Designing Economic Evaluations in Clinical Trials

Henry Glick
University of Pennsylvania

Society for Medical Decision Making Short Course #AM7

Phoenix, Az
October 20, 2012

Good Value for the Cost

- Economic data collected as secondary (or primary??) endpoint 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

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 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, net monetary benefit, or the probability that the ratio is acceptable
- Potential hypothesis: Cost per quality-adjusted life year saved 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

Design Issues Not Unique To Trials

- A number of design issues apply equally to economic evaluations in clinical trials and to other economic evaluations:
  - The type of analysis that will be conducted
  - The types of costs that will be included
  - The study perspective
- Issues 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 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
  - Don't always expect to affect 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 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
  - E.g., collecting 6700 blood gas tests that accounted for 1.8% of lab costs vs 420 cardiac studies that represented 4.3%

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% ($p = 0.006$)

Blinded Vs Unblinded Studies

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

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

#2a. 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

#2b. 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 isn't 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

#2c. 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
    - e.g., 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"
- 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 need 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 testing in usual care, but it can't be observed if everyone routinely receives the test
- Protocol induced testing may bias treatment cost and outcome in an unknown direction
  - Trial's extra testing may lead to:
    - Detection and treatment of outcomes that wouldn't have been detected or treated in usual care
    - Earlier detection and treatment of problems when they are less severe and easier to treat
- Adjustment requires assumptions about what would or wouldn't have been detected in usual 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., we may hypthesize that the point estimate for the cost-effectiveness ratio will be 20,000 per QALY
  - May want to design an experiment that provides 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

Basic 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)

Complexities

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

Sample Size / Power Formulas

- Sample Size

$$n = \frac{2\left(z_{\alpha}+z_{\beta}\right)^2\left(sd_c^2 + \left(W\,sd_q\right)^2 - \left(2\,W\,\rho\,sd_c\;sd_q\right)\right)}{\left(W\Delta Q - \Delta C\right)^2}$$

- Power

$$z_{\beta} = \sqrt{\frac{n * \left(W\Delta Q - \Delta C\right)^2}{2\left(sd_c^2 + \left(W\,sd_q\right)^2 - \left(2\,W\,\rho\,sd_c\;sd_q\right)\right)}} - z_{\alpha}$$

  - e.g., if z_{β} = -1.96 = 2.5% power; -0.84 = 20% power; 0 = 50% power; .84 = 80% power; 1.28 = 90%

Correlation of the Difference

- 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 positive correlation than when they 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

Sample Size Tables, SD

- We commonly construct sample size tables for different values of ΔC, ΔQ, the standard deviations for C and Q, and W

SD_c	N/Group	SD_q	N/Group
2500	306	0.1	114
5000	340	0.2	340
7500	389	0.3	710
10,000	455	0.4	1224
15,000	634	0.6	2685

ΔC=250; ΔQ=.0.05; unless otherwise specified, sd_c=5000; sd_q=.2; ρ=-.1; α=.05; β=.8

"Typical" Sample Size Table, W

	Sample Size Per Group
WTP	Exp 1 *
20,000	321
30,000	273
50,000	234
75,000	214
100,000	204
150,000	194

* ΔC=-120; ΔQ=0.015; sd_c=1000; sd_q=.05; ρ= -.8; α=.05; 1-β=.8

Sample Size Can Increase with Increasing W

	Sample Size Per Group	
WTP	Exp 1 *	Exp 2 *
20,000	321	36
30,000	273	42
50,000	234	68
75,000	214	92
100,000	204	108
150,000	194	127

* ΔC=-120; ΔQ=0.015; sd_c=1000; sd_q=.05; ρ= .8; α=.05; 1-β=.8

Sample Size Not Necessarily Monotonic with W

	Sample Size Per Group		
WTP	Exp 1 *	Exp 2 *	Exp 3 *
20,000	321	36	178
30,000	273	42	158
50,000	234	68	***151***
75,000	214	92	154
100,000	204	108	156
150,000	194	127	160

* ΔC=-120; ΔQ=0.015; sd_c=1000; sd_q=.05; ρ= 0; α=.05; 1-β=.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.

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

Further reading:

Glick HA, Doshi JA, Sonnad SS, Polsky D. Economic Evaluation in Clinical Trials. Oxford: Oxford University Press, 2007.

Glick HA. Sample size and power for cost-effectiveness analysis (part 1). Pharmacoeconomics. 2011;29;189-98.

Glick HA. Sample size and power for cost-effectiveness analysis (part 2). The effect of maximum willingness to pay. Pharmacoeconomics. 2011;29:287-96.