Sampling Uncertainty for Cost-Effectiveness Analysis: Where We are Going Wrong

Henry Glick
University of Pennsylvania
AcademyHealth
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

- Confidence statements about clinical outcomes are based on whether or not the confidence interval for odds ratios and relative risks include 1 or whether or not the confidence interval for risk differences include 0
  - Refer to 1 and 0 as “decision thresholds”
- Confidence statements about economic outcomes are also based on whether or not the confidence interval for the economic outcome includes its decision threshold
  - For cost-effectiveness, W, our maximum willingness to pay
  - For NMB calculated by use of W, 0
  - For the acceptability curve, at W is the acceptability curve less than α/2 or greater than 1-(α/2)

Economic Inferences

- We draw cost-effectiveness 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

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

Measuring Economic Sampling Uncertainty

- In 1994, O’Brien et al. published “In search of power and significance: issues in the design and analysis of stochastic cost effectiveness studies in health care” (Med Care. 1994;32:150-63)
- Since then, our field has come a long way:
  - Developed 3 dependably accurate methods for evaluating sampling uncertainty for cost-effectiveness analysis
  - Understand how to interpret the results of these methods
  - Introduced probabilistic sensitivity analysis, a method for quantifying sampling uncertainty, into decision analysis
Common Mistakes in the Literature

- But assessment of sampling uncertainty still has not reached a mature stage where it routinely is correctly reported in published cost-effectiveness analyses and where everyone understands it
- What are some of the common mistakes?
  - Failing to report sampling uncertainty
  - “Point-estimate” decision making, by which I mean reporting sampling uncertainty and then ignoring it when making adoption recommendations
  - Mistakenly overstating the confidence in their results
  - Not knowing how to calculate the CI

1) Failing to Report Sampling Uncertainty

- Although we know how to report sampling uncertainty, many investigators are still not doing so
- Doshi and colleagues* have reported that of the economic assessments in trials published in 2003 that reported both costs and effects, only 57% (24 of 42) reported sampling uncertainty for cost-effectiveness


Failing to Report Sampling Uncertainty (2)

- An additional 38 studies that reported only cost differences may have done so because they found no significant difference in effects (i.e., cost-minimization)
  - This historical recommendation is no longer appropriate, because it doesn’t distinguish underpowered trials from adequately powered trials that demonstrate no difference

Inappropriate Cost-Minimization and Effect Maximization

Would have needed 1497/group to have 80% power to be 95% confident for a willingness to pay of 50,000

Would have needed 1197/group to have 80% power to be 95% confident for a willingness to pay of 50,000

2) “Point Estimate” Decision Making

- Many investigators calculate and report sampling uncertainty, but then ignore it when they make recommendations about value
- For example, in one paper the authors found that the therapy under investigation:
  - Significantly increased direct costs (€449, p=0.01)
  - Had no significant effect on total direct and indirect costs (€332, p=0.2)
  - Had no significant effect on life years gained (0.04, p=0.4)
- The resulting cost-effectiveness ratios were €10,993 (Direct) and €8127 (Total)

Reported Acceptability Curve
“The predicted cost-effectiveness ratios were well below the threshold values generally considered cost-effective…[The therapy] appeared to be cost-effective….”

Point Estimate Decision Making (2)

- Depending on whether we consider direct or total cost, the therapy significantly increased cost or had no effect on cost.
- The therapy had no significant effect on effect.
- Under the old decision rules, we’d have said either the therapies don’t differ or we’d reject the new therapy.
- The acceptability curve approaches 80%; thus, we can be at most 60% confident the therapy is good value even if we are willing to pay as much as ε144,000 per life year gained.
- The authors ignored all of these findings, and instead made their overly optimistic recommendation based on the magnitude of the point estimate.

3) Mistakenly Overstate the Confidence in Their Results

- When the reported confidence limits don’t allow us to be confident of good value, investigators commonly use language that appears like a confidence statement, but which overstates normal confidence levels.

Pressure Mattresses for Prevention of Pressure Ulcers

- The authors of a study that compared alternating pressure mattresses vs alternating pressure overlays for the prevention of pressure ulcers found:
  - “The differences in health benefits and total costs…were not statistically significant.”
  - They did not report the fact, but the 95% CI for the CER was undefined.
  - They added the following: "however a cost-effectiveness acceptability curve indicated … [that pressure mattresses] were associated with an 80% probability of being cost saving."
  - They concluded: “Alternating pressure mattresses…are more likely to be cost effective…”

Statistical Restatement

- It probably would have been more obvious that this was simply an underpowered trial that did not allow us to differentiate between the therapies if they’d used common statistical language to report their result:
  - “The differences in health benefits and total costs…were not statistically significant,” and p=0.4 that pressure mattresses saved money.

4) Not Knowing How to Calculate the CI

- A number of investigators don’t know how to calculate confidence intervals for the hard cases (and probably wouldn’t know how to interpret the results if they did).
- A recent study reported the following results:

<table>
<thead>
<tr>
<th>Therapy</th>
<th>Cost (SD)</th>
<th>Heroin-Free Days (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Therapy 1</td>
<td>1415 (968)</td>
<td>6.84 (10.9)</td>
</tr>
<tr>
<td>Therapy 2</td>
<td>1729 (1291)</td>
<td>5.27 (9.96)</td>
</tr>
<tr>
<td>Difference</td>
<td>-314</td>
<td>1.57</td>
</tr>
</tbody>
</table>
CI for the CER

- While generally considered inappropriate for cost-effectiveness analysis, the authors used a test of medians and reported no significant difference in either cost or effect.
  - They failed to note that the difference in mean cost was significant.
- They tried to use a bootstrap percentile method and reported that the confidence interval for the cost-effectiveness ratio ranged from -2069 to 1809.
  - Interpretation: We can be confident of value so long as our WTP per heroin-free day is greater than 1809.
- This result is unlikely given the lack of significance of the difference in either mean or median heroin-free days.

The Authors' Mistake

- The authors ignored the fact that the bootstrapped distribution of the difference in cost and effect crossed the y axis and used the naïve ordering of most negative to most positive.

The Authors' Mistake (2)

- Had they plotted the data on the cost-effectiveness plane, they would have seen that they had constructed an incorrect interval.

Correct Ordering

Interpretation: Good value if WTP < 650/H-F day

Summary

- As a field, we seem to feel that making decisions with confidence is an unnecessary burden.
  - Unlike other fields, “We have to make decisions.”
- As many as 40% of studies aren’t reporting measures of sampling uncertainty for the difference in cost and effect.
- Among studies that report them, substantial numbers:
  - Ignore them when they discuss the adoption decision.
  - Use 2-tailed confidence for the assessment of the difference in cost and difference in effect, but without comment switch to 1-tailed confidence for cost-effectiveness.
  - Report mistaken measures of sampling uncertainty.
- We still have substantial room for improvement!