

Designing Economic Evaluations in Randomized Trials

Henry Glick

January 2007

www.ups.upenn.edu/dgimhsr



Good Value for the Cost

- Cutting edge evaluation of the value for the cost is supported by use of an evaluation from a randomized trial that included economic outcomes as primary or secondary endpoints
 - 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, outcomes, and the comparison of costs and effects
 - Impact of sensitivity analysis on the comparison of costs and effects judged by its impact on both the point estimates and the confidence intervals of the ratios



Example

Analysis	Point Estimate	95% CI
Incremental Costs	-713	-2123 to 783
Incremental QALYs	0.13	0.07 to 0.18
Cost-Effectiveness Ratios		
Principal Analysis	Dominates	Dom to 6650
Survival benefit		
-33%	Dominates	Dom to 9050
+33%	Dominates	Dom to 5800
Hospitalization costs		
-50%	Dominates	Dom to 5300
+50%	Dominates	Dom to 8400
Drug costs		
-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
- Strategic issues



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

- 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 analysis



Ideal Economic Evaluation Within a Trial

- An ideal economic evaluation within a 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



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 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

- 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 new therapy 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 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



6 Strategic Issues

- What medical service use should one collect?
- How should one value medical service use
- What is the appropriate sample size?
- How should one interpret results from multicenter studies?
- Is there a treatment-by-time interaction?
- How naturalistic should the study design be?



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

- Real/perceived problems
 1. Don't have sufficient resources to track all medical service use
 2. (In some cases), 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"
 - Why not limit data to disease-related services?



Limited Data Collection Resources

- Access to billing data may obviate resource limitation associated with 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

- The best approach is to measure as many services as possible
 - There are no a priori guidelines about how much data are enough, nor are there data on the incremental value of specific items in the economic case report form




Measure as Much as Possible (II)

- Decisions improved if one documents 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
- Must 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 angiocardionpneumographies 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 attributes AEs to the intervention, even when participants received vehicle/placebo
- Much of medical practice is 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" were reduced by 30% (combined endpoint, death and HF hospitalization, $p < 0.0001$); simultaneously, hospitalizations for noncardiovascular reasons were 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



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
- 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, comedications



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.



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

- Availability of billing data may simply valuation
- If billing data aren't available, collect price weights for a selected set of medical services from a selected set of countries
 - For international studies, most often derived from a national data or a single center per country
- Sample sources of data:
<http://www.dh.gov.uk/PolicyAndGuidance/OrganisationPolicy/FinanceAndPlanning/NHSReferenceCosts/fs/en>
<http://new.cms.hhs.gov/PhysicianFeeSched/PFSRVF/list.asp#TopOfPage>
<http://new.cms.hhs.gov/AcuteInpatientPPS/FFD/list.asp>



Price Weights from Which Centers / Countries

- The centers/countries from which price weights are collected might be ones:
 - That enroll a large number of patients
 - That represent the spectrum of economic conditions
 - In which regulators require a submission
 - For which price weights are readily available
 - In which the sponsor wishes to make economic claims



Estimating Missing Price Weights

- Eventually, one will need to identify price weights for all medical services recorded in the case report form
- Because collecting price weights for all services may be expensive, one commonly:
 - Collects price weights for service use that 1) occurs most frequently in the trial, 2) is considered likely to be affected by the intervention, or 3) has particularly high or low costs
 - Develops a method of imputation (e.g., by use of national DRG weights or weights derived from fee schedules, etc) to estimate price weights that haven't been collected



More Price Weights/Fewer Centers or the Reverse?

- Presuming one is using a reliable method for imputing price weights (e.g. DRG weights), do we know anything about how we should trade-off number of centers/countries sampled versus number of price weights per center/country?



More Price Weights/Fewer Centers/Countries or the Reverse (II)

- In simulations based on data from 4 countries:
 - If the number of price weights one plans to collect is fixed:
 - It is better to sample a smaller number of price weights in more centers than it is to sample a larger number of price weights in fewer centers
 - e.g., in our simulations the imputation error was smaller when 12 price weights were collected in each of 4 countries than it was when 47 were collected in a single country

Glick HA, Orzol SM, Tooley JF, Polsky D, Mauskopf JM. Design and Analysis of Unit Cost Estimation Studies: How Many Hospital Diagnoses? How Many Countries? Health Economics. 2003;12:517-27.



Center/Country-Specific vs Averaged Price Weights

- Once one has a number of different sets of price weights (e.g., weights from multiple centers that participated in the trial), how should they be used to construct the cost outcome of the trial?



Center/Country-Specific vs Averaged Price Weights
(II)

- Ideal: Because relative prices can affect quantities of services provided, where ever feasible, multiply center-specific price weights times times center-specific counts of medical services
- For centers for which price weights aren't available:
 - Use (averages of) price weights from similar centers
 - e.g., in a trial that enrolls patients in community and tertiary care hospitals, one might average price weights from community hospitals to value service use in a community hospital, but wouldn't want to use this average for tertiary care hospitals



Center/Country-Specific vs Averaged Price Weights
(III)

- Corollary: If one has a set of price weights for each center that participated in the trial, one should not average them and use this average for all services measured in the trial
 - The most common reasons suggested for such a strategy are that 1) reducing variability in the price weights will reduce variability in the estimated costs and 2) an average set of price weights may be more representative



Center/Country-Specific vs Averaged Price Weights
(IV)

- However:
 - Empirically, use of a single set of price weights need not reduce variance
 - If substitution effects are strong, this strategy may introduce bias in the estimates of cost differences
 - Why is it more "representative" to use a set of price weights that no one faces?



Issue #3. What is the Appropriate Sample Size to Address Economic Questions?

- Prior to the literature that described confidence intervals for cost-effectiveness ratios, sample size was commonly based 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
- 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)



Sample Size Formula

- Sample size for NMB uses the following formula:

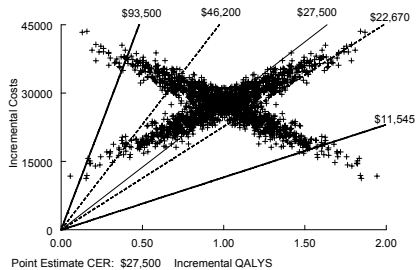
$$n = \frac{(\alpha + \beta)^2 (2 \text{sd}_c^2) + (2 \text{rc}^2 \text{sd}_q^2) - (2 \text{rc} \rho (2 \text{sd}_c^2)^{0.5} (2 \text{sd}_q^2)^{0.5})}{\Delta \text{NMB}^2}$$

where n equals n/group; sd = the standard deviation for costs (c) and effects (q); rc equals the maximum willingness to pay one wishes to rule out; and ρ equals the correlation of the difference in cost and effect; and ΔNMB equals (rc Δq) – Δc

<http://www.uphs.upenn.edu/dgimhsr/stat%20samps.htm>



Correlation Between Costs and Effects



Correlation Between Costs and Effects

- Win/Lose (positive) correlation: As the effectiveness (cost) increases, the cost (effectiveness) increases
- Win/Win (negative) correlation: As the effectiveness (cost) increases, the cost (effectiveness) decreases
- Correlation between costs and effects can have dramatic effects on the confidence interval for the cost effectiveness ratio/NMB and thus on the sample size required to demonstrate value for the cost



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
 - Potential sources
 - Medical charts of administrative data sets
 - Patient logs of their health care resource use
 - Asking patients and experts about the kinds of care received by those with the condition under study
 - In addition, at least one study has suggested that the simple correlation between costs and effects observed in these data may be an adequate proxy for the measure of correlation used for estimating sample size



Obtaining Data for Novel Therapies

- For novel therapies, information about the magnitude of the incremental costs and outcomes may not be available
 - May need to be generated by assumption
 - Data on the standard deviations for those who receive usual care/placebo may be obtained from feasibility studies or patient records
 - One may assume that the standard deviation will apply equally to both treatment groups, or one may make alternative assumptions about their relative magnitudes
 - The correlation also would be obtained from such data



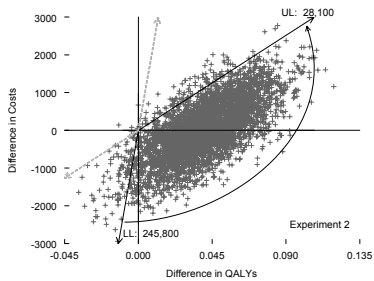
Power and the Joint Outcome of Cost and Effect

- What would your clinical colleagues conclude about this therapy's clinical effectiveness?
- Your economic colleagues about whether it saves costs

Variable	Mean	SE	T	P-value
Cost	35	777.5	0.26	0.96
QALY	.04	.0224	1.786	0.07



Power and the Joint Outcome (II)



- 95% CI for the cost-effectiveness ratio indicates that for willingnesses to pay ranging between \$28,200 and \$245,200, one can be 95% confident of adoption



Maximum Willingness to Pay 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
- In many medical specialties, researchers use disease specific outcomes
- 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



Issue #4. How Should One Interpret Results From Multicenter (Multinational) Trials?

- The Problem:
 - There has been growing concern that the pooled (i.e., average) economic results from multinational trials may not be reflective of the results that would be observed in individual countries that participated in the trial
 - Similar issues arise for any subgroup of interest in the trial (e.g., more and less severely ill patients)



Common Sources For Concern

- Transnational differences in morbidity/mortality patterns; practice patterns (i.e., medical service use); and absolute and relative prices for this service use (i.e., price weights)
- Thus decision makers may find it difficult to draw conclusions about the value for the cost of the therapies that were evaluated in multinational trials



Bad Solutions

- Use trial-wide clinical results, trial-wide medical service use, and price weights from one country
 - e.g., to tailor the results to the U.S., just use U.S. price weights, and conduct the analysis as if all participants were treated in the U.S.
- Use trial-wide clinical results and use costs derived from the subset of patients treated in the country
- Ignore the fact that clinical and economic outcomes may influence one another (cost affects practice which affects outcome; practice affects outcome which affects cost)



Impact of Price Weights vs Other Variation

Country	Trial-Wide Effects		Country Specific Costs and Effects
	Price Weights	Country Specific Costs †	
1	46,818	5,921	11,450
2	57,636	91,906	60,358
3	53,891	90,487	244,133
4	69,145	93,326	181,259
5	65,800	**	**
Overall	45,892	45,892	45,892



Two Analytic Approaches To Transferability

- Two approaches -- which rely principally on data from the trial to address these issues -- have made their way into the literature
 - Hypothesis tests of homogeneity (Cook et al.)
 - Multi-level random-effects model shrinkage estimators



Hypothesis Tests Of Homogeneity

- Evaluate the homogeneity of the results from the different countries
 - If there is no evidence of heterogeneity (i.e., a nonsignificant p-value for the test of homogeneity), and if one believes the test was powerful enough to rule out economically meaningful differences in costs, then one cannot reject that the pooled economic result from the trial applies to all of the countries that participated in the trial
 - If there is evidence of heterogeneity, then the method indicates one should not use the pooled estimate to represent the result for the individual countries, but this method is less clear about the result that should be used instead



Estimation

- Multi-level random-effects model shrinkage estimation assesses whether observed differences between countries are likely to have arisen simply because one has divided the trial-wide sample into subsets or whether they are likely to have arisen due to systematic differences between countries
 - Borrows information from the mean estimate to add precision to the country-specific estimates
 - These methods have the potential added advantage of providing better estimates of the uncertainty surrounding the pooled result than naive estimates of the trial-wide result



Issue #5. 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, one should consider the likelihood that the cost-effectiveness ratio observed in the trial will describe longer term therapy
- To address this uncertainty 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)



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 one decides that the primary analysis will be a projection beyond the period of observation, one should still 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

- To investigate whether the cost-effectiveness ratio is homogeneous with respect to time, one 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, one should make projections for different time horizons
 - Even if the longest time horizon in a lifetime projection is 30-40 years, one may observe that projected long-term cost-effectiveness reaches equilibrium after only 5 or 10 years of projection
- To add face validity to the trajectory of the projected cost-effectiveness ratios:
 - 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.)



CER And CI Within the Trial and Projected

Years of Follow-up	Point Estimate	95% CI
Within the trial		
1	Dominated	168,884 to Dominated
2	282,857	45,577 to Dominated
3	73,529	Dominates to Dominated
4	12,074	Dominates to Dominated
5	15,258	Dominates to 122,772
Longer term projection		
10	12,246	Dominates to 42,263
15	8,578	Dominates to 26,721
20	7,320	681 to 21,841



Issue #6. 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, in terms of participants, analysis based on the intention to treat, and limitation of loss to follow-up, the more likely 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 benefits 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 subjects 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



Naturalism: Loss to Follow-up

- Investigators should also ensure that:
 - follow-up continues until the end of the study period
 - data collection not be discontinued simply because a subject 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 (II)

- 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



Naturalism: Protocol-Induced Costs and Effects (III)

- Adjustment for this extra detection -- potentially by use of decision analytic models -- may be difficult, because information usually is not available from the trial about whether active therapy avoided complications that would have been detected and treated in routine practice or whether it avoided those that never would have been detected



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