South Afr J Anaesth Analg. 2022;28(5 Suppl 1):S72-79
https://doi.org/10.36303/SAJAA.2022.28.5.2914

South Afr J Anaesth Analg
ISSN 2220-1181 EISSN 2220-1173

FCA REFRESHER COURSE

The principles of evidence-based medicine

B Biccard

Department of Anaesthesia and Perioperative Medicine, Groote Schuur Hospital, University of Cape Town, South Africa
Corresponding author, email: *bruce.biccard@uct.ac.za*

Keywords: evidence-based medicine, principles

This Refresher Course summarises the key points from the Evidence-based Medicine book of Guyatt and Colleagues.[1]

What is evidence-based medicine?

The three fundamental principles of evidence-based medicine are:

- An awareness of the best available evidence,
- The ability to decide the trustworthiness of the evidence, and
- The consideration of the values and preferences for the patient.

Evidence-based medicine and the theory of knowledge

The epistemological principles of evidence-based medicine are to:

- Examine the totality of the evidence,
- Understand that not all evidence is equal, and
- Acknowledge that evidence is necessary but not sufficient. It must include values and preferences in decision making.

How to define a question

A clinician must be able to;

- Distinguish between 'background' and 'foreground' questions.

Background questions are the established general knowledge of a subject, and foreground questions are the problem-solving or research questions.

- Understand the concepts of 'aim', 'objectives', 'hypotheses' and 'outcomes'.

'Aim' is a broad summary statement of purpose. It is the 'what' that the research question hopes to answer.

'Objectives' are the operationalised steps describing how the aim(s) will be reached. It is how we "quantify" or "determine" the aim.

'Hypotheses' are the testable opposing statements.

'Outcomes' usually have one primary outcome for which the study is powered. The secondary outcomes are essentially descriptive or explorative, i.e. hypothesis-generating for future research.

To frame a question we use the 'PICO' format: Patients or Population, Intervention(s) or Exposure(s), the Comparator and the Outcome.

There are five fundamental types of clinical questions (according to study design): Therapy, Harm, Differential diagnosis, Diagnosis and Prognosis.

How to find the current best evidence

Make the librarian your friend. They have tremendous skills. Bookmark web resources that make finding the current best evidence easier. Bookmark websites according to the levels of evidence as follows:

Summaries and guidelines

- https://www.evidence.nhs.uk/
- https://g-i-n.net/library/international-guidelines-library
- https://www.ahrq.gov/gam/index.html
- https://www.uptodate.com/contents/search
- https://bestpractice.bmj.com/info/

Pre-appraised resources

- https://hiru.mcmaster.ca/hiru/HIRU_McMaster_PLUS_projects.aspx
- https://www.nyam.org/library/collections-and-resources/databases/
- https://www.cochrane.org/
- https://www.crd.york.ac.uk/CRDWeb/
- https://www.essentialevidenceplus.com/content/poems

Non pre-appraised resources

- https://pubmed.ncbi.nlm.nih.gov/
- https://www.embase.com/login

Federated resources (these cross all the above resource groups)

- https://www.accessss.org/
- https://www.tripdatabase.com/
- https://hiru.mcmaster.ca/hiru/HIRU_McMaster_PLUS_projects.aspx
- https://www.epistemonikos.org/

Table I: The six basic study designs

Ecological study	Observational	Retrospective	Occurrence and associations in groups
Case study/case series	Observational	(Usually) Retrospective	Descriptive
Cross-sectional study	Observational	Snapshot*	Descriptive, analytical, diagnostic
Case-control study	Observational	Retrospective	Analytical (cannot describe occurrence)
Cohort study	Observational	Longitudinal (retrospective, concurent, prospective)	Descriptive and analytical
Randomised controlled trial	Experimental	Prospective	Interventional and analytical

* Loss of temporal precedence. The exposure and outcome of interest is measured concurrently. Also true for ecological studies.

My personal favourite resources are:

- Best federated resource: https://www.accessss.org/
- PubMed
- The HIRU Hedges Filters: https://hiru.mcmaster.ca/hiru/HIRU_McMaster_PLUS_projects.aspx

The HIRU filters allow us to set up personalised filters for PubMed (https://hiru.mcmaster.ca/hiru/HIRU_Hedges_home.aspx). This is useful for limiting searches which are prohibitively large.

To conduct a decent literature search, you need to learn how to:

- use MeSH (Medical Subject Headings) in PubMed (https://www.ncbi.nlm.nih.gov/),
- save a search on PubMed,
- import search 'hits' into Endnote,[2] and
- export 'hits' from Endnote into Excel, so you can screen 'hits' for inclusion/exclusion in review. There are good YouTube videos on doing this.

Types of study designs

The various types of study designs are shown in Figure 1 and Table I.

The study design is also associated with the hierarchy of evidence, as shown in Figure 2 from Guyatt et al.[1]

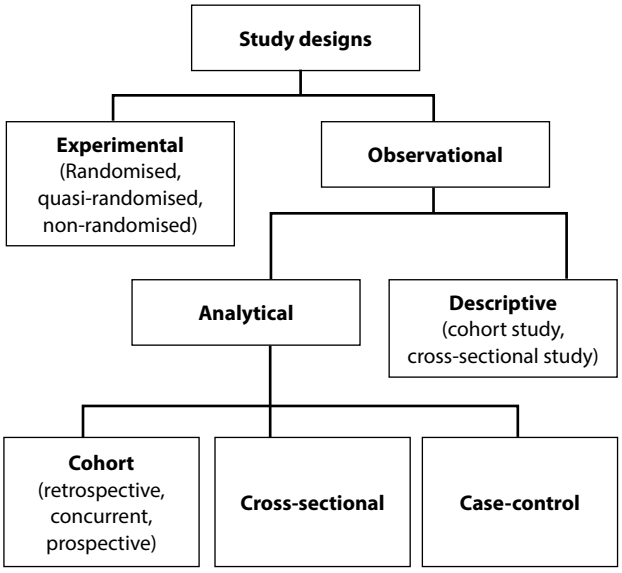

Figure 1: Types of study designs

What are the factors that contribute to bias in studies?

A study of an intervention can be biased if the:

- Intervention and control groups may be different at the start,
- Intervention and control groups become different as the study progresses, and
- Intervention and control groups differ, independent of treatment at the end of the study.

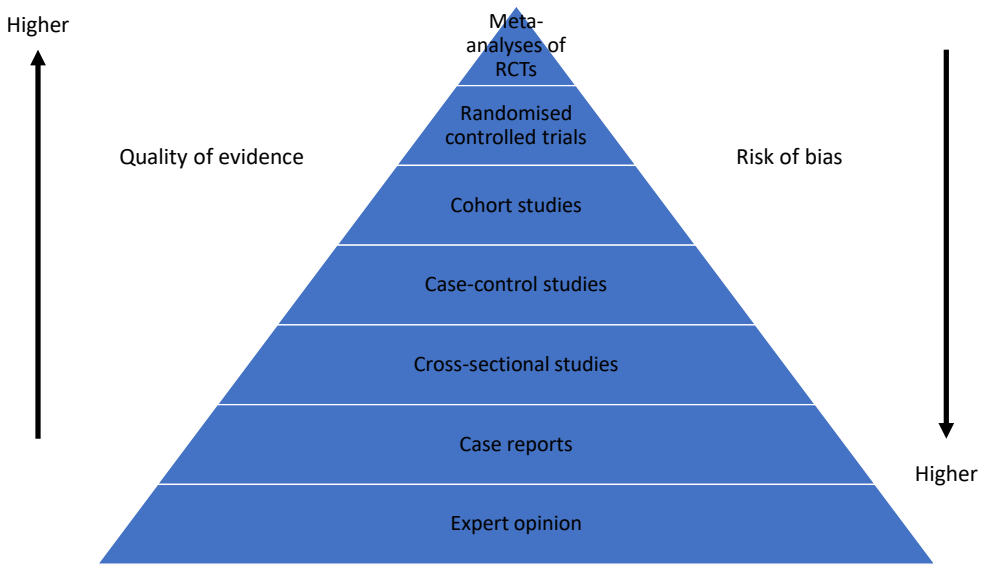

Figure 2: The hierarchy of evidence

Table II: Strategies to reduce the bias in studies of therapy and harm

Source of bias	Therapy: strategy to reduce bias	Harm: strategy to reduce bias
Intervention and control groups may be different at the start		
Treatment and control patients differ in prognosis	Randomisation	Statistical adjustment of prognostic factors
	Randomisation with stratification	Matching
Intervention and control groups become different as the study progresses		
Placebo effects	Blinding of patients	Outcomes associated with less subjective effects, e.g. mortality
Cointervention	Blinding of caregivers	Document treatment differences and statistically adjust
Bias in assessment	Blinding of assessors of outcomes	Document treatment and statistically adjust
Intervention and control groups differ, independent of treatment at the end of the study		
Loss to follow-up	Ensure complete follow-up	Ensure complete follow-up
Stop study early because of large effect	Complete study as initially planned	Not applicable
Omitting patients who did not receive assigned treatment	Include all patients in the arm to which they were randomised	Not applicable

Confidence intervals and statistical and clinical significance

The broad concept of statistical versus clinical importance is summed up as follows:

- the concept of the *p*-value is essentially meaningless for clinical significance, and
- clinical significance is dependent on i) our definition of what clinical difference constitutes clinical significance, and ii) the fragility of the result, i.e. the fragility of the *p*-value.

The fragility[3] is a measure of the robustness (or fragility) of the results of a clinical trial. The fragility index is a number indicating how many patients would be required to convert a trial from being statistically significant to not significant ($p \geq 0.05$) result. The larger the fragility index, the better (more robust) a trial's data are. There are examples of fragility in perioperative medicine and critical care studies/trials.[4,5] A useful website to conduct fragility analyses for studies is https://clincalc.com/Stats/FragilityIndex.aspx

How to assess studies of therapy (or randomised trials)

When assessing an article about therapy, it is important to consider the following three areas:

- How serious was the risk of bias?
  - Did the intervention and control groups start with the same prognosis?
  - Were the patients randomised?
  - Was randomisation concealed?
  - Were patients in the study groups similar with respect to known prognostic factors?
  - Was prognostic balance maintained as the study progressed?
  - To what extent was the study blinded?
  - Were the groups prognostically balanced at the study's completion?
  - Was the follow-up complete?
  - Were the patients analysed in the groups to which they were randomised?
  - Was the trial stopped early?
- What are the results?
  - How large was the treatment effect?
  - How precise was the estimate of the treatment effect?
- How can I apply the results to patient care?
  - Were the study patients similar to my patient?
  - Were all patient-important outcomes considered?
  - Are the likely treatment benefits worth the potential harm and costs?

It is important to consider the sample size. Was the study sufficiently powered for the primary outcome to ensure we do not have a type 2 error, i.e. a false negative result. Useful websites for sample sizes include i) clincalc.com (use the post hoc function), and https://www.sealedenvelope.com/.

How to interpret treatment effect

Once a study is complete, it is important to understand the study results. For binary outcomes, this is presented by a 2x2 contingency table.

Table III: The 2x2 contingency table

Exposure	Outcome	
	Yes	**No**
Yes	a	b
No	c	d

From the 2x2 table, we can define the different variables necessary to describe the outcomes.

Risk with exposure = $\frac{a}{(a+b)}$

Risk without exposure = $\frac{c}{(c+d)}$

Odds with exposure = $\frac{a}{b}$

Odds without exposure = $\frac{c}{d}$

Relative risk = $\frac{a}{(a+b)} / \frac{c}{(c+d)}$

Odds ratio = $\frac{a}{b} / \frac{c}{d}$ = (a x d)/(c x b)

Absolute risk reduction (ARR) = $\frac{c}{(c+d)} - \frac{a}{b}$

Number needed to treat (NNT) = 100/(ARR expressed as a %)

The importance of absolute risk reduction (ARR) is that a higher ARR is associated with a lower NNT.

Considerations when composite endpoints are used

A composite endpoint is often used in a trial. To ensure that it is appropriate to consider the various outcomes together in the composite endpoint, we should consider the following:

- Are the component endpoints of the composite endpoint of similar importance?
- Did the more and less important endpoints occur with similar frequency?
- Can one be confident that the component endpoints are similar enough that one would expect similar relative risk reductions?
- Are the point estimates of the relative risk reductions similar, and are the confidence intervals (CIs) sufficiently narrow?
- To the extent that one can answer yes to these questions, one can feel confident using the treatment effect on the combined endpoint as the basis for decision making.
- To the extent one answers no to these questions, one should look separately at the treatment effect on the component endpoints as the basis for decision making.

How authors may mislead the presentation of clinical trial results

Clinical trial results can be presented in a way that is misleading. To avoid being misled, I would advocate the following:

- Read methods and results: bypass the discussion section.
- Read the summary structured abstract published in evidence-based secondary publications (i.e. pre-appraised resources).
- Beware large effects in trials with only a few events.
- Beware faulty comparators.
- Beware small treatment effects and extrapolation to very low-risk patients.
- Beware uneven emphasis on benefits and harms, and
- Wait for the overall results to emerge; do not rush to respond to early preliminary presentations of the results.

The reasons for being cautious in adopting new interventions based on initial studies are the following:

- Initial studies may be biased by inadequacies in concealment, blinding, loss to follow-up, or stopping early.
- Initial studies are susceptible to reporting bias.
- Initial studies are susceptible to dissemination bias; markedly positive studies are likely to receive disproportionate attention.
- Initial studies may overestimate effects by chance (particularly if effects are large and the number of events is small).
- There is substantial probability (20%) that serious adverse effects will emerge subsequently.
- On rare occasions, research results will prove to have been fraudulent.

Be on the outlook for faulty comparators:

- Comparison with placebo when effective agents are available.
- Comparison with less effective agents when more effective comparators are available.
- Comparison with too low a dose (or inadequate dose titration) of an otherwise effective comparator, leading to misleading claims of effectiveness, and
- Comparison with a too high (and thus toxic) dose (or inadequate dose titration) of an otherwise safe comparator, leading to misleading claims of lower toxicity.

Authors may use strategies to make a treatment effect appear larger than it is:

- Use relative rather than absolute risk; a 50% RRR may mean a decrease in risk from 1% to 0.5% (ARR of 0.5%, NNT of 200).
- Express risk during a long period; the reduction in risk from 1% to 0.5% may occur during 10 years.
- For visual presentations, make sure the x-axis intersects the y-axis well above 0. If the x-axis intersects the y-axis at 60%, you can make an improvement from 70–75% appear as a 33% increase in survival.
- Include a few high-risk patients in a trial of predominantly low-risk patients; even though most events occur in high-risk individuals, claim important benefits for a large number of low-risk patients in the general population.
- Ignore the lower boundary of the CI; when the lower boundary of the CI around the relative risk reduction approaches 0, declare significance and henceforth focus exclusively on the point estimate, and
- Focus on statistical significance; when a result achieves statistical significance but both relative and absolute effects are small; highlight the statistical significance and downplay or ignore the magnitude.

How to evaluate a study of harm (observational studies)

When considering studies about harm, evaluate the following:

- How serious was the risk of bias?
  - In a cohort study, aside from the exposure of interest, did the exposed and control groups start and finish with the same risk for the outcome?
  - Were the patients similar for prognostic factors that are known to be associated with outcome (or did statistical adjustment address the imbalance)?

- Were the circumstances and methods for detecting the outcome similar?
- Was the follow-up complete?
- In a case-control study, did the cases and the control group have the same risk for being exposed in the past?
- Were cases and controls similar with respect to the indication or circumstances that would lead to exposure (or did statistical adjustment address the imbalance)?
- Were the circumstances and methods for determining exposure similar for cases and controls?

- What are the results?
  - How strong is the association between exposure and outcome?
  - How precise was the estimate of risk?
- How can I apply the results to patient care?
  - Were the study patients similar to my patient?
  - Was follow-up sufficiently long?
  - Is the exposure similar to what might occur in my patient?
  - What is the magnitude of the risk?
  - Are there any benefits that are known to be associated with the exposure?

How to interpret a study of prognosis

When evaluating an article about prognosis, you should consider the following:

- How serious is the risk of bias?
  - Was the sample of patients representative?
  - Were the patients classified in prognostically homogenous groups?
  - Was follow-up sufficiently complete?
  - Were the outcome criteria objective and unbiased?
- What are the results?
  - How likely are the outcomes over time?
  - How precise are the estimates of likelihood?
- How can I apply the results to patient care?
  - Were the study patients and their management similar to those in my practice?
  - Was follow-up sufficiently long?
  - Can I use the results in the management of patients in my practice?

Systematic reviews and meta-analysis

A systematic review is a summary of research that addresses a focused clinical question in a systematic and reproducible manner. A meta-analysis is a statistical pooling of results from different studies to provide a single best estimate of effect.

The reasons why we should conduct systematic reviews

1. Single studies may be unrepresentative of the total body of evidence.

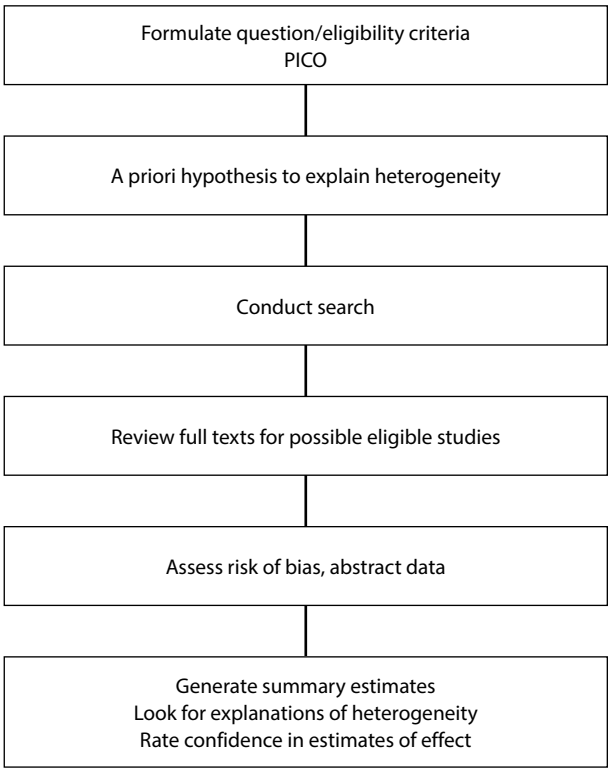

Figure 3: The process of a systematic review

2. An accompanying meta-analysis will provide the best estimate of effect, and increase the precision of that estimate of effect. This data aid clinical decision-making.
3. A systematic review provides information to inform our confidence in the current evidence.

The process of a systematic review

The process of a systematic review is shown in Figure 3.

The credibility of the effect estimates

The two fundamental problems which may adversely affect the credibility of the effect estimates in a meta-analysis are:

1. The credibility of the review, i.e. to what extent did the design and conduct of the review protect against misleading results, i.e. what are the methodological standards of the review process, and
2. The individual studies may include studies with a high risk of bias which will decrease confidence in the estimates.

The rest of this text will address the strategies adopted to either minimise and/or understand whether either of these points has a significant effect on our interpretation of the effect estimates from a meta-analysis.

The credibility of the systematic review process

A systematic review has eight strategies to increase the credibility of the review.

Was the review prospectively registered?

To ensure credibility and remove the ability of the authors from biasing the results, registration of the systematic review prospectively with PROSPERO (https://www.crd.york.ac.uk/prospero/) is recommended. We need to ensure that there is no reporting bias, i.e. where the reviewers report the experimental intervention associated most strongly with the favourable outcome.

Did the review explicitly address a sensible clinical question?

One needs to consider if it is appropriate to aggregate the various studies together in the systematic review. It is important to consider if the underlying biology suggests that across the range of interventions aggregated, one expects a similar treatment effect. Appropriate eligibility criteria for study inclusion in the systematic review is important:

a) Are the results likely to be similar across the range of included patients?

b) Are the results likely to be similar across the range of studied interventions?

c) Are the results likely to be similar across the range of ways in which the outcome was measured, e.g. duration of follow-up?

Explicit eligibility criteria will ensure that the authors' own biases are less likely to influence which studies are included.

Was the search for relevant studies exhaustive?

The literature search needs to be exhaustive, covering a number of biographic databases, e.g. MEDLINE, EMBASE, Cochrane Central Register of Controlled Trials, etc. The reference lists of included articles need to be scrutinised for other studies which may have been missed. Other strategies include the review of abstracts of scientific meetings, and databases of ongoing trials. The appendix of the systematic review should include the exact search strategy used, including search terms for each database.

To limit reporting bias, attempts to identify unpublished studies should also be made, through 'grey literature' searching. Ideally, the full reports of unpublished studies (as opposed to an abstract) should be included.

Was the risk of bias of the primary studies assessed?

Studies with less rigorous methodology are more likely to overestimate the effectiveness of the intervention. A classic example is trials stopped early for efficacy.[6]

The determinants of bias are dependent on the type of study (i.e. therapy, diagnosis, harm, or prognosis). Key factors associated with limited bias are listed below:

a) Therapy: randomisation, was complete follow-up complete?

b) Diagnosis: Is the patient sample representative, was the diagnosis verified by credible criteria?

c) Harm: Adjusted for known determinants of outcome, was follow-up sufficiently complete?

d) Prognosis: Was there a representative sample of patients, was follow-up sufficiently complete?

There are a number of tools to assess bias. These include i) For randomised trials (RoB2, https://methods.cochrane.org/bias/resources/rob-2-revised-cochrane-risk-bias-tool-randomized-trials), and ii) for mixed methods studies (Mixed Methods Appraisal Tool [MMAT][7]). A list of tools can be found in this paper,[8] and the EQUATOR network has lists of reporting guidelines for each type of study (https://www.equator-network.org/).

Did the review address possible explanations of between-study differences in the results?

The explanations for expected differences in outcomes should be stated a priori, that is, in the systematic review protocol. Subgroup analyses where differences are expected, may include age cohorts or specific comorbidities, amongst others.

Did the review present results that are ready for clinical application?

For binary outcomes, the results are presented as proportions (e.g. outcomes of death, myocardial infarction, etc.), and the 'relative' efficacy of an intervention should generally be consistent across the entire cohort. Therefore, the preference is to present the relative effects of the intervention, i.e. relative risk (RR), odds ratio (OR) or hazards ratio (HR). To understand a specific patient's risk, we would then need to estimate the patient's baseline risk, and then calculate the patient's absolute risk difference from the relative risk. Using this information, one could calculate the NNT.[9]

For continuous outcomes (e.g. walking disease, forced expiratory volume, etc.), we usually present the weighted mean difference (WMD) and standardised mean difference (SMD). The SMD is the mean difference divided by the standard deviation. The SMD is used for continuous data where different measurement instruments have been used to assess a similar outcome between studies.

The effect size between SD units is important to understand clinical effect; 0.2 SD is small, 0.5 SD is moderate, and 0.8 SD is large. A difference of 0.5 is generally considered to be of clinical importance. To understand the impact of the intervention, it would be possible to calculate the NNT for the number of patients who achieve a 'clinically important' threshold.

Were the selection and assessment of studies reproducible?

Data extraction should be conducted in duplicate by two independent reviewers. The reason for this is that two reviewers extracting data prevents mistakes (i.e. random errors) and bias (i.e. systematic errors). Good agreement between reviewers (e.g.

chance-corrected agreement, such as the κ statistic) should also be reported to establish the agreement between the independent reviewers. The appendix should document the search strategy, data extraction plan and assessment of data extraction.

Did the review address confidence in effect estimates?

Addressing bias can increase the confidence in the effect estimates. A meta-analysis would decrease the imprecision (by decreasing the width of the CI) and document any inconsistencies through the heterogeneity between study results. Authors need to make an explicit assessment of the confidence in the estimates of effect.

The credibility of the individual studies

Interpretation and understanding the effect estimate of a meta-analysis includes assessing the credibility of the individual studies. An example is the 'risk of bias' tool (RoB2, https://methods.cochrane.org/bias/resources/rob-2-revised-cochrane-risk-bias-tool-randomized-trials).

The forest plot

Different studies have different weightings, based on i) size of the study, and ii) number of events. Studies with a higher weighting have narrower CIs, and the point estimate is represented by a square, which is larger due to the increased weighting.

The pooled estimate of the individual studies is shown as a diamond, with the width showing the CI. There is a vertical line of no effect. If the CI crosses the line of no effect, it is uncertain whether there is a difference between interventions.

Presentation of outcomes

Outcomes for dichotomous outcomes are reported as RR or OR, and continuous data as WMD or SMD, as discussed above.

Assessing the confidence in the estimates and GRADE recommendations

Based on the assessment of the quality of the individual studies, the consistency of study results, and the local applicability, the GRADE (Grading of Recommendations, Assessment, Development and Evaluations) provides a transparent framework for developing and presenting summaries of evidence and provides a systematic approach for making clinical practice recommendations.[10]

The GRADE confidence in estimates of effect has four categories: high, moderate, low and very low. The lower the confidence, the more likely it is that the observed effect estimate is substantially different from the true effect. Confidence in the effect estimate is determined by:

The study design

Randomised controlled trials are assumed to have higher confidence, and observational studies are low. These initial estimates of confidence are further modified by:

Risk of bias

This is systematic rather than random error. It may be due to inappropriate or suboptimal i) randomisation sequence, ii) allocation concealment, iii) blinding of patients, caregivers or study personnel, and iv) lost to follow-up. Risk of bias can be assessed by the Cochrane Collaboration risk of bias assessment tool. Bias can lead to 'no change' assigned confidence ratings, or to a 1 or 2 level downgrading.

Inconsistency

The assumption is that the treatment effect applies to a broad range of patients. However, this may not be the case across different groups of patients, and may require an a priori defined subgroup analysis. Consistency is evaluated by:

i) The visual assessment of variability. Visual assessment would show point estimates (on the same side of the line of no effect) and the CIs of the various studies overlapping if the studies provided consistent findings. Causes for concern or inconsistency between studies would be associated with study point estimates which are far apart, and CIs which do not overlap.

ii) Yes or no statistical test of heterogeneity. The Cochran Q is a chi-squared test which assumes the difference between studies is due to chance. A significant finding therefore suggests significant 'inconsistency' of results between studies. A word of caution for studies with large sample sizes, as these may generate a statistically significant result, although there may be no clinically important heterogeneity.

iii) Magnitude of heterogeneity (variability). The I^2 statistic focuses on the magnitude of variability as opposed to the statistical significance. At about 25%, we would be getting concerned about the consistency of the findings between studies, and at 75%, we would consider the findings inconsistent between studies.

When the between-study variability is large, one needs to consider factors which may have contributed to this. These may include different population effects, e.g. ill versus less ill patients, differences in the intervention between studies, e.g. different doses, and differences between comparators, e.g. control receiving other treatment, versus control receiving placebo. A test of interaction for these subgroups is necessary to determine whether this occurred by chance. A significant finding suggests that the differences in effect estimates cannot be attributed to chance alone, and these may be real differences between the subgroups. Remember, if these are post hoc analyses, then these data are only 'hypothesis-generating'. Inconsistency would lead one to consider whether it is appropriate to include these studies in a meta-analysis. Any residual inconsistency would require

downgrading of the confidence in the estimates in the GRADE recommendations.

Imprecision

Precision is dependent on the width of the CI. If our clinical decision-making remains consistent across the 95% CI, i.e. lower and upper boundary, then this would increase our confidence in the effect estimate. If our clinical decision-making would change from the lower to upper boundary, then we would have less confidence in the effect estimate, and this is due to the imprecision of the findings. To test this, we would need to determine the absolute risk difference, and NNT at the lower and upper boundary. This would determine the clinical utility of the intervention.

Indirectness

Directness means that the research applies to our population of interest, that the interventions are appropriate in our population, and the outcomes are important to our population. Indirectness would include studies where the populations differ from ours, interventions which are tested against a placebo and not our standard of care, and outcomes which are surrogates for the real outcomes of interest.

Publication bias

This is most likely when negative studies are not published (reporting bias), when specific outcomes are reported (selective outcome reporting), and reporting in less prominent journals (dissemination bias). Reporting bias can be assessed by a funnel plot, where we would expect studies to be symmetrically arranged around the summary estimate, with the larger studies closer to the summary estimate, and all quadrants populated with studies. If this is not the case, reporting bias may be a concern. Selective outcome reporting can be identified by assessing the registration protocols of studies and the listed primary outcome. Studies registered late or unregistered should raise concerns about reporting bias.

Effect size

A larger effect size should increase confidence. However, remember if it is implausible, or due to studies of low quality, then this may be the reason for the large effect size.

Conclusion

Once a systematic review and meta-analysis has been conducted, ideally an evidence-based summary of the findings should be produced as an evidence profile. This allows for knowledge translation and communication with patients concerning informed choices about care. A good example is the 'Living WHO guideline on drugs for COVID-19' (https://www.bmj.com/content/370/bmj.m3379).

ORCID

B Biccard https://orcid.org/0000-0001-5872-8369

References

1. Guyatt GH, Rennie D, Meade MO, et al. Users' guides to the medical literature: A manual for evidence-based clinical practice. 3rd ed. New York: McGraw-Hill Education; 2015.
2. London S, Gurdal O, Gall C. Automatic export of PubMed citations to EndNote. Med Ref Serv Q. 2010;29(2):146-53. https://doi.org/10.1080/02763861003723317.
3. Walsh M, Srinathan SK, McAuley DF, et al. The statistical significance of randomized controlled trial results is frequently fragile: a case for a fragility index. J Clin Epidemiol. 2014;67(6):622-8. https://doi.org/10.1016/j.jclinepi.2013.10.019.
4. Ridgeon EE, Young PJ, Bellomo R, et al. The FRAGILITY INDEX in multicenter randomized controlled critical care trials. Crit Care Med. 2016;44(7):1278-84. https://doi.org/10.1097/CCM.0000000000001670.
5. Bertaggia L, Baiardo Redaelli M, Lembo R, et al. The fragility index in peri-operative randomised trials that reported significant mortality effects in adults. Anaesthesia. 2019;74(8):1057-60. https://doi.org/10.1111/anae.14656.
6. Montori VM, Devereaux PJ, Adhikari NK, et al. Randomized trials stopped early for benefit: a systematic review. JAMA. 2005;294(17):2203-9. https://doi.org/10.1001/jama.294.17.2203.
7. Pluye P, Gagnon MP, Griffiths F, Johnson-Lafleur J. A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in mixed studies reviews. Int J Nurs Stud. 2009;46(4):529-46. https://doi.org/10.1016/j.ijnurstu.2009.01.009.
8. Ma L-L, Wang Y-Y, Yang Z-H, et al. Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? Mil Med Res. 2020;7(1):7. https://doi.org/10.1186/s40779-020-00238-8.
9. Barratt A, Wyer PC, Hatala R, et al. Tips for learners of evidence-based medicine: 1. Relative risk reduction, absolute risk reduction and number needed to treat. CMAJ. 2004;171(4):353-8. https://doi.org/10.1503/cmaj.1021197.
10. Siemieniuk R, Guyatt G. What is GRADE? London: BMJ Best Practice; 2021. Available from: https://bestpractice.bmj.com/info/toolkit/learn-ebm/what-is-grade/. Accessed 15 Sept 2021.