the decision should be specified.

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.”

## Task force Insights

The format for the “Task force Insights” includes 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 a recent published systematic review.”*

*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.”*

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

The summary documents (CoSTR template or task force insights if no CoSTR), are approved by the TF and the SAC rep prior to posting on ILCOR.org for public comment. It is recommended that the SAC rep seek independent SAC review prior to posting.

# **Final document prepared for posting on ILCOR.org and annual ILCOR publication of consensus on science**

The task force will incorporate the information obtained from public comments into their final summary documents (CoSTR template or taskforce insights if no CoSTR) for posting on ILCOR.org. There is a guidance document for how to address public comments on ilcor.org. The final version of the summary documents (CoSTR or Task force insight if no CoSTR) will then be able to be incorporated into the relevant annual ILCOR publication of consensus on science.

# **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.

Schünemann HJ, Wiercioch W, Brozek J, Etxeandia-Ikobaltzeta I, Mustafa RA, Manja V, Brignardello-Petersen R, Neumann I, Falavigna M, Alhazzani W, Santesso N. **GRADE Evidence to Decision (EtD) frameworks for adoption, adaptation, and de novo development of trustworthy recommendations: GRADE-ADOLOPMENT**. Journal of clinical epidemiology. 2017 Jan 1;81:101-10.

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

Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, Moher D, Tugwell P, Welch V, Kristjansson E, Henry DA. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. Bmj. 2017 Sep 21;358:j4008.

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.

Key websites <https://gdt.gradepro.org/app/handbook/handbook.html> [https://gradepro.org](https://gradepro.org/) (requires log in or sign up) [http://www.gradeworkinggroup.org](http://www.gradeworkinggroup.org/)

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

# Appendix A: guide to updating an Evidence Profile table

## 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 updated “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 the TFA team is 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, e.g. 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 B. 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 (e.g. 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.

Appendix B: Estimating the Optimal Information Size



