Sampling Uncertainty / Sample Size for Cost-Effectiveness Analysis

Cost-Effectiveness Evaluation in Addiction Treatment Clinical Trials

Henry Glick
University of Pennsylvania

www.uphs.upenn.edu/dgimhsr

Good Value for the Cost

• A common goal of an economic analysis is to identify when we can be confident that one therapy is good value compared to another
• A threat to such confidence arises because the economic results from experiments are derived from single samples and thus may not truly reflect the result in the population
• This form of uncertainty is referred to as sampling (or stochastic) uncertainty
• Methods for estimating sampling uncertainty for economic results have much in common with methods for estimating sampling uncertainty for clinical results, but there are also differences

Sampling Uncertainty and Clinical Outcomes

• Confidence statements about clinical outcomes are based on whether or not the confidence interval for the clinical outcome includes what I refer to as its decision threshold
  – e.g., for odds ratios and relative risks, the decision threshold equals 1.0
  – for risk differences or differences in outcomes such as cognition, the decision threshold equals 0
Sampling Uncertainty and Economic Outcomes

- Confidence statements about economic outcomes are also based on whether or not the confidence interval for the economic outcome includes its decision threshold.
- One of the differences between clinical and economic confidence statements is that more methods are available for assessing confidence of economic conclusions:
  - Confidence intervals for cost-effectiveness ratios
  - Confidence intervals for net monetary benefits
  - Acceptability curve
- All 3 methods are transformations of one another and all 3 yield the same confidence statements:
  - For simplicity, I focus on confidence intervals for cost-effectiveness ratios.

Sampling Uncertainty and Economic Outcomes (2)

- When expressing sampling uncertainty by use of confidence intervals for cost-effectiveness ratios:
  - We estimate the point estimate and 95% confidence interval for the cost-effectiveness ratio.
  - We determine whether we can be 95% confident that a therapy is good value by comparing the confidence interval to the economic decision threshold: Maximum Willingness to Pay.

Economic Inferences

- We draw economic inferences as follows:
  - If maximum willingness to pay is included within the confidence interval, we CANNOT be confident that the two therapies differ in their cost-effectiveness.
  - If maximum willingness to pay is excluded from / outside the interval, and the point estimate is less than the maximum willingness to pay, we CAN be 95% confident that the therapy with the larger point estimate for effectiveness is cost-effective.
  - If maximum willingness to pay is excluded from / outside the interval, and the point estimate is greater than the maximum willingness to pay, we CAN be 95% confident that the therapy with the smaller point estimate for effectiveness is cost-effective.
Maximum Willingness to Pay

- Maximum willingness to pay is to cost-effectiveness ratios as 1 is to odds ratios or relative risks:
  - It is the decision threshold we are trying to rule out if we are to be confident about value
- A second difference between clinical and economic confidence statements is that while we rarely consider comparing an odds ratio or a relative risk to a decision threshold other than 1, we often choose the maximum willingness to pay we want to rule out, because in countries like the US, there is no consensus on what its value is
  - It is expected that the maximum willingness to pay can differ among decision makers, particularly in different decision making jurisdictions

What is the Maximum Willingness to Pay?

- Traditionally, therapies with cost-effectiveness ratios less than $40,000 to $50,000 per quality-adjusted life-year (QALY) saved have been considered good value
- Little analytic attention has been given to identifying a “social” maximum willingness to pay
  - Even countries that claim to know what their willingness to pay is have usually derived it based on past administrative decisions rather than by systematic elicitation
- There has been a growing debate about whether the maximum willingness to pay in the U.S. has increased (e.g., at a minimum to $100,000 per QALY)
- The debate about what our maximum willingness to pay is complicates our ability to generate convincing information about the value of new therapies

Sample Size for Cost-Effectiveness Analysis

- A goal of sample size and power calculation for cost-effectiveness analysis is to identify the likelihood that an experiment will allow us to be confident that a therapy is good or bad value when we adopt a particular willingness to pay
  - e.g., we may expect that the point estimate for the cost-effectiveness ratio will be 20,000 per QALY and want to design an experiment that will provide an 80% chance (i.e., power) to be 95% confident that the therapy is good value when we are willing to pay at most 75,000 per QALY
Sample Size Formula, Common SDs

- Sample size is calculated by use of the following formula:

\[ n = \frac{2 (z_\alpha + z_\beta)^2 (\text{sd}_C^2 + \text{sd}_Q^2) - (2 W \rho \text{sd}_C \text{sd}_Q)}{(WQ - C)^2} \]

where \( n \) = sample size/group; \( z_\alpha \) and \( z_\beta \) = z-statistics for \( \alpha \) (e.g., 1.96) and \( \beta \) (e.g., 0.84) errors; \( \text{sd} \) = standard deviation for cost (\( \text{sd}_C \)) and effect (\( \text{sd}_Q \)); \( W \) = maximum willingness to pay we wish to rule out; \( \rho \) = correlation of the difference in cost (\( C \)) and effect (\( Q \))

Correlation

- When increasing effects are associated with decreasing costs, a therapy is characterized by a negative (win/win) correlation between the difference in cost and effect  
  – e.g., asthma care
- When increasing effects are associated with increasing costs, a therapy is characterized by a positive (win/lose) correlation between the difference in cost and effect  
  – e.g., life-saving care
- All else equal, fewer patients need to be enrolled when therapies are characterized by a positive correlation than when therapies are characterized by negative correlation

Willingness to Pay

- Moving willingness to pay “nearer to” or “further away from” the expected point estimate of the cost-effectiveness ratio increases or reduces the sample size we need to be confident of value  
  – Caution: “Nearer” and “further away” are not measured on the real number line

Implication: Sample size need not always decrease as willingness to pay increases
Sample Size Tables

- When we have input into sample size decisions, we commonly calculate the sample size per group needed to rule out different values of willingness to pay

<table>
<thead>
<tr>
<th>W</th>
<th>Sample Size/G</th>
<th>ΔC=250; ΔQ=0.05; sd_c=5,000; sd_q=0.2; ρ=-0.10; α=.05; β=.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>20,000</td>
<td>1256</td>
<td></td>
</tr>
<tr>
<td>30,000</td>
<td>673</td>
<td></td>
</tr>
<tr>
<td>50,000</td>
<td>419</td>
<td></td>
</tr>
<tr>
<td>75,000</td>
<td>340</td>
<td></td>
</tr>
<tr>
<td>100,000</td>
<td>310</td>
<td></td>
</tr>
</tbody>
</table>

Sample Size Often More Sensitive to SD_q than to SD_c

\[
\frac{2(z + z_{\beta})^2 \left( sd_c^2 + (W^2 sd_q^2) \right)}{\Delta NMB^2} - (W^2 \rho (2 sd_c^2)^2 + (2 sd_q^2)^2) \rho^2
\]

- Except for the enlarged terms above, the sample size formula is generally symmetric for the SDs of cost and effect
- Changes in the square of the QALY SD are weighted by the square of W; changes in the square of the cost SD are unweighted
- **When W is substantially greater than SD for cost**, percentage changes in the QALY SD will have a substantially greater effect on sample size than will equivalent percentage changes in cost SD

Power Formula, Common SDs

\[
z_{\beta} = \sqrt{\frac{n \times \Delta NMB^2}{2 (sd_c^2 + (W^2 sd_q^2)) - (2 W \rho sd_c sd_q)}} \cdot z_{\alpha}
\]

- Unlike the sample size equation, where the result = N, result of formula is \( z_{\beta} \), not power
- To estimate power, use the normal distribution table to identify the fraction of the tail that is to the left of \( z_{\beta} \)
Power Tables

- When sample size per group is fixed, we commonly calculate the power for multiple values of $W$

<table>
<thead>
<tr>
<th>$W$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>20,000</td>
<td>.308</td>
</tr>
<tr>
<td>30,000</td>
<td>.512</td>
</tr>
<tr>
<td>50,000</td>
<td>.714</td>
</tr>
<tr>
<td>75,000</td>
<td>.801</td>
</tr>
<tr>
<td>100,000</td>
<td>.835</td>
</tr>
</tbody>
</table>

$\Delta C=250; \Delta Q=.05; sd_c=5,000; sd_e=.2; \rho=.1; \alpha=.05; $  
Sample size per group=340

Confidence about $\Delta C$ and $\Delta Q$

- Before the development of the literature about confidence intervals for cost-effectiveness ratios, it was generally accepted that if neither $\Delta C$ nor $\Delta Q$ were significant, then there was no reason to assess cost-effectiveness
- But $\Delta C$ and $\Delta Q$ can be considered a joint outcome just as differences in nonfatal CVD events and all cause mortality are often combined into a joint outcome
- In the same way that a trial can be “clinically” successful when the differences in the individual outcomes are not significant but the difference in the joint outcome is, a trial can be “economically” successful when neither the difference in cost nor the difference in effect is significant, but the joint outcome of the difference in costs and effects is

Economic Vs Clinical Sample Sizes

- While the sample size required to answer economic questions tends to be larger than the sample size required to answer clinical questions, it need not be for all experiments
- As previously mentioned, the sample sized needed to answer the economic question is more likely to be smaller than that needed to answer the clinical question when:
  - The correlation of cost and effect is positive
  - We have more power for the joint outcome of difference in cost and effect than we do for either outcome alone
Summary

- Goal of sample size and power calculation for cost-effectiveness analysis is to identify the likelihood that an experiment will allow us to be confident that a therapy is good or bad value when we adopt a particular willingness to pay
- Sample size and power depend on the difference in cost and effect, the SD of cost and effect, the correlation of the difference, our willingness to pay, and our target confidence level
- When WTP is substantially greater than the SD for cost, changes in the SD for effect generally have greater impact on sample size than do changes in the SD for cost

Summary (cont.)

- Positive correlations generally decrease sample size / increase power
- Sample size required to answer economic questions is often larger than that needed to answer clinical questions
  - but sample size required for the comparison of cost and effect can be smaller than that required for the comparison of cost or effect alone