SELF-INSTRUCTIONAL MODULE FOR SYSTEMATIC REVIEWS or META-ANALYSIS

Citation: Macpherson H, Pipingas A, and Pase MP. Multivitamin-multimineral supplementation and mortality: a meta-analysis of randomized controlled trials. Am J Clin Nutr 2013; 97:437-44.

I. APPRAISING DIRECTNESS

Does the study provide a direct enough answer to your clinical question in terms of type of patients (P), exposure/intervention (E) used and disease or outcome (O) of interest?

Yes. See PEO table below.

Possible follow-up questions:

a. What is the difference between a review, a systematic review and a meta-analysis?

A review discusses the results of 2 or more studies. A systematic review makes sure these studies were obtained using an objective and efficient strategy. A meta-analysis comes up with a statistical summary of the results of these different studies.

b. What are the components of a focused clinical question?

P=the patient population of interest; E=the exposures being compared; O=the outcomes being measured.
Note: M (methodology) is not part of the question.

c. Compare the clinical PEO with the Research PEO in a directness table.

Comparison Table (Ask group to generate):

	Clinical Question	Research Question
P	Adult male with parental hx of cancer	Adults in primary & secondary prevention care; (Healthy, at risk, or with concomitant illness which is not terminal)
E	Multivitamin-multimineral supplementation	Multivitamin-multimineral supplementation
O	Prolong life	All-cause, cancer, & vascular mortality

d. In what way can there be a mismatch in P?

The target population may not be exactly the same:
1) disease spectrum might be different (eg - mild vs severe disease),
2) the general population may be different (eg – adults vs. children), or
3) the degree of uncertainty in the diagnosis may be different (eg – angiographically confirmed coronary disease vs. a diagnosis based purely on clinical grounds).

Note: Avoid getting too specific with "P" at this point or there will never be a perfect match. The goal of questioning directness is just to decide if the paper might be useful for the reader.

e. In what ways can there be a mismatch in the E's assessed?

The drug(s) studied may not be exactly the same:
1) same class but different molecule (sometimes useful, eg – when the drug effect is a "class effect")
2) same molecule but different route (eg – IV study but drug is only available orally)
3) unfair comparisons (seldom useful clinically, eg - maximum does of one drug vs usual dose of another)

When can we be convinced there is class effect?

When previous studies show a consistent effect of different members of a class of drugs.

f. In what ways can there be a discrepancy in O?

The outcomes that researchers are interested in may be different:
1) Researchers often measure surrogate outcomes; clinicians and patients want clinical outcomes.
2) Researchers often monitor composite endpoints; clinicians and patients want to know the effect of treatment on individual components.

Why do researchers use surrogate endpoints?

Shorter study, smaller sample size, treatment is new and they want to see if it might work.

When can we believe surrogate outcomes?

When previous studies show a strong and consistent relationship between the surrogate and the clinical outcome. (Almost never.)

Why do researchers use composite endpoints?

To decrease sample size requirements.

When can we believe composite endpoints?

When components are of approximately equal importance; when results for each component are in the same direction.

II. APPRAISING VALIDITY

1. Were the criteria for inclusion of studies appropriate?

Yes. Please see p. 438 under Study Selection. "Randomized and controlled primary or secondary prevention trials were considered…the following inclusion criteria were enforced: the trials must have been randomized and controlled…the participants must have been supplemented with a daily multivitamin-multimineral (MVMM) formulation in at least one study arm…each trial must have reported on the number of deaths in both…"

Possible follow-up questions:

a. Why is it important to include details regarding the P, E, O and methodologic criteria for inclusion of studies?
To make sure that study content and quality (rather than conclusions) are the basis for study selection (which is often the case in non-systematic reviews)

2. Was the search for eligible studies thorough?

Please see p. 437 under Data Sources and Searches. "We searched Medline…The Cochrane Database of Systematic Reviews…Cochrane Central Register of Controlled Trials…CINAHL…Database of Abstracts of Reviews of Effects…Searching was performed by using the following string of search terms with truncation…search was not limited to any specific years of publication…Scopus was used to search the references of all studies included in the review." The search for published studies was thorough; however, there were no explicit statements on searching for unpublished studies.

Possible follow-up questions:

a. What are the usual methods for identifying relevant articles?
The usual methods for identifying relevant articles include: 1) electronic databases, 2) textbooks, 3) cross-references, 4) hand-searches through relevant journals, 5) communication with experts/researchers, 6) drug company files, 7) files in regulatory agencies.

b. Why should effort be exerted to search for unpublished articles?
Authors should show that every effort was exerted to search for unpublished articles to make sure that the review has included every possible article for inclusion.

What articles tend to be published more – the ones with positive or negative results? Why?
Studies with positive results tend to get published perhaps because of the following: 1) journal editors want positive results which are more dramatic, 2) drug companies want to hide negative results which may be bad for business, and 3) researchers themselves are disappointed when the null hypothesis is correct.

If positive articles tend to be published more than negative trials, how will this affect meta-analyses of treatment interventions?
Treatment will look better (label this "publication bias").

How do we search for unpublished trials?
Unpublished trials can be searched through: 1) correspondence with researchers and manufacturers, 2) trial registries, and 3) conference proceedings.

3. Was the validity of the included studies assessed?

Yes. Please see p. 438, paragraph 2, under Data Extraction and Quality Assessment. "In accordance with Cochrane guidelines, each trial was accessed across the following criteria: allocation sequence generation, allocation concealment, blinding, complete outcome data reporting, selective outcome data reporting, and other apparent biases."

Possible follow-up questions:

a. Why should authors appraise the validity of individual studies?
The strength of conclusions from a systematic review depends on the validity of included studies (GIGO principle: garbage in, garbage out). Differences in quality may explain differences in results.

b. What are the criteria for validity?
There are validity criteria for various claims (effectiveness, accuracy of a test, causation and prognosis). One can utilize the appropriate guide.

4. Were the assessments of the studies reproducible?

Yes. Please see p. 438, paragraph 3, under Data Extraction and Quality Assessments. "Two researchers…independently completed database searching and assessed the suitability of each trial for

inclusion…These researchers were blinded to the study aim of assessing mortality…Discrepancies were resolved in consultation with the principal study investigators…who also rechecked each relevant study against the inclusion criteria…independently assessing each trial's risk of bias and completing data extraction…Discrepancies resulting in this process were resolved by mutual consensus."

Possible follow-up questions:

a. Why do we want several assessors of quality?
   Several assessors of quality minimize the chance of bias. Statistically speaking, if agreement is great, the possibility of bias is less. There will always be subjectivity in assessing study quality therefore 2 reviewers should always evaluate the quality of included studies. High agreement between 2 reviewers reinforces credibility. It should be emphasized though that the 2 reviewers should make assessments independently.

b. How do we settle disagreement among authors?
   Disagreements are usually settled by 1) third party decision or 2) discussion and consensus.

III. APPRAISING RESULTS

1. What are the overall results of the review?

Please refer to Figure 2.
MVMM supplementation on all-cause mortality - RR 0.98; 95%CI: 0.94, 1.02 (all pooled studies), RR 0.94; 95% CI: 0.89, 1.00 (13 primary prevention trials), RR 1.04; 95% CI: 0.98, 1.11 (4 secondary prevention trials), RR: 0.94; 95% CI: 0.84, 1.05 (4 secondary prevention of cancer trials)

Please refer to Figure 3.
MVMM supplementation and mortality due to cancer – RR: 0.96; 95% CI: 0.88, 1.04 (9 studies)
MVMM supplementation and mortality of vascular causes – RR: 1.01; 95% CI: 0.93–1.09 (10 studies)

Possible follow-up questions:

a. What information can you derive from the forest plot?
   1) how many studies there were
   2) how many were positive
   3) how many were negative
   4) which was the largest
   5) which was the smallest
   6) which had the most events
   7) which had the least events
   8) what was the overall result
   9) what were the CI's around the overall results
   10) were the results similar from study to study (and this brings us to the next question)

2. Were the results similar from study to study?

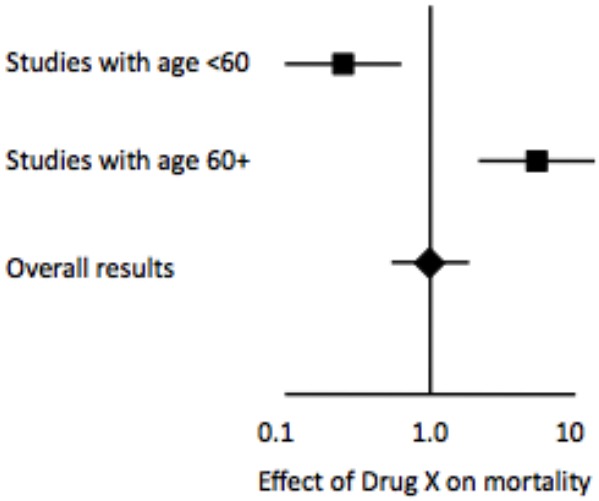
Yes. Please see pages 439 & 441 text and figures of each outcome. "There was little evidence of heterogeneity."

Possible follow-up questions:

a. What does heterogeneity mean?
   That the results are significantly different between studies.

b. Why do we worry about heterogeneity?
   Because we may be combining results that should not be combined.

Suggested approach: draw this hypothetical meta-analysis and ask – would it be appropriate to conclude that drug X has no effect on mortality?



Proper conclusion is that drug X saves lives in young patients, and causes deaths amongst the elderly. To conclude that it has no effect on mortality would be totally wrong.

c. How do we detect heterogeneity?
Examine the forest plot then 1) Examine the inclusion criteria and see if the research question was reasonably focused, that is, it avoided combining studies that had diffrent questions, 2) Look at the forest plots - when CIs do not overlap, there is heterogeneity, 3) check what statistical tests were done to determine heterogeneity (chi square $p<0.10$ implies heterogeneity, I^2 statistic > 50% implies substantial heterogeneity).

c. If all studies satisfied the inclusion criteria (P, E, O, M) why should they have heterogeneous results?
There may be subtle differences in P, E, O or M.

d. What can authors do if there is heterogeneity?
1) Look for the culprit article and see if excluding it will eliminate heterogeneity (risky), 2) Identify the source of heterogeneity (P,E,O, or M) and try to subgroup studies into homogeneous categories, or if all of these fail, 3) abandon attempts at statistical combination (just do a systematic review).

3. How precise were the results?
The results were precise. Please see in Figures 2 & 3, confidence intervals of overall results (random & fixed).

Possible follow-up questions:

a. What is the difference between accuracy and precision?
Accuracy refers to whether the results are close to the truth, while precision refers to repeatability. To determine the accuracy of a weighing scale, you check its weight reading of a mass of known weight. To determine its precision, you measure an object several times and see if the measurements are close together

b. Why are 95% CI's considered measures of repeatability?
While caregivers see 95% CI's as best and worst case scenarios of the truth, statisticians see them differently – as the range of possible results in 95 studies, if you repeated the study 100x.

IV. ASSESSING APPLICABILITY: Are the results applicable to the patients you see?

1. Are there biologic issues that may affect applicability of treatment?
No biologic issues noted.

Possible follow-up questions:

a. What biologic issues should be considered in assessing applicability? Are any of these issues of concern in this study?
Biologic issues: sex, comorbidity, race, age, and pathology.

2. Are there socio-economic issues affecting applicability of treatment?
People might not understand that they are purchasing or taking expensive multivitamin-multimineral supplements that they really do not need and therefore resources are wasted.

Possible follow-up questions:

a. What socio-economic issues should be considered in assessing applicability? Are any of these issues of concern in this study?
Socio-economic issues: provider compliance and patient compliance

b. What is the difference between directness and applicability?

Directness	**Applicability**
Should I read this article?	Can I use the results in my patients?
Did they study the disease of interest?	Are there factors that can modify the treatment effect (RR)?
Broad criteria (P, E, O)	Specific criteria (SCRAP and socio-economic issues)
Like deciding what clothes you will buy.	Like trying on the clothes before buying them.
Others listed by participants	Others listed by participants

Note: these issues affect applicability if they could potentially lower or increase the RR or RRR, ie, if they can enhance or reduce the effectiveness of treatment. Changes in baseline risk for the outcome of interest should not affect applicability (but may affect the individualized ARR).

3. If the overall results of the review are not directly applicable to your patient, are there credible subgroup analyses that you could use?
Yes. Please see p. 438 under Data Synthesis and Analysis. "With respect to all-cause mortality, subanalyses were performed to examine the association between MVMM treatment and mortality across 1) trials with and without a high risk of bias, 2) primary and secondary prevention trials separately, and 3) trials conducted in high- and low-income countries separately…"

Possible follow-up question:
a. What are the dangers of a subgroup analysis?
If there are too many subgroup analyses, one of them can become positive by chance alone. (This is called data dredging. If you torture the data, it will confess to anything).

Suggested example: ISIS-2 an unpublished subgroup analysis (Do we believe the subgroups?):

	ASA N=8587	Placebo N=8800	RRR	p
Overall Mortality	9.4%	11.8%	23%	P<0.05
Gemini or Libra	11.1%	10.3%	-9%	NS
Other Astrology	9.26%	12.0%	28%	P<0.05

b. Would you believe the above subgroup analysis? Why?
No. Because 1) these don't make sense and 2) they could have happened by chance.

c. When can you believe a subgroup analysis?
1) it should be planned a priori, 2) there shouldn't be too many, 3) there should be statistical adjustments for p 4) the conclusions should make biologic sense.

V. INDIVIDUALIZING THE RESULTS

1. Estimate the individualized NNT for your patient(s).

Step 1. Estimate baseline risk of your patient (Rc)	5%
Step 2. Estimate RR with treatment	0.96 (mortality from cancer – trend to benefit)
Step 3. Estimate post-treatment risk (Rt = Rc x RR)	Rt = 0.05 x 0.96 = 0.048
Step 4. Estimate ARR (Rc-Rt)	0.05 – 0.048 = 0.002
Step 5. Estimate Individualized NNT	1/0.002 = 500

Possible follow-up questions:
a. How do we estimate baseline risk?
A risk calculator is available for many diseases, or a risk scoring system, cohort studies on untreated patients, and the control group of the study (last choice because it is the least generalizeable)

2. Would you offer the treatment to your patients?
Answer will differ between caregivers, nurses, even patients.

Encourage debate. There are no right or wrong answers here. In the end, we may agree on the numbers but disagree in action plans because we value the numbers differently. Use this to illustrate that EBM is not just about evidence.