SMAA for benefit-risk analysis

Gert van Valkenhoef

Department of Epidemiology, University Medical Center Groningen (NL), Faculty of Economics and Business, University of Groningen (NL)

1st MeHTA study group meeting
9 June 2011, Kuopio, Finland
Outline

1. Introduction
2. SMAA
3. SMAA for BR
4. MTC/SMAA for BR
5. ADDIS Example
6. Discussion
Outline

1. Introduction
2. SMAA
3. SMAA for BR
4. MTC/SMAA for BR
5. ADDIS Example
6. Discussion

Ask questions any time!
Problem

- Current statistical methods insufficient to get integrated overview of the alternatives and criteria:
  - Meta-analysis (evidence pooling) on single criterion
  - Only pair-wise comparisons.
- BR analysis is unstructured
  - No pre-specified criteria or models
- BR analysis is non-transparent
  - Evidence basis not (sufficiently) explicit
  - Measurements and value judgments not separated
Multi-Criteria Decision Analysis (MCDA) methods allow
- to evaluate multiple alternatives
- in terms of multiple criteria

Mixed Treatment Comparison (MTC) models enable
- indirect comparisons
- between ≥ 2 alternative treatments

We put MCDA together with MTC to
- systematically assess the (relative) benefits and risks
- of any number of alternative treatments
- on the relevant criteria
- take into account + quantify uncertainty
- explicitly based on clinical evidence
Other proposed MCDA models don’t take uncertainty into account.

The Lynd & O’Brien model is limited to 2x2 problems.

Stochastic Multi-criteria Acceptability Analysis (SMAA) allows \( m \times n \) problems:

- \( m \) alternatives
- evaluated on \( n \) criteria
- performance of alternative \( i \) on criterion \( j \): \( C_{i,j} \sim f(c_{i,j}) \)
SMAA BR analysis

SMAA models for benefit-risk:

- Can be based on a single trial (published)
- Or (network) meta-analysis (under review)
- And is implemented in ADDIS

Stochastic Multi-criteria Acceptability Analysis (SMAA)

- SMAA is a multi-criteria decision aiding (MCDA) method for ranking
  - a set of $m$ alternatives
  - based on a set of $n$ criteria
- Evaluation of alternative $x$ on criterion $y$
  - may be uncertain: specify a probability distribution
- Preference information:
  - a weight vector (optional) and
  - a value function (usually linear)
- SMAA is based on Multi-Attribute Value Theory (MAVT)
SMAA: forward / inverse approach

Criteria measurements $x$ → Decision model $M(x, w)$ → Best solution

DMs’ valuations $w$
SMAA: forward / inverse approach

Criteria measurements $x$ 

Decision model $M(x, w)$ 

Favorable valuations 

Prospective solution
Weight space

- Total lack of preference information is represented by a uniform joint probability distribution of the weight space.

- If some preference information is available, the weight space can be restricted with linear constraints.
SMAA decision aiding metrics

**Rank acceptability index** share of weights and measurements making an alternative have ranks 1, . . . , \( m \) (most preferred, second most, etc.).

**Central weight vector** center of gravity of the favourable weight space: “Which preferences support an alternative to be the most preferred one?”

**Confidence factor** probability for an alternative to be preferred when preferences equal its central weight vector: “Are the measurements are sufficiently precise?”
SMAA for Benefit-Risk Assessment

- $m$ alternative treatments are evaluated with respect to efficacy and $n - 1$ most important ADRs
- Based on a single clinical trial
- All measurements (efficacy and ADRs) are absolute risk
  - Assumed beta distribution
  - Fitted to incidences in trial
- Implemented in ADDIS v1.2

Case study: anti-depressants

- Placebo-controlled randomized clinical trial:
  - Fluoxetine
  - Venlafaxine
  - Placebo

- Criteria (selected by expert):
  - Benefit: efficacy (treatment response)
  - Risks: nausea, insomnia, anxiety
Example: measurements

Measurements were Beta distributions for (absolute) risk.

**Table:** Incidence rates of HAM-D responders and three ADRs, with their Risk Differences (RD) versus Placebo

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Placebo</th>
<th>Fluoxetine</th>
<th>RD (95% CI)</th>
<th>Venlafaxine</th>
<th>RD (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficacy</td>
<td>37/101</td>
<td>45/100</td>
<td>0.08 (-0.05, 0.22)</td>
<td>51/96</td>
<td>0.16 (0.03, 0.30)</td>
</tr>
<tr>
<td>Nausea ADRs</td>
<td>8/102</td>
<td>22/102</td>
<td>0.14 (0.04, 0.23)</td>
<td>40/100</td>
<td>0.32 (0.21, 0.43)</td>
</tr>
<tr>
<td>Insomnia ADRs</td>
<td>14/102</td>
<td>15/102</td>
<td>0.01 (-0.09, 0.11)</td>
<td>22/100</td>
<td>0.08 (-0.02, 0.19)</td>
</tr>
<tr>
<td>Anxiety ADRs</td>
<td>1/102</td>
<td>7/102</td>
<td>0.06 (0.01, 0.11)</td>
<td>10/100</td>
<td>0.09 (0.03, 0.15)</td>
</tr>
</tbody>
</table>
Example: value functions

- The criteria value functions were rescaled
  - Approximate 0–1 scale, 1 being best
  - 95% confidence interval hull of measurements
- Limit region on which preferences are assessed
  - Protects assumption of preference linearity
  - Is it needed in this case? Risk is naturally $[0, 1]$ and linear?

Table: Criteria, preference directions, and scaling vectors

<table>
<thead>
<tr>
<th>Name</th>
<th>Preference direction</th>
<th>$c'_k$</th>
<th>$c''_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficacy</td>
<td>↑</td>
<td>0.28</td>
<td>0.63</td>
</tr>
<tr>
<td>Nausea ADRs</td>
<td>↓</td>
<td>0.50</td>
<td>0.04</td>
</tr>
<tr>
<td>Insomnia ADRs</td>
<td>↓</td>
<td>0.31</td>
<td>0.08</td>
</tr>
<tr>
<td>Anxiety ADRs</td>
<td>↓</td>
<td>0.17</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Example: scenarios

- We considered 3 scenarios:
  1. Health policy decision making with no preferences
  2. Prescription for mild depression
  3. Prescription for severe depression

- Ordinal swing weighting for prescription decisions
Example: preference-free analysis

Figure: Rank acceptability indices
Example: preference-free analysis

**Figure:** Central weights and confidence factors
Example: severe depression

Figure: Rank acceptability indices
Example: mild depression

Figure: Rank acceptability indices
Case study:
- Despite lack of ‘significant’ results, there are trade-offs
- Still a lot of uncertainty, small sample size
Discussion

- (+) Account for uncertainty in inputs and outputs
- (+) Inverse approach and partial preferences $\rightarrow$ low effort
- (+) Inverse approach: identify scenarios for use
- (+) Separate measurements and preferences $\rightarrow$ transparency
- (-) Based on a single trial
- (-) There may be additional criteria not measured in trials
- (-) Not tested in other therapeutic areas
MTC/SMAA for Benefit-Risk Assessment

- \( m \) alternative treatments are evaluated with respect to efficacy and \( n - 1 \) most important ADRs
- Based on
  - \( n \) network meta-analyses (normal distr. log-odds ratio)
  - \( n \) baseline models (normal distr. log-odds)
- All measurements (efficacy and ADRs) are absolute risk
  - Sampled from the log-odds ratio
  - Conditional on the baseline log-odds
- Implemented in ADDIS v1.6

G. van Valkenhoef, T. Tervonen, J. Zhao, B. de Brock, H.L. Hilleg, D. Postmus, Multi-criteria benefit-risk assessment using network meta-analysis. Submitted manuscript.
MTC/SMAA process

1. Identify or perform systematic review
2. Choose criteria
   - \( k := 1 \)
3. Select criterion \( k \)
4. Run inconsistency model
5. Investigate
   - [inconsistency explained]
6. Run consistent model
7. Estimate baseline
8. [unexplained]
9. \([k < n]\)
10. [All criteria done \((k = n)\)]
11. Construct SMAA model
Why use network meta-analysis?

An earlier model used pair-wise meta-analysis. Problems:

- Choice of common comparator has unknown influence on model
  - Selection bias: arbitrary exclusion of evidence
  - Sensitivity analysis with different comparators?
- Only applicable when common comparator available
  - Not the case in many clinical domains

And we want to offer an automated solution!
Measurement scales

- Meta-analysis results in relative measurements
  - E.g. odds ratio, mean difference
  - Statistically more robust
  - Hard to interpret clinically

- For decision making, we need absolute measurements
  - E.g. risk, change from baseline
  - Choose a baseline treatment and estimate absolute effect
  - Sample effects of other treatments conditional on that

- Problem: how to estimate absolute effect?
  - No general answer, lots of options
Case study: anti-depressants

- Based on existing systematic review
- Alternatives (those with most data):
  - Fluoxetine
  - Paroxetine
  - Sertraline
  - Venlafaxine
  - Placebo
- Criteria (selected by expert):
  - Benefit: efficacy (treatment response)
  - Risks: diarrhea, dizziness, headache, insomnia, nausea
- Same scenarios
Example: network meta-analysis

Figure: Evidence network (25 studies total)
Example: relative measurements

(a) Fluoxetine

(b) Paroxetine

(c) Sertraline

(d) Venlafaxine
**Example: baseline measurements**

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Log-odds</th>
<th>Risk (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAM-D</td>
<td>$-0.17 \pm 0.11$</td>
<td>0.46 (0.40, 0.51)</td>
</tr>
<tr>
<td>Diarrhea</td>
<td>$-2.19 \pm 0.21$</td>
<td>0.10 (0.07, 0.14)</td>
</tr>
<tr>
<td>Dizziness</td>
<td>$-2.23 \pm 0.61$</td>
<td>0.10 (0.03, 0.26)</td>
</tr>
<tr>
<td>Headache</td>
<td>$-1.20 \pm 0.29$</td>
<td>0.23 (0.15, 0.35)</td>
</tr>
<tr>
<td>Insomnia</td>
<td>$-2.61 \pm 0.19$</td>
<td>0.07 (0.05, 0.10)</td>
</tr>
<tr>
<td>Nausea</td>
<td>$-2.02 \pm 0.19$</td>
<td>0.11 (0.08, 0.16)</td>
</tr>
</tbody>
</table>
Example: preference-free analysis

Figure: Rank acceptability indices
Example: preference-free analysis

Figure: Central weights and confidence factors
Example: severe depression

Figure: Rank acceptability indices
Example: mild depression

Figure: Rank acceptability indices
Example: discussion

Case study:
- Preference-free analysis gives insight in trade-offs
- Anti-depressents warranted for severe depression
  - But not for mild depression
- Fluoxetine is unlikely to be the best
- There is a lot of residual uncertainty
  - May be due to individual differences in response
  - Choosing between anti-depressants is difficult
- SMAA analysis more informative than just meta-analysis
Discussion

Relative to only using SMAA:

- (+) Take into account all relevant evidence
- (+) Clinically relevant (absolute) scales
- (-) Network meta-analysis labour intensive, difficult
- (-) Scale conversion needs more work
SMAA model based on network meta-analysis.
SMAA example (ADDIS)

Measurements (input distributions).
SMAA example (ADDIS)

Model without preference information.
SMAA example (ADDIS)

Model without preference information.
SMAA example (ADDIS)

Preferences for severe depression.
Severe depression results.
**SMAA example (ADDIS)**

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Scale</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAM-D Responders</td>
<td>OR: [0.90 - 1.74] Risk: [0.51 - 0.67] RD: 0.16 NNT 6.35</td>
<td>5</td>
</tr>
<tr>
<td>Diarrhea</td>
<td>OR: [0.32 - 1.00] Risk: [0.03 - 0.10] RD: 0.07 NNH 15.32</td>
<td>4</td>
</tr>
<tr>
<td>Dizziness</td>
<td>OR: [0.90 - 4.69] Risk: [0.06 - 0.26] RD: 0.20 NNH 5.04</td>
<td>1</td>
</tr>
<tr>
<td>Headache</td>
<td>OR: [0.34 - 1.59] Risk: [0.07 - 0.26] RD: 0.19 NNH 5.31</td>
<td>3</td>
</tr>
<tr>
<td>Nausea</td>
<td>OR: [0.90 - 2.37] Risk: [0.19 - 0.38] RD: 0.19 NNH 5.25</td>
<td>2</td>
</tr>
</tbody>
</table>

Preferences for mild depression.
SMAA example (ADDIS)

Mild depression results.
Relevance: EMA BR methodology project

<table>
<thead>
<tr>
<th>Approach/method</th>
<th>Relevance to regulators</th>
<th>Usefulness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probabilistic simulation</td>
<td>Can illuminate the risk/benefit trade-off when uncertainty is a major feature of a regulatory decision.</td>
<td>Medium</td>
</tr>
<tr>
<td>Bayesian statistics</td>
<td>Can integrate evidence and its uncertainty, both pre- and post-approval, with multiple criteria in decision models.</td>
<td>High</td>
</tr>
<tr>
<td>MCDA</td>
<td>Multi-criteria decision analysis extends decision theory to accommodate multiple, conflicting objectives. Provides common units of value for both benefits and risks.</td>
<td>High</td>
</tr>
</tbody>
</table>

Table: MTC/SMAA integrates 2 of 2 quantitative approaches rated 'High' on usefulness, and 1 rated 'Medium'.