Comparative Effectiveness Research Collaborative Initiative (CER-CI)

PART 1: INTERPRETING OUTCOMES RESEARCH STUDIES FOR HEALTH CARE DECISION MAKERS

HEALTH EVIDENCE FOR DECISION MAKING: ASSESSMENT TOOL FOR MODELING STUDIES

J. Jaime Caro, MDCM, FRCPC, FACP
Task Force Chair
CER-CI: Interpreting Modeling Studies for Health Care Decision Makers Task Force

Forum: Tuesday November 6th

ISPOR 15th EUROPEAN CONGRESS, BERLIN, GERMANY

Agenda

- Background
- Objectives
- Process
- Draft Assessment Tool
- Next steps
- Q&A
Background

• Decision makers need to determine which intervention(s) to support
• Growing interest in measures to ensure decisions better informed by relevant evidence
• Available interventions & evidence increasing
• RCTs simultaneously comparing all interventions of interest almost never available
• Modeling helps assemble & synthesize evidence
• Models are perceived as black boxes that decision-maker cannot adequately evaluate.

AMCP/ISPOR/NPC CER Collaborative Initiative

• **Objective:** Enhance usefulness of CER to improve patient health outcomes
  – Guidance and practical tools to help P&T critically appraise CER studies to inform decision making
  – Guidance to industry on what kinds of evidence payers want to see and how evidence will be considered in decision making
  – Provide greater uniformity and transparency in use & evaluation of CER for coverage decisions.
Critical Appraisal and Coverage Determinations

What evidence exists?
- Lit Review
- Systematic Search
- DERP NICE reviews
- Guidelines
- Data on file/poster

Is a piece of evidence good?
Critical Appraisal of a single study

What does the evidence say?
Evidence Synthesis

What is the decision?
- Intervention
- Organization
- Cost
- Value
- Affected Population

CER-CI, PART 1

- Create ‘Assessment Tools’ to evaluate credibility & relevance of CER studies:
  - Prospective observational
  - Retrospective observational
  - Modeling
  - Network meta-analyses
- Develop questions pertinent to each type
- Promote consistency across the tools.
CER-CI Interpreting Modeling Studies For Health Care Decision Makers Task Force

Task Force Members:

AMCP
Cheryl Kaltz, RPh, MBA
Lead Clinical Pharmacist
University of Michigan,
Prescription Drug Plan
Ann Arbor, MI, USA

Bimal Patel MS, PharmD
Director, Outcomes and PE Clinical Research Dept.
MedImpact Healthcare Systems, Inc.
San Diego, CA, USA

Cheryl Kaltz
Lead Clinical Pharmacist
University of Michigan,
Prescription Drug Plan
Ann Arbor, MI, USA

Bimal Patel
MS, PharmD
Director, Outcomes and PE Clinical Research Dept.
MedImpact Healthcare Systems, Inc.
San Diego, CA, USA

ISPOR
John Penrod, PhD
Associate Director, HEOR
Bristol-Myers Squibb

Hong J. Kan, PhD, MA, MPP
Director, Immuno-Inflammation
US Health Outcomes
GlaxoSmithKline
Research Triangle Park, NC, USA

Andrew Briggs, DPhil
William R. Lindsay Chair of Health Economics
Health Economics & Health Technology Assessment
Institute of Health & Wellbeing
University of Glasgow
Glasgow, Scotland UK

John Penrod
PhD
AMCP Meeting
April 2013

• Web-based tools
• Educational materials

Develop Questionnaires Based on ISPOR GRPs
• September 2011–June 2012

Health Evidence for decision making: Assessment Tools Status update

Tool first draft
• June–July 2012

Self testing
• Aug–Sep 2012

User testing
• Sep–Oct 2012

Members review
• Dec 2012

Final product
• Nov 2012

AMCP Meeting
April 2013

• Web-based tools
• Educational materials
Key Decisions

• Level of expertise needed
• Granularity
• Organization
• Scoring.

Level of Expertise Required

• Basic understanding of modeling concepts
• Some jargon ok in tool
  – To gain acceptance for a wide range of users, including experts
  – To allow assessors to become familiar with the ‘language of modeling’
  – Expect that, after some practice, non-experts will understand the key terms used in the instrument
  – Supporting document explaining the concepts in a non-technical manner.
Granularity

• Tool needs to be sufficiently specific to help decision-makers judge relevance & credibility
• More specific questions will help non-experts in their assessment
• Too many questions will compromise the practicality of the tool.

Two main questions common across all four tools

• Is the study relevant?
  – Extent to which results, if study is credible, apply to decision-maker’s setting
  – Addresses population, comparators, endpoints, timeframe
• Is the study credible?
  – Extent to which model accurately responds to the question(s) it is designed to answer
  – Addresses issues of Validity
    • Face
    • Verification
    • Predictive
  – Also addresses reporting comprehensiveness.
Tool Characteristics

• Web-based
  – Graphical indicators
  – Hyperlinked explanations
  – Companion document to explain use
• Credibility questions organized by importance
• Many items have sub-questions that can be utilized or not depending on reviewer expertise.
### Aldosterone Antagonists

The cost-effectiveness of Aldosterone Antagonists

**Journal Reference:** VALUE IN HEALTH 15 (2011) 420 - 428

<table>
<thead>
<tr>
<th>Open</th>
<th>Study Title</th>
<th>Study Type</th>
<th>Assessor/Dates</th>
<th>Relevance</th>
<th>Validity</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Cost Effectiveness of Aldosterone Antagonists to the Treatment of</td>
<td>Evidence-Based Medicine</td>
<td>9/2012 to 2015</td>
<td>21/2012</td>
<td>High</td>
<td>Published</td>
</tr>
</tbody>
</table>

#### Study Relevance - Step 2 Questions

- **Domain: Relevance**
  - **Relevance:** Based on your answers to the question in this domain, please provide an overall judgment of strength. Is it strong, neutral, or weak?
  - **Answer:** Can't answer
  - **Check:** Can't answer

1. **1.1** Are the demographics (e.g., age, gender, race/ethnicity) similar? [688]
   - **Answer:** Can't answer
   - **Check:** Can't answer

2. **1.2** Are the risk factors (e.g., blood pressure, cholesterol, weight) similar? [689]
   - **Answer:** Can't answer
   - **Check:** Can't answer

3. **1.3** Are the behaviors (e.g., smoking) similar? [691]
   - **Answer:** Can't answer
   - **Check:** Can't answer

4. **1.4** Were the medical conditions similar? [692]
   - **Answer:** Can't answer
   - **Check:** Can't answer

5. **1.5** Are the co-morbidities similar? [693]
   - **Answer:** Can't answer
   - **Check:** Can't answer
Instrument
7 Categories, 14 Main Questions

• Relevance (4)
• Credibility
  – Validation (3)
  – Design (1)
  – Data (1)
  – Analysis (2)
  – Reporting (1)
  – Interpretation (1)
  – Conflict of interest (1)

“Scoring” of Tool

• Relevance & credibility assessed separately
  – Does the analysis have sufficient relevance?
  – Is it sufficiently credible?
• Each question worded for yes/no (can’t answer)
• After answering the items in each credibility domain, it is rated as
  – Strength
  – Weakness
  – Neutral
• Some questions trigger a “fatal flaw”.
Step 2 - Relevance

1. Is the population relevant? Y = ✅
2. Are any relevant interventions missing? N = ✗
3. Are any relevant outcomes missing? N = ✗
4. Is the context (settings & circumstances) applicable? Y = ✅

2nd Level Questions

- Are the demographics (e.g., age, gender, race/ethnicity) similar?
- Are risk factors (e.g., blood pressure, cholesterol, weight) similar?
- Are behaviors (e.g., smoking) similar?
- Was the medical condition similar?
- Are co-morbidities similar?
Step 3 - Credibility

**Domain: Validation**
1. Is the model sufficiently well externally validated to make its results credible for your decision? $N = \text{✗}$
2. Is the model sufficiently well internally verified to make its results credible for your decision? $N = \text{✗}$
3. Does the model have sufficient face validity to make its results credible for your decision?

**Domain: Design**
4. Is the design of model adequate for your decision problem? $N = \text{✗}$

**Domain: Data**
5. Are the data used in populating the model suitable for your decision problem?

**Domain: Analyses**
6. Were the analyses of the model adequate to inform your decision problem?
7. Was there an adequate assessment of uncertainty?

**Domain: Reporting**
8. Was the reporting of the model adequate to inform your decision problem?
Step 3 - Credibility

Domain: Interpretation
9. Was the interpretation of results fair and balanced?

Domain: Conflict of interest
10. Was the study free of conflict of interest?
Questions and Comments
Welcome!
Comparative Effectiveness Research Collaborative Initiative (CER-CI)
PART 1: INTERPRETING OUTCOMES RESEARCH STUDIES FOR HEALTH CARE DECISION MAKERS

HEALTH EVIDENCE FOR DECISION MAKING: ASSESSMENT TOOL FOR NETWORK META ANALYSIS STUDIES

Jeroen P Jansen, PhD
Interpreting Indirect Treatment Comparison Studies For Health Care Decision Makers Task Force Chair
VP- Health Economics & Outcomes Research, MAPI Consultancy, Boston, MA, USA

Forum: Monday November 5th
ISPOR 15th EUROPEAN CONGRESS, BERLIN, GERMANY

AMCP/ISPOR/NPC

Agenda

• Background
• Objective
• Process
• Draft Assessment Tool
• Next steps
• Q&A
AMCP, NPC, ISPOR Comparative Effectiveness Research Collaborative Initiative (CER-CI)

• The aim is enhancing usefulness of CER to improve patient health outcomes:
  – Guidance and practical tools to help P&T members critically appraise CER studies to inform decision making
  – Guidance to industry on what kinds of evidence payers want to see and how evidence will be considered in decision making
  – Provide greater uniformity and transparency in the use and evaluation of CER for coverage and decision making

Comparative Effectiveness Research Collaborative Initiative (CER-CI), PART 1

• Create a set of ‘Assessment Tools’ to evaluate the credibility, and relevance of non experimental studies
  – Prospective observational studies
  – Retrospective observational studies
  – Modeling studies
  – Network meta-analysis

• Develop specific questions pertinent to the type of study

• Promote consistency across the different tools
Why are these tool potentially useful to decision makers?

- Provides a critical appraisal tool to help identify credible evidence to help inform decision-making
- Promote uniformity when appraising the quality of evidence
- Can be used as an educational tool to increase sophistication in the critical review of CE information

Network Meta-Analysis: Background

- Clinicians, patients, and health-policy makers often need to decide which treatment is “best”.
- Growing interest in measures/methodologies to help ensure healthcare decision making is better informed by results of relevant evidence.
- Available treatments tend to increase over time.
- Unfortunately, robustly designed RCTs that simultaneously compare all interventions of interest are almost never available.
- *Network meta-analysis* is an extension of standard pairwise meta-analysis by including multiple pairwise comparisons across a range of interventions.
- Network meta-analysis provides indirect comparisons for interventions not studied in a head-to-head fashion.
Evidence Networks

Network Meta-Analysis: Background

- There is an increase in the number of indirect treatment comparisons and network meta-analysis published.

- Given the relevance of network meta-analysis to inform healthcare and coverage decision-making, it is of great interest to improve understanding of these studies by decision-makers and increase their use.

Objective

• To develop a measurement tool for the assessment of *credibility* and *relevance* of a network meta-analysis.
Interpreting Indirect Treatment Comparison Studies
For Health Care Decision Makers Task Force

Chair: Jeroen Jansen, PhD
VP- Health Economics & Outcomes Research, MAPI Consultancy, Boston, MA, USA

Members:

AMCP
Sherry Andes, BSPharm, PharmD, BCPS, BCPP, PAHM
Clinical Pharmacist - Drug Information Services, informedRx, an SXC Company
Louisville, Kentucky, USA

Jessica Daw, PharmD, MBA
Director, Clinical Pharmacy
UPMC Health Plan
Pittsburgh, PA, USA

NPC
Vijay Joish BS Pharm, MS, PhD
Deputy Director HEOR, Bayer Pharmaceuticals, Wayne, New Jersey, USA

Joseph C. Cappelleri PhD, MS, MPH
Senior Director
Pfizer Inc
New London, Connecticut, USA

ISPOR
Thomas Trikalinos MD, PhD
Research Associate in Health Services Policy & Practice, Public Health-Health Services Policy & Practice, Brown University, Providence, Rhode Island, USA

Georgia Salanti PhD
Assistant Professor, Department of Hygiene and Epidemiology
University of Ioannina, School of Medicine University Campus, Ioannina, Greece

Steps for Development of Tool

1. Outlining requirements (“Desiderata”)
   - Relevant for decision-making
   - Total evidence base
   - Appropriate statistical methods
   - Transparency
   - Understandable and easy to use

2. Draft items/questions along with glossary

3. Self and user testing

4. Modification (if needed)
Key Decisions

• Level of expertise needed
• Granularity
• Organization
• Scoring

Level of Expertise Required of Assessor

• Basic understanding of study design and elementary epidemiological concepts is required.

• Some jargon in tool
  – To gain acceptance for a wide range of users, including experts
  – To allow assessors to become familiar with the ‘language of ITC/NMA’
  – Expected that after some practice, non-experts will understand the key terms used in the instrument
  – Glossary and supporting document explaining the concepts in a non-technical manner
Granularity

- The tool needs to be sufficiently specific to help decision-makers judge validity and relevance.
- More specific questions will help non-experts in their assessment.
- Too many questions will compromise the practicality of the tool.

Organization

- Divided into domains
- Questions organized by journal article format
- Sub-questions that can be utilized depending on reviewer expertise
- Web-based with hyperlinked explanation of terms
Two Main Domains

• **Relevance**
  - Relevance addresses the extent to which the results of the study, if accurate, apply to the setting of interest to the decision-maker.
  - Addresses issues of population, interventions, endpoints and policy-relevant differences.

• **Credibility**
  - Internal validity
  - The extent to which the study appropriately answers the question it is designed or intended to answer.
  - The extent to which the results are affected by bias due to design of the study and conduct of analysis.
  - Addresses issue of reporting comprehensiveness

Instrument

6 Categories, 26 Questions

• **Relevance**

• **Credibility**
  - Evidence base
  - Analysis
  - Reporting Quality & Transparency
  - Interpretation
  - Conflict of interest
Scoring

• Credibility and Relevance scored separately
  – Does the study have sufficient credibility or relevance based on the answers to the questions in each domain?

• Questions are grouped by domain and each question is worded as a yes/no. After answering the items in a domain, it is then rated by the user as
  – Strength
  – Weakness
  – Neutral

• When some questions are answered in a particular way, they are considered fatal flaws
Relevance

1. Is the population relevant?
2. Are any critical interventions missing?
3. Are any relevant outcomes missing?
4. Is the context (e.g. settings and circumstances) applicable to your population?
**Credibility: Evidence base**

5. **Was the body of randomized controlled trials concerning relevant interventions identified?**
   - Was the literature search strategy defined in such a way that available randomized controlled trials concerning the relevant interventions could be identified?
   - Were multiple databases searched (e.g. MEDLINE, EMBASE, Cochrane)?
   - Were study selection criteria defined in such a way that relevant randomized controlled trials were included if identified with the literature search?

6. **Are the treatments that are being compared with the network meta-analysis all part of one connected network of only randomized controlled trials?**

**Decision Set vs. All Evidence Set**

- **Decision Set**
  - Interested in interventions A, B and C AB and AC studies

- **All Evidence Set**
  - Interested in interventions A, B and C AB, AC, AD, CD, and BD studies provide relevant evidence
Credibility: Evidence base (cont’d)

7. Is it likely that bias was induced by including poor quality studies?

8. Is it likely that bias was induced by selective reporting of outcomes in the studies?

9. Are there systematic differences in treatment effect modifiers across the different treatment comparisons in the network?

10. If yes, were imbalances in effect modifiers across the different treatment comparisons identified prior to reviewing individual study results?

Imbalance in Treatment Effect Modifiers Cause Bias in Indirect Comparisons

Credibility: Analysis

11. Were statistical methods used that synthesize results of trials without “breaking randomization”? (No naïve comparisons)

12. If direct and indirect comparisons are available for pairwise contrasts (i.e. closed loops), was agreement (i.e. consistency) evaluated or discuss?

13. In the presence of consistency between direct and indirect evidence for pairwise contrasts, were both direct and indirect evidence included in the network meta-analysis?

14. Were any systematic differences in potential treatment effect-modifiers across comparisons and (unexplained) inconsistency taken into account in the network meta-analysis?

Imbalance in Treatment Effect Modifiers Cause Inconsistency in Mixed Treatment Comparisons

\[ d_{BC}^{\text{direct}} \neq d_{BC}^{\text{indirect}} = d_{AC}^{\text{direct}} - d_{AB}^{\text{direct}} \]

- Severe Biased Mild Moderate

\[ d_{AC}^{\text{direct}} \neq d_{AC}^{\text{indirect}} = d_{AB}^{\text{direct}} + d_{BC}^{\text{direct}} \]

- Mild Biased Moderate Severe

\[ d_{AB}^{\text{direct}} \neq d_{AB}^{\text{indirect}} = d_{AC}^{\text{direct}} - d_{BC}^{\text{direct}} \]

- Moderate Biased Mild Severe

Inconsistency, all estimates biased
Adjustment for Imbalance in Treatment Effect Modifiers in Indirect Comparisons

15. Was a valid rationale provided for the use of random effects or fixed effects models?

16. If a random effects model was used, were assumptions about heterogeneity explored or discussed?

17. If there are indications of heterogeneity, were subgroup analyses or meta-regression analysis with pre-specified covariates performed?
Credibility: Reporting Quality & Transparency

18. Is a graphical or tabular representation of the evidence network provided with information on the number of RCTs per direct comparison?

19. Are the individual study results reported?

20. Are direct results reported separately?
Credibility: Reporting Quality & Transparency (cont’d)

21. Are all pairwise contrasts between interventions as obtained with network meta-analysis reported along with measures of uncertainty?

22. Is a ranking of interventions provided given the reported treatment effects and its uncertainty by outcome?

23. Is the impact of important patient characteristics on treatment effects reported?

### Probabilistic Interpretation of the Uncertainty

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Odds ratio and credible interval</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspirin + dipyridamole</td>
<td>0.66</td>
<td>0.58</td>
<td>0.75</td>
</tr>
<tr>
<td>Thiopurines + aspirin</td>
<td>0.73</td>
<td>0.61</td>
<td>0.88</td>
</tr>
<tr>
<td>Thiopurines</td>
<td>0.79</td>
<td>0.70</td>
<td>0.91</td>
</tr>
<tr>
<td>Aspirin</td>
<td>0.85</td>
<td>0.77</td>
<td>0.94</td>
</tr>
<tr>
<td>Placebo</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
</tbody>
</table>

Credibility: Interpretation

24. Are the conclusions fair and balanced?

Credibility: Conflict of interest

25. Were there any potential conflicts of interest?
26. If yes, were steps taken to minimize these?

Timelines

Tool first draft developed
• June – July 2012

Self testing & modification
• Aug – Sep 2012

User testing & modification
• Sep – Oct 2012

Members' review
• November 2012

Final product
• December 2012

AMCP Meeting
April 2013
• Web-based tools
• Educational materials

Develop Questionnaires Based on ISPOR GRPs
• September 2011 – June 2012
Thank you

http://www.ispor.org/TaskForces/InterpretingIndirectTreatmentComparisonStudiesTF.asp

Questions and comments welcome!