**University of Utah Internal Medicine Journal Club Template:**

Updated: Brian Locke 10/31/2019

**SYSTEMATIC REVIEW AND META-ANALYSIS FACILITATOR’S GUIDE**

Systematic Review and Meta-analysis Assigned to Resident**:**

* Article Title
* Study question: PICO (Patient Intervention Comparison Outcome) Format. Why is this question important?
* **Systematic Review Portion:**
* How did they identify relevant studies? (Was the search for studies described? Exhaustive? Reproducible?) Did they state the criteria that studies had to meet for inclusion?
* Was risk of validity and bias in the primary studies assessed? How? Were there specific sources of bias?
* **Meta-analysis portion:**
* How did they pool study data? Did they pool individual patient data or study-level?
* How heterogenous were the studies included? How heterogenous were the results? E.g. did they give a Forest plot? Did they offer an explanation for any studies that varied significantly?
* Results: Are the results usable? How large is the relative and absolute effect in our target population? Is it clinically important?  
  Critique: Were benefits, harms, and costs considered?
* Can I apply the results to my patient? How?

Facilitator points to consider:

**Question:**

-Did the authors review address an important/concretely defined question?

-Why did they do a systematic review and/or meta-analysis? Advantages of a systematic review? Single study = possibly unreliable or unrepresentative. Doesn’t include as many patients in combination. Clinicans reviewing all the individual studies is time consuming. Meta-analysis can provide more statistical power.

-Is the scope appropriate? (meaning, given what we know about biology.. do we expect any all the interventions to result in a similar treatment effect). For example

Study Design Example: (from <https://www.cmaj.ca/content/172/5/661.long>)

Consider the following 3 hypothetical systematic reviews. For which of these systematic reviews does it make sense to combine the primary studies?

* A systematic review of all therapies for all types of cancer, intended to generate a single estimate of the impact of these therapies on mortality.
  + No, overly broad question. significant variability in the pathophysiology of different cancers and in the mechanisms of action of different cancer therapies
* A systematic review that examines the effect of different antibiotics, such as tetracyclines, penicillins and chloramphenicol, on improvement in peak expiratory flow rates and days of illness in patients with acute exacerbation of obstructive lung disease, including chronic bronchitis and emphysema.[**7**](https://www.cmaj.ca/content/172/5/661.long#ref-7)
  + Maybe = 4 key elements of study design: the patients, interventions, outcomes and methods of the primary studies. Combining results is appropriate when the biology is such that across the range of patients, interventions, outcomes and study methods, one can anticipate more or less the same magnitude of treatment effect. The range of characteristics of the primary studies across which it is sensible to combine results is a matter of judgment based on the researcher's understanding of the underlying biology of the disease.
* A systematic review of the effectiveness of tissue plasminogen activator (tPA) compared with no treatment or placebo in reducing mortality among patients with acute myocardial infarction.[**8**](https://www.cmaj.ca/content/172/5/661.long#ref-8)
  + Yes - expect more or less the same magnitude of treatment effects across the range of patients, interventions and outcomes that the investigators have included in their systematic review)

**Search strategy:**

-Was the process credible? Or is there risk of bias introduced by excluding certain types of studies? (Best: use bibliographic databases such as MEDLINE, Cochrane, EMBase. Language barriers are the most common reason relevant literature might not be identified at all)

–Would certain types of important literature be likely to be excluded? – such as if medication side effects are import, were we restricted to randomized trials that may underreport long term, drug interactions, and other things that only become apparent in actual use (e.g. spironolactone hyperkalemia) - these often require cohort / observational datasets that are larger/longer.

-did they have to include studies that may not be optimal?, as the optimal types of studies haven’t been done? e.g. was it mostly observational studies addressing a treatment question – such as in Pneumonia treatment

**Validity and Bias of included studies:**

-Garbage in (individual studies) = garbage out

-Did the authors comment any particular threats to validity or sources of bias? Can we infer others from what we know about the literature (e.g. changing or vague definitions)

-Consider if this applies to the subject:

* if there is a failure to report negative results, this will give a falsely large estimate of the treatment effect. Analogous change can also occur w/ adverse effects.
* Another common bias is reporting bias: version of this is selective outcome reporting bias – if 1 group has a more thorough assessment of outcomes in the underlying studies. This is ameliorated in an individual patient data meta-analysis.

|  |
| --- |
| -From Dartmouth: Was the methodological quality of each study assessed using predetermined criteria appropriate to the type of study? (e.g. randomization, allocation concealment, and follow-up for randomized controlled trials) |

**Pooling:**

-Should these studies be pooled (also stated, is it appropriate to do a meta-analysis on the results of this systematic review)? The meta-analysis portion of a systematic review is only possible if the studies are similar: they should each have roughly the same research question, with similar populations and treatments. Ideally, things like selection process, confounding variables, and attrition rates should align also.

-Did they use Individual or Study level pooling?: often, authors won’t be able to obtain the patient level data. Advantages of individual patient level allows for true ITT and subgroup analysis.

-Did they use a fixed effects (=analysis with the assumption we are measuring 1 underlying effect and all differences are due to random chance, which requires homogenous studies, generally I2 statistic <25%) or random effects (=works for higher heterogeneity, assumes that each study can have a different “true” effect size – which is true if studies differ in design) modeling?. See below for details on heterogeneity and I2. Ignore the stats of how this is actually done

**Heterogeneity:**

-What is heterogeneity? potentially important differences in the results of individual studies being considered for a meta-analysis. Assessment for heterogeneity occurs within 4 key elements of the design of the original studies: the patients, interventions, outcomes and methods.

Goal: is it sensible to pool these studies? (=low heterogeneity in study design)

Is it useful once combined? (=low heterogeneity in results)

Interpreting heterogeneity on the forest plot:

-difference of magnitude or direction of effect between studies is not, in itself, a problem if confidence intervals are overlapping.. however, if they are not overlapping, that is – by definition – statistically unlikely and so begs the question of whether there is some underlying heterogeneity that explains the difference in effect.

-If the studies have non-overlapping confidence intervals, did the authors give a reason why? ideally authors will have generated a priori hypotheses to explain the heterogeneity in magnitude of effect across studies that they are liable to encounter.

-How can you tell how much heterogeneity there is in the included studies? Heterogeneity can be eyeballed on the forst plot (how many confidence intervals are overlapping – which side of the null effect line they are on doesn’t matter per-se) or formally with stats- see box 1 at the end of this section

**Results:**

-How did they present the results, and why? Note that relative risks (or OR – or in continuous variables a statistic called the weighted mean difference or WMD) are more consistent across studies (as they reflect the treatment effect independent of the population risk) but absolute risks are what we’re interested in as providers / patients (because they also reflect the underlying risk of the illness in the population). Meta-analyses should generally present RELATIVE risks, to account for different baseline risks in the different studies.

-Are there other possible ouseful results besides the estimate of magnitude of treatment effect? Quantifying the amount of disagreement among studies in the field, giving insight into which study out of several to ‘trust’

How do you interpret the forest plot? – usually presented with the following format: (from CEBM)

-Individual studies are represented by a black square and a horizontal line (whiskers), which corresponds to the point estimate and 95% confidence interval of the odds ratio. Longer line = less precision in estimate.

-The size of the black square reflects the weight of the study in the meta-analysis. The solid vertical line corresponds to ‘no effect’ of treatment - an odds ratio of 1.0.

- When the confidence interval includes 1 it indicates that the result is not significant at conventional levels (P>0.05).

-The diamond at the bottom represents the combined or pooled odds ratio of all five trials with its 95% confidence interval. (Image below from <https://pubmed.ncbi.nlm.nih.gov/16894442/>)

A close up of a map

Description automatically generated

