ECONOMIC ASSESSMENT IN CONTROLLED CLINICAL TRIALS

Henry Glick, Ph.D.,
University of Pennsylvania Health System

CDER Seminar Series
May 8, 2002

http://www.uphs.upenn.edu/dgimhsr/


ECONOMIC MESSAGES

- Economic studies may convey a number of potential messages:
  - Primary economic message:
    * The diagnostic technology or the therapy is “good value for the cost”
  - Other messages:
    * The disease poses a substantial burden to society in terms of morbidity, mortality, costs, and quality of life (QOL)
    * Therapy reduces this burden

STUDY DESIGNS FOR SUPPORTING ECONOMIC MESSAGES

- Many different potential study designs are available to support these economic messages
  - Clinical trials, decision analytic models, and observational studies can be used to support messages about value for the cost
  - Decision analytic models and observational studies to support messages about disease burden

RECENT HISTORY, PRIMARY ECONOMIC MESSAGE: THE TECHNOLOGY IS “GOOD VALUE FOR THE COST”

- 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
  * Little or none 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 economic impact of therapy
  - 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
RECENT HISTORY, PRIMARY MESSAGE (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.

EXAMPLE OF TYPICAL MID ‘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>
<tr>
<td>Cost-Effectiveness Ratios</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Principal Analysis</td>
<td>Dominates</td>
<td>NA</td>
</tr>
<tr>
<td>Survival benefit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-33%</td>
<td>Dominates</td>
<td>NA</td>
</tr>
<tr>
<td>+33%</td>
<td>Dominates</td>
<td>NA</td>
</tr>
<tr>
<td>Intervention costs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-50%</td>
<td>Dominates</td>
<td>NA</td>
</tr>
<tr>
<td>+50%</td>
<td>Dominates</td>
<td>NA</td>
</tr>
<tr>
<td>Hospitalization costs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-50%</td>
<td>Dominates</td>
<td>NA</td>
</tr>
<tr>
<td>+50%</td>
<td>Dominates</td>
<td>NA</td>
</tr>
<tr>
<td>Discount rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0%</td>
<td>Dominates</td>
<td>NA</td>
</tr>
<tr>
<td>7%</td>
<td>Dominates</td>
<td>NA</td>
</tr>
</tbody>
</table>
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>
<tr>
<td>Cost-Effectiveness Ratios</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Principal Analysis</td>
<td>Dominates</td>
<td>Dom to 6650</td>
</tr>
<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>
**STEPS IN ECONOMIC EVALUATION**

- Step 1: Quantify the costs of care
- Step 2: Quantify outcomes
- Step 3: Assess whether and by how much average costs and outcomes differ among the treatment groups
- Step 4: Compare magnitude of differences in costs and outcomes and evaluate "value for cost" (e.g., by reporting a cost effectiveness ratio or the probability that the ratio is acceptable)
  - A hypothesis that might be tested in such a study is that the ratio of the cost per quality-adjusted life year saved is significantly less than $60,000
- Step 5: Perform sensitivity analyses

**DIFFICULTIES ACHIEVING AN IDEAL EVALUATION**

- Potential difficulties in meeting these goals within trials
  - Settings often controlled; comparator isn’t always the most commonly used therapy; investigators haven’t always learned fully how to use the therapies under study
  - In some cases, sample size required to answer economic questions is greater than sample size required for clinical questions
  - In some cases, ideal length of follow-up required to answer economic questions is longer than accepted follow-up accepted for answering clinical questions
- However, these trials may be the only source of information needed for important early decisions about the adoption and diffusion of the therapy
- **TRADEOFF:** ideal vs best feasible study

**FOUR STRATEGIC ISSUES**

- Addressing stochastic uncertainty and sample size for economic questions
- Addressing the study’s generalizability
- Follow-up
  - Identifying an appropriate length of follow-up
  - Addressing uncertainty about projection of results

**IDEAL ECONOMIC EVALUATION WITHIN A TRIAL**

- An ideal economic evaluation within a clinical trial is:
  - Conducted in naturalistic settings; compares the therapy with other commonly used therapies; and studies the therapy as it would be used in usual care
  - Performed with adequate power to assess the homogeneity of results in the wide range of clinical settings and among the wide range of clinical indications in which the therapy will be used
  - Designed with an adequate length of follow-up to assess the full impact of the therapy
  - Conducted within a time frame that allows the resulting information to inform important decisions in the adoption and dissemination of the therapy
STRATEGIC ISSUE #1. ADDRESSING STOCHASTIC UNCERTAINTY AND SAMPLE SIZE FOR ECONOMIC QUESTIONS

- 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 stochastic uncertainty for the evaluation of questions about value for the cost of new medical technologies
  - Fieller’s theorem (potentially undefined)
  - Nonparametric bootstrap (potentially undefined)
  - Acceptability curves
  - Net benefits

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 relatively new composite measure (part cost-effectiveness, part cost benefit analysis), usually expressed in dollar terms, which is derived by rearranging the following decision rule:
  \[ CR > \frac{(\text{Costs}_1 - \text{Costs}_2)}{(\text{Outcomes}_1 - \text{Outcomes}_2)} \]
  where CR = ceiling ratio (e.g., $40,000)
- Most commonly expressed as what may be called net monetary benefits
  \[ (CR \times (\text{Outcomes}_1 - \text{Outcomes}_2)) - (\text{Costs}_1 - \text{Costs}_2) > 0 \]
- 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
  \[ \text{VAR}(\text{NMB}) = s_{\text{C}}^2 + R_c^2 s_{\text{E}}^2 - 2R_c p s_{\text{C}} s_{\text{E}} \]
  where \( s_{\text{C}} \) equals the standard error of the mean difference in costs; \( s_{\text{E}} \) equals the standard error of the mean difference in effects; and \( R_c \) equals the ceiling ratio (e.g., $40,000 per QALY)
IMPLICATIONS FOR SAMPLE SIZE DETERMINATION

- Recent changes in the methods for estimating sample size for economic assessments in trials reflect changes in the economic question being asked.

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

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

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.
WHAT DO WE MEAN BY CORRELATION (I)?

- Positive Correlation: As the effectiveness (cost) increases, the cost (effectiveness) increases
  - Hypothetical example: Intervention that avoids death for a disease where death reduces overall healthcare costs (e.g., interventions for the acute treatment of stroke)
  - As the effectiveness increases (i.e., when there are greater decreases in mortality), the costs increase (due to the more costly, but greater survival)

WHAT DO WE MEAN BY CORRELATION (II)?

- Negative Correlation: As the effectiveness (costs) decreases, the cost (effectiveness) increases
  - Hypothetical example: Intervention that avoids morbidity, which generally is treated with costly medical interventions (e.g., interventions for asthma; sleep apnea?)
  - As the effectiveness increases, there is less morbidity, thus the costs decrease; when effectiveness decreases, there is more morbidity, thus costs increase

CORRELATION BETWEEN COSTS AND EFFECTS

- Correlation between costs and effects can have dramatic effects on the confidence interval for the cost effectiveness ratio
CEILING RATIO AND IDENTIFICATION OF AN APPROPRIATE OUTCOME MEASURE

- The sample size calculations described above assume that we have an idea about what we would like to pay to obtain a unit of outcome (e.g., $50,000 per quality-adjusted life year saved)

- In many medical specialties, researchers use disease specific outcomes
  - e.g., cases detected; abstinence days; etc.

- While one can calculate a cost-effectiveness ratio for any outcome one wants (e.g., cost/case detected or cost/additional abstinence day), to be convincing that a new, more costly and more effective therapy is good value, the outcome must be one for which we have recognized benchmarks of cost effectiveness
  - Argues against use of too disease-specific an outcome for economic assessment

STRATEGIC ISSUE #2. ADDRESSING A STUDY’S GENERALIZABILITY

- There has been growing concern that the pooled (i.e., average) clinical and economic results from multicenter and multinational trials may not be reflective of the results that would be observed in individual centers and countries that participated in the trial
  - Thus it may be difficult for decision makers in specific centers and countries to draw useful conclusions about the value for the cost of the therapies that were evaluated in the trials

TRADITIONAL APPROACHES TO GENERALIZABILITY

- Traditional approaches to generalizing the economic results from multicenter and multinational trials to individual centers and countries include:
  - Using trial-wide clinical results with costs based on trial wide utilization but using unit prices of the center/country in question
  - Using trial-wide clinical results and costs based on patients treated in the center/country

- These approaches ignore the fact that clinical and economic outcomes may influence one another (differences in costs may affect practice patterns, which in turn may affect outcomes; differing practice patterns may affect outcomes, which in turn may affect costs)

- Others have addressed the issue by use of decision analytic models (e.g., Drummond et al. IJTAHC. 1992;8:671-82)
TEST OF HOMOGENEITY

- Given the fact that cost-effectiveness ratios are made up of absolute differences in costs and effects, yet homogeneity tests of effects (and sometimes costs) generally relate to relative differences (e.g., odds ratios):
  - Estimate center/country-specific net monetary benefits (cost effectiveness ratios) and evaluate their homogeneity directly

- One would use the more precise pooled (average) net monetary benefits (ratio) for the overall study to represent these centers'/countries' ratios only if:
  1) It appears that there is no center/country-by-net monetary benefit and (center/country-by-ratio) interaction
  2) The minimum detectable difference was small enough to be economically important

TYPES OF INTERACTION

- Gail and Simon have proposed two tests of interaction to determine whether results are inconsistent in both direction and magnitude or whether they are consistent in direction but not in magnitude
  - A qualitative or crossover interaction occurs if the treatment effect is positive (e.g., has positive net monetary benefits) in some centers/countries, and negative (e.g., has negative NMB) in others (i.e., inconsistent in both direction and magnitude)
  - A non-crossover or quantitative interaction occurs if there is variation in the magnitude of the effect (e.g., variation in the NMB), but not in its direction (e.g., when the treatment has positive NMB in all centers/countries)

QUALITATIVE INTERACTION

- The formal test for qualitative interactions of the center/country-by-outcome effects uses estimates of the treatment effect and its variance for each of the countries being evaluated
  - For cost effectiveness ratios, the null hypothesis is that all of the NMB are greater than 0 (ratios are acceptable), whereas the alternative hypothesis is that some NMB are greater than 0 (ratios are acceptable) and some are less than 0 (ratios are unacceptable)
  - The statistical test is based on a likelihood ratio, with critical values of the test given in Gail and Simon
QUANTITATIVE INTERACTION

- The test for quantitative interaction is based on the sum of squared errors of the country-specific treatment effects and the variance of these effects

- The center/country-specific NMB (ratios) and their variance estimates are used to compute the test statistic

- A weighted mean is used in estimating the errors rather than the arithmetic mean

- The test statistic is compared to critical values of the $\chi^2$ distribution with one less degree of freedom than there are centers/countries being evaluated

MAGNITUDE ESTIMATION

- Employ empirical Bayes methodology

- Effectively provides an estimate of the per-center/country NMB (ratios) while adjusting for the higher than expected variance when dividing the data

- Can be estimated with a random effects model where the random effect is specified across centers/countries

- Adjusts for the fact that differences between sub-groups could have arisen purely by chance

- This approach has the effect of 'shrinking' the raw estimates of cost-effectiveness in the individual centers/countries back toward the pooled estimate of cost-effectiveness for the entire trial

STRATEGIC ISSUE #3. IDENTIFYING APPROPRIATE LENGTH OF FOLLOW-UP

- Economic assessments conducted as part of randomized trials are meant to allow decision makers to use the results of the trial to reach conclusions about the economic benefits of the therapy under investigation

- One design issue that may limit the interpretability of the economic data collected within the trial is the study time horizon

- Although clinical efficacy may be demonstrated when a difference in clinical endpoints is observed between study arms, from an economic perspective the appropriate time horizon for a trial would include all (or a substantial portion) of the time when there is resource use related to the illness under study

- The economic time horizon that would best inform decision makers about the value for the cost of a therapy thus need not be the same as the one adopted for answering the clinical question

Migraine Costs
Pennsylvania Medicaid Data

![Graph of Migraine Costs](image)
FDA-ACCEPTED ENDPOINTS AND ECONOMIC EVALUATION

- Not always in harmony
  - 28 day mortality endpoint for sepsis trials
  - Treatment failure for antibiotic trials

- Once the FDA agrees to an endpoint, it may be difficult to convince companies to adopt a longer period of follow-up

STRATEGIC ISSUE #4. ADDRESSING UNCERTAINTY ABOUT PROJECTION OF RESULTS

- When therapies are studied with short- or intermediate length trials but have long term effects, uncertainty exists about whether the relative magnitude of costs and effects observed within the trials will be descriptive of the relative magnitudes that would be observed had the trials continued for the patients’ lifetime

- To address these uncertainties one should:
  - Evaluate what was observed during the trial (within-trial analysis)
  - Develop decision analytic models to make projections beyond the period of observation (projection)

- The two types of analyses have offsetting strengths/weaknesses:
  - We are more certain of the data from the trial, but they may fail to capture the most important impacts of the therapy
  - We are less certain about the projections beyond the trial, but they may attempt to quantify the most important impacts of the therapy

WITHIN-TRIAL ANALYSIS

- Even if one decides that the primary analysis will be a projection beyond the period of observation, one should also evaluate the costs and outcomes that were observed during the trial
  - In such a “within-trial” evaluation, one should maintain the same time horizons for costs and outcomes observed in the trial (e.g., if follow-up for the trial was for one year, then costs and effects should be measured for one year)
  - Not always easy to demonstrate cost-effectiveness in a within-trial analysis
    - e.g., no within-trial analysis of cholesterol-modifying therapy has demonstrated reasonable cost-effectiveness
LONGER-TERM PROJECTION

- Because the relative magnitude of incremental costs and outcomes observed during the trial may not be reflective of the relative magnitude that would have been observed had the trial been continued until all study subjects discontinued therapy or died, one should also project the results for longer periods.

- For projection: Maintain a common time horizon for both costs and effects.
  - A number of studies have used the cost difference observed within the trial; argued that the benefits of the therapy extend beyond the trial; and incorporated the benefits from beyond the trial.
  - E.g., West of Scotland Coronary Prevention Study.
    * If the therapy has downstream benefits that have not been adequately captured during the trial, it most likely has downstream costs that also have not been adequately captured during the trial.

- Given that the longer the projection, the less certain the results, make projections for different time horizons.

COST EFFECTIVENESS RATIO AND 95% CI, 4S FOR DIFFERENT LENGTHS OF FOLLOW-UP/PROJECTION

<table>
<thead>
<tr>
<th>Years of Follow-up</th>
<th>Point Estimate</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dominated</td>
<td>$168,884 to Dominated</td>
</tr>
<tr>
<td>2</td>
<td>$282,857</td>
<td>$45,577 to Dominated</td>
</tr>
<tr>
<td>3</td>
<td>$73,529</td>
<td>Dominates to Dominated</td>
</tr>
<tr>
<td>4</td>
<td>$12,074</td>
<td>Dominates to Dominated</td>
</tr>
<tr>
<td>5.5</td>
<td>$15,258</td>
<td>Dominates to $122,772</td>
</tr>
<tr>
<td>10</td>
<td>$12,246</td>
<td>Dominates to $42,263</td>
</tr>
<tr>
<td>15</td>
<td>$8,578</td>
<td>Dominates to $6,721</td>
</tr>
<tr>
<td>20</td>
<td>$7,320</td>
<td>$681 to $21,841</td>
</tr>
</tbody>
</table>
OTHER CONTENTIOUS STRATEGIC ISSUES

- What proportion of the total resource use data should be collected?
- In what delivery settings should data be collected, and how should utilization in nonstudy sites be identified?
- Among which patients should data be collected?
- Should follow-up be discontinued if participants fail therapy?
- Which unit cost estimates should be used for the study?

SUMMARY

- Many opportunities exist for incorporating economic assessments into the randomized trials assessing medical therapies.
- Clinical trials may provide the best opportunity for developing information about a medical therapy’s value early in its product life.
- When appropriate types of data are collected and when they are analyzed appropriately, these evaluations can provide data about uncertainties related to the assessment of the value for the cost of new therapies that may be used by policy makers, drug manufacturers, health care providers and patients when the therapy is first introduced in the market.