Cost Effectiveness Analysis Within an RCT

Henry Glick
University of Pennsylvania

www.uphs.upenn.edu/dgimhsr

Pelvic Floor Disorders Network
10/11/12

Outline

- (Very) Brief introduction to economic evaluation
- (Very) Brief description of the ideal economic evaluation in a clinical trial
- 5 issues in designing economic evaluations in clinical trials
  - What Medical Service Use Should We Collect?
  - How Should We Value Medical Service Use?
  - What Effectiveness Measure Should We Use?
  - How Naturalistic Should the Study Be?
  - What is the Right Sample Size?

Types of Analysis

- Cost identification
- Cost-effectiveness
- Cost-benefit
- Cost-utility
- Net monetary benefit

Cost-Identification / Cost-Minimization

- Estimates costs of an intervention, but not benefits
- Appropriate when two therapies of equal efficacy are compared
- Introduction of sampling uncertainty to the evaluation of cost-effectiveness tends to undermine use of cost-identification analysis
  - When effects don't differ significantly, often can't determine whether one therapy costs more, does more, and is good value or whether the alternative therapy costs less and does more

Cost-Effectiveness Analysis

- Estimates costs and outcomes of intervention, but the two are measured in different units
- Results meaningful in comparison with other interventions or a predetermined threshold / cut-off for willingness to pay (WTP)

Cost-Benefit Analysis

- Estimates costs and benefits in the same (usually monetary) units
- Analysis based on the difference in costs and the difference in benefits
- Don't need to compare results with other interventions or a predetermined threshold / cut-off

Other Types of Analyses

- Cost-utility analysis
  - Form of cost-effectiveness analysis in which the effectiveness measure is expressed in terms of utility (e.g., quality-adjusted life years)
- Net monetary benefits
  - Part cost-effectiveness, part cost-benefit
  - Multiply difference in effectiveness by threshold WTP and subtract costs
    - Results greater than zero indicate that the value of the difference in effects is greater than the difference in costs

Review

- Investigators compared 2 treatments, “LessCost” and “MoreCure”
- They found that “LessCost” was less expensive and recommended its adoption by physicians
  - 100,000 vs 300,000
- What type of economic analysis are the investigators carrying out?
- Do you agree with their conclusion?

Example 2

- Investigators compared 2 treatments, “LessCost” and “MoreCure.” They observed the following:

	LessCost	MoreCure	Difference
Cost	100,000	300,000	-200,000
Benefit	100,000	500,000	400,000

- The authors concluded that MoreCure is net beneficial.
- What type of economic analysis are the investigators carrying out?
- Do you agree with their conclusion?

Example 3

- Investigators compared 2 treatments, “LessCost” and “MoreCure.” They observed that MoreCure cost 200,000 more than LessCost and provided 8 additional QALYs, i.e., 25,000 per QALY
- The authors recommended that MoreCure was good value for the cost
- What type of economic analysis are the investigators carrying out?
- Do you agree with their conclusion?

Types of Costs

- Direct: medical or nonmedical
- Time costs: Lost due to illness or to treatment
- Intangible costs
- Types of costs included in an analysis depend on:
  - What is affected by illness and its treatment
  - What is of interest to decision makers
    - e.g., a number of countries’ decision makers have indicated they are not interested in time costs

Study Perspective

- Economic studies should all adopt 1 or more “perspectives”
  - Societal
  - Payer (often insurer)
  - Provider
  - Patient
- Perspective helps identify what services should be included in the analysis and how these services should be cost out
  - e.g., patient out-of-pocket expenses may be excluded from insurer perspective
  - Not all payments may represent costs from the societal perspective

Good Value for the Cost

- Economic data collected as secondary (or primary) endpoints in randomized trials commonly used in the evaluation of the value for the cost
  - Short-term economic impacts directly observed
    - Within-trial analysis
  - Longer term impacts potentially projected by use of decision analysis
    - Long term projection
  - Reported results: point estimates and confidence intervals for estimates of:
    - Incremental costs and outcomes
    - Comparison of costs and effects

Sample Results Table

Analysis	Point Estimate	95% CI
Incremental Cost	-713	-2123 to 783
Incremental QALYs	0.13	0.07 to 0.18
Cost-Effectiveness Analysis		
Principal Analysis	Dominates	Dom to 6650
Survival Benefit		
-33%	Dominates	Dom to 9050
+33%	Dominates	Dom to 5800
Hospitalization Cost		
-50%	Dominates	Dom to 5300
+50%	Dominates	Dom to 8400
Drug Cost		
-50%	Dominates	Dom to 4850
+50%	Dominates	Dom to 8750
Discount rate		
0%	Dominates	Dom to 6350
7%	Dominates	Dom to 7000

Steps in Economic Evaluation

Step 1: Quantify 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 ratio of the cost per quality-adjusted life year saved is significantly less than $75,000

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)

- 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
- 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

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 fully learned 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 needed to answer economic questions may be longer than follow-up needed 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 problem: Don't have sufficient resources to track all medical service use

Limited Data Collection Resources

- Availability of administrative data may reduce costs of 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
- While accounting for the expense of collecting particular data items

Document Likely Service Use During Trial Design

- Can improve decisions by documenting types of services used by patients who are similar to those 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

Limit Data to Disease-Related Services?

- Little if any evidence about the accuracy, reliability, or validity of such judgments
- Easy for judgments to be flawed
- 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% ($p = 0.006$)

General Recommendations

- General Strategy: Identify a set of medical services for collection, and assess them any time they are used, independent of the reason for their use
- Decision to collect service use independent of the reason for use does not preclude ADDITIONAL analyses testing whether designated “disease-related” costs differ

Issue #2. How Should We Value Medical Service Use?

- Availability of billing data may simplify valuation
- If billing data aren’t available, common strategy is to measure service use in the trial and identify price weights (unit costs) to value this use

Common Sources of Price Weights

- Hospital care
  - Hospital bills adjusted by Federal cost-to-charge ratios
  - DRG payments
  - National inpatient sample
    - Calculator or dataset
  - Other administrative databases that include patient-level clinical and cost information
- Physician services
  - Medicare fee schedule
  - Other administrative databases

Common Sources (2)

- Laboratory tests
  - Clinical Diagnostic Laboratory Fee Schedule
- Durable equipment
  - Medicare Durable Good Fee Schedule
- Pharmaceuticals
  - Federal Supply Schedule
  - Adjusted AWP

Concomitant Medications

- Common to be very precise when costing study medications
- Greater problems posed by costing out concomitant medications
  - Number of agents / routes of administation / dosages / # of doses
- In many studies, investigators simplify the process:
  - Categorize drugs into classes
  - Identify 1 or 2 representatives of the class (including route / dosage / # of doses)
  - Cost out representative drugs and use their cost as the cost for all members of the class

Issue #3. What Effectiveness Measure?

- Can calculate a ratio for any outcome
  - Cost per toe nail fungus day averted
- To be an informative cost-effectiveness ratio, must be one for which we know what we are willing to pay
  - In many jurisdictions, the quality-adjusted life year (QALY) represents the recommended outcome of cost-effectiveness analysis
- Some resistance to this outcome, particularly from Congress
  - [PCORI] “shall not develop or employ a dollars per quality adjusted life year (or similar measure that discounts the value of a life because of an individual’s disability) as a threshold to establish what type of health care is cost effective or recommended”

QALYs

- Economic outcome that combines preferences for both length of survival and its quality into a single measure
- Help us decide how much we should pay for:
  - A therapy that saves fully functional lives/life years

    VS
  - A therapy that saves less than fully functional lives/life years (e.g., a drug for heart failure that extends survival, but patients spend the extra time in NYHA class III)

    VS
  - A therapy that doesn't save lives/life years but improves patients' functioning (e.g., patients with heart failure spend most of their remaining years in NYHA class I instead of NYHA class III)

QALY Scores

- QALY or preference scores generally range between 0 (death) and 1 (perfect health)
  - For example, a health state with a preference score of 0.8 indicates that a year in that state is worth 0.8 of a year with perfect health
  - There can be states worse than death with preference scores less than 0

Prescored Health State Classification Instruments

- Dominant approach for QALY measurement uses prescored health state classification instruments (indirect utility assessment)
- Participants' report their functional status across a variety of domains
- Preference scores derived from scoring rules that have usually been developed by use of samples from the general public

Prescored Instruments

- A number of prescored instruments are currently available for the measurement of preference scores for current health
  - The EuroQol instrument (EQ-5D), 3 and 5 level
  - The Health Utilities Index Mark 2 (HUI2)
  - The Health Utilities Index Mark 3 (HUI3)
  - The SF-6D
- Most of these instruments ask participants or their proxies to report on the health status of the patient

The EQ-5D, the HUI2, and the HUI3

- The EQ-5D, the HUI2, and the HUI3 are three of the most commonly used prescored preference assessment instruments
- All three of these instruments share features of ease of use
  - e.g., high completion rates and the ability to be filled out in 5 min or less
- All have been used to assess preferences for a wide variety of diseases

Comparison of Instruments

- With the exception of 1 large study, there is substantial evidence that the HUI2 yields significantly higher scores than either the EQ-5D 3 level or the HUI3
- Evidence about the EQ-5D 3L and HUI3 less consistent
  - Most studies found no significant differences
  - One large study found that the EQ-5D scores were significantly lower than HUI3 scores
  - A second large study found that the EQ-5D scores were significantly higher than the HUI3 scores

Superiority?

- Widespread direct comparison of the instruments doesn't provide an answer about which instrument should be used in which circumstances
- Most studies evaluating correlations between the preference scores find good correlation
- Most studies evaluating correlations between preference scores and convergent validity criteria find good correlation
- Most studies evaluating responsiveness find all instruments to be responsive
- Most studies have concluded that there is little evidence that one instrument was superior to another

Issue # 4. How Naturalistic?

- 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

#4a. Intention to Treat

- 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

#4b. 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

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

#4c. Protocol-Induced Costs and Effects

- Common concerns:
  - Standardization of care in clinical trial protocols often means that care delivered in trials differs from usual care
    - Protocol may require substantial number of investigations and diagnostic tests that would not be performed under normal clinical practice
  - Protocols often prescribe aggressive documentation and treatment of potential adverse effects that differ from usual care
- Omit these costs???

Omission of Protocol-Induced Costs

- Criterion for including costs should NOT be "Would the services have been provided in usual care
- Instead should be: "Could the services have affected care / outcomes (and thus costs)
- No problem omitting services that cannot affect care / services
  - e.g., Cost of genetic samples that will not be analyzed until after follow-up is completed
- More problematic to omit services that can change treatment and affect outcome
  - "Cadillac" costs may yield "Cadillac" outcomes
  - Would have to adjust both costs and their effects on outcomes

Biases?

- Protocol-induced testing may bias the testing cost to the null
  - There might be a difference in this testing in usual care, but it can't be observed if everyone routinely receives the test
- Protocol induced testing may bias cost and outcome it in an unknown direction
  - Trial's extra testing may detect and lead to treatment of outcomes that in usual care would not have been detected or treated

Issue #5. What is the Right 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 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

- At the most basic level, sample size for cost-effectiveness is calculated using the same formula as the sample size for a difference in any continuous variable:

$$n = \frac{2\,(z_{\alpha}+z_{\beta})^2\ \ sd^2_{nmb}}{\Delta nmb^2}$$

where n = sample size/group; z_{α} and z_{β} = z-statistics for α (e.g., 1.96) and β (e.g., 0.84) errors; sd = standard deviation for cost (sd_c) and effect (sd_q)

Sample Size Formula (2)

- Complexities arise because 1) the difference that is being assessed is the difference in NMB (WΔQ – ΔC) and 2) the standard deviation of NMB is a complicated formula
- Data needed to calculate sample size include:
  - Difference in cost
  - SD, difference in cost
  - Difference in effect
  - SD, difference in effect
  - Z_{α} and Z_{β}
  - Correlation of the difference in cost and effect
  - Willingness to pay

Correlation of the Difference

- When increasing effects are associated with decreasing costs, a therapy is characterized by a negative (win/win) correlation between the difference in cost and effect
  - e.g., asthma care
- When increasing effects are associated with increasing costs, a therapy is characterized by a positive (win/lose) correlation between the difference in cost and effect
  - e.g., life-saving care
- 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

Effect of SD_q VS SD_c on Sample Size

- Commonly thought that sample size for cost-effectiveness driven more by the standard deviation for cost than it is by SD for effect
  - If not, why would we need a larger sample for the economic outcome than you do for the clinical outcome?
- However, if willingness to pay is substantially greater than the standard deviation for cost, percentage changes in QALY SD can have a substantially greater effect on sample size than will equivalent percentage changes in cost SD

Economic Vs Clinical Sample Sizes

- Sample size required to answer economic questions often larger than the sample size required to answer clinical questions
  - But it need not be
- ΔC and ΔQ are a joint outcome just as differences in nonfatal CVD events and all cause mortality are often combined into a joint outcome
- In the same way that we can have more power for the joint cardiovascular outcome than either individual outcome alone, we can have more power for cost-effectiveness than we do for costs or effects alone

Willingness to Pay and Identification of an Appropriate Outcome Measure

- Sample size calculations require stipulation of willingness to pay for a unit of outcome
- In many medical specialties, researchers use disease specific outcomes
- Can calculate a cost-effectiveness ratio for any outcome (e.g., cost/case detected; cost/abstinence day), to be informative, 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

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 data are analyzed appropriately, trial-based evaluations may 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