EVALUATING SAMPLING UNCERTAINTY IN COST-EFFECTIVENESS ANALYSIS

Daniel Polsky
University of Pennsylvania
http://www.uphs.upenn.edu/dqimhsr/

ISPOR
Philadelphia
May 21, 2006

Sampling Uncertainty

• The point estimates of cost and effect differences observed in studies are the result of a single sample drawn from a population
• Had one repeated the experiment with a new sample from the population, one would have obtained different point estimates
• Sampling uncertainty is the degree to which point estimates will differ as a result of the fact that the estimate is based on a sample from a population

Outline

• Provide detailed steps on how to use a sample estimate of a cost effectiveness ratio in the presence of sampling uncertainty from a decision making perspective
  – Point estimates
  – Decision threshold
  – Confidence intervals and p-values
• The objective is to improve the understanding of the various methods of expressing sampling uncertainty for the purposes of decision making
  – I will not focus on the technical aspects of estimation.
  – Computer code is available on my website.
STEPS WHEN USING ESTIMATES FROM A SAMPLE FOR DECISION MAKING

1. Determine point estimate in the sample
2. Compare against decision threshold
3. Quantify statistical uncertainty relative to decision threshold
   - Confidence intervals
   - P-values

Common Example

A drug company claims that their new drug (drug A) reduces blood pressure more than placebo (drug B). They randomized 100 subjects to group A and 100 subjects to group B. They measured systolic blood pressure after 6 months.

<table>
<thead>
<tr>
<th></th>
<th>Drug A</th>
<th>Drug B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean effect</td>
<td>125.8</td>
<td>135.2</td>
</tr>
<tr>
<td>st. dev.</td>
<td>22.5</td>
<td>22.5</td>
</tr>
</tbody>
</table>

1. What is the point estimate?
2. What is the decision threshold?
3. How would we express the uncertainty?

COMMON EXAMPLE: ANSWERS

1. What is the point estimate? \( SBP_b - SBP_a = 135.2 - 125.8 = 9.4 \)

2. What is the decision threshold? 0

3. How would we express the uncertainty?

Confidence interval: \( 9.4 \pm (t_{tail} \times s.e.) \)

\[ 95\% \text{ confidence interval } = (3.16, 15.64) \]

P-value: \( .0035 \)
INTERPRETING 95% CI

1) Not confident A differs from B (\(A - B\))

Less Decision threshold Greater

2) Confident A greater than B (\(A - B\))

Less Decision threshold Greater

3) Confident A less than B (\(A - B\))

Less Decision threshold Greater

In our example, we are in category 2.

IMPLICATIONS

• If the confidence interval INCLUDES the decision threshold, we CANNOT be confident that the alternatives differ from one another

• If the confidence interval EXCLUDES the decision threshold, we CAN be confident that the alternatives differ from one another

COST EFFECTIVENESS EXAMPLE

• A pharmaceutical company has conducted an economic evaluation of two drugs. Their new drug (drug A) is cost effective relative to placebo (drug B).

<table>
<thead>
<tr>
<th>Drug A</th>
<th>Drug B</th>
<th>A-B</th>
<th>s.e. of diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Cost</td>
<td>2000</td>
<td>1000</td>
<td>1000</td>
</tr>
<tr>
<td>Mean QALY</td>
<td>0.86</td>
<td>0.85</td>
<td>0.01</td>
</tr>
</tbody>
</table>

• (The correlation between costs and effects=−.7102 and \(n=250\) in each treatment group)
• They claim that their new drug is cost effective relative to drug B.

STEPS:
1. What is the point estimate?
2. What is the decision threshold?
3. How would we express the uncertainty?
1. POINT ESTIMATE OF COST EFFECTIVENESS RATIO (CER)

- NW: Treatment Dominated
- SE: Cost-Effectiveness Ratio Treatment Dominates
- (-) Difference in Effects (+) (-) Difference in Costs (+)

CER = $100,000 / QALY

2. DECISION THRESHOLD FOR THE COST EFFECTIVENESS RATIO

- The decision threshold for a CER is the maximum willingness to pay (WTP) for the marginal effects of the therapy
  - SE quadrant: less costly and more effective:
    - Therapy would meet any WTP
  - NW quadrant: more costly and less effective:
    - Therapy would not meet any WTP
  - NE quadrant: More costly and more effective:
    - A tradeoff between costs and effects
    - Whether the effects are worth the cost depends on whether the CER is less than the maximum WTP for those effects

λ: WILLINGNESS TO PAY (WTP)

- Is there a value that could be applied across an entire population?
  - If so, what is it?
  - Is it $50,000 per QALY?

- There is no single value of λ that applies to all decision makers
  - When estimating effect differences, there is consensus that an effect greater than zero is a reasonable decision threshold. No comparable certainty will ever exist for a decision threshold for cost effectiveness

- For cost-effectiveness analysis to be relevant for decision making, results must have meaning for decision makers with different decision thresholds (i.e., different WTP)
DECISION RULE

- NEW THERAPY COST EFFECTIVE IF:
  \[ \text{CER} < \lambda \]
- NEW THERAPY NOT COST EFFECTIVE IF:
  \[ \text{CER} > \lambda \]

NET MONETARY BENEFITS

- An alternative method for expressing the results of a cost-effectiveness analysis
- NMB is a rearrangement of the cost-effectiveness decision rule:
  \[ \text{CER} = \frac{(\Delta C)}{(\Delta E)} < \lambda \]
  \[ \text{NMB} = \lambda \times (\Delta E) - (\Delta C) > 0 \]
- The decision threshold for NMB: Always 0
- Decision rule: Interventions are cost-effective if NMB > 0

CAN ONLY NUMERICALLY EXPRESS NMB FOR A SINGLE \( \lambda \)

- Given that \( \lambda \) might vary, this expression is limited in its usefulness to decision makers

A GRAPH CAN EXPRESS NMB ACROSS A RANGE OF \( \lambda \)

NMB GRAPH

[Graph showing net monetary benefits across different willingness to pay values]
3. EXPRESSING UNCERTAINTY IN COST-EFFECTIVENESS ANALYSIS

1. 95% Confidence Intervals
   - CER
   - NMB

2. P-values
   - Acceptability Curves

CONFIDENCE INTERVALS FOR COST EFFECTIVENESS RATIOS

Lines through the origin that each exclude $\alpha/2$ of the distribution of the difference in costs and effects

CONFIDENCE INTERVAL FOR EXAMPLE CER

Recall data from example:
$\Delta C = 1000; \text{SEC} = 324; \Delta Q = 0.01; \text{SEQ} = 0.001929; p = .7102; \text{DOF} = 498$
CONFIDENCE INTERVAL OF CER RELATIVE TO DECISION THRESHOLD

1) Not confident A differs from B

Maximum WTP

2) Confident A cost-effective compared to B

Maximum WTP

3) Confident B cost-effective compared to A

Maximum WTP

Confidence regarding whether the estimated cost-effectiveness ratio is good value for the money depends on the decision threshold (i.e., the maximum willingness to pay (WTP) for a unit of effectiveness).

DISCONTINUITIES IN THE DISTRIBUTION OF THE COST-EFFECTIVENESS RATIO

- The distribution of a CER is discontinuous on the real line
  - Because the denominator of a ratio can equal 0 (i.e., the ratio can be undefined)
- The standard techniques for estimating uncertainty can be problematic

CONFIDENCE INTERVAL OF NMB

Must be estimated at a particular $\lambda$.

Using the example, the estimate the confidence interval of NMB at a $\lambda$ of $50,000/\text{QALY}$ is:

\[ \text{NMB}=-500 \]
\[ \text{CI of NMB} = (-1283, 283) \]
CONFIDENCE INTERVAL OF NMB RELATIVE TO ITS DECISION THRESHOLD OF 0

1) $\lambda = 50,000$: Not confident A differs from B ($A - B$)
2) $\lambda = 250,000$: Confident A net beneficial compared to B
3) $\lambda = 10,000$: Confident B net beneficial compared to A

One must calculate separate CI for each policy-relevant $\lambda$ (i.e., at the decision threshold for CER)

The confidence interval derived for a single $\lambda$ may have little in common with the confidence interval - and resulting policy inference - derived for other $\lambda$.

CONFIDENCE FRONTIER OF NMB

What conclusions would you draw about one's confidence in the differences between drugs A and B given that CI for NMB includes its decision threshold of 0?

INTERPRETATION OF NMB GRAPH

- the Y-axis value at A = $-\Delta C$
  - also the upper and lower NMB limits cut the vertical axis at the CI for $\Delta C$
- the X-axis value at C = CER
  - the X-axis value at B and D = CI for CER
QUANTIFYING UNCERTAINTY USING P-VALUES / ACCEPTABILITY CURVES

PROBABILITY OF ACCEPTABILITY

- The probability that the estimated ratio falls below a specified ceiling ratio (i.e., a particular WTP or λ).
- This is analogous to (1-p-value) for a CER at a decision threshold of that particular λ.
- Given that there is no agreed upon ceiling ratio, we routinely estimate the probability of acceptability over the range of possible positive values.

ALTERNATIVE ACCEPTABILITY CRITERIA

- In this example, the ratio of $245,200 per QALY saved has the equivalent of a one-tailed p-value less than 0.025
  - i.e., 2.5% of the distribution falls above the $245,200 line.
The acceptability curve is the plot of the probability of acceptability across values of WTP. The curve's value on the y-axis is the probability that therapy A is cost-effective. This is analogous to 1 - P-value (one sided).

Y-axis value at A = p-value for ΔC
Y-axis value where C approaches = p-value for ΔE
X-axis value at B = CER
X-axis values at D=(1-2X)% confidence interval

Confidence interval for CER
CER CI: (28,300 to 245,200)

Confidence Frontier of NMB
notice same CER CI:
(28,300 to 245,200)

Acceptability Curve
notice same CER CI:
(28,300 to 245,200)
BOTTOM LINE

If the goal is to provide information for decision makers, ultimately all three methods provide the same information. The same decision should be made under each method! The difference is in the way the information is expressed. The choice of method should be based on the most effective way to express the point estimate and the uncertainty to decision makers rather than basing the choice on the statistical properties of each estimator.

CASE #2: WHERE EXPRESSION OF CONFIDENCE IS NOT STRAIGHT FORWARD

- What happens when there is a possibility that the cost effectiveness ratio of a sample from a population may fall into three of the quadrants on the cost effectiveness plane?

Case #2 Example: Confidence intervals for CER
Case #2 Example:
Confidence Frontier of NMB

Case #2 Example:
Acceptability Curve

Case #2: Example
Confidence interval for CER
CER CI: (19,740 [SW] to 5,045 [NE])

Confidence Frontier of NMB
notice same CER CI:
(19,740 [SW] to 5,045 [NE])

Acceptability Curve
notice same CER CI:
(19,740 [SW] to 5,045 [NE])
THIS EXAMPLE HAS A NON STANDARD CONFIDENCE LIMIT PATTERN

- Observed when:
  - Confidence interval for CI includes the Y axis (DE = 0)
  - One NMB confidence limit crosses the decision threshold twice; the other limit never crosses the decision threshold
  - The acceptability curve crosses α/2 or 1-α/2 twice

PATTERN #2

<table>
<thead>
<tr>
<th>Ceiling Ratio</th>
<th>One cannot be confident the two therapies differ</th>
<th>One can be confident that one of the therapies is good value</th>
<th>One cannot be confident the two therapies differ</th>
</tr>
</thead>
<tbody>
<tr>
<td>-∞</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ceiling Ratio</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+∞</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CASE #2

BOTTOM LINE

If the goal is to provide information about whether there is an acceptable level of confidence relative to the decision threshold, all three methods still provide the same information.

CASE #3: No information

- The CI for the CER is undefined
- The NMB CI contain the decision threshold for all ceiling ratios
- The acceptability curve is always above α/2 and below 1-α/2

PATTERN #3

<table>
<thead>
<tr>
<th>Ceiling Ratio</th>
<th>One cannot be confident the two therapies differ from one another</th>
</tr>
</thead>
<tbody>
<tr>
<td>-∞</td>
<td></td>
</tr>
<tr>
<td>Ceiling Ratio</td>
<td></td>
</tr>
<tr>
<td>+∞</td>
<td></td>
</tr>
</tbody>
</table>

• One cannot be confident that one therapy differs from another
RECOMMENDATIONS

- The joint distribution of $\Delta C$ and $\Delta E$ is best expressed graphically given that meaning comes from the joint distribution (two dimensions).
- Each method has advantages and disadvantages for communicating the results of a cost-effectiveness analysis.
- When collapsing the decision problem onto the real line (one dimension) the CER and the CI for CER provides the most information for decision making with the fewest numbers.
  - The NMB is only relevant for a particular $\lambda$.
- The cases when statistical issues arise for CERs expressed on a real line, interpretation of results relative to decision threshold is still possible.
  - In all cases, CER, NMB and acceptability curves provide the same information.
- When reporting uncertainty in a single dimension, one must avoid potential pitfalls of Cases #2 and #3.

A UNIQUE ADVANTAGE OF THE NET MONETARY BENEFIT

- NMB can be estimated on a per-patient basis.
- In this case, the transformation is made prior to analysis, and the analysis directly assesses differences in net monetary benefits.
- Results are equivalent to estimation of NMB using group means.

ESTIMATING UNCERTAINTY IN COST-EFFECTIVENESS ANALYSIS

1. Confidence intervals for CERs (Fieller’s theorem)
2. Confidence intervals for NMB
3. Acceptability curves

See programs online:
http://www.uphs.upenn.edu/dqmhsr/stat%20cicer.htm
Estimation of Confidence intervals for CER: Fieller’s theorem

- Fieller’s theorem method: A parametric method based on the assumption that the differences in costs and the differences in effects follow a bivariate normal distribution
  - i.e., the expression $R \Delta E - \Delta C$ is normally distributed with mean zero (where $R$ equals $\Delta C / \Delta E$ and $\Delta E$ and $\Delta C$ denote the mean difference in effects and costs, respectively)
  - Standardizing this statistic by its standard error and setting it equal to the critical value from a normal distribution generates a quadratic equation in $R$
- The roots of the quadratic equation give the confidence limits

Fieller’s Theorem Formula

Lower limit (LL): $(M - [M^2 - NO]) / N$

Upper limit (UL): $(M + [M^2 - NO]) / N$

Where:
- $M = \Delta E \Delta C - (t_{\alpha/2}^2) \rho \sigma_{\Delta E} \sigma_{\Delta C}$
- $N = \Delta E^2 - (t_{\alpha/2}^2) \sigma_{\Delta E}^2$
- $O = \Delta C^2 - (t_{\alpha/2}^2) \sigma_{\Delta C}^2$

- $\Delta E$ and $\Delta C$ denote mean difference in effect and cost;
- $\sigma_{\Delta E}$ and $\sigma_{\Delta C}$ denote estimated standard errors for the diff in costs and effects;
- $\rho$ equals the estimated Pearson correlation coefficient between the difference in costs and effects;
- $t_{\alpha/2}$ is the critical value from the $t$-distribution

Data Required for Fieller’s Theorem

- The data needed to calculate these limits can be obtained from most statistical packages by:
  - Testing the difference in costs (and obtaining $\Delta C$ and $s_{\Delta C}$)
  - Testing the difference in effects (and obtaining $\Delta E$ and $s_{\Delta E}$)
  - Estimating the correlation between the difference in costs and effects
STATA code: Fieller

http://www.uphs.upenn.edu/dgimhsr/stat%20cicer.htm

Variable | Obs | Mean | Std. Dev. | Min | Max
-------------+--------------------------------------------------------
cost | 500 | 25000 | 3655.213 | 15127.93 | 36227.73
treat | 500 | .5 | .5005008 | 0 | 1

.ipinput cost qaly treat

Difference in cost: 999.99
SE, difference in cost: 324.18
Difference in effect: .01
SE, difference in effect: .0019
Correlation of differences: -.71
Degrees of freedom: 498

.fielleri 999.99 324.18 .01 .0019 -.71 498 0.95

STATA results: Fieller

• Point Estimate (pe): 100000
• Quadrant: NE

• Fieller 95 % Confidence Interval
  • Lower limit: 28295
  • Upper limit: 245263

Estimation of NMB Confidence Intervals

• Use formula for a difference in two normally distributed continuous variables

  NMB CI = NMB ± t_{α/2} SE_{NMB}

• Standard error for NMB equals:

  \[ SE_{NMB} = s_{E}^2 + R^2 s_{C}^2 - 2 R \rho s_{E} s_{C} \]

• where sE and sC denote estimated standard errors for the difference in costs and effects; \( \rho \) equals the estimated Pearson correlation coefficient between the difference in costs and effects; \( t_{α/2} \) is the critical value from the normal distribution
STATA code:
NMB confidence intervals

```
.ipinputs cost qaly treat

Difference in cost:        999.99
SE, difference in cost:        324.18
Difference in effect:          .01
SE, difference in effect:       .0019
Correlation of differences:    -.71
Degrees of freedom:            498
```

```
.nmbi 999.99 324.18 .01 .0019 -.71 498 .95
```

STATA Results: NMB

```
<table>
<thead>
<tr>
<th>Ac</th>
<th>NMB</th>
<th>95 % Lower Limit</th>
<th>95 % Upper Limit</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>-81404</td>
<td>-1416</td>
<td>-1963</td>
<td>-979</td>
<td>0.0010</td>
</tr>
<tr>
<td>-3833</td>
<td>-1283</td>
<td>-1843</td>
<td>-717</td>
<td>0.0010</td>
</tr>
<tr>
<td>15240</td>
<td>-898</td>
<td>-1498</td>
<td>-117</td>
<td>0.0020</td>
</tr>
<tr>
<td>24024</td>
<td>-738</td>
<td>-1494</td>
<td>-27</td>
<td>0.0420</td>
</tr>
<tr>
<td>39777</td>
<td>-650</td>
<td>-1415</td>
<td>15</td>
<td>0.0420</td>
</tr>
<tr>
<td>201172</td>
<td>-154</td>
<td>-1578</td>
<td>2761</td>
<td>0.0220</td>
</tr>
<tr>
<td>234064</td>
<td>-68</td>
<td>2741</td>
<td>0.0420</td>
<td></td>
</tr>
<tr>
<td>240153</td>
<td>54</td>
<td>3024</td>
<td>0.0420</td>
<td></td>
</tr>
<tr>
<td>287956</td>
<td>272</td>
<td>3024</td>
<td>0.0420</td>
<td></td>
</tr>
<tr>
<td>1194574</td>
<td>5959</td>
<td>15972</td>
<td>0.0010</td>
<td></td>
</tr>
<tr>
<td>1465543</td>
<td>10966</td>
<td>5959</td>
<td>0.0010</td>
<td></td>
</tr>
</tbody>
</table>
```

Estimation of Acceptability Curves

- One means of deriving parametric acceptability curves is by estimating the 1-tailed probability that the net monetary benefits, calculated by use of the ceiling ratios defined on the X-axis, are greater than 0.
STATA code:
Acceptability Curves
. ipinputs cost qaly treat

Difference in cost: 999.99
SE, difference in cost: 324.18
Difference in effect: .01
SE, difference in effect: .0019
Correlation of differences: -.71
Degrees of freedom: 498

. accepti 999.99 324.18 .01 .0019 -.71 498

STATA Results:
Acceptability Curves

<table>
<thead>
<tr>
<th>No</th>
<th>% Accept</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>-7655559</td>
<td>0.03780</td>
<td>0.0756</td>
</tr>
<tr>
<td>-664168</td>
<td>0.03705</td>
<td>0.0751</td>
</tr>
<tr>
<td>-356187</td>
<td>0.03825</td>
<td>0.0747</td>
</tr>
<tr>
<td>18299</td>
<td>0.01000</td>
<td>0.0200</td>
</tr>
<tr>
<td>19340</td>
<td>0.01100</td>
<td>0.0230</td>
</tr>
<tr>
<td>24224</td>
<td>0.02100</td>
<td>0.0420</td>
</tr>
<tr>
<td>3897</td>
<td>0.03100</td>
<td>0.0620</td>
</tr>
<tr>
<td>221172</td>
<td>0.95900</td>
<td>0.0820</td>
</tr>
<tr>
<td>234664</td>
<td>0.96900</td>
<td>0.0620</td>
</tr>
<tr>
<td>254013</td>
<td>0.97900</td>
<td>0.0420</td>
</tr>
<tr>
<td>287956</td>
<td>0.98900</td>
<td>0.0220</td>
</tr>
<tr>
<td>2509921</td>
<td>0.96415</td>
<td>0.0717</td>
</tr>
<tr>
<td>3403528</td>
<td>0.96365</td>
<td>0.0727</td>
</tr>
<tr>
<td>6834078</td>
<td>0.96315</td>
<td>0.0737</td>
</tr>
</tbody>
</table>

Summary

http://www.uphs.upenn.edu/dgi/mhsr/stat%20cicer.htm