DESIGNING ECONOMIC EVALUATIONS IN RANDOMIZED TRIALS

Henry Glick
University of Pennsylvania Health System

Statistical Considerations In Economic Evaluations
ISPOR 10th ANNUAL INTERNATIONAL MEETING
May 15 2005


“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 of the trial
  - Short-term economic impacts of the therapy are directly observed; longer term impacts potentially are be projected by use of decision analysis
  - Reported results include point estimates and confidence intervals for estimates of incremental costs, outcomes, and the comparison of costs and effects
  - The impact of sensitivity analysis on the comparison of costs and effects would be judged by its impact on both the point estimates and the confidence intervals of the ratios

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Point Estimate</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incremental Costs</td>
<td>-713</td>
<td>-2123 to 783</td>
</tr>
<tr>
<td>Incremental QALYs</td>
<td>0.13</td>
<td>0.07 to 0.18</td>
</tr>
<tr>
<td>Cost-Effectiveness Ratios</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Principal Analysis</td>
<td>Dominates</td>
<td>Dom to 6650</td>
</tr>
<tr>
<td>Survival benefit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-33%</td>
<td>Dominates</td>
<td>Dom to 9050</td>
</tr>
<tr>
<td>+33%</td>
<td>Dominates</td>
<td>Dom to 5800</td>
</tr>
<tr>
<td>Hospitalization costs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-50%</td>
<td>Dominates</td>
<td>Dom to 5300</td>
</tr>
<tr>
<td>+50%</td>
<td>Dominates</td>
<td>Dom to 8400</td>
</tr>
<tr>
<td>Drug costs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-50%</td>
<td>Dominates</td>
<td>Dom to 4850</td>
</tr>
<tr>
<td>+50%</td>
<td>Dominates</td>
<td>Dom to 8750</td>
</tr>
<tr>
<td>Discount rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0%</td>
<td>Dominates</td>
<td>Dom to 6350</td>
</tr>
<tr>
<td>7%</td>
<td>Dominates</td>
<td>Dom to 7000</td>
</tr>
</tbody>
</table>
OUTLINE

- Steps in economic evaluation
- The gold standard and its tensions
- 5 strategic issues
  1. What medical service use should one collect?
  2. How should one value the medical service use that is collected?
  3. What is the appropriate sample size for economic questions?
  4. How should one interpret results from multicenter (multinational) 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?

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

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 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

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
- 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

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 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

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, comedinations

BETTER APPROACH

- 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 THE MEDICAL SERVICE USE THAT IS COLLECTED?

- Availability of billing data may simplify 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 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
* Sample sources of national data:
  


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 in the trial
  - That represent the spectrum of economic conditions among centers/countries that participated in the trial (e.g., secondary vs. tertiary hospitals or developing/developed countries)
  - In which 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 centers/countries sampled vs number of price weights per center/country?

- In simulations based on data from 4 countries, we have found that:
  - If the number of price weights one plans to collect is fixed, and
  - If one has a reliable method for cost imputation (e.g., DRG weights),
  - Then, 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


CENTER/COUNTRY-SPECIFIC VS. AVERAGED PRICE WEIGHTS

- Once one has a number of different sets of price weights (e.g., weights from 5 of the 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
CENTER/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 reasons cited for such a strategy are that 1) reducing variability in price weights will reduce the variability in estimated costs, and 2) an average set of price weights might be more representative.

- 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 difference between the therapies under evaluation.
  * Why is it more "representative" to use a set of price weights that no one faces?

ISSUE #3. WHAT IS THE APPROPRIATE SAMPLE SIZE FOR ECONOMIC QUESTIONS?

Prior to the development of the literature that described confidence intervals for cost-effectiveness ratios, a common approach for estimating sample size was 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 following formula *

\[
n = \frac{\left(\alpha + \beta\right)^2 \left(2 \text{sd}_c^2 + (2 \text{rc} \cdot \text{sd}_c^2) - (2 \text{rc} \cdot 2 \text{sd}_c \cdot 2 \text{sd}_q - 2 \text{sd}_c \cdot 2 \text{sd}_q \cdot \text{D}^2)\right)}{\Delta \text{NMB}^2}
\]

where \(n\) equals n/group; \(\text{sd} = \) the standard deviation for costs (\(c\)) and effects (\(q\)); \(\text{rc} = \) the ceiling ratio one wishes to be able to rule out; and \(\rho\) equals the correlation of the difference in cost and effect.

- Basic data for such calculations include the magnitude of the incremental costs and outcomes (in the formula, they appear in the NMB term); the standard deviations for costs and outcomes; and the correlation between costs and outcomes.

- See attached STATA-based sample size / power programs.
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?

- 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:

<table>
<thead>
<tr>
<th>Costs ($&gt;0$)</th>
<th>QALYS Saved $&lt;0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>122</td>
<td>134</td>
</tr>
</tbody>
</table>

- 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 MULTICENTER (MULTINATIONAL) TRIALS?

- The Problem:
  - There has been growing concern that the pooled (i.e., average) clinical and economic results from multicenter (or multinational) trials may not be reflective of the results that would be observed in individual centers 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
  - 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 multicenter 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 tertiary care hospitals, just use price weights from these centers, and conduct the analysis as if all participants were treated in tertiary care centers

- Use trial-wide clinical results and use costs derived from the subset of patients treated in the tertiary care centers

- 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 PRICE WEIGHTS VS OTHER VARIATION

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

<table>
<thead>
<tr>
<th>Country</th>
<th>Price Weights</th>
<th>Country Specific Costs †</th>
<th>Country Specific Costs and Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>46,818</td>
<td>5,921</td>
<td>11,450</td>
</tr>
<tr>
<td>2</td>
<td>57,636</td>
<td>91,906</td>
<td>60,358</td>
</tr>
<tr>
<td>3</td>
<td>53,891</td>
<td>90,487</td>
<td>244,133</td>
</tr>
<tr>
<td>4</td>
<td>69,145</td>
<td>93,326</td>
<td>181,259</td>
</tr>
<tr>
<td>5</td>
<td>65,800</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Overall</td>
<td>45,892</td>
<td>45,892</td>
<td>45,892</td>
</tr>
</tbody>
</table>

† Country-specific resource use × Country-specific price weights

** New therapy dominates

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 centers
  - 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 centers, but this method is less clear about the result that should be used instead


ESTIMATION

- The second method uses multi-level random-effects model shrinkage estimation to provide more precise estimates of the center-specific results than are yielded by separate -- and naive -- analysis of each center's costs and effects
  - Borrow information from the mean estimate to add precision to the center-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 center-specific results aren't substantially different from the pooled 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 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 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.)
COST EFFECTIVENESS RATIO AND 95% CI, 5.5-YEAR CHOLESTEROL LOWERING TRIAL

<table>
<thead>
<tr>
<th>Years of Follow-up/Projection</th>
<th>Point Estimate</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within the trial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Dominated</td>
<td>$168,884 to Dominated</td>
</tr>
<tr>
<td>2</td>
<td>$282,857</td>
<td>$45,577 to Dominated</td>
</tr>
<tr>
<td>3</td>
<td>$73,529</td>
<td>Dominates to Dominated</td>
</tr>
<tr>
<td>4</td>
<td>$12,074</td>
<td>Dominates to Dominated</td>
</tr>
<tr>
<td>5.5</td>
<td>$15,258</td>
<td>Dominates to $122,772</td>
</tr>
<tr>
<td>Projection beyond the trial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>$12,246</td>
<td>Dominates to $42,263</td>
</tr>
<tr>
<td>15</td>
<td>$8,578</td>
<td>Dominates to $6,721</td>
</tr>
<tr>
<td>20</td>
<td>$7,320</td>
<td>$681 to $21,841</td>
</tr>
</tbody>
</table>

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
* These programs estimate sample sizes needed and power to detect
* NMB differences that are greater than 0. They yield results that
* are identical to those derived from the NHB formula in:
* Willan AR. Analysis, sample size, and power for estimating
* incremental net health benefit from clinical trial data.

*****

capture program drop cess1i
program define cess1i

* SAMPLE SIZE PROGRAM

* This program is used when one assumes there are common
* standard deviations for costs and effects (SDs, not SEs
* for the difference in costs and effects).

* 1: Difference in costs
* 2: Difference in effects
* 3: Standard deviation, costs (assumed the same for both groups)
* 4: Standard deviation, effects (assumed the same for both groups)
* 5: Correlation, difference in costs and effects
* 6: Ceiling ratio
* 7: Two-tailed alpha level
* 8: One-tailed beta level

tempname nmb alpha beta ss

scalar `nmb'=(`6'*`2')-`1'
scalar `alpha'=invnorm(1-(`7'/2))
scalar `beta'=invnorm(`8')

scalar `ss'=((`alpha'+`beta')^2)*(((2*(`3'^2))+((`6'^2)*(2*(`4'^2))))-(2*`6'*`5'*((2*(`3'^2))^.5)*((2*(`4'^2))^.5)) )/(`nmb'^2))
scalar `ss'=round(`ss',1)
display " "
display "SAMPLE SIZE CALCULATION (Common SD Costs and Effects)"
display " "
display "Assumptions"
display " "
display "Difference in costs:                             ",`1'
display "Difference in effects:                           ",`2'
display "Standard deviation, costs:                       ",`3'
display "Standard deviation, effects:                     ",`4'
display "Correlation, difference in costs and effects:    ",`5'
display " "
display "Ceiling ratio:                                   ",`6'
display "Two-tailed alpha level:                          ",`7'
display "One-tailed beta level:                           ",`8'
display " "
display "*** SAMPLE SIZE PER GROUP ***                    ",`ss'
display " "
end

*****

capture program drop cepow1i
program define cepow1i

* POWER CALCULATION PROGRAM

* This program is used when one assumes there are common
* standard deviations for costs and effects (SDs, not SEs
* for the difference in costs and effects).

* 1: Difference in costs
* 2: Difference in effects
* 3: Standard deviation, costs (assumed the same for both groups)
* 4: Standard deviation, effects (assumed the same for both groups)
* 5: Correlation, difference in costs and effects

* 6: Ceiling ratio
* 7: Two-tailed `alpha' level
* 8: Sample size per group

tempname nmb alpha beta

scalar `nmb'=(`6'^2)-`1'
scalar `alpha'=invnorm(1-(`7'/2))

scalar `beta'=invnorm(`10')
scalar `beta'=round((norm(`beta')),.001)

display " **
display "SAMPLE SIZE CALCULATION (Different SD, Costs and Effects)"
display " **

end

* SAMPLE SIZE PROGRAM

* This program is used when one assumes there are Rx-specific
  * standard deviations for costs and effects (SDs, not SEs
  * for the difference in costs and effects).

* Calculations for group 1 - group 0

* 1: Difference in costs
* 2: Difference in effects
* 3: Standard deviation, costs, group 0
* 4: Standard deviation, costs, group 1
* 5: Standard deviation, effects, group 0
* 6: Standard deviation, effects, group 1
* 7: Correlation, difference in costs and effects
* 8: Ceiling ratio
* 9: Two-tailed alpha level
* 10: One-tailed beta level

tempname nmb alpha beta ss

scalar `nmb'=(`8'^2)-`1'
scalar `alpha'=invnorm(1-(`9'/2))
scalar `beta'=invnorm(`10')

scalar `ss'=((`alpha'+`beta')^2)*((((`3'^2)+(`4'^2))+((`8'^2)*((`5'^2)+(`6'^2)))-(2*rc*corrcq*(((`3'^2)+(`4'^2))^.5)*(((sdq0'^2)+(sdq1'^2))^.5)))/(`nmb'^2))

scalar `ss'=round(`ss',1)

display " **
display "ASSUMPTIONS disadvantages"
display " **
display "" display "Difference in costs: " 1 display "Difference in effects: " 2 display "" display "Standard deviation, costs, group 0: " 3 display "Standard deviation, costs, group 1: " 4 display "Standard deviation, effects, group 0: " 5 display "Standard deviation, effects, group 1: " 6 display "Correlation, difference in costs and effects: " 7 display "" display "Ceiling ratio: " 8 display "Two-tailed alpha level: " 9 display "One-tailed beta level: " 10 display "" display **** SAMPLE SIZE PER GROUP ***  ss display "" end 

****
capture program drop cepow2i
program define cepow2i
* POWER CALCULATION PROGRAM
* This program is used when one assumes there are Rx-specific
* standard deviations for costs and effects (SDs, not SEs
* for the difference in costs and effects).
* Calculations for group 1 - group 0
* 1: Difference in costs
* 2: Difference in effects
* 3: Standard deviation, costs, group 0
* 4: Standard deviation, costs, group 1
* 5: Standard deviation, effects, group 0
* 6: Standard deviation, effects, group 1
* 7: Correlation, difference in costs and effects
* 8: Ceiling ratio
* 9: Two-tailed alpha level
* 10: Sample size
tempname nmb alpha beta
scalar `nmb'=(`8'*`2')-1'
scalar `alpha'=invnorm(1-(`9'/2))
scalar `beta'=((`10'/((((`3'^2)+(`4'^2))+((`8'^2)*((`5'^2)+(`6'^2)))-(2*`8'*`7'*(((`3'^2)+(`4'^2))^.5)*(((`5'^2)+(`6'^2))^.5)))/(`nmb'^2)))^0.5)-`alpha'
scalar `beta'=round((norm(`beta')),.001)

display "" display "POWER CALCULATION (Different SD, Costs and Effects)"
display "" display "Assumptions"
display "" display "Difference in costs: " 1 display "Difference in effects: " 2 display "" display "Standard deviation, costs, group 0: " 3 display "Standard deviation, costs, group 1: " 4 display "Standard deviation, effects, group 0: " 5 display "Standard deviation, effects, group 1: " 6 display "Correlation, difference in costs and effects: " 7 display "" display "Ceiling ratio: " 8 display "Two-tailed alpha level: " 9 display "Sample Size: " 10 display "" display **** POWER TO DETECT DIFFERENCE *** beta display "" end