

ECONOMIC EVALUATION IN RANDOMIZED TRIALS

Henry Glick
University of Pennsylvania Health System
<http://www.uphs.upenn.edu/dgimhsr>

World Wide Outcomes Research Forum 2004
Pfizer, Inc.
New York, New York
April 21, 2004

Further reading: Glick HA, Polsky D, Schulman K. Trial-based Economic Evaluations: An Overview of Design and Analysis. Chapter 6 in *Economic Evaluation in Health Care. Merging Theory and Practice*, ed. By Michael Drummond and Alistair McGuire, Oxford: Oxford University Press, 2001, 113-140

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 differences in costs and outcomes and evaluate "value for cost" (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: Evaluate stochastic uncertainty and perform sensitivity analysis

IDEAL ECONOMIC EVALUATION WITHIN A TRIAL

- An ideal economic evaluation within a clinical 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 reason for the costs
- Most feasible when:
 - Easy to identify when services are provided
 - Service / cost data already being collected
 - Ready access to data

IDEAL ECONOMIC EVALUATION WITHIN A TRIAL (III)

- A number of design issues apply equally to economic evaluations that are incorporated within clinical trials as well as 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
- However, these trials may be the only source of information needed for important early decisions about the adoption and diffusion of the therapy
- TRADEOFF: ideal vs best feasible study

OUTLINE

- Could devote days to discussing issues related to design and analysis of economic assessments in trials
- Focus on 5 major sets of issues:
 1. What medical service use should one collect?
 2. How should this medical service use be valued?
 3. Will the trial be powered to yield economic results in which one can be confident?
 4. How should one interpret results from multinational (multicenter) trials?
 5. What should one do if the full benefits (and costs) of therapy aren't expected to be observed during the period of observation in the trial?
- End with practical advice about implementing economic assessments in trials

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 (I)

- Access to billing data may obviate resource limitations associated with tracking all medical service use
- If administrative data 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 the services that go unmeasured reduces the likelihood that differences among them will lead to biased estimates
 - * Provides a measure of overall variability of costs

LIMITED DATA COLLECTION RESOURCES (II)

- 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
- Decisions about the services to measure will be improved if, when designing a study, 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

LIMITED DATA COLLECTION RESOURCES (II)

- 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 angiocardiopneumographies 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:
 - 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 (positively or negatively) 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?
- Potential biases more of a problem in unblinded studies, but need not "balance out" in double-blinded studies

DATA COLLECTION 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 was 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

DATA COLLECTION RECOMMENDATIONS (II)

- 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. VALUING MEDICAL SERVICE USE

- Collect price weights for a selected set of medical services from a selected set of countries
 - Most often, derived from a single source per country
 - * Sometimes based on national data, sometimes on data from a single center
 - * The number of sources (e.g., centers) may affect the results of the study

PRICE WEIGHTS FROM WHICH COUNTRIES?

- The countries from which price weights are collected might be ones:
 - That enroll a large number of patients in the trial
 - That represent the spectrum of economic development among countries that participated in the trial
 - In which the countries' regulators require a submission for reimbursement
 - For which price weights are readily available
 - In which the study 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 using national DRG weights or weights derived from fee schedules, etc) to estimate price weights that haven't been collected

MORE PRICE WEIGHTS/FEWER COUNTRIES VS MORE COUNTRIES/FEWER PRICE WEIGHTS

- Presuming one is using a reliable method for imputing price weights, do we know anything about how we should trade-off number of countries sampled vs number of price weights per country?
- In simulations we performed to evaluate whether it was better to collect fewer price weights in more countries or more price weights in fewer countries, we found:
 - Imputation error decreased as the number of medical services and countries sampled increased, but the rate of reduction in this error shrank
 - Error was minimized by obtaining estimates for fewer types of hospitalization -- in our experiment, at least 25 -- from as many countries as is feasible, rather than by obtaining estimates for more types of hospitalization in fewer countries
 - We did not evaluate such a strategy, but it may be more efficient to measure a core group of services in all countries plus a set of services that differ for different countries

COUNTRY-SPECIFIC VS. AVERAGED PRICE WEIGHTS

- Once one has a number of different sets of price weights (e.g., weights from 5 countries that participated in the trial), how should they be used to construct the cost outcome of the trial?
 - Ideal: Because relative prices can affect quantities of services provided, where ever feasible, one should multiply country-specific price weights times country-specific counts of medical service use
 - For countries for which price weights are not available
 - * Use (averages of) price weight estimates from similar countries
 - * e.g., in a trial that enrolls patients in Western Europe, Eastern Europe and South America, one might average price weights from the UK, Sweden, and France to value medical service use in Germany, but one wouldn't want to do so to value it in Eastern Europe and Latin America

COUNTRY-SPECIFIC VS. AVERAGED PRICE WEIGHTS (II)

- Corollary: If one has a set of price weights for each country that participated in the trial, one should not average them and use this average for all services measured in the trial
 - The most common reason cited for such a strategy is the idea that it minimizes variance
 - However:
 - * Empirically, this strategy need not reduce variance
 - * If substitution effects are strong, this strategy may introduce bias in the estimates of cost difference between the therapies under evaluation

ISSUE #3. POWER AND ECONOMIC ANALYSIS

- Prior to the development of the literature that described confidence intervals for cost-effectiveness ratios, a common approach for estimating sample size to base it 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 standard formula for continuous variables

$$n = \frac{2 \sigma_{\text{NMB}}^2 (1-R^2) (z_{\alpha/2} + z_{\beta})^2}{\Delta \text{NMB}^2}$$

where ΔNMB^2 equals the square of the difference in mean NMB between the therapies; σ^2 equals the variance of NMB; and $R^2 =$ variance explained in OLS (assuming OLS is used to predict the difference)

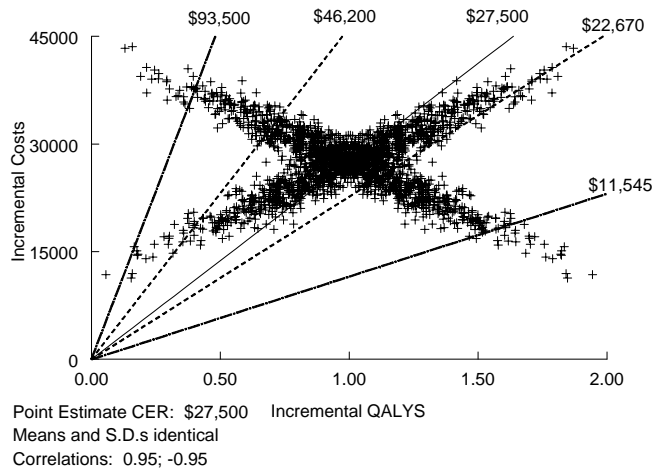
INFORMATION REQUIRED TO ESTIMATE SAMPLE SIZE FOR ECONOMIC EVALUATION

- While the sample size equation above is similar to that for any continuous variable, its error term is complicated by the fact that it includes variances and covariances for costs and effects

$$se^2(\text{NMB}) = s_{\Delta C}^2 + R_c^2 s_{\Delta E}^2 - 2 R_c \rho s_{\Delta C} s_{\Delta E}$$

- Basic data for such calculations include the magnitude of the incremental costs and outcomes; the standard errors for the difference in costs and outcomes; and **the correlation between costs and outcomes**

CORRELATION BETWEEN COSTS AND EFFECTS



- 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?

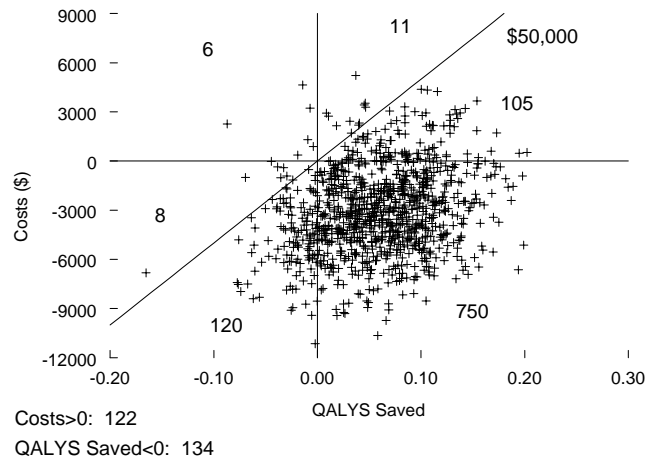
- If both therapies are already in use, expected differences in outcomes and standard deviations can be derived from feasibility studies or from records of patients like those who will be enrolled in the trial
 - Potential sources
 - * Medical charts of administrative data sets
 - * Patient logs of their 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

WHERE TO OBTAIN THE NECESSARY DATA? (II)

- 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

RELATIVE SAMPLE SIZE FOR CLINICAL AND ECONOMIC QUESTIONS

- Generally accepted that economic results have less power than do clinical results
- Exceptions exist:



- Results from the fact that the joint outcome can be more powerful than the individual outcomes
- Would you accept such a finding?

ISSUE #4. HOW SHOULD ONE INTERPRET RESULTS FROM MULTINATIONAL TRIALS?

- The Problem:
 - There has been growing concern that the pooled (i.e., average) clinical and economic results from multinational (or multicenter) 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; absolute and relative prices for medical service use (i.e., price weights); and practice patterns (i.e., medical service use)
- Thus decision makers may find it difficult to draw useful 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 center
 - 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
- These approaches ignore the fact that clinical and economic outcomes may influence one another (differences in costs may affect practice patterns, which in turn may affect outcomes; differing practice patterns may affect outcomes, which in turn may affect costs)

IMPACT OF UNIT COSTS VS OTHER VARIATION

Tirilazad Mesylate for subarachnoid hemorrhage, cost per death averted (subanalysis using data from 5 countries) *

| Country | Trial-Wide Effects | | Country Specific Costs and Effects |
|---------|--------------------|--------------------------|------------------------------------|
| | Unit Costs | 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 |

† Country-specific resource use × Country-specific unit costs

** New therapy dominates

* Willke RJ, et al. Estimating Country-Specific Cost-Effectiveness from Multinational Clinical Trials. Health Economics. 1998;7:481-93.

TWO APPROACHES TO TRANSFERABILITY

- Two approaches -- which rely principally on data from the trial to address these issues -- are currently making their way into the literature
 - Hypothesis tests of homogeneity
 - 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 can conclude 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

Cook JR, Drummond M, Glick H, Heyse JF. Assessing the appropriateness of combining economic data from multinational clinical trials. Statist. Med. 2003; 22:1955–76.

ESTIMATION

- The second method uses multi-level random-effects model shrinkage estimation to provide more precise estimates of the country-specific results than are yielded by separate -- and naive -- analysis of each country's costs and effects
 - Borrow 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
- MY PROGNOSTICATION: The strongest evidence for adopting a therapy will be derived from some version of the pooled result (and evidence that country-specific results aren't substantially different from the pooled result)

ISSUE #6. ESTIMATING LONG TERM EFFECTS

- 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 investigate 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)

ESTIMATING LONG TERM EFFECTS (II)

- 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 observed during the trial is likely to represent the ratio of longer-term therapy, one should also project the results for longer periods
- For projection: Maintain a common time horizon for both costs and effects
 - a number of 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 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 during the trial

TIME HORIZON FOR PROJECTION

- Given that the longer the projection, the less certain the results, one should project 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.)

COST EFFECTIVENESS RATIO AND 95% CI, 5.5-YEAR CHOLESTEROL LOWERING TRIAL

| Years of Follow-up/ Projection | 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 |
| 5.5 | \$15,258 | Dominates to \$122,772 |
| Projection beyond the trial | | |
| 10 | \$12,246 | Dominates to \$42,263 |
| 15 | \$8,578 | Dominates to \$6,721 |
| 20 | \$7,320 | \$681 to \$21,841 |

PRACTICAL COMMENTS

ORGANIZATIONAL ISSUES

- Undertaking an economic assessment requires a substantial effort by those organizing the studies, the clinical investigators, and the study subjects in the trial
- This effort includes collecting data before a trial is begun and investing in the trial itself
- a successful economic assessment requires cooperation and coordination among everyone involved in the trial, including the organizers, the clinical team from the company, the clinical researchers, and the study subjects. Disinterest or dissatisfaction among any of these groups can lead to collection of flawed data and can undermine the study

IMPLEMENTATION

- Cooperation and coordination
- Procedure manuals
- Instruction for data collectors
- Timely follow-up / answering of questions
- Timely monitoring of the data and source documentation
- Timely data entry, error detection and correction
- Investigator newsletters

COMMITMENT

- Be willing to invest sufficient resources in the study
 - What's acceptable vs. what's true
- Inform clinical investigators early in the planning
- Expect patient participation

RISK TAKING

- Investing resources early in a product's life
 - Potential clinical and economic loser
- Clinical winner but economic loser
 - Economic adverse events

a PRIORI DECISIONS NOT TO PERFORM AN ECONOMIC ASSESSMENT WITHIN THE TRIAL?

- When the economic advantage cannot be demonstrated within the trial, nor projected from it
 - Does not exist
 - Exists, but no evidence of it will be observable (even nonsignificantly) during the period of observation
- When it won't be needed
 - e.g., therapy so effective that people won't worry about its cost
 - * The early economic assessments of viagra generally were not funded by Pfizer
 - e.g., therapy so novel that clinicians will use it even if it isn't cost effective

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