Sampling Uncertainty and Patient-Level Cost-Effectiveness Analysis (Part 1)

Henry A. Glick, Ph.D.
HCMG 901
October 15, 2015

Confidence About Value for the Cost

• Common goal of economic analysis: identify when we can be confident that a therapy is good value compared to another
• Threat to confidence: economic result observed in experiment may not reflect result in the population
  – Single sample drawn from population
• Referred to as sampling (or stochastic) uncertainty
• Methods for estimating sampling uncertainty for economic outcomes have much in common with methods used for clinical findings

Outline

• Describe methods for identifying when we can and cannot be confident about a therapy’s value
  – Acceptability curves
  – CI for ICER
  – CI for NMB
• Goal: demonstrate quantification and interpretation of sampling uncertainty
• Don’t focus on technical aspects of estimation
  – Stata programs available at: www.uphs.upenn.edu/dgimhsr
Joint Distribution of Cost and Effect

- Bivariate normal curves ($\Delta c$, $SE_c$, $\Delta q$, $SE_q$, $\rho$) (left)
- Bootstrap of patient level data (right)
- Second order Monte Carlo (decision analysis with variables represented as distributions) (right)

Joint Distribution of Cost and Effect (2)

- Mean cost difference, $4600, SE, 1803$
- Mean QALY difference, 0.2090, SE, 0.2430
- Correlation of difference, -0.045
- ICER Point estimate = 22,010 (4600 / 0.2090)

Information from the Plane

- Cost-effectiveness plane provides information about point estimates, confidence intervals and p-values for:
  - Difference in effect
  - Difference in cost
  - Cost-effectiveness analysis

In Which Experiment(s) is $\Delta Q$ Significant?

Red and blue (because all of their densities fall on one side of 0 on X-axis)

Black triangles not significantly different (because too large a density falls on each side of 0)
In Which Experiment(s) is \( \Delta C \) Significant?

Red and blue
(because all of their densities fall on one side of 0 on Y-axis)

Black triangles not significantly different
(because too large a density falls on each side of 0)

Value and the Cost-Effectiveness Plane

Ratios refer to old Rx
Ratios refer to new Rx
In Which Experiments Can We Be Confident of Value?

Red, blue, and cyan (because all of their densities fall on one side of WTP)

Black triangles not confident because large fractions of density fall on both sides of WTP

For red, blue and cyan, what confidence statements can we make?
What Can We Conclude About $\Delta C$, $\Delta Q$, Value?

What if there is No Single WTP?

Provide a statistic that allows decision makers to determine if – based on their own WTP – they can be confident of value.
Confidence Intervals

- Graphs above provide examples of 0 (for differences in means, including NMB), 1 (for OR and RR), or willingness to pay (W) (for CI for CER) falling either well inside or fully outside distribution of results
- Don’t typically require that results be fully outside distribution to conclude they differ from 0, 1, or W
  - Parametrically never happens
- Usual strategy: Identify a tolerance – e.g., 2.5% for 95% confidence – for the maximum fraction of results that can fall on one side of 0, 1, or W
- Conclude with 95% confidence that result excludes 0, 1, or W if 0, 1, or W fall outside 95% CI

Can be 95% confident of a difference for red and blue (because 0 on X-axis does not fall within the 95% CI)

Can’t be 95% confident of difference for black triangles (because 0 on X-axis falls within 95% CI)
Can be 95% confident of a difference for red and blue (because 0 on Y-axis does not fall within the 95% CI)

Can't be 95% confident of difference for black triangles (because 0 on Y-axis falls within 95% CI)
**Construction of CI for Difference, OR, or RR**

- **Common algorithm**
  - Develop distribution of difference (e.g., NMB, OR, or RR)
    - e.g., create empiric distribution from bootstrap or assume a distribution such as normal or log normal
  - Order distribution from smallest to largest
  - Construct 95% CI by identifying 2.5th and 97.5th percentiles of rank-ordered distribution
    - Either by counting (nonparametric) or estimating density (parametric)
  - Values of outcome that bound these percentiles represent the 95% confidence limits
- Works well for differences, OR, or RR

**Construction of CI for ICER**

- To use same algorithm for construction of CI for CER:
  - Develop joint distribution of difference in C and Q and calculate ratios
  - Order ratios from smallest to largest ("naïve ordering")
  - For 95% CI, identify 2.5th and 97.5th percentiles of rank-ordered ratios
  - Values of ratio that bound 2.5th and 97.5th percentiles represent 95% confidence limits
Pluses and Minuses of Ordering for CI for ICER

- In many cases, method works well for ICER
- But conditions when it fails well defined (e.g., Q: p>.05)
- CI for CER technically NOT an “order statistic”
  - Instead defined by lines through origin of CE plane that each exclude α/2% of joint distribution
  - (Independent of whether lower limit is a larger or smaller number than upper limit.) on CE plane, interval stretches counter-clockwise from lower (clockwise) limit to upper (counter-clockwise) limit

"Naïve" or "Smart" Ordering Can Work

All Replicates in Lower and Upper Right Quadrants

- Naïve ordering (smallest to largest ratio) works
Replicates in Upper Right and Left Quadrants

- Smart – but not naïve – ordering works

• Order from upper right to upper left quadrants

Replicates in 3 Quadrants

- Smart – but not naïve – ordering works

• Order from upper left to lower left to lower right quadrants

Will Smart Ordering Work?

Authors: ΔC, -1818, 95%CI -1540 to -1295

- Elsewhere in paper, authors’ report 79% of distribution below X-axis

- Too much density on both sides of X-axis to conclude ΔC significantly differs from 0
  
  (If 21% above X-Axis, p=0.58)

  95% CI cannot equal -1540 to -1295
Authors: \( \Delta Q, 0.074, 95\% \text{CI} 0.066 \text{ to } 0.082 \)

- At least 21\% of distribution is to left of Y-axis

Too much density on both sides of Y-axis to conclude \( \Delta Q \) significantly differs from 0

(If 21\%+ to left of Y-Axis, \( p>0.58 \))

95\% CI cannot equal 0.066 to 0.082

Authors: \( \Delta C / \Delta Q, -19155, 95\% \text{CI} -23,815 \text{ to } $2044 \)

If authors’ are correct that costs significantly reduced AND QALYs significantly increased, interval CI should indicate dominance (e.g., -23,815 to -$2044 \))
Based on scatter plot, cannot identify line through origin that excludes 2.5%

No 95% CI can be defined!

When \( p > 0.05 \) for \( \Delta Q \), lower limit of CI for CER can never be smaller number than upper limit

Conclusion

Something very wrong with either Brown et al.’s data plotted on CE plane or with Brown’s reported statistics

CE Plane does not confirm any statistical conclusions reported in their Table V

Sampling Uncertainty Issues

- # of methods available
  - Acceptability curve
  - CI for ICER
  - CI for NMB
- What is threshold, maximum willingness to pay?
  - Differs across jurisdictions
  - Differs within jurisdictions
- Should we be 95% confident?
  - A lot of economists claim not
First Example:
(Nonparametrically) all replicates on one side of Y-axis and naïve ordering works

Experiment 1
• Therapy A vs Therapy B (A – B)
• $\Delta$cost = 1000  (SE: 324.9, $p=0.002$)
• $\Delta$QALYs = 0.01  (SE: 0.001925, $p=0.000$)
• A is significantly more costly but significantly more effective
  – CER = 1000 / 0.01 = 100,000 / QALY gained
• 250 participants in each arm of the trial
• Correlation between difference in cost and effect is - 0.71015

Distribution of Results
“Counting” Method 1: Acceptability Curve

• Previously said usually identify a tolerance – e.g., 2.5% for 95% confidence – for the maximum fraction of results that can fall on one side of 0, 1, or W
• Can determine fraction that falls on one side by counting / estimating density of results distribution falling on each side of W
• Referred to as acceptability curve

Parametric or Nonparametric Construction

• Can be constructed nonparametrically or parametrically
• Nonparametric construction usually derived by counting bootstrap replicates
  – Does not assume bivariate normality
  – Particularly for acceptability curve and CI for CER, calculating fraction falling on each side of exact same lines through origin
• Parametric construction generally based on (rearrangement of) Fieller’s theorem formula for CI for CER (i.e., transformation of same formula)
  – Assumes difference in costs and effects distributed bivariate normal

Acceptability Curve

• Acceptability criterion defined on cost-effectiveness plane as a line passing through origin with slope equal to WTP
• Proportion of distribution of difference in cost and effect below and to right of line is “acceptable” (i.e., has positive NMB)
  – Proportion acceptable for one therapy = 1-proportion acceptable for alternative therapy
• Proportion above and to left of line is “unacceptable”
  – Proportion unacceptable for one therapy = 1-proportion unacceptable for alternative therapy
Constructing the Acceptability Curve

Acceptability Curve

% Acceptable, W = 245,200

Experiment 1

-0.005 0.000 0.005 0.010 0.015 0.020
difference in QALYs

-100 0 100 300

difference in costs

4000 Replicates; 100 = 2.5%

Experiment 1

-500 0 500 1000

difference in costs

4000 Replicates; 100 = 2.5%

Experiment 1

0 100000 200000 300000 400000

Willingness to Pay

0.00 0.25 0.50 0.75 1.00

Proportion

28,200 245,200

245,200: 3900, .975

179,600: 3600, .90

127,700: 2800, .70

76,800: 1200, .30

49,100: 400, .10

10,000: 16, .004
"Common" Conclusions from Acceptability Curves

<table>
<thead>
<tr>
<th>W</th>
<th>What is often said</th>
</tr>
</thead>
<tbody>
<tr>
<td>28,200</td>
<td>&quot;97.5% chance Rx A not good value&quot;</td>
</tr>
<tr>
<td>76,800</td>
<td>&quot;70% chance Rx A not good value&quot;</td>
</tr>
<tr>
<td>100,000</td>
<td>&quot;50% chance either therapy good value&quot;</td>
</tr>
<tr>
<td>127,700</td>
<td>&quot;70% chance Rx A good value&quot;</td>
</tr>
<tr>
<td>245,200</td>
<td>&quot;97.5% chance Rx A good value&quot;</td>
</tr>
</tbody>
</table>

- Common to adopt 1-tailed interpretation of acceptability curve
- Ignores fact that 50% – not 0% – represents no information

2-tailed Confidence Statements

- Two-tailed confidence statements
  - (For heights > 0.5) Confidence level: 
    \[(2 \times \text{Height}) - 1\]
  - e.g., if height of curve is 0.975 for \(W = 50,000\),
    \[(2 \times \cdot975) - 1 = 95\% \text{ confident that therapy is acceptabe} / \text{cost-effective}^\text{a}\]
  - (For heights < 0.5) Confidence level: 
    \[1 - (2 \times \text{Height})\]
  - e.g., if height of curve is 0.025 for \(W = 50,000\),
    "95\% confident alternative therapy is acceptable / cost-effective"^\text{b}

Additional Information

- a, curve height at intersection with x axis = 1-tailed p, \(\Delta C\)
- b, point-estimate for ICER
- c, curve height at \(\infty\) approaches 1-tailed p, \(\Delta Q\)
- D, \(W\) where curve height is \(y/2\) and \[1 - (y/2) = (1-y)% \text{ CL}\]
Observable Acceptability Curves for WTP > 0

Two Basic Acceptability Curve Patterns

Observed Shape Depends on Location of 0 Line
Observed Shape Depends on Location of 0 Line

Acceptability Curve

If Means and SE Were as Suggested...

"Counting" Method 2: CI for ICER

- Can also determine fraction of results that fall on one side of W by identifying slopes of 0, 1, or 2 lines through the origin that exclude α/2 of distribution
  - Identification by either counting/estimating distribution of results falling on each side of lines through origin
    - Slopes of lines that have 2.5% of distribution on 1 side and 97.5% on other define 95% CL for CER
    - Slopes of same lines define values of W for which acceptability curve has heights of 2.5% and 97.5%
- Referred to as confidence interval for cost-effectiveness ratio

Confidence Intervals for ICER

- Reiteration: common suggestion for constructing ratios:
  - Order ratios from smallest to largest
  - Identify 2.5th percentile (e.g., 26th observation out of 1000) and 97.5th percentile (e.g., 975th observation out of 1000)

Lower 95% Confidence Limit (Same as Slide 49)

Green: (W * ΔQ) - ΔC > 0

4000 Replicates; 100 = 2.5%
CI for ICER Not Tangent to 95% Ellipse

Tangencies With 85.5% Confidence Ellipse

Confidences Statements for CI for CER

- Confident of value if:
  - LL < UL < W (confident of good value)
  - W < LL < UL (confident of bad value)
  - UL < W < LL (confident of good value if PE<W; confident of bad value if PE>W)

- Not confident of value if:
  - CI is undefined
  - LL < W < UL
  - W < UL < LL
  - UL < LL < W
Confidences Statements for Current Experiment

- Can be confident of value when W not included in confidence interval
- When lower limit is a smaller number than upper limit
  - Interval ranges between lower and upper limit
    - 28,200 to 245,200
  - Confident of value if WTP is either smaller than lower limit or greater than upper limit
    - Confident of bad value if WTP < 28,200
      - Because at least 97.5% of samples have ratios greater than 28,200
    - Confident of good value if WTP > 245,200
      - Because at least 97.5% of samples have ratios less than 245,200

“Common” Conclusions, CI for CER

<table>
<thead>
<tr>
<th>W</th>
<th>What is often said</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;28,200</td>
<td>“95% confident Rx A not good value”</td>
</tr>
<tr>
<td>76,800</td>
<td>Can’t be 95% confident value of Rxs differs</td>
</tr>
<tr>
<td>100,000</td>
<td>Can’t be 95% confident value of Rxs differs</td>
</tr>
<tr>
<td>127,700</td>
<td>Can’t be 95% confident value of Rxs differs</td>
</tr>
<tr>
<td>&gt;245,200</td>
<td>“95% confident Rx A good value”</td>
</tr>
</tbody>
</table>

- Usually employ 2-tailed interpretation of CI for CER

“Counting” Method 3: CI for NMB

- Finally, can determine if W falls inside or outside distribution by constructing distribution of NMB for specified W and identifying whether 0 falls within interval
- As for any difference, construct interval by ordering distribution of NMB and identifying values of NMB that define the 2.5th and 97.5th percentiles
- In contrast with acceptability curve and CI for CER, not (typically) defining lines through the origin of CE plane
  - But lines through origin have same meaning as for acceptability curves and CI for CER
NMB Recap

\[ \text{NMB} = (W \Delta Q) - \Delta C \]

- For a WTP of 50,000, NMB for experiment 1:
  \[(50,000 \times 0.01) - 1000 = -500\]
- Study result a difference in means of net benefits, not a ratio of means, and is always defined (i.e., no odd statistical properties like ratio) and continuous
- Unlike cost-effectiveness ratio, standard error of net benefits is always defined
- Given not all decision making bodies have agreed upon maximum willingness to pay, routinely estimate net benefit over range of policy relevant values of willingness to pay

### Net Benefit Graphically

- For a given W, can calculate value of NMB for every point on CE plane
- Formula: \( \text{NMB} = W \Delta Q - \Delta C \)
- If \( W = 50,000 \), the following points all fall on same line and have same value of NMB

<table>
<thead>
<tr>
<th>( \Delta C )</th>
<th>( \Delta Q )</th>
<th>NMB</th>
</tr>
</thead>
<tbody>
<tr>
<td>-500</td>
<td>0</td>
<td>(50,000 * 0) - (-500) = 500</td>
</tr>
<tr>
<td>49,500</td>
<td>1</td>
<td>(50,000 * 1) - (49,500) = 500</td>
</tr>
<tr>
<td>99,500</td>
<td>2</td>
<td>(50,000 * 2) - (99,500) = 500</td>
</tr>
<tr>
<td>149,500</td>
<td>3</td>
<td>(50,000 * 3) - (149,500) = 500</td>
</tr>
</tbody>
</table>

- Value of NMB for lines with 50,000 slope = -intercept
  - e.g., \(-(-500) = 500\)

### Net Benefit Graphically (2)

- Defined on cost effectiveness plane using a family of lines
- Each line represents a single value of NMB and equals -intercept (because when \( \Delta Q = 0 \), \( W \Delta Q \) drops out of equation
- Slope of all lines equal to \( W \)
- 95% CI for NMB defined by identifying 2 NMB lines that each omit 2.5% of distribution
Constructing CI for NMB, WTP=28.2K

• We’ve seen line defining upper CI before!

Constructing CI for NMB, WTP=100K

Constructing CI for NMB, WTP=245.2K

• We’ve seen line defining lower CI before!
Confidences Statements for CI for NMB

- If both confidence limits negative, 95% confident therapy is bad value
  - i.e., for values of WTP ≤ 28,200
- If both confidence limits positive, 95% confident therapy is good value
  - i.e., for values of WTP ≥ 245,200
- If one confidence limit positive and one negative, cannot be 95% confident value of 2 therapies differs
  - i.e., for values of WTP > 28,200 and < 245,200
Similarities and Differences

• For magnitude estimation for a single value of W, NMB provides information that is NOT shared by acceptability curve or CI for ICER
  – i.e., generally isn’t identifying lines through origin as are acceptability curve and CI for ICER
• NMB provides information that IS shared for meta-question, “For what values of W can we or can’t we be confident therapy is good value compared to another?”
  – Nonparametrically, identification of whether CI for NMB includes or excludes 0 relies on same lines through origin as acceptability curve and CI for ICER
  – Parametrically, CI for NMB and acceptability curve use transformation of Fieller’s theorem equation for CI for ICER

Acceptability & CI for CER

• Acceptability curve plots confidence intervals for the cost-effectiveness ratio
  – e.g., the value of WTP where the height of the acceptability curve equals 0.025 and/or 0.975 represent the 95% confidence limits for the cost-effectiveness ratio
  • In current example, 95% CL = 28,200 and 245,200
Acceptability & CI for NMB

- Acceptability curves also report values of WTP for which one of NMB confidence limits equals 0
  - e.g. if we calculate NMB using values of WTP where height of acceptability curve equals 0.025 and/or 0.975, one of 95% confidence limits for NMB will equal 0
  - If we calculate NMB using values of WTP where height of the acceptability curve equals 0.25 and/or 0.75, one of 50% confidence limits for NMB will equal 0

Review of Results for Experiment 1

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

- Confidence frontier for NMB
  CI intersect decision threshold (0) at 28,200 and 245,200

"Pattern 1" Findings

- Refer to findings like experiment 1’s as pattern 1 findings
- Occur when difference in effect is significant
- Know we are observing pattern 1 finding when:
  - Confidence interval for cost-effectiveness ratio excludes Y axis (i.e., LL < PE < UL)
  - Both NMB confidence limits curves intersect decision threshold (0) once
  - Acceptability curve intersects horizontal lines drawn at both 0.025 and 0.975
Region of Acceptability Related to Pattern 1

- For this curve, widest pattern 1 finding is 78.81% CI

3 Ranges of WTP for Pattern 1 Findings

- In cases where some of boundaries between regions occur at negative values of willingness to pay, may not always observe all 3 regions on acceptability curve or NMB plot

Confidence vs Value of Information

- Requiring statistical significance (i.e., confidence) prior to the adoption of a new therapy that maximizes NMB runs counter to expected utility theory
  - Said to impose opportunity costs on patients
Quality of the Evidence

- Rejection of significance tests for cost-effectiveness ratios/NMB does not imply that decisions should be made using point estimates alone
  - Particularly if a decision can be made to collect more information
- "Value of information" represents difference in expected value of outcome given current decision and expected value of outcome that would result if we had perfect information (EVPI)
  - Determined based on probability decision is wrong and costs of wrong decision if it occurs

Per-Person Expected Value of Perfect Information

\[
EVPI_{\text{PP}} = \min \{ V_j, V_k \}
\]

where

\[
V_j = \forall_j NMB_j > 0: \frac{N}{N_j} \sum NMB_j
\]

\[
V_k = \forall_k NMB_k < 0: \text{abs} \left( \frac{N}{N_k} \sum NMB_k \right)
\]

Calculating Per-Person EVPI, 28,200

4000 Replicates; 100 = 2.5%
Calculating Per-Person EVPI, 70,000

- EVPI = 60.38
- N = 3019
- Cond Mean = 246.207
- Mean = 60.38

Calculating Per-Person EVPI, 100,000

- EVPI = 191.55
- N = 1988
- Cond Mean = 380.82
- Mean = 191.55

Calculating Per-Person EVPI, 150,000

- EVPI = 58.38
- N = 765
- Cond Mean = 690.44
- Mean = 58.39
Calculating Per Person EVPI, 245,200

Experiment 1

N = 86
Cond Mean = -320.83
Mean = -6.90

N = 3914
Cond Mean = 1490.96
Mean = 1458.91

EVPI = 6.90

Per-Person EVPI Graph

Two Basic Parametric EVPI Curve Patterns
Per Person EVPI

- Can be large because either there is a lot of uncertainty or because cost of mistakes (i.e., W) is large
  - e.g., might already be very certain (e.g., 99.99% confident), but if cost of mistakes is extremely high might want even greater certainty
- Can be small because either there is a lot of certainty or because costs of mistakes are small
  - e.g., might be very uncertain (e.g., only 10% confident), but if the cost of mistakes is extremely low, might not need greater certainty

Total EVPI

- Total EVPI = N * EVPI_{pp}
  - where N = number of people for whom treatment is indicated
- Net EVPI = Total EVPI - Cost of gathering additional information
- Given additional research is unlikely to yield perfect information, net EVPI at best provides upper bound on how much additional research should be funded
  - Need to focus on value of expected change in information
- Can also be used to evaluate particular uncertainties for which research is needed: expected value of perfect information for a parameter (EVPPI)

Potential VOI Caveat

- “…value of information methods require consideration of the totality of the evidence base.....”
- “…may not therefore be appropriate to simply base value of information estimates on the sampling variability from a single study where other studies exist.”
**Asserted EVPI Advantages**

- Quantitative measure of when we have enough information to make a decision
- Avoids inference
- Avoids temptation to use 'need for evidence' to delay decision making
- Recognises information gathering is not costless
- Can distinguish value of different types of information which might guide study design

**Rejection of Inference Applies to Everything**

- Nothing different about economic decisions and other decisions
  - If we adopt an EVPI decision criterion – i.e., reject an inference criterion – for making economic decisions about therapies, should do same for other decisions
    - FDA should stop requiring significance for drug adoption decisions
  - Economics (theoretical) vs medical (life and death) decision making?
- Significance testing may be transactionally efficient
  - Assuming there are costs of switching therapies, interpret significance tests as a mechanism for limiting switching and reducing these costs
  - Can build these (and other costs) into EVPI

**Acceptability Curves When More Than 2 Therapies**
Acceptability Curves When More than 2 Therapies

- When comparing more than 2 therapies, common to graph one curve per therapy with curves representing proportion of time therapy is best value (e.g., for Rx1: fraction that Rx1 > Rx2 AND Rx1 > Rx3)

Violation of Independence of Irrelevant Alternatives

- Best criterion violates Independence of irrelevant alternatives (IIA)
  - IIA a ubiquitous assumption in welfare economics / social choice theory
- IIA: Choice between alternatives x and y depends on preferences for x and y only (and is not affected by preferences for z)
  - e.g., if Rx 1 is chosen over Rx2 and Rx3, Rx1 must be both better than Rx2 and better than Rx3
- Focusing solely on fraction of time a therapy is best throws away information about the preference between 2 therapies (e.g., x and y) when a third therapy (e.g., z) is best

Fraction of Time Best

- Suppose making choice for 7 people between 3 mutually exclusive modes of travel. Choose single mode for all 7
  - buses (B), cabs (C), or walking trails (W)
- Suppose most preferred choices are as follows:

<table>
<thead>
<tr>
<th>Obs</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pref</td>
<td>W</td>
<td>W</td>
<td>W</td>
<td>B</td>
<td>B</td>
<td>C</td>
<td>C</td>
</tr>
</tbody>
</table>

- If basing decision solely on first preferences, heights of “multi-way” acceptability curves would equal:
  - walk, 3/7; bus, 2/7; cab, 2/7
  - i.e., walking is “best”
Fraction of Time Better Value
• Suppose people who prefer cabs or buses least prefer walking; people who prefer walking least prefer cabs.
• Rank-ordered preferences would be:

<table>
<thead>
<tr>
<th>Obs</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>W W W B B C C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd</td>
<td>B B B C C B B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd</td>
<td>C C C W W W W</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

• B is preferred to both W (4/7) and C (6/7)
• C is preferred to W (4/7), but not B (2/7)
• W is least preferred (3/7 against both B and C)

Summary
• According to best rule, W is best and indifferent between B and C
• If instead consider complete set of preferences:
  – B preferred to both W and C
  – C preferred to W

What’s the Alternative
• Return to use of multiple pairwise comparisons
  – Strong tradition in economic choice theory, e.g., basis of Arrow impossibility theorem
• Analog to “best” algorithm is to select therapy that in pair-wise comparison is better than all other therapies
  – ??? Significantly better ???
What’s the Alternative (2)

- For each value of WTP plot lowest percentage acceptable against all other therapies
  - If B better than W 4/7 of time and better than C 5/7, height of B curve = 4/7
  - If C better than B 3/7 of time and better than W 4/7, height of C curve = 3/7
  - If W better than both B and C 3/7 of time, height of W curve = 3/7
- Best alternative has highest curve
  - i.e., select B because it is better than other 2 options at least 4/7 of the time
- Note, sum of heights of curves >1

Example For Single Value of W

- Assume 4 Rx, 1–4; WTP = 1900

  Fraction of times NMB for Rx (column 1) exceeds NMB for other Rxs

<table>
<thead>
<tr>
<th></th>
<th>Rx 1</th>
<th>Rx 2</th>
<th>Rx 3</th>
<th>Rx 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rx 1</td>
<td>--</td>
<td>0.188</td>
<td>0.574</td>
<td>0.737</td>
</tr>
<tr>
<td>Rx 2</td>
<td>0.812</td>
<td>--</td>
<td>0.892</td>
<td>0.921</td>
</tr>
<tr>
<td>Rx 3</td>
<td>0.426</td>
<td>0.108</td>
<td>--</td>
<td>0.782</td>
</tr>
<tr>
<td>Rx 4</td>
<td>0.262</td>
<td>0.079</td>
<td>0.218</td>
<td>--</td>
</tr>
</tbody>
</table>

Simulation

<table>
<thead>
<tr>
<th></th>
<th>Rx1</th>
<th>Rx2</th>
<th>Rx3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5000)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(10,000)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(16000)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QALY</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>-0.9</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td>-0.9</td>
<td>0.9</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1</td>
<td>0.9</td>
<td>-0.9</td>
<td>-0.9</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q2</td>
<td>-0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>-0.9</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td>-0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>-0.9</td>
<td>0.9</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Multi-way Curve Simulation, Best Curves

- Rx1 “best” for W between 0 and 97,500 (red dashed line)
- Rx2 never “best” (green dashed line)
- Rx3 “best” for W greater than 97,500 (blue dashed line)

Multi-way Curve Simulation, Better Frontier

- While Rx2 never “best”, between 53K and 60K it is better (green solid line) than both Rx1 and Rx3

Multi-way Curve Simulation, Better Frontier (2)

- While Rx1 “best” for W up to 97,500 (red dashed line), Rx3 (solid blue line) better than both Rx1 and Rx2 for W>60K
In Usual Practice...

- While example suggests differences can be dramatic, for typical kinds of results, 2 approaches probably have similar recommendations over wide ranges of $W$
  - However:
    - Can observe differences around boundaries between therapies
    - Compared to “Best” algorithm, “Better” algorithm yields more appropriate measure of magnitude of probability therapy is better than alternative

Typical Kinds of Results, Best Curves

Typical Kinds of Results, Better Curves