Lessons learned in protocol development and implementation of multi-endpoint systematic reviews

Daniele Wikoff, PhD
Health Sciences Practice Leader
ToxStrategies, Inc
Toxicologist, ~10 years in consulting
Consulting firm that provides toxicology and risk assessment services to private and public organizations
  – ~30 scientists (toxicologists, engineers, statisticians)
Responsible for integrating systematic review into our health sciences practice – generally a “user” of what has already been established in EBT
NRC 2014. FIGURE S-1 Systematic review in the context of the IRIS process. The committee views public input and peer review as integral parts of the IRIS process, although they are not specifically noted in the figure.
Lessons learned from a (slightly) different perspective

Protocol development
• Developing SR team (and roles)
• PECO – multiple outcomes (and multiple endpoints)
• Publication

Implementation
• Literature searching and screening
• Individual study assessment

Practical lessons
• General conduct of SRs
• Resource allocation and time
The term “systematic review” gets used very loosely in the field of toxicology.

There are many “correct” ways to do systematic review, as long as the key elements are maintained (particularly emphasis on problem formulation, protocol development, transparency, and documentation).

Systematic review is an excellent tool but it does not eliminate scientific judgments (and subjectivity), rather it provides a platform to integrate and evaluate scientific judgment and data, and present findings in a systematic way.
Background on projects used as case examples
Lessons come from a variety of projects

<table>
<thead>
<tr>
<th>Project</th>
<th>Streams</th>
<th>Hazard</th>
<th>Risk</th>
<th>Client Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment of adverse effects of caffeine</td>
<td>Human (including TK)</td>
<td>X</td>
<td></td>
<td>Non-profit scientific foundation</td>
</tr>
<tr>
<td>Derivation of inhalation toxicity value for environmental compound</td>
<td>Human, animal, mechanistic</td>
<td>X</td>
<td>X</td>
<td>Private</td>
</tr>
<tr>
<td>Assessment of compound in common consumer product</td>
<td>Human, animal, mechanistic (including TK)</td>
<td>X</td>
<td></td>
<td>Government</td>
</tr>
</tbody>
</table>
Analytic framework for the adverse effects of caffeine

- PECO – based on updating Nawrot et al. 2003:
  - For [population], is caffeine intake above [dose], compared to intakes [dose] or less, associated with adverse effects on [endpoint]?
Project overview

- 5 SRs conducted using IOM framework
  - OHAT Handbook and RoB Tool
- Team: SAB + ToxStrategies
- Sponsor (ILSI-North America Caffeine Working Group)
- FDA awareness
- ~2 years in duration (scoping to publication)

**PHASE 1. Initiate Systematic Review**
1. Define project team & SAB
2. Determine project objectives
3. Develop a systematic review protocol
4. Select appropriate tools for review implementation

**PHASE 2. Literature Search and Screening**
1. Conduct comprehensive systematic literature search
2. Screen & select studies; document data collection
3. Report findings of literature search
4. Refine outline (i.e., topic areas)
5. Develop specific systematic review process
6. Obtain literature

**PHASE 3. Individual Study and Body of Evidence Assessment**
1. Systematically assess individual studies and the body of evidence (by topic area)
2. Conduct a qualitative synthesis (by topic area)

**PHASE 4. Report Systematic Review**
1. Prepare draft manuscript
2. Prepare updated draft for submission (based on SAB comments)
3. Peer-review journal submission and publication
Lessons learned in problem formulation and protocol development
Develop SR team and roles

Establishing the Review Team (per IOM):

“The review team is composed of individuals who will manage and conduct the review. The objective of organizing the review team is to pull together a group of researchers as well as key users and stakeholders who have the necessary skills....”

• Multidisciplinary (SR methodology, information specialists, technical experts) and balanced

Developing roles

• Type (e.g., team leader, evidence analysts, technical experts, SAB, etc.)

• Application/tasks (e.g., protocol development, screening, data extraction, individual study assessment)
Sponsor Input and Role

Provided at each phase of the project, per IOM:

• “While an SR should respond to the sponsor’s questions, the sponsor should not overly influence the SR process. The relationship between the sponsor and the SR review team needs to be carefully managed to balance the competing goals of maintaining the scientific independence of the SR team and the need for oversight to ensure the quality and timeliness of their work.”

• “Sponsors should not be allowed to delay or prevent publication of an SR in a peer-reviewed journal and should not interfere with the journal’s peer review process”

ILSI North America Caffeine Working Group is the Sponsor

This work was supported by the North America Branch of the International Life Sciences Institute (ILSI N.A) and through unrestricted grants to ILSI N.A. by the American Beverage Association (ABA) and the National Coffee Association (NCA) ILSI NA is a public, non-profit foundation that provides a forum to advance understanding of scientific issues related to the nutritional quality and safety of the food supply by sponsoring research programs, educational seminars, and workshops, and publications. ILSI NA receives support primarily from its industry membership. ABA is the national trade association that represents the U.S. non-alcoholic beverage industry. NCA is the national trade association that represents the U.S. coffee industry. The opinions expressed herein are those of the authors and do not necessarily represent the views of the funding organization.

ILSI NA mission: scientific integrity
# Case Study: Caffeine - Roles

<table>
<thead>
<tr>
<th>Entity</th>
<th>Description</th>
<th>Roles</th>
</tr>
</thead>
<tbody>
<tr>
<td>ToxStrategies</td>
<td>Scientists with a range of expertise (caffeine, toxicology, epidemiology, systematic review, literature searching, etc.)</td>
<td>A, D, F, I, P Develop and perform the SR (consistency in application of SR process, independent assessment, documentation)</td>
</tr>
<tr>
<td>SAB</td>
<td>Multidisciplinary experts (systematic review, behavior, cardiovascular, bone &amp; calcium, acute, pharmacokinetics – PhD’s and MDs from academic, private, and clinical practices)</td>
<td>I, A Provide input, review, and approval</td>
</tr>
</tbody>
</table>
| Sponsor      | Members of the ILSI-North America Caffeine working group                    | I, A*  
*Budgetary                                                                                     |

A = approve; D = develop; F = facilitate; I = input; P = perform; X = not involved

---

**Lessons Learned**

- *Diverse team strengthens the SR (but also requires flexibility and influences timeline)*
  - *Refined and clear roles = smooth(er) implementation*
  - *Establishes independence of the scientific assessment*
Word choice can be difficult – every word counts
  • PECO “guides” all decisions – when in doubt, revisit the PECO

Important to have contextual key questions

Determination of issues and rationale is best informed by a multidisciplinary team

PECO structure varies depending on application and objectives – hazard, risk, candidate datasets/values, etc.
Case study: PECO for multiple endpoints

- For [population], is caffeine intake above [exposure], compared to intakes [comparator] or less, associated with adverse effects on [outcome]?

<table>
<thead>
<tr>
<th>Population</th>
<th>Healthy Adults</th>
<th>Healthy Pregnant Women</th>
<th>Healthy Adolescents</th>
<th>Healthy Children</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exposure</strong></td>
<td>&gt; # mg/day</td>
<td>&gt; # mg/day</td>
<td>&gt; # mg/kg-day</td>
<td>&gt; # mg/kg-day</td>
</tr>
<tr>
<td><strong>Comparator</strong></td>
<td>≤ # mg/day</td>
<td>≤ # mg/day</td>
<td>≤ # mg/kg-day</td>
<td>≤ # mg/kg-day</td>
</tr>
<tr>
<td><strong>Outcome</strong></td>
<td>Adverse effect (five outcomes)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Issue: Determining inclusion/exclusion based on healthy populations

• Compromised physical function; caffeine as a therapeutic
• Physicians and SR expert played a key role in helping with this determination

1. Define healthy
• Subjects who were not specifically described as hospitalized, diagnosed with disease, and/or receiving medical treatment for a disease at the time of the study

2. Impact on SR
• Many abstracts/papers excluded based on definition (retained if included healthy control arm or similar)
  – E.g., diabetes, Parkinson’s

Example:

Habitual consumption of caffeine, a non-selective adenosine receptor (AR) antagonist, has been suggested to be beneficial in Parkinson's and Alzheimer's diseases. Experimental evidence support that ARs play a role in Huntington's disease (HD) raising the hypothesis that caffeine may be a life-style modifier in HD. To determine a possible relationship between caffeine consumption and age at onset (AAO) in HD, we retrospectively assessed caffeine consumption in 80 HD patients using a dietary survey and determined relationship with AAO. Following adjustment for gender, smoking status and CAG repeat length, caffeine consumption greater than 190mg/day was significantly associated with an earlier AAO. These data support an association between habitual caffeine intake and AAO in HD patients, but further studies are warranted to understand the link between these variables.
Issue: Nawrot et al. 2003 update to adverse effects

1. Define adverse
   - Use effects in Nawrot?
   - Subjective, difficult to draw a line
   - Decision: be conservative and comprehensive

2. Impact on SRs
   - Pilot searches utilized “adverse” term (and related) – too restrictive
   - Extracted information based on author conclusion as well as analyst conclusion
   - Characterization of effects in subgroups (e.g., physiological/clinical, order in progression of effect, etc.)
Outcome: “adverse” – an added layer of complexity

Issue: Caffeine exposure has been associated with benefits and adverse effects

- Adverse effects in studies that are evaluating benefits? (e.g., RCTs)
  - Controlled exposure, etc.

- OR lack of adverse effects in studies evaluating benefits?
  - Null findings on potentially adverse outcomes (e.g., heart rate)

Decision

- Include: studies reporting data associated with adverse effects within a benefit/therapy study.

- Exclude: Studies assessing only beneficial or therapeutic endpoints or outcomes following exposure to caffeine.
Outcome: multiple endpoints, multiple outcomes

Issue: Outcome of each review associated with many endpoints
Decision: Comprehensive and inclusive (consistent w/ characterization of hazards)
Impact: Frameworks and implementation activities need to be accommodating

<table>
<thead>
<tr>
<th>Cardiovascular: Endpoint Keywords Identified From Abstracts (281)</th>
</tr>
</thead>
<tbody>
<tr>
<td>acute stiffness (pulse wave velocity)</td>
</tr>
<tr>
<td>arrhythmia</td>
</tr>
<tr>
<td>atrial fibrillation</td>
</tr>
<tr>
<td>blood pressure</td>
</tr>
<tr>
<td>cardiovascular disease</td>
</tr>
<tr>
<td>cerebral blood flow</td>
</tr>
</tbody>
</table>
Protocol publication

Issue: publication is the “gold standard”

• Considerations
  • Level of transparency (proprietary investigations)
  • Platform for publication

ToxStrategies minimum standard: internal comprehensive documentation (including SOPs)

Decisions implemented:

• Publication on PROSPERO
• Methods/supplementary to publication
• Appendix to reports
Case study: Protocol publication in PROSPERO

PROSPERO International prospective register of systematic reviews

Citation

http://www.crd.york.ac.uk/PROSPERO/
Lessons learned in protocol implementation
Role of librarian/information specialist

Per IOM, review team includes a librarian/information specialists trained in searching bibliographic databases

In our experience, we have utilized librarian/information specialists assists with:

- Conducting preliminary searches
- Development of overall search strategy – including selection of databases
- Development of search strings (syntax specific to individual databases)
- Assist with refinements and rationale to protocol (particularly as part of pilot)

Practical application note: not all of us have “easy” access to such specialists
Databases - everyone uses PubMed, right?

Starting point for most health-related searches (though terms/syntax are rarely reported)

Is one database sufficient?

How choose (other) databases? Depends on many things…

- Objectives (e.g., turn over every stone, only publicly available, etc.)
- PECO (e.g., some databases specific/unique outcomes [DART, ReproEXPERT], exposures/chemicals [SciFinder], document types [ExPub, CEBS], etc)
- Practical considerations (e.g., facilitation with software, resources/access, overlap in aggregation)

Per the IOM:

The appropriate sources of information for an SR depend on the research question, analytic framework, outcomes of interest, study population, etc.
Embase vs. Medline (content)

- >30 million records (>8,500 journals, grey literature)
  - Over 2700 journals not indexed on MEDLINE
- Emtree Life Science thesaurus, which has over twice as many terms as the MEDLINE thesaurus (MeSH)
- Lesson: not an “either/or”
  - Includes most of Medline
  - Lag for indexing Medline

![Venn Diagram]

- **Embase Unique**
  - 5.9m
  - 2700
- **Embase & Medline**
  - 11.8m
  - 3000
- **Medline on Embase**
  - 8.3m
  - 2500

[embase]/lim NOT [medline]/lim
[embase]/lim AND [medline]/lim
[medline]/lim NOT [embase]/lim
Developing database-specific syntax (some examples)

<table>
<thead>
<tr>
<th>Operators</th>
<th>Syntax example</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AND</strong></td>
<td>‘caffeine’ AND ‘addiction’</td>
<td>Tends to focus or narrow a search.</td>
</tr>
<tr>
<td>Both words or phrases in the record (e.g. title, abstract, keywords, etc)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>OR</strong></td>
<td>‘caffeine’ OR ‘coffee’</td>
<td>Tends to broaden or make a search more inclusive. Useful when terms are related.</td>
</tr>
<tr>
<td>At least one word or phrase in the record (e.g. title, abstract, keywords, etc)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>NOT</strong></td>
<td>‘caffeine’ NOT ‘adverse’</td>
<td>Tends to focus a search. May be useful for exclusion terms.</td>
</tr>
<tr>
<td>Word or phrase after NOT must be excluded</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Wildcards</strong></td>
<td>‘caffeine’ AND cardio*</td>
<td>Broad search on the root, though can be too broad (example gave only 43 results and an error message)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MeSH</strong></td>
<td>‘caffeine’ AND cardiovascular[MeSH]</td>
<td>Established indexed groupings, focused results (example gave 3080 results)</td>
</tr>
</tbody>
</table>

⚠️ Wildcard search for ‘cardio*’ used only the first 600 variations. Lengthen the root word to search for all endings.
Of 105 Medline search strategies examined, 63 were assessed; 31 were excluded because they were inadequately reported.

Most (90.5%) of the assessed search strategies contained > or = 1 errors...

The most common search errors were missed MeSH terms (44.4%), unwarranted explosion of MeSH terms, and irrelevant MeSH or free text terms (28.6%)

Missing spelling variants, combining MeSH and free text terms in the same line, and failure to tailor the search strategy for other databases occurred with equal frequency (20.6%)

Logical operator error occurred in 19.0% of searches
Tools to facilitate the SR

Many tools available

- Selection dependent on many variables
  - objectives, team members, and platforms for communication and reporting, etc.

- Vary in capabilities and transparency
  - Protocols, literature screening and selection, library/source management, data extraction, text mining, meta-analysis, etc.

- Vary substantially in price

http://systematicreviewtools.com/index.php
Case example: caffeine – librarian role

1. Selected three databases to be included in the search
   • Considerations for the voluminous body of literature, endpoints of interest, etc.

2. Crafted search syntax for each database
   • Involved many iterations

3. Result: outcome-specific search strings

Lessons Learned

• Librarians/information specialists significantly enhance the literature search strategy – but it is critical that the (technical) project team be highly engaged in decisions regarding the search strategy
Case example: caffeine – searching multiple databases for multiple outcomes

**Issue:** Independent searches resulted in large number of hits across databases

**Solution:** Run concatenated outcome search, upload all findings in DistillerSR, and categorize during screening (more efficient)

- “Automatic” de-duplication
- Identify articles that address multiple outcomes

<table>
<thead>
<tr>
<th>Combination</th>
<th># Hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual</td>
<td></td>
</tr>
<tr>
<td>Acute Toxicity</td>
<td>2337</td>
</tr>
<tr>
<td>Behavior</td>
<td>3942</td>
</tr>
<tr>
<td>Bone</td>
<td>622</td>
</tr>
<tr>
<td>Cardio</td>
<td>1058</td>
</tr>
<tr>
<td>PK</td>
<td>1254</td>
</tr>
<tr>
<td>Reproductive</td>
<td>875</td>
</tr>
<tr>
<td>Sum of individual searches</td>
<td>10088</td>
</tr>
<tr>
<td>Concatenated search string (all outcomes)</td>
<td>4942</td>
</tr>
</tbody>
</table>
Case example: caffeine-screening pilot

Extensive multi-day pilot conducted with all screeners present
  - Consistent with IOM framework

Initiated prior to final search; tested using preliminary and final search strings

Developed, tested, and modified DistillerSR screening forms

Developing and refined internal “SOP”
  - Minor additions/clarifications to protocol

Result: homogeneity in responses across reviewers
  - Builds confidence

Lessons Learned

1. **Pilot is critical for consistency; plan for refinements to the protocol and consider “SOPs”**

2. **Documentation can be cumbersome – but it’s critical**

3. **SR software (with audit log) ensures transparency and completeness**
Challenges in individual study assessment

Design and execution of data extraction

• Generic templates (Distiller SR forms) designed to cover all outcomes
• Include author information and conclusions as well as topic/project specific fields that involve interpretation in order to apply to the PECO (and to easily apply in the body of evidence synthesis)

Selection of “grading” criteria

• Applicable to all outcomes and streams (requires flexibility) and/or aspects of PECO
  – Going beyond RoB?
• Most appropriate to assess overall project objectives (e.g., hazard, risk, causation) as well as data types (including contextual data)
  – May require multiple systems/framework or tiered approach
Challenges in individual study assessment (cont.)

Application of “grading” criteria

- Interpretation of criteria (not always directly applicable, or well described etc.)
  - Tailor guidance/interpretation to PECO (less so in “broad” applications”)
  - Requires extensive pilot and generation of internal SOPs
- Use of results in subsequent analyses
  - Inclusion/exclusion, subgroup analyses, etc.
  - Identification of candidate datasets for development of toxicity values
- Available frameworks still seem to require expert judgment when applying to a body of evidence/conclusions (as well as identifying candidate datasets)
Practical Lessons Learned
Overall “practical” lessons

1. Every SR is different; each SR has many, many moving parts

2. Problem formulation and protocol development are not only fundamental, but are vital to the SR (and can be difficult)

3. SR approaches vary within EBT – particularly for hazard, risk, causation, etc.
   • Broad PECO (and multi-endpoint assessments) (may) require a (slightly) different application of SR methodologies

4. Selection of appropriate software tools greatly enhances efficiency (and compliance with SR methodologies)
   • Documentation can be cumbersome, but it’s critical

5. The field of EBT is evolving
   • We are gaining resources, but we are still learning how to consistently apply such

6. Cannot ignore the role of resources and time
### Overall “practical” lessons (cont.) - Time

<table>
<thead>
<tr>
<th>Phase</th>
<th>Time</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem formulation and protocol development</td>
<td>Variable (not measured in minutes or hours)</td>
<td>Includes preliminary searching, collaborations, publication etc.</td>
</tr>
<tr>
<td>Literature searching and database generation</td>
<td>Variable</td>
<td>Number of databases, software compliance, de-duplication, hand-searching</td>
</tr>
<tr>
<td>Pilot screening</td>
<td>1-5+ minutes/hit</td>
<td>Discussion, documentation, protocol refinements, form development/refinement; # of outcomes</td>
</tr>
<tr>
<td>Title and abstract Review*</td>
<td>~1-2 minutes + conflict/group review</td>
<td>Specificity of SOP + inclusion/exclusion; Level of documentation</td>
</tr>
<tr>
<td>Pilot study assessment</td>
<td>&gt;2 hours/study</td>
<td>Discussion, documentation, protocol refinements, form development/refinement</td>
</tr>
<tr>
<td>Individual study assessment*</td>
<td>1.5-3 hours (with outliers)</td>
<td>(obtain/manage papers), data extraction, grading; # of endpoints, complexity</td>
</tr>
</tbody>
</table>

*Discussed in OHAT 2015
Questions and Discussion

Thank you!