Good Value for the Cost

• Economic data collected as primary or secondary endpoints in randomized trials are commonly used in the evaluation of the value for the cost of medical therapies
  – Short-term economic impacts directly observed
  – Longer term impacts potentially projected by use of decision analysis
  – Reported results: point estimates and confidence intervals for estimates of:
    • Incremental costs and outcomes
    • Comparison of costs and effects
  – Impact of sensitivity analysis judged by its impact on both the point estimates and the confidence intervals of the ratios

Example

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Point Estimate</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incremental Cost</td>
<td>$-173$</td>
<td>$-2153$ to $763$</td>
</tr>
<tr>
<td>Incremental QALYs</td>
<td>$0.13$</td>
<td>$0.07$ to $0.18$</td>
</tr>
<tr>
<td>Cost-Effectiveness Analysis</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 Cost</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 Cost</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>
Outline

• Steps in economic evaluation
• The gold standard and its tensions
• 3 Strategic issues
  – What medical service use should we collect?
  – How naturalistic should the study design be?
  – What is the appropriate sample size?

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 difference in costs and outcomes and evaluate “value for costs” (e.g. by reporting a cost-effectiveness ratio or the probability that the ratio is acceptable
  – Potential hypothesis: The cost per quality-adjusted life year saved is significantly less than $60,000
Step 5: Perform sensitivity analysis

Ideal Economic Evaluation Within a Trial

• Conducted in naturalistic settings
  – 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
• Designed with an adequate length of follow-up
  – Allows the assessment of the full impact of the therapy
• Timely
  – Can inform important decisions in the adoption and dissemination of the therapy
**Ideal Economic Evaluation Within a Trial (II)**

- Measures 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
- Independent of the reasons for the costs
- Most feasible when:
  - Easy to identify when services are provided
  - Service/cost data already being collected
  - Ready access to data

**Design Issues Not Unique To Trials**

- A number of design issues apply equally to economic evaluations that are incorporated within clinical trials and to other economic evaluations:
  - The type of analysis that will be conducted (e.g. cost-benefit, cost-effectiveness, or cost-minimization analysis)
  - The types of costs that will be included (e.g. direct medical, direct nonmedical, productivity, and intangible)
  - The perspective from which the study will be conducted
- These issues have been well addressed in the literature

**Difficulties Achieving an Ideal Evaluation**

- Settings often controlled
- Comparator isn’t always the most commonly used therapy or the currently most cost-effective
- 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 may be longer than follow-up necessary to answer clinical questions
Trade-off

• These trials may be the only source of information needed for important early decisions about the adoption and diffusion of the therapy

• TRADE-OFF: Ideal vs best feasible

Issue #1: What Medical Service Use Should We Collect?

• Real/perceived problems
  1. Don’t have sufficient resources to track all medical service use
  2. Don’t always 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”
  → Why not limit data to disease-related services?

Limited Data Collection Resources

• Availability of administrative data may reduce costs related to tracking all medical service use
• If administrative data are unavailable:
  – 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

• Best approach: measure as many services as possible
  – No a priori guidelines about how much data are enough
  – Little to no data on the incremental value of specific items in the economic case report form

Document Likely Service Use During Trial Design

• Decisions improved if we document the types of services used by patients who are similar to ones who will be enrolled in the trial
  – Review medical charts or administrative data sets
  – Survey patients and experts about the kinds of care received
  – Have patients keep logs of their health care resource use
• Guard against possibility that new therapy will induce medical service use that differs from current medical service use

Account for Data Collection Expense

• Decisions about the services to measure should take into account the expense of collecting particular data items
  – e.g., frequently performed, low cost items?
    • 6,700 blood gas tests equaled 1.8% of procedure and diagnostic test costs
    • 420 angiocardio pneumographies equaled 4.3%
Limit Data to Disease-Related Services?
• Little if any evidence exists about the accuracy, reliability, or validity of such judgments
• Easy for judgments to be flawed

Limit Data to Disease-Related Services (II)
• Investigators routinely attribute AEs to the intervention, even when participants received vehicle/placebo
• Medical practice often multifactorial: modifying disease in one body system may affect disease in another body system
  – In the Studies of Left Ventricular Dysfunction, hospitalizations "for heart failure" and death reduced by 30% (p<0.0001)
  – Hospitalizations for noncardiovascular reasons reduced 14% (p = 0.006)
• 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?

Limit Data to Disease-Related Services (III)
• Potential biases more of a problem in unblinded studies, but need not "balance out" in double-blinded studies
Other Types of Costs?

- Other types of costs that sometimes are documented within economic evaluations include:
  - Time costs: Lost due to illness or to treatment
  - Intangible costs

- Types of costs that should be included in an analysis depend on:
  - What is affected by illness and its treatment
  - What is of interest to decision makers

  - e.g., the National Institute for Clinical Excellence (U.K.) and the Australian Pharmaceutical Benefits Scheme has indicated they are not interested in time costs

General Recommendations

- General Strategy: Identify a set of medical services one will collect, and assess them any time they are used, independent of the reason for their use

- Decision to collect service use independent of their reason does not preclude ADDITIONAL analyses testing whether designated “disease-related” costs differ

General Recommendations (2)

- If data collection is limited to a single page in the CRF:
  - First impression: Collect big-ticket items, (e.g., hospitalization, long term care, etc); don't sweat smaller ticket items

  - Heart failure: hospitalization costs, number of outpatient visits

  - Hospitalized infections: ICU, stepdown, and routine care days; major procedures

  - Asthma: ER visits, Hospitalizations, comediations
Better Approach

• Prior to the study, invest in determining which services will likely make up a large portion of the difference in costs between the treatment groups
  – If the therapy is likely to affect the number of hospitalizations, collect information that will provide a reliable estimate of the cost of these hospitalizations
  – If the therapy is likely to affect days in the hospital and location in the hospital, collect this information
  – If the therapy is principally likely to affect outpatient care, collect measures of outpatient care, etc.

Specific Recommendations, Which Services?

• Identify common patterns of medical service use in centers/countries that will participate in the trials
  – Speak with experts in multiple centers/countries
  – Focus groups, etc.
• Design case report forms to collect important, common medical service use
• Collect the services independent of the reason for their use
• Pilot test the forms
• Consider collecting costs other than medical service use

Issue #2: How Naturalistic Should The Study Design Be?

• Primary purpose of cost-effectiveness analysis:
  Inform real-world decision-makers about how to respond to real-world health care needs
• Greater naturalism, in terms of participants, analysis based on the intention to treat, and limitation of loss to follow-up, implies greater likelihood that the data developed within the trial will speak directly to the decision question
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 to minimize the occurrence of missing data
  - Study designs should include plans to aggressively pursue participants and data throughout the trial
  - Strategies may include:
    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

Naturalism: Loss to Follow-up (2)

- Investigators should also ensure that:
  - Follow-up continues until the end of the study period
  - Data collection is not discontinued simply because a participant reaches a clinical or treatment stage such as failure to respond (as often happens in 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 likely biases the results
Clinical trial protocols often try to standardize the care of patients in the trial – may require substantial number of investigations and diagnostic tests that would not be performed under normal clinical practice.

Protocols also tend to prescribe aggressive documentation and treatment of potential adverse effects observed in the trial.

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.
- E.g., in trials of prophylaxis for DVTs in elective hip replacement surgery, repeated testing for DVTs may identify a number of cases that never would have been detected or treated in usual practice.

Adjustment for this extra detection – potentially by use of decision analytic models – may be difficult.

- Does the therapy avoid complications that would have been detected and treated in routine practice or does it avoid those that never would 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 techniques 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 #3: What is the appropriate sample size?

- Sample size and power calculations allow us to conduct experiments with an expected likelihood that at the conclusion of the experiment we will be able to be confident in the resulting comparison of costs and effects
  - e.g., may hypothesize that the point estimate for the cost-effectiveness ratio will be 20,000 per quality-adjusted life year (QALY)
  - May want to identify a sample size 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

- Assuming equal standard deviations for cost and effect and equal sample sizes, the sample size formula is:
  \[ n = \frac{2(z_{\alpha} + z_{\beta})^2 sd_c^2 + (W sd_q)^2}{(W \Delta Q - \Delta C)^2} \]
  where \( n \) = sample size/gro8p; \( z_{\alpha} \) and \( z_{\beta} \) = z-statistics for \( \alpha \) (e.g., 1.96) and \( \beta \) (e.g., 0.84) errors; \( sd \) = standard deviation for cost (c) and effect (q); \( W \) = maximum willingness to pay one wishes to rule out; and \( p \) = correlation of the difference in cost and effect

www.uphs.upenn.edu/dgimhsr/stat-samps.htm
Similarities With Clinical Sample Size Formulas

\[
\text{Error Rate} \quad \text{NMB Variance}
\]

\[
n = \frac{2 (z_\alpha + z_\beta)^2 \left( \text{sd}_c^2 + (W^2 \text{sd}_c^2) - (2 W \rho \text{sd}_c \text{sd}_e) \right)}{\Delta \text{NMB}^2}
\]

\[
n = \frac{2 (z_\alpha + z_\beta)^2 \left( \text{sd}_e^2 \right)}{\Delta Q^2}
\]

Differences in Formulas

\[
\text{Var}_{\text{NMB}} = \text{sd}_c^2 + (W^2 \text{sd}_c^2) - (2 W \rho \text{sd}_c \text{sd}_e)
\]

- Variance of NMB more complicated than variance for usual continuous clinical differences
  - Includes \( \rho \), the correlation of the difference between cost and effect
  - Includes \( W \), the decision threshold we are trying to rule out

Correlation

- The correlation of the difference in cost and effect indicates how changes in the difference in cost are related to changes in the difference in effect
  - Negative (win/win) correlation: increasing effects are associated with decreasing costs
    - e.g., asthma care
  - Positive (win/lose) correlation: increasing effects are associated with increasing costs
    - e.g., life-saving care
  - All else equal, fewer patients need to be enrolled when therapies are characterized by a positive correlation than when they are characterized by negative correlation
Sample Size Tables, SD

- We commonly construct sample size tables for different values of \( \Delta C, \Delta Q \), the standard deviations for \( C \) and \( Q \), and \( W \)

<table>
<thead>
<tr>
<th>SD( _C )</th>
<th>N/Group</th>
<th>SD( _Q )</th>
<th>N/Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>2500</td>
<td>306</td>
<td>0.1</td>
<td>114</td>
</tr>
<tr>
<td>5000</td>
<td>340</td>
<td>0.2</td>
<td>340</td>
</tr>
<tr>
<td>7500</td>
<td>389</td>
<td>0.3</td>
<td>710</td>
</tr>
<tr>
<td>10,000</td>
<td>455</td>
<td>0.4</td>
<td>1224</td>
</tr>
<tr>
<td>15,000</td>
<td>634</td>
<td>0.6</td>
<td>2685</td>
</tr>
</tbody>
</table>

\( \Delta C=250; \Delta Q=0.05; \) unless otherwise specified, \( sdc=5000; sdq=.2; \rho=-1; \alpha=.05; \beta=.8 \)

“Typical” Sample Size Table, W

<table>
<thead>
<tr>
<th>WTP</th>
<th>Sample Size Per Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp 1</td>
<td>Exp 2 *</td>
</tr>
<tr>
<td>20,000</td>
<td>321</td>
</tr>
<tr>
<td>30,000</td>
<td>273</td>
</tr>
<tr>
<td>50,000</td>
<td>234</td>
</tr>
<tr>
<td>75,000</td>
<td>214</td>
</tr>
<tr>
<td>100,000</td>
<td>204</td>
</tr>
<tr>
<td>150,000</td>
<td>194</td>
</tr>
</tbody>
</table>

* \( \Delta C=-120; \Delta Q=0.015; sdc=1000; sdq=.05; \rho=-.8; \alpha=.05; 1-\beta=.8 \)

Sample Size Can Increase with Increasing W

<table>
<thead>
<tr>
<th>WTP</th>
<th>Sample Size Per Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp 1</td>
<td>Exp 2 *</td>
</tr>
<tr>
<td>20,000</td>
<td>321</td>
</tr>
<tr>
<td>30,000</td>
<td>273</td>
</tr>
<tr>
<td>50,000</td>
<td>234</td>
</tr>
<tr>
<td>75,000</td>
<td>214</td>
</tr>
<tr>
<td>100,000</td>
<td>204</td>
</tr>
<tr>
<td>150,000</td>
<td>194</td>
</tr>
</tbody>
</table>

* \( \Delta C=-120; \Delta Q=0.015; sdc=1000; sdq=.05; \rho=.8; \alpha=.05; 1-\beta=.8 \)
Sample Size Not Necessarily Monotonic With W

<table>
<thead>
<tr>
<th>WTP</th>
<th>Exp 1</th>
<th>Exp 2</th>
<th>Exp 3 *</th>
</tr>
</thead>
<tbody>
<tr>
<td>20,000</td>
<td>321</td>
<td>36</td>
<td>178</td>
</tr>
<tr>
<td>30,000</td>
<td>273</td>
<td>42</td>
<td>158</td>
</tr>
<tr>
<td>50,000</td>
<td>234</td>
<td>68</td>
<td>151</td>
</tr>
<tr>
<td>75,000</td>
<td>214</td>
<td>92</td>
<td>154</td>
</tr>
<tr>
<td>100,000</td>
<td>204</td>
<td>108</td>
<td>156</td>
</tr>
<tr>
<td>150,000</td>
<td>194</td>
<td>127</td>
<td>160</td>
</tr>
</tbody>
</table>

* \( \Delta C = -120; \Delta Q = 0.015; sd_c = 1000; sd_q = 0.05; p = 0; \alpha = 0.05; 1-\beta = 0.8 \)

Where to Obtain the Necessary Data?

- When therapies are already in use: Expected differences in outcomes and standard deviations can be derived from feasibility studies or from records of patients
- Simple correlation between observed costs and effects may be an adequate proxy for the measure of correlation used for estimating sample size
- For novel therapies, information may need to be generated by assumption
  - e.g., sd from usual care will apply to new therapy, etc.

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:

