Goal of Economic Evaluation in Clinical Trials

- To evaluate the value for the cost of medical therapies
- Uses economic data collected as primary or secondary endpoints in the randomized trial.
  - Could evaluate using directly observed short term impact or using projected long term impact

Why Include an Economic Evaluation within a Randomized Trial?

- Can provide economic rationale
  - for adoption decision
  - For reimbursement
    - Reimbursement can promote diffusion
    - Payers often more reluctant to reimburse behavioral treatments
- Economic information may not be timely if economic outcomes are modeled after effectiveness has been proven
- Can establish the value of a potentially expensive treatment that may reduce (and therefore the cost of managing) morbid endpoints related to disease
- Estimates economic endpoints with strong internal validity
Study Objectives

Estimate the incremental cost-effectiveness of two alternative interventions for a given medical condition

\[
CER = \frac{\bar{C}_A - \bar{C}_B}{\bar{E}_A - \bar{E}_A} = \frac{\Delta C}{\Delta E}
\]

COST EFFECTIVENESS EXAMPLE

- Consider a behavioral intervention A versus the standard of care B.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>A-B</th>
<th>s.e. of diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Cost</td>
<td>2000</td>
<td>1000</td>
<td>1000</td>
<td>3254</td>
</tr>
<tr>
<td>Mean QALY</td>
<td>0.86</td>
<td>0.85</td>
<td>.01</td>
<td>0.001629</td>
</tr>
</tbody>
</table>

POINT ESTIMATE OF COST EFFECTIVENESS RATIO (CER)

\[CER = \frac{\bar{C}_A - \bar{C}_B}{\bar{E}_A - \bar{E}_A} = \frac{\Delta C}{\Delta E}\]

CER = $100,000 / QALY
Ideal Economic Evaluation Within a Trial

- Naturalistic setting
  - compares the therapy with other commonly used therapies
  - studies the therapy as it would be used in usual care
- Well powered for
  - Average effects
  - Subgroup effects
- Adequate length of follow-up
  - Should be able to assess the full impact of the therapy
- Timely
  - Can inform important decisions in the adoption and dissemination of the therapy

Difficulties Achieving an Ideal Evaluation

- Settings often controlled; comparator isn’t always the most commonly used therapy; investigators haven’t always learned fully how to use the new therapy under study
- Sample size required to answer economic questions may be greater than sample size required for clinical questions
- Ideal length of follow-up required to answer economic questions is longer than follow-up necessary to answer clinical questions
- Design involves tradeoffs between ideal and feasible
  - These trials may be the only source of information needed for important early decisions about the adoption and diffusion of an important new therapy

Typical Design Issues

- Data collection
  - Clinical trials do not traditionally collect information on costs or even resource use
  - Infrequently include measures of effectiveness necessary for cost-effectiveness
    - Survival is the one exception
- Naturalistic setting
  - Control arm should be the most effective or standard available therapy rather than placebo
  - Intent-to-treat
- Sample size may need to be larger to power for economic endpoints
Outline

• Collecting needed cost data
• Naturalistic Setting
• Sample size

Design: collecting needed data
COSTS - methods

• Method of calculation: resource costing
  – Identify resources consumed during course of care
  – Assign values or prices to each resource
  – Sum the products

• Measure all costs of all participants prior to randomization and for the duration of follow-up
  – Costs after randomization—cost outcome
  – Costs prior to randomization—potential predictor

COSTS – perspective

• What gets included depends on perspective
  – Society, Payer, Patient, Provider, …
• Societal perspective includes all resources consumed
  – Direct medical care (intervention plus follow-up care)
  – Nonmedical care related to treatment
• Travel costs
  – Value of time in caring for patient
  – Value of patient’s time in treatment or productivity loss
COSTS – resource use

- Tracking resource use
  - Key items often in collection forms designed for trial
  - Modifications may be necessary
    - Hospital stays, outpatient visits
    - Behavioral services outside of trial sites
  - Forms should be filled out frequently by coordinators or the subjects
  - Billing data or insurance claims may be helpful if
    - Records easily obtainable
    - Comprehensive and consistent

Design: collecting needed data
COSTS - prices

- Prices weights should aim to estimate economic costs
  - Distinguish between charges, reimbursements, and true costs
  - True cost: the amount that parties in a well functioning market are willing to pay for the resource
  - Appropriate price weights depend on the study’s perspective

- Finding prices
  - National sources often helpful
    - MEPS, HCUP, RBRVS, DRG
    - www.cms.hhs.gov/FeeScheduleGenInfo/
    - www.cms.hhs.gov/ProspMedicareFeeSvcPmtGen/
    - www.cms.hhs.gov/Mcrpartbdrugavgsalesprice/02.aspfiles.asp

COSTS – Common issues

1. Don’t have sufficient resources to track all medical service use
2. At what level should medical service use be aggregated?
3. Should the focus be on disease-related services or all medical services?
COSTS – Issues

1. Limited Data Collection Resources
   • Measure services that make up a large portion of the difference in treatment between patients randomized to the different therapies under study
     – Provides an estimate of the cost impact of the therapy
   • Measure services that make up a large portion of the total bill
     – Minimizing unmeasured services reduces the likelihood that differences among them will lead to biased estimates
     – Provides a measure of overall variability
   • Measure as much as possible

COSTS – Issues

2. At What Level Should Medical Service Use Be Aggregated?
   - For example for inpatient care, should we count:
     • Hospitalizations?
     • Days in the hospital?
     • Days in the hospital stratified by location in the hospital?
     • Days in the hospital stratified by location plus individual services provided in the hospital?
     • Micro-costing of every service within hospitalization

Costs - Issues

2. At What Level Should Medical Service Use Be Aggregated?
   - Do we expect the intervention to affect:
     - The number of hospitalizations
     - The length of stay of a hospitalization
     - The intensity of medical services utilized
     - Or is the hospitalization secondary
       • e.g. intervention in low risk outpatient setting
   - Will more or less aggregated information affect study result and/or cost of data collection
   - Resulting decisions affect the price weight estimates required for the calculation of cost
COSTS – Issues

3. Should data be limited to Disease-Related Services?

Don’t expect to affect all medical service use, just that related to the disease in question.

Implication: given sample size in trial, collection of all medical services, independent of the reason for these services, may swamp the “signal” with “noise.”

However:
- Investigators routinely attributes AEs to the intervention, even when participants received vehicle/placebo
- Judgments often lack data or are flawed
- Modifying disease in one body system may affect disease in another body system
- If a patient has an automobile accident, how does the clinician determine whether or not it was due to a hypotensive event caused by therapy?
- Potential biases more of a problem in unblinded studies, but need not “balance out” in double-blinded studies

Issue:
How Naturalistic Should The Study Design Be?

- The primary purpose of cost-effectiveness analysis is to inform real-world decision-makers about how to respond to real-world health care needs
- Thus, the more naturalistic the trial, the more relevant the results for decision making
  – analysis based on the intention to treat
  – limit loss to follow-up
  – Minimize protocol induced costs and effects

Naturalism: Intention-To-Treat Analysis

- Economic questions relate to treatment decisions (e.g., whether to prescribe a therapy), not whether the patient received the drug prescribed nor whether, once they started the prescribed drug, they were switched to other drugs
  – Implication: costs and effects associated with these later decisions should be attributed to the initial treatment decision
- Thus, trial-based cost-effectiveness analyses should adopt an intention-to-treat design
Naturalism: Loss to Follow-up

- Trials should be designed in such a way that they minimize the occurrence of missing data
  - For example, study designs should include plans to aggressively pursue study participants and data throughout the trial
  - One recent long-term study of treatment for bipolar disorder was designed from the outset to respond to missed interviews by:
    1) intensive outreach to reschedule the assessment, followed by
    2) telephone assessment, followed by
    3) interview of a proxy who had been identified and consented at the time of randomization

- Investigators should also ensure that:
  - Follow-up continues until the end of the study period
  - Data collection not be discontinued simply because a participant reaches a clinical or treatment stage such as failure to respond (as, for example, happens in some antibiotic, cancer chemotherapy, and psychiatric drug trials)
  - Given that failure often is associated with a change in the pattern of costs, discontinuation of these patients from the economic study is likely to bias the results of an economic evaluation that is conducted as part of the trial

Naturalism: Protocol-Induced Costs and Effects

- Clinical trial protocols often try to standardize the care of patients in the trial
  - They may require substantial number of investigations and diagnostic tests that would not be performed under normal clinical practice
- Trials also tend to prescribe aggressive documentation and treatment of potential adverse effects observed in the trial
Naturalism: Protocol-Induced Costs and Effects (2)

- These requirements for diagnostic testing may bias the evaluation:
  - Use and cost of tests may be biased towards the null hypothesis of no difference
  - Diagnosis and treatment cost may be increased because of detection in the trial of outcomes that in usual care would not have been detected

Specific Recommendations, Naturalism

- Use the intention to treat sample for the economic analysis
- Be aggressive in maintaining follow-up, including continuing to collect data on those who fail or switch therapy
- Use appropriate analytic methods to address missing data if and when they occur
- To the extent possible in a registration trial, minimize the effect of the protocol on patient care

Issue: What is the Appropriate Sample Size?

- 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 quality-adjusted life year (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:

\[ n = \frac{2(z_{\alpha} + z_{\beta})^2 (sd^2_c + sd^2_q) - (2 W \rho sd_c sd_q))}{(WQ - C)^2} \]

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; \( sd \) = standard deviation for cost (\( sd_c \)) and effect (\( sd_q \)); \( W \) = maximum willingness to pay we wish to rule out; \( \rho \) = correlation of the difference in cost (\( C \)) and effect (\( Q \)).

Parameters needed for power calculation

- Consider a behavioral intervention A versus the standard of care B.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>A-B</th>
<th>s.e. of diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Cost</td>
<td>2000</td>
<td>1000</td>
<td>1000</td>
</tr>
<tr>
<td>Mean QALY</td>
<td>0.86</td>
<td>0.85</td>
<td>0.01</td>
</tr>
</tbody>
</table>

- (The correlation between costs and effects is estimated at\(=-.7102 \) and the trial is powered for the clinical outcome with an \( n=250 \) in each treatment group)

Correlation

- 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

- As effects increase, costs decrease.
  - Negative correlation
  - e.g., asthma care

- As effects increase, costs increase.
  - Positive correlation
  - e.g., life-saving care such as stroke
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

Willingness to Pay and Identification of an Appropriate Outcome Measure

- The sample size calculations depend on knowing what we would like to pay to obtain a unit of outcome
- In many medical specialties, researchers use disease specific outcomes
- A convincing story of cost-effectiveness depends on a recognized benchmarks of cost effectiveness
  - Argues against use of too disease-specific an outcome for economic assessment

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>
</tr>
</thead>
<tbody>
<tr>
<td>20,000</td>
<td>1256</td>
</tr>
<tr>
<td>30,000</td>
<td>673</td>
</tr>
<tr>
<td>50,000</td>
<td>419</td>
</tr>
<tr>
<td>75,000</td>
<td>340</td>
</tr>
<tr>
<td>100,000</td>
<td>310</td>
</tr>
</tbody>
</table>

$\Delta C=250; \Delta Q=.05; \text{sd}_c=5,000; \text{sd}_q=.2; \rho=-.10; \sigma=.05; \beta=.8$
Power Formula, Common SDs

- Power is calculated by use of the following formula:

\[
Z_{\beta} = \sqrt{\frac{n \cdot \Delta \text{NMB}^2}{2 \left( \sigma_d^2 + (\Delta W \sigma_d^2) - \left(2 W \rho \sigma_d \sigma_q\right)\right)}} \cdot Z_{\alpha}
\]

- 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>

\( \Delta C = 250; \Delta Q = .05; \sigma_d = 5,000; \sigma_q = 2; \rho = -1; \alpha = .05; \)
Sample size per group = 340

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 size 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

- Clinical trials may provide the best opportunity for developing information about a medical therapy's value for the cost 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.

Further reading:


Extra design topic
Issue: What Is The Likelihood That The Cost-effectiveness Ratio Observed In The Trial Describes Longer Term Therapy?

• When the trial observes cost-effectiveness for a time-limited period (e.g., 2 or 3 years), but the therapy will be taken for lifetime, we should consider the likelihood that the cost-effectiveness ratio observed in the trial will describe longer term therapy
  – Referred to as a time by treatment interaction

Likelihood of a Treatment-by-Time Interaction

• Time by treatment interaction less likely to be substantial when the intervention’s cost and outcome begin at approximately the same time and continue to be incurred together over time (e.g., drug therapy for heart failure)
  – Interaction more likely to be substantial when:
    – Treatment cost and outcome incurred over time, but outcome delayed for a number of years (e.g., risk reduction from cholesterol-modifying therapy) OR increasing with time
    – Treatment cost incurred initially (e.g., surgical removal of tumor) and outcome (e.g., survival) accrued over time

Addressing a Time By Treatment Interaction

• Evaluate what was observed during the trial (within-trial analysis)
• Develop decision analytic models to make projections beyond the period of observation (projection)
Strengths and Weaknesses

- Within-trial analysis and longer term projections have opposing strengths/weaknesses:
  - We are more certain of what was observed during the trial, but follow-up may be too short to capture the most important impacts of the therapy.
  - We are less certain about the projection beyond the trial, but this projection attempts to quantify what may be the most important impacts of the therapy.

Within-Trial Analysis

- Even if we decide that the primary analysis will be a projection beyond the period of observation, we should still evaluate the costs and outcomes that were observed during the trial.
- In such a within-trial evaluation, we 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

- To investigate whether the cost-effectiveness ratio is homogeneous with respect to time, we should also project the results for longer periods.
- For projection: Maintain a common time horizon for both costs and effects.
  - some 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 but not costs 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.
Time Horizon for Projection

• Given that the longer the projection, the less certain the results, we should make projections for different time horizons
  – Even if the longest time horizon in a lifetime projection is 30-40 years, we may observe that projected long-term cost-effectiveness reaches equilibrium after only 5 or 10 years of projection
  – If there is sufficient follow-up during the trial, make estimates for differing lengths of follow-up during the trial (e.g., the first year, the first 2 years, etc.)
    • Adds face validity to the trajectory of the projected cost-effectiveness ratios:

<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</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</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 26,721</td>
</tr>
<tr>
<td>20</td>
<td>7,320</td>
<td>681 to 21,841</td>
</tr>
</tbody>
</table>

Specific Recommendations, Treatment by Time Interaction

• Evaluate whether a strong treatment by time interaction is expected
• If yes, consider development of a decision model to assess the potential magnitude of the interaction
  – Substantial amounts of the data used for the decision model should be derived from the trial
  – Where necessary, augment data from the trial with epidemiologic data on long term outcomes, etc.
Design: collecting needed data

OUTCOMES

• Health outcomes represent the ascribed benefits of an evaluated intervention
• While clinical trials often use multiple outcome measures, CEA necessitates use of a single measure for the denominator of the cost-effectiveness ratio
• Cost-effectiveness outcome could be treatment specific or could be a common measure that extends across different diseases and treatments

OUTCOMES – treatment specific

• Examples of treatment specific measures
  – Drug free months
  – Sight-years gained
  – Symptom-free time
  – Cases of disease prevented
• Advantages of specific measures
  – Quantified more precisely
  – More sensitive to the treatment
• Disadvantages
  – Can not be easily compared across treatments particularly for different diseases
  – Threshold of acceptable level of cost-effectiveness is unknown
OUTCOMES – common measure

• Outcomes measures for CEA that are common
  – Life extension
  – Health related quality of life
  – Quality-adjusted survival – QALY

• Advantages:
  – Allows for a common decision rule for resource allocation
  – Health policy makers with a fixed budget could prioritize and fund projects less than a given threshold of acceptability

• Disadvantages:
  – Assessment can be complex
  – Less precisely measured
  – Harder for clinical audiences to grasp than treatment specific measures

OUTCOMES – assessing quality of life preferences

• Gold standard is direct elicitation of health state preferences
  – Time trade-off
  – Standard Gamble
  – Rating scale

• Prescored health state classification instruments are well accepted and easier to implement in trials
  – Participants report functional status across a variety of domains
  – Preference scores are derived from scoring rules developed using samples from the general public
  – Examples: EQ-5D, HUI2, HUI3, SF-6D, Quality of Well-Being Scale (QWB)