COST EFFECTIVENESS ANALYSIS AND NEW METHODOLOGICAL APPROACHES

Henry Glick
University of Pennsylvania

www.uphs.upenn.edu/dgimhsr

March 29, 2004

BACKGROUND

• In response to increasing health care costs, regulators, providers, payers, and patients have begun to question the value for the cost of individual medical therapies

• One of the growing trends in this evaluation has been the incorporation of economic evaluations within randomized controlled trials of medical therapies

• Most frequently these evaluations are incorporated into the drug development process
  - Prior to approval in phases II and III: during which a drug's safety and efficacy are evaluated prior to regulatory approval
  - After approval in phase IV

• To a lesser extent, they are conducted within trials of other medical therapies (e.g., surgical procedures, behavioral interventions, etc.)

• A number of national regulatory bodies have indicated they are comfortable with economic evidence derived from trials, although they may ask for it to be “tailored” to their country

DEMONSTRATING THAT A TECHNOLOGY IS “GOOD VALUE FOR THE COST” (I)

• 10-15 years ago, most likely would have supported this message by use of a decision analytic model such as a decision tree or a Markov model
  - Few if any of the economic results would have been directly observed
    * e.g., the clinical evidence about the therapy would be that it reduced blood pressure
    * These data often would be combined with epidemiologic data relating blood pressure to death and disability to project the likely impact of therapy on outcomes, and thus a therapy's economic impact
  - Reported results would have included point estimates of incremental costs, outcomes, and comparison of costs and effects from a "principal" or "base-case" analysis as well as the results of sensitivity analysis

"GOOD VALUE FOR THE COST" (II)

• In the mid-90’s, for a cutting edge evaluation, the message would have been supported by use of data from a randomized trial that included economic outcomes as primary or secondary endpoints of the trial
  - Short-term economic impacts of the therapy would be directly observed; longer term impacts potentially would be projected by use of decision analysis
  - Reported results would have included point estimates and confidence intervals for estimates of incremental costs and outcomes as well as point estimates and results of sensitivity analysis for the comparison of costs and effects
"GOOD VALUE FOR THE COST" (III)

- By the end of the 90's, for a cutting edge evaluation, the message would have been supported by use of an evaluation from a randomized trial that included economic outcomes as primary or secondary endpoints of the trial.

- Short-term economic impacts of the therapy would be directly observed; longer term impacts potentially would be projected by use of decision analysis.

- Reported results would include point estimates and confidence intervals for estimates of incremental costs, outcomes, and the comparison of costs and effects.

- The impact of sensitivity analysis on the comparison of costs and effects would be judged by its impact on both the point estimates and the confidence intervals of the ratios.

EXAMPLE OF TYPICAL LATE '90'S RESULTS

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Point Estimate</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incremental Costs</td>
<td>-713</td>
<td>-2123 to 783</td>
</tr>
<tr>
<td>Incremental QALYS</td>
<td>0.13</td>
<td>0.07 to 0.18</td>
</tr>
</tbody>
</table>

Cost-Effectiveness Ratios

<table>
<thead>
<tr>
<th>Principal Analysis</th>
<th>Dominates</th>
<th>Dom to 6650</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survival benefit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-33%</td>
<td>Dominates</td>
<td>Dom to 9050</td>
</tr>
<tr>
<td>+33%</td>
<td>Dominates</td>
<td>Dom to 5800</td>
</tr>
<tr>
<td>Hospitalization costs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-50%</td>
<td>Dominates</td>
<td>Dom to 5300</td>
</tr>
<tr>
<td>+50%</td>
<td>Dominates</td>
<td>Dom to 8400</td>
</tr>
<tr>
<td>Drug costs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-50%</td>
<td>Dominates</td>
<td>Dom to 4850</td>
</tr>
<tr>
<td>+50%</td>
<td>Dominates</td>
<td>Dom to 8750</td>
</tr>
<tr>
<td>Discount rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0%</td>
<td>Dominates</td>
<td>Dom to 6350</td>
</tr>
<tr>
<td>7%</td>
<td>Dominates</td>
<td>Dom to 7000</td>
</tr>
</tbody>
</table>
CURRENT ISSUES

- In the last decade we have addressed, and principally solved, the issues related to stochastic uncertainty in the comparison of costs and effects.

- Attention has now turned to a number of other issues that need attention, if we want our analyses to report valid and reliable results:
  - Censored/missing data
  - Cost estimation
  - Net monetary benefit estimation
  - Representativeness of the data
    * Transferability/subgroup analysis
    * Design issues
      - Sample selection
      - Protocol
  - Integration of trial-based and observational data

STOCHASTIC UNCERTAINTY AND THE COMPARISON OF COSTS AND EFFECTS

- Prior to early 1990's, clinical economists did not have an answer to questions about stochastic uncertainty related to cost-effectiveness ratios.
  - Could express this uncertainty for the numerators and denominators of the ratios separately
  - For the ratio, however, we usually said that we evaluated uncertainty using sensitivity analysis.

- Since that time there has been rapid development of methods for assessing this uncertainty.
  * Confidence intervals for cost-effectiveness ratios
    - Fieller’s theorem (potentially undefined)
    - Nonparametric bootstrap (potentially undefined)
  * Other approaches
    - Acceptability curves (evaluates the probability that the ratio is acceptable)
    - Net benefits (transforms the ratio in a way that its ratio is always defined)
CURRENT STATE OF THE ART

- Develop and test hypotheses about cost-effectiveness ratios
  - E.g., the incremental ratio of therapy X compared with therapy Y will be lower than $Z per QALY (where Z represents one’s estimate of the acceptable upper limit for the confidence interval, referred to as the ceiling ratio)
- Test these hypotheses by determining whether the net monetary benefits (NMB) calculated using a ceiling ratio of Z are significantly greater than 0 (or whether the limits of the confidence interval around the cost effectiveness ratio are acceptable)

NET MONETARY BENEFITS

- A composite measure (part cost-effectiveness, part cost-benefit analysis), usually expressed in dollar terms, which is derived by rearranging the following decision rule:
  \[ \text{CR} > \frac{\text{Costs}_1 - \text{Costs}_2}{\text{Outcomes}_1 - \text{Outcomes}_2} \]
  
  where CR = ceiling ratio (e.g., $60,000)
- Most commonly expressed as what may be called net monetary benefits
  \[ (\text{CR} \times [\text{Outcomes}_1 - \text{Outcomes}_2]) - (\text{Costs}_1 - \text{Costs}_2) > 0 \]
- All else equal, one should adopt programs with net monetary benefits that are greater than 0 (i.e., programs with incremental cost effectiveness ratios that are less than the ceiling ratio)

ADVANTAGES OF USE OF NMB FOR EVALUATING STOCHASTIC UNCERTAINTY

- NMB a continuous variable like age, height, or weight, with a relatively normal distribution
- When each patient contributes an estimate of total costs and outcome to the analysis (e.g., when one is not using techniques to account for censored data), NMB can be calculated -- and tested statistically -- as a difference in per-patient means
- Variance (and thus the CI) of NMB is well defined
  \[ \text{VAR}(\text{NMB}) = s_{\text{C}}^2 + R_C^2 s_{\text{E}}^2 - 2 R_C \rho s_{\text{C}} s_{\text{E}} \]
  
  where \( s_{\text{C}} \) and \( s_{\text{E}} \) equal the standard error of the mean difference in costs and effects, respectively; \( R_C \) equals the ceiling ratio (e.g., $40,000 per QALY); and \( \rho \) is the correlation between the difference in costs and effects

IMPLICATIONS FOR SAMPLE SIZE DETERMINATION

- The ability to develop and test hypotheses about economic value of new therapies has led to changes in methods for estimating sample size for such evaluations
- Prior to the development of the literature that described confidence intervals for cost-effectiveness ratios, a common approach was to base sample size on the larger of the sample sizes needed for estimating pre-specified cost and effect differences
  - i.e., what sample size was required to identify a $1000 difference in costs, and what was required to identify a 10% reduction in mortality
SAMPLE SIZE FOR ECONOMIC QUESTIONS

- Once the literature on confidence intervals developed, however, it became clear that the goal of economic evaluations in trials was to determine the likelihood that the therapy represented good value for the cost.

- Current sample size methods base their calculations on the number of study subjects needed to rule out unacceptably high upper confidence limits for the cost-effectiveness ratio (equivalently, to rule out that the net monetary benefits of the intervention are less than 0).

- E.g., sample size for NMB uses the standard formula for continuous variables

\[ n = \frac{2 \sigma^2 (1-R^2) (z_{\alpha/2} + z_{\beta})^2}{\Delta \text{NMB}^2} \]

where \( \Delta \text{NMB} \) equals the difference in mean NMB between the therapies; \( \sigma^2 \) equals the variance; and \( R^2 = \) variance explained by other covariates in OLS (assuming OLS is used to predict the difference).

INFORMATION REQUIRED TO ESTIMATE SAMPLE SIZE FOR ECONOMIC QUESTION

- The newer methods generally require more information than is needed for estimating sample sizes for clinical outcomes or for cost differences alone.

- Basic data for such calculations include the magnitude of the incremental costs and outcomes one expects to observe in the trial; the standard deviations for costs and outcomes in each of the treatment groups; and the correlation between costs and outcomes:
  - Win/lose (positive) correlation
  - Win/win (negative) correlation

CORRELATION BETWEEN COSTS AND EFFECTS

- Correlation between costs and effects can have dramatic effects on the confidence interval -- and thus the sample size -- required for the cost effectiveness ratio.
WHERE TO OBTAIN THE NECESSARY DATA?

- If both therapies are already in use, expected differences in outcomes and standard deviations can be derived from feasibility studies or from records of patients like those who will be enrolled in the trial
  - Potential sources
    * Medical charts of administrative data sets
    * Patient logs of their health care resource use
    * Asking patients and experts about the kinds of care received by those with the condition under study
  - In addition, at least one study has suggested that the correlation between costs and effects observed in these data may be an adequate proxy for the measure of correlation used for estimating sample size

WHERE TO OBTAIN THE NECESSARY DATA? (II)

- For novel therapies, information about the magnitude of the incremental costs and outcomes may not be available
  - May need to be generated by assumption
  - Data on the standard deviations for those who receive usual care/placebo may be obtained from feasibility studies or patient records
    * One may assume that the standard deviation will apply equally to both treatment groups, or one may make alternative assumptions about the relative their magnitudes
  - The correlation also would be obtained from such data

CENSORED (MISSING) DATA (I)

- The problem: Censored data pose threats to estimation of costs of medical therapies and differences in these costs
  - Censoring mechanisms can be:
    * Completely at random: the censored data represent a random sample of all of the data observed in the experiment
    * At random: the censored data represent a random sample of a predictable subsample of the data observed in the experiment
    * Non-ignorable: the censored data are not a random sample of either all of the data or a predictable subsample of the data (i.e., additional data -- most likely from outside the experiment -- are needed to estimate the missing data)
Recent methodologic developments - A number of authors have proposed methods for addressing issues posed by missing data

- General strategy: Identify observations without censored data that are "similar" to observations with censored data, and use data from the former to represent (censored) data from the latter

- Two of the most cited approaches have been provided by Lin and colleagues


  Lin DY. Linear regression analysis of censored medical costs. Biostatistics. 2000;1:35-47. *Describes methods for data that are missing at random*

- An alternative approach for imputing data that are missing at random is described by:


The future

- It's not clear that we've come to closure on the methods for use when data are missing completely at random or at random

  * Alternative methods may still be proposed

- As with most other fields, we have made little or no headway in addressing issues related to data that are nonignorably missing

  * Strategy will probably continue to be to attempt to identify observations without censored data that are "similar" to observations with censored data; however, will probably need to look outside the experiment for the similar observations and their data

- For example, it may be that participants who experience "catastrophes" are the ones who become censored, and none of the participants who are uncensored experience such catastrophes. One might collect data about the costs and effects of such catastrophes outside of the experiment
COST ESTIMATION (I)

- Multivariable estimates of costs generally thought to be better than univariate estimates because:
  - Practice pattern differences by provider, center, or country may have a large influence on costs and the randomization may not account for all imbalances between groups
  - Variations in economic conditions often not controlled for in a randomized trial, therefore multivariable analysis takes on added importance
  - Improves efficiency of estimation/inference (by explaining variation due to other causes)
  - Helps explain what is observed (e.g., coefficients for other variables should make sense economically)

COST ESTIMATION (II)

- The problem: The distribution of costs -- that is, the fact that they cannot be negative and usually have long, heavy right tails -- may make what have been common approaches to cost estimation either biased or inefficient
  - Recent methodologic developments: Use of OLS with or without log transformation for the analysis of costs has been giving way to use of GLM ("fitting a gamma with a log link")
  - Little if any evidence exists in the literature that demonstrates that the results of correctly applied methods differ dramatically from one another

COST ESTIMATION (III)

- Future: Develop better guidelines for helping analysts choose an appropriate method for the analysis of costs
  - It is most likely the case that no single model will always be most appropriate for estimating cost differences associated with medical therapies
  - Guidance for selection of the most appropriate model may be derived from theory, but it also may be derived by empirical estimation within the dataset (e.g., by use of split sample experiments)
  - The most familiar criteria used for choice between models are based on observed vs. predicted values for individual observations within the sample (MSE, absolute error, etc.)
  - It may not be the case that the method that minimizes error for individual predictions is the same method that provides the best out-of-sample estimate of the difference in means between two treatment groups (e.g., doing better with the tails does not necessarily mean doing better with difference in means)

NET MONETARY BENEFIT (RΔE-ΔC) ESTIMATION (I)

- Originally proposed as a means of avoiding the poorly behaved statistical properties of CER
  - With net monetary benefit (NMB), the study result is a difference in means, not a ratio of means, and its standard error is always defined (i.e., no odd statistical properties like the ratio)
- The problem: The same issues that arise for the estimation of costs may arise for the estimation of NMB. However, in addition, there are issues related to where in the estimation process one makes the NMB transformation
NET MONETARY BENEFIT ($\Delta E - \Delta C$) ESTIMATION (II)

- Recent methodologic developments:
  - Three major strategies for the analysis of NMB
    * Direct estimation of patient-level NMB (e.g., net monetary benefit regression)
      - Potential advantages: Single estimation that maintains correlation between cost and outcome; can utilize advances in the estimation of costs when estimating NMB
      - Potential disadvantages: Explanatory power of separate predictions of costs and outcomes may be greater than explanatory power of prediction of direct estimate of NMB; may be easier to address censoring of costs and effects separately than it is to address censoring of NMB

NET MONETARY BENEFIT ($\Delta E - \Delta C$) ESTIMATION (III)

- Recent methodologic developments (cont)
  * Independent estimation of costs and effects
    - (Current) potential advantages: Allows use of any estimation procedure for costs and effects (including GLM, etc.)
    - Potential disadvantages: Ignores correlation structure between costs and effects in the estimation; results of estimation don't provide measure of correlation between the difference in costs and effects (necessary for parametric estimation of CI)

NET MONETARY BENEFIT ($\Delta E - \Delta C$) ESTIMATION (IV)

- Recent methodologic developments (cont)
  * Use of seemingly-unrelated regression (sure command, Stata 8.0) to separately predict costs and effects
    - Potential advantages: Uses correlation structure between costs and effects in the estimation of each of the two outcomes; provides a measure of the correlation between the difference in costs and effects
    - (Current) potential disadvantages: May limit the functional forms one can use in the estimation of the outcomes

- The future
  - Currently a rapidly progressing area of research
  - Anticipate we will develop methods that will combine most if not all of the strengths of the current methods
TRANSFERABILITY (I)

- Differences in morbidity and mortality patterns, genetic susceptibilities, practice patterns, unit costs, etc., may lead decision makers to question whether the pooled economic results from multinational or multi-center trials represent the economic results that would be observed in their country or center.

- Similar issues arise when considering whether the pooled estimate represents the economic results that would be observed for any subgroup of interest in the trial (e.g., more and less severely ill patients).

- Recent methodologic developments:
  - Two approaches -- which rely principally on data from the trial to address these issues -- are currently making their way into the literature (other approaches include decision analytic models that attempt to incorporate trial-based and observational data).

TRANSFERABILITY (II)

- Hypothesis testing: One of the two methods tests the homogeneity of the economic results observed in the trial.
  - If there is no evidence of heterogeneity (i.e., a nonsignificant p-value for the test of homogeneity), and if one believes the test was powerful enough to rule out economically meaningful differences in costs, then one can conclude that the pooled economic result from the trial applies to all of the countries that participated in the trial.
  - If there is evidence of heterogeneity, then the method indicates one should not use the pooled estimate to represent the result for the individual countries, but this method is less clear about the result that should be used instead.


TRANSFERABILITY (III)

- Estimation: The second method uses multi-level random-effects modeling shrinkage estimators to provide more precise estimates of the country-specific results than are yielded by separate -- and naive -- analysis of each country’s costs and effects.
  - These methods have the potential added advantage of providing better estimates of the uncertainty surrounding the pooled result than naive estimates of the trial-wide result.

- My prior: The strongest evidence for adopting a therapy will be derived from some version of the pooled result (and evidence that country-specific results aren’t substantially different from the pooled result).
MORE GENERAL IMPLICATIONS OF SHRINKAGE ESTIMATORS

- If we begin to use shrinkage estimators to address transferability, will have to return to more general methods for estimating uncertainty (e.g., confidence intervals or acceptability curves) and incorporate shrinkage estimators there as well.

- Same issues arise for estimating uncertainty around the clinical endpoints.

CONCLUSIONS

- Don't want to come across like the enlightenment philosophers who pictured the human condition as a never ending spiral towards perfection.

- On the other hand, the methodologic underpinnings of economic analysis of medical therapies have been strengthened in the past 10 years, and the current active research agenda should improve them further.

- In our field, as in almost every other, one of the greatest problems is dissemination of what we have learned.

- While some of that problem will only be addressed when we incorporate the methodologic advances into readily available, easy to use software packages, I would hope that the attendees here will provide one of the mechanisms for such dissemination.