Comparative Effectiveness Research Collaborative Initiative (CER-CI)
PART 1: INTERPRETING OUTCOMES RESEARCH STUDIES FOR HEALTH CARE DECISION MAKERS

ASSESSING NETWORK META ANALYSIS STUDIES: A PROPOSED MEASUREMENT TOOL FOR HEALTH CARE DECISION MAKERS

Interpreting Indirect Treatment Comparison Studies For Health Care Decision Makers Task Force

Forum: Monday June 4th

ISPOR 17th International Meeting, Washington DC

AMCP/ISPOR/NPC

Assessing Network Meta Analysis Studies: A Proposed Measurement Tool For Health Care Decision Makers

Moderator and Speakers

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

Jessica Daw, PharmD, MBA
Member of Interpreting Indirect Treatment Comparison Studies for Health Care Decisions Task Force
Director, Clinical Pharmacy UPMC Health Plan
Pittsburgh, PA, USA

AMCP/ISPOR/NPC CER-CI Part 1
Forum- June 4, 2012
CER_CI: Interpreting Indirect Treatment Comparison Studies for Health Care Decision Task Force

**Agenda**

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

**Background**

Over the next 2 years, how likely is your plan to invest more in CER?

<table>
<thead>
<tr>
<th>Likelihood</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Likely</td>
<td>6%</td>
</tr>
<tr>
<td>2</td>
<td>29%</td>
</tr>
<tr>
<td>3</td>
<td>31%</td>
</tr>
<tr>
<td>4</td>
<td>31%</td>
</tr>
<tr>
<td>Very Likely</td>
<td>4%</td>
</tr>
</tbody>
</table>

N = 49

Over the next 2 years, how likely is your plan to request more CER from industry?

<table>
<thead>
<tr>
<th>Likelihood</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Likely</td>
<td>6%</td>
</tr>
<tr>
<td>2</td>
<td>29%</td>
</tr>
<tr>
<td>3</td>
<td>31%</td>
</tr>
<tr>
<td>4</td>
<td>31%</td>
</tr>
<tr>
<td>Very Likely</td>
<td>4%</td>
</tr>
</tbody>
</table>

N = 49

Source: Online Survey of Xcenda’s Managed Care Network. September 2010.
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.

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 (validity)* and *relevance* of a network meta-analysis.

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

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

- Specific questions pertinent to the type of study
- Consistency across the different tools

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, BPharm, 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. Modified Delphi approach

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.
Two Main Questions

• Credible?
  – 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.

• Relevant?
  – 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.
Organization

• By domain?
  – Credibility (Internal Validity)
  – Relevance

• By journal article format / reporting sequence?
  – Abstract
  – Background
  – Methods
  – Findings
  – Interpretation

• Other?

Draft Instrument
3 Categories, 27 Questions

Relevance for the decision-problem
- 5 questions (1 on validity; 4 on relevance)
- Focus areas: population, intervention, outcomes

Assessment of adequacy of methodology and validity
- 14 questions on validity
- Focus areas: good search practices; quality of included studies; outcome selection; evidence network; systematic differences in effect modifiers across comparisons; consistency; appropriate statistical methods; explaining heterogeneity

Completeness of Reporting
- 8 questions on relevance
- Focus areas: presentation of network; individual study results; results of NMA for all comparisons; direct effects separate from indirect effects; ranking; subgroup effects; conclusions in line with results and limitations
Relevance for the Decision-Problem

1. Has the target population been clearly defined?
2. Is the target population relevant for the decision-problem?
3. Are all the relevant interventions for the decision-problem been included? (Decision Set)
4. Have additional interventions been included beyond those relevant for the decision-problem? (Evidence Set)
5. Have relevant outcomes been included?

Decision Set vs. All Evidence Set

Interested in interventions A, B and C AB and AC studies

Interested in interventions A, B and C AB, AC, AD, CD, and BD studies provide relevant evidence
Assessment of Adequacy of Methodology and Validity

6. Were good systematic literature search and study selection practices performed?

7. Have all randomized controlled trials involving at least two treatments of the Evidence Set for the population of interest been included?

8. Is there a possibility of bias induced by compromised quality of included studies?

9. Is there a possibility of bias induced by selection of outcomes?

Assessment of Adequacy of Methodology and Validity (cont’d)

10. Are the treatments that are being compared with the network meta-analysis all part of one connected network of trials?

11. Are there systematic differences in treatment effect modifiers across the different types of treatment comparisons in this network?*

Was this justified by the researchers before seeing the results of the individual studies?
Imbalance in Treatment Effect Modifiers Cause Bias in Indirect Comparisons


Source: Jansen JP. Network meta-analysis of individual and aggregate level data. (under review)
Assessment of Adequacy of Methodology and Validity (cont’d)

12. Were statistical methods used that synthesize results of trial comparisons? (No “naïve comparisons”)

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

14. Were in the presence of consistency between direct and indirect evidence, both included in the analysis?

15. 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

Inconsistency, all estimates biased

\[
\begin{align*}
\delta_{BC}^{direct} \neq \delta_{BC}^{indirect} &= \delta_{AC}^{direct} - \delta_{AB}^{direct} \\
\delta_{AC}^{direct} \neq \delta_{AC}^{indirect} &= \delta_{AB}^{direct} + \delta_{BC}^{direct} \\
\delta_{AB}^{direct} \neq \delta_{AB}^{indirect} &= \delta_{AC}^{direct} - \delta_{BC}^{direct}
\end{align*}
\]

Severe  Biased  Mild  Moderate

Mild  Biased  Moderate  Severe

Moderate  Biased  Mild  Severe
Assessment of Adequacy of Methodology and Validity (cont’d)

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

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

18. If there are indications of heterogeneity, were subgroup analyses or meta-regression analysis with pre-specified covariates performed?

Completeness of Reporting

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

20. Are relevant study and patient characteristics (e.g. effect modifiers) reported or available?

21. Are the individual study results reported?

22. Are direct results reported separately?
**Completeness of Reporting (cont’d)**

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

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

25. Are subgroup effects or the impact of important patient characteristics on treatment effects reported?

26. Are conclusions in line with research questions, results and validity/limitations of analyses?

---

**Probabilistic Interpretation of the Uncertainty**

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Odds ratio and credible interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>Aspirin + dipyridamole</td>
<td>0.66</td>
</tr>
<tr>
<td>Thienopyridines + aspirin</td>
<td>0.73</td>
</tr>
<tr>
<td>Thienopyridines</td>
<td>0.79</td>
</tr>
<tr>
<td>Aspirin</td>
<td>0.85</td>
</tr>
<tr>
<td>Placebo</td>
<td>Reference</td>
</tr>
</tbody>
</table>

---

Scoring

• Each question will be scored as Yes / No / Unclear
  – with annotation, if necessary

• No weighting of items or quantitative scoring

• Total score for validity (credibility) and a total score for relevance based on number of items scored ‘Yes’.

• ‘Fatal flaws’ will be based on the scoring of some of the validity items.

Next Steps

• Finalize glossary and supporting document
• Circulate to external experts in NMA for feedback
• Circulate to wider audience of decision-makers and ISPOR community
• Write manuscript
Questions?

Thank you

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