Overview and Comparisons of Risk of Bias and Strength of Evidence Assessment Tools: Opportunities and Challenges of Application in Developing DRIs

January 9, 2017

Mei Chung, PhD, MPH
Assistant Professor, Department of Public Health and Community Medicine, School of Medicine
Tufts University
Outline

• Terminology
• Overview and comparisons of risk of bias (ROB) and strength of evidence (SOE) assessment tools
• Applications in Developing DRIs and DGA
• Remaining challenges and possible solutions
Terminology

• **Risk of bias (ROB)** assessment = Internal validity of a study
  – Whether results of a primary study should be believed

• **Strength of (a body of) evidence (SOE)** = Quality of a body of evidence
  – The certainty level in the conclusions (answers to the key question) drawn from the synthesis of the body of evidence

• ROB of a systematic review (with or without meta-analysis)
  – Not ROB of a primary study nor a SOE rating
  – Only needed when a systematic review (SR) integrate existing SRs
Basic Steps of a Systematic Review (SR)

- SR is not a clinical/nutrition guideline or recommendation
- SR may include more than 1 key questions (e.g., more than 1 outcomes)
  - Each key question is usually defined according to PI(E)CODS
- SR may integrate existing SRs but currently no guidance on how to synthesize, grade the SOE, and present bodies of evidence composed of primary studies and existing systematic reviews*

*Robinson et al. Integration of existing systematic reviews into new reviews: identification of guidance needs. Systematic Reviews 2014; 3:60
SOE ≠ Grading of recommendations

Source:
Terminology

- **Grading of recommendations** = Strength of a recommendation
  - To what extent one can be confident that adherence to the recommendation will do more good than harm?
  - To make these judgments, one needs to know:
    - The quality of evidence across studies for each important outcome = SOE
    - Which outcomes are critical to a decision
    - The overall quality of evidence across these critical outcomes = overall rating across multiple SOE ratings
    - The balance between benefits and harms

Goals of ROB Assessment

• Risk of bias (ROB) assessment
  – To avoid “Garbage in, garbage out”
  – To assess the confidence in the validity of study findings

• ROB in the results of each study is one of several factors that must be considered when judging strength of evidence (SOE)
Tools for Risk of Bias Assessment

• Many tools, but few “validated” (rigorously developed or tested for validity and reliability) tools*
  – Cochrane risk of bias assessment: Intervention studies http://bmg.cochrane.org/assessing-risk-bias-included-studies
  – The Newcastle-Ottawa Scale: Cohort or case-control studies http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

• No well-accepted nutrition specific ROB assessment tools**

---


**Salam et al. Systematic reviews on selected nutrition interventions: descriptive assessment of conduct and methodological challenges. BMC Nutrition 2015; 1:9
Risk of Bias Tools in SRs Used to Inform DRIs or DGA

• Cochrane ROB tool for RCTs and Newcastle-Ottawa Scale for cohort studies, both supplemented with “nutrition-specific items” in consultation with the technical expert panel
  – Some nutrition-specific items are topic dependent, e.g., 25(OH)D assay methods
  – Some are more generic, e.g., compliance issue for RCTs and measurement errors/biases in dietary assessment methods
    • Accurate estimates of “doses” and dose-response relationships are very important for developing DRIs

• Nutrition Evidence Library Bias Assessment Tool (BAT)*
  http://www.nel.gov/topic.cfm?cat=3384
  – Based on Cochrane ROB domains: Selection Bias; Performance Bias; Detection Bias; Attrition Bias
  – Tailored by study design, with different sets of questions applying to RCTs (14 questions), non-randomized controlled trials (14 questions), and observational studies (12 questions)
  – No “nutrition-specific items”
Tools for Strength of Evidence Assessment

- No SOE ratings in the SR that was used to inform Vit D and Ca DRIs

### USDA Nutrition Evidence Library Conclusion Statement Evaluation Criteria

<table>
<thead>
<tr>
<th>Elements</th>
<th>Grade I: Strong</th>
<th>Grade II: Moderate</th>
<th>Grade III: Limited</th>
<th>Grade IV: Grade Not Assignable*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Risk of bias</strong> (as determined using the NEL Bias Assessment Tool)</td>
<td>Studies of strong design free from design flaws, bias, and execution problems</td>
<td>Studies of strong design with minor methodological concerns OR only studies of weaker study design for question</td>
<td>Studies of weak design for answering the question OR inconclusive findings due to design flaws, bias, or execution problems</td>
<td>Serious design flaws, bias, or execution problems across the body of evidence</td>
</tr>
<tr>
<td><strong>Quantity</strong></td>
<td>Several good quality studies; Large number of subjects studied; Studies have sufficiently large sample size for adequate statistical power</td>
<td>Several studies by independent investigators; Doubts about adequacy of sample size to avoid Type I and Type II error</td>
<td>Limited number of studies; Low number of subjects studied and/or inadequate sample size within studies</td>
<td>Available studies do not directly answer the question OR no studies available</td>
</tr>
<tr>
<td><strong>Consistency of findings across studies</strong></td>
<td>Findings generally consistent in direction and size of effect or degree of association, and statistical significance with very minor exceptions</td>
<td>Some inconsistency in results across studies in direction and size of effect, degree of association, or statistical significance</td>
<td>Unexplained inconsistency among results from different studies</td>
<td>Independent variables and/or outcomes are too disparate to synthesize OR single small study unconfirmed by other studies</td>
</tr>
<tr>
<td><strong>Impact</strong></td>
<td>Studied outcome relates directly to the question; Size of effect is clinically meaningful</td>
<td>Some study outcomes relate to the question indirectly; Some doubt about the clinical significance of the effect</td>
<td>Most studied outcomes relate to the question indirectly; Size of effect is small or lacks clinical significance</td>
<td>Studied outcomes relate to the question indirectly; Size of effect cannot be determined</td>
</tr>
<tr>
<td><strong>Generalizability to the US population of interest</strong></td>
<td>Studied population, intervention and outcomes are free from serious doubts about generalizability</td>
<td>Minor doubts about generalizability</td>
<td>Serious doubts about generalizability due to narrow or different study population, intervention or outcomes studied</td>
<td>Highly unlikely that the studied population, intervention AND/OR outcomes are generalizable to the population of interest</td>
</tr>
</tbody>
</table>

*Standard conclusion statement is used to communicate that there is either insufficient evidence or no evidence available to answer the question.*
Tools for Strength of Evidence Assessment

• *Grading the Strength of a Body of Evidence when Comparing Medical Interventions* in AHRQ Methods Guide 2009* [has been updated]
  
  – SOE ≠ evidence hierarchies (based on only study designs)
  – Similar to GRADE but *does not make clinical or practice recommendations*
  – Required domains: risk of bias, consistency, directness, and precision
  – Additional domains: dose-response association, existence of confounders that would diminish an observed effect, strength of association (i.e., magnitude of effect), and publication bias
  – Applicability of the evidence

• GRADE approach to evaluating the quality of evidence (=SOE)**
  
  – Also include guidance for *going from evidence to recommendation* (=grading or strength of recommendations)

*This chapter has also been published in edited form: Owens et al. J Clin Epidemiol 2010; 63, 513-523.

Summary of Tools

• ROB of primary studies $\rightarrow$ SOE of outcomes (one SOE for each outcome) $\rightarrow$ Grading/strength of recommendations
  – Subjective judgments are needed regardless of what tool is used
  – Subjectivity in each step is incorporated into the next step
  – Key to minimize subjectivity is transparency

• ROB of a systematic review (with or without meta-analysis)
  – Two tools: AMSTAR* (assessing mostly “reporting quality”) & ROBIS** (new tool and have questions addressing ROB in a systematic review)
  – Not ROB of a primary study nor a SOE rating
  – Not all SRs performed SOE assessment
  – It is important to perform ROB assessment of SRs when existing SRs are integrated in to a new SR but many challenges remained (see Extra Material I: Current Guidance and Areas that Need Future Guidance)

What is the role of basic molecular and mechanistic studies in assessing the evidence on dietary intake and disease prevention?

- Basic molecular and mechanistic studies are usually not included in systematic reviews
  - "Biological plausibility" of the SR key question should have been predetermined
  - Biological plausibility is one of the essential components for causal inference (Hill Criteria)

- SR can include and analyze any study design but
  - Evidence from different study designs may have different “weights” in grading SOE

**FIGURE 4-1** Ranking study designs: Ranking is shown in descending order of quality from top to bottom; the length of bars is arbitrary and indicates the relative strength of a study design.

Figure source: IOM. Dietary Reference Intakes for Calcium and Vitamin D. 2011
“Hierarchy of Evidence”

Figure 1 The proposed new evidence-based medicine pyramid.
(A) The traditional pyramid
(B) Revising the pyramid: (1) lines separating the study designs become wavy (GRADE), (2) systematic reviews are ‘chopped off’ the pyramid.
(C) The revised pyramid: systematic reviews are a lens through which evidence is viewed (applied).

Figure source: Murad et al. New evidence pyramid. Evidence-Based Medicine. Available from http://ebm.bmj.com/content/early/2016/06/23/ebmed-2016-110401.full
Do the tools need to be modified for use in nutrition? **YES!!!**

- **See** “Extra Materials II: Challenges of conducting nutrition vs. medical intervention Systematic Reviews” for the reasons

- **What specific modifications are needed for DRIs?**
  - ROB tools need to include “nutrition-specific items,” e.g., compliance issue for RCTs and measurement errors/biases in dietary assessment methods
  - SOE tools need to include specific instructions on **how to integrate different study designs to assess dose-response relationships**
    - Accurate estimates of “doses” and dose-response relationships are very important for developing DRIs

- **What new tools need to be developed for DRIs?**
  - Framework to select and prioritize “critical” outcomes for decision-making
  - New ROB tools for basic molecular and mechanistic studies?
  - Guidance on grading or strength of recommendations for DRIs
  - How to integrate existing SRs? Need guidance on how to synthesize, grade the SOE, and present bodies of evidence comprising primary studies and existing systematic reviews
Comparing cohort studies and RCTs in nutrition

**Prospective cohort studies**
- Mostly in general populations
- Foods (primarily) and/or supplements
- Wide ranges of nutrient intakes (but often much lower than supplement doses)
- Many concerns re: nutrition observational studies (see next slide)

**Randomized controlled trials**
- Mostly in populations with high risk for a disease
- Supplement vs. placebo but unlike pharmaceuticals, the control group is exposed to some level of dietary intake*
- Narrow ranges of supplement doses (but often much higher than dietary intake levels)

*Therefore, the real question being examined in nutrition RCTs is what are the health benefits or risks of higher versus lower levels of intake. In addition, background exposure level may affect the response to supplementation.
Concerns about nutrition observational studies

• Measurement: Errors and biases in dietary assessments
• Confounding
• Terminology: no generally agreed upon definitions for some foods, drinks, and dietary constituents (e.g. dietary patterns)

FFQ is widely used.
• Generally consider inappropriate to use FFQ data to estimate quantitative parameters
  o Ranking of dietary exposures
  o Dietary patterns

Illustration of the problems: Vitamin E and CVD

“true” exposure levels are often not known in RCTs
Ranges of Exposures and Comparisons are Important but often ignored in SR or MA

Table source: Am J Epidemiol 2004; 160:1005-1010

Possible Solutions

• DRI committees need to reach consensus on a standardized and transparent approach (creating a “DRI method guide” for both evidence synthesis and for developing DRIs):
  – Establish a framework to select and prioritize “critical” outcomes for decision-making
    • Different set of “critical” outcomes for EAR? UL? AI?
  – Determine the role of basic molecular and mechanistic studies in DRI development. If needed, develop new ROB tools for basic molecular and mechanistic studies
  – Modify GRADE SOE approach for DRIs
    • How to integrate existing SRs? Need guidance on how to synthesize, grade the SOE, and present bodies of evidence composed of primary studies and existing systematic reviews
  – Determine whether a new approach to grade strength of recommendations is needed for DRIs
    • DRIs are a set of recommended intake values so criteria for judging “balance of benefit and harms” of DRIs based on chronic disease endpoints need to be established.
Thank you for your attention.

Look forward to our discussions
EXTRA MATERIALS
Extra Materials I:
Current Guidance and Areas that Need Future Guidance for Integration of Existing SRs into a new SR
Ways to integrate existing SRs into a new SR

- Assuming that at least one relevant existing review has been identified that is considered of “acceptable quality,” there are several ways to integrate existing SR(s) into a new SR:
  - Use review without modifying or adding new studies
  - Use review and add new studies
  - Use review with new or modified analysis
  - Use selected elements of review
Figure 1 Methodological steps in using existing systematic reviews (SRs).

1. **Step 1. Locate existing SR(s)**
   - Existing SR(s)?
     - Yes
     - **Step 2. Assess relevance**
       - Questions
       - Methods
       - Search dates
     - Relevant SR(s)?
       - Yes
       - **Step 3. Assess quality of existing SR(s)**
         - Sufficient Quality?
           - Yes
           - **Step 4. Determine appropriate use and incorporate existing SR(s)**
             - AND
             - **Step 5. Report methods and results from using existing SR(s)**
               - Stop. Proceed with SR of primary evidence
               - Use “almost” relevant SRs to frame and provide context (Contextual Use)
               - Scan References of “almost” relevant SRs to check new search results
               - Scan references, check new search results
               - Use existing search
               - Use existing data abstraction, study-level risk of bias assessments and/or synthesis
               - Use complete review
### Table 1: Strength of evidence for **KQ1**: maternal outcomes (gestational length, preterm birth, SGA/IUGR, low birth weight, birth weight, antenatal and/or postnatal depression)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Intervention/exposure</th>
<th>Population</th>
<th>SoE Grade</th>
<th>Design No. Studies</th>
<th>Study Limitations</th>
<th>Consistency</th>
<th>Precision</th>
<th>Other Issues</th>
<th>Finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gestational length</td>
<td>n-3 FA supplementation</td>
<td>Healthy pregnant women</td>
<td>Moderate</td>
<td>RCT: 14 (original report); 14 Obs intake: 3 Obs biomarkers: 1</td>
<td>Low</td>
<td>RCT: Consistent Obs intake: Consistent Obs biomarkers: NA All: Consistent</td>
<td>Precise</td>
<td>Large heterogeneity in the meta-analysis</td>
<td>Original report: Mixed findings Meta-analysis of 12 RCTs in update: WMD 0.33 (95% CI 0.04, 0.6) weeks</td>
</tr>
<tr>
<td>Gestational length</td>
<td>DHA or DHA-rich fish oil</td>
<td>Healthy pregnant women</td>
<td>Moderate</td>
<td>RCT: 3 (original report); 11 Obs intake: 1 Obs biomarkers: 0</td>
<td>Low</td>
<td>RCT: Inconsistent Obs intake: NA Obs biomarkers: NA All: Inconsistent</td>
<td>Precise</td>
<td>Large heterogeneity in the meta-analysis</td>
<td>Original report: Mixed findings Meta-analysis of 11 RCTs in update: WMD 0.34 (95% CI 0.02, 0.67) weeks</td>
</tr>
<tr>
<td>Gestational length</td>
<td>Fish oil or EPA+DHA</td>
<td>Healthy pregnant women</td>
<td>Low</td>
<td>RCT: 11 (original report); 2 Obs intake: 2 Obs biomarkers: 1</td>
<td>Low</td>
<td>RCT: Consistent Obs intake: Consistent Obs biomarkers: NA All: Consistent</td>
<td>Imprecise</td>
<td>A few studies excluded preterm infants</td>
<td>Original report: No effects No effects</td>
</tr>
</tbody>
</table>

Challenges

• Areas in need of additional guidance include:
  – Providing templates or advisory considerations for construction of evidence tables for reviews combining primary and secondary (systematic review-level) evidence
  – Reporting guidelines for clearly communicating the methods for locating, selecting, and deciding how best to utilize existing systematic reviews
  – Methods that limit the potential for bias in selecting reviews to incorporate from among multiple existing reviews
  – Guidance on methods that limit the potential for bias in incorporating selected portions of a review
  – Qualitative and quantitative methods for summarizing bodies of evidence that include a systematic review as the only or as one source of evidence
  – More robust means for quality rating of existing systematic reviews (beyond AMSTAR)
  – Specific methods to grade strength of evidence for bodies of evidence that include a systematic review as the only or as one source of evidence
References


• Robinson et al. Integration of existing systematic reviews into new reviews: identification of guidance needs. Systematic Reviews 2014; 3:60
Extra Materials II:
Challenges of conducting nutrition vs. medical intervention
Systematic Reviews
Baseline/Background Exposure

• All people have some level of exposure to the nutrient or dietary substance of interest (sources: food, supplements or endogenous synthesis)
  o Randomization may not distribute the baseline/background exposure equally (often it was not measured in RCTs and assumed no change during the study)
  o No true “placebo” or unexposed group
Nutrient status

• The nutrient status (e.g., deficiency, adequacy, or toxicity) of an individual or population can affect the response to supplementation
Factors that influence bioavailability

• Some nutrients interact with each other (nutrient-nutrient interaction)
  – E.g., Competing for the same enzyme in the metabolic pathway (n-3 and n-6 FAs)
• Drug-nutrient interaction
  – E.g., statins + n-3 FA supplement
• Life stage – nutrient interaction
  – E.g., physiological state such as pregnancy may affect the utilization of the nutrient
  – E.g., endogenous synthesis of 25(OH)D decreased with aging
Multiple and interrelated biological functions of a nutrient

• Primarily affecting the scope of a SR
  – Expect large number of outcomes (and number of included studies)
  – Expect complicated analytic framework
  – Expect heterogeneous data

• No well-established approach to prioritize

• How to identify and include new nutrient-outcome hypotheses in an update SR is challenging
Nature of nutrient or dietary substance intervention

• Food-based interventions require detailed documentation of the approaches taken to assess nutrient or dietary substance intake.

• Blinding and allocation concealment may not be possible
Uncertainties in assessing dose-response relationships

• Errors and biases in self-reported dietary assessment
• Assay methods for biomarkers of intake
• Information on baseline/background exposure may not be available in RCTs
  – Randomization does not necessary “distribute” the exposure equally
• Trials or real life (observational studies): often provide non-comparable data in terms of exposure ranges
References


Advancing the Role of Systematic Reviews in Nutrition Research and Applications

• Volume 1: Application of Systematic Review Methodology to the Field of Nutrition.


• Volume 3: Reporting of Systematic Reviews of Micronutrients and Health: A Critical Appraisal.

• Volume 4: Effects of Eicosapentanoic Acid and Docosahexanoic Acid on Mortality Across Diverse Settings: Systematic Review and Meta-Analysis of Randomized Trials and Prospective Cohort.

• Volume 5: Comparison of Translational Patterns in Two Nutrient Disease Associations

• Volume 6: Concordance Between the Findings of Epidemiological Studies and Randomized Trials in Nutrition

http://www.ahrq.gov/research/findings/evidence-based-reports/tr17-series.html