Sample Size and Power for the Cost-Effectiveness Analysis

Henry Glick
Epi 550
April 15, 2011

Outline
- Introduce sample size for cost-effectiveness analysis
  - Goal, formulas, differences from sample size for clinical outcomes
- Usual types of tables
- Effects of varying the parameters on sample size and power
  - Special emphasis: effect of varying willingness to pay (W)

Goal of Sample Size and Power Calculation
- 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 hypothesize that the point estimate for the cost-effectiveness ratio will be 20,000 per quality-adjusted life year (QALY) and want to design an experiment that will provide an 80% chance (i.e., power) to be 95% confident that the therapy is good value when we are willing to pay at most 75,000 per QALY
Sample Size Formula, Common SDs

- Assuming equal SDs and sample sizes, the sample size formula is:

\[ n = \frac{2 \left( z_\alpha + z_\beta \right)^2 \left( sd_c^2 + (W sd_c) + (2 W \rho sd_c sd_q) \right)}{(W Q - C)^2} \]

where \( n \) = sample size/group; \( z_\alpha \) and \( z_\beta \) = z-statistics for \( \alpha \) (e.g., 1.96) and \( \beta \) (e.g., 0.84) errors; \( sd \) = standard deviation for cost (c) and effect (q); \( W \) = maximum willingness to pay we wish to rule out; and \( \rho \) = correlation of the difference in cost and effect.

www.uphs.upenn.edu/dgimhsr/stat-samps.htm

Null Hypothesis, NMB

- This formula identifies a sample size that provides a \( 1-\beta\% \) chance to have \( 1-\alpha\% \) confidence for the rejection of the null hypothesis that the net monetary benefit (NMB = WQ – C) calculated by use of W equals 0.

  - If our assumptions about C, Q, sdc, sdq, and \( \rho \) are correct and if \( \alpha=0.05 \) and \( \beta=0.2 \), then
    - In approximately 800 of 1000 repeated experiments, the lower limit of the 95% confidence interval for the difference in NMB will be greater than 0.
    - In approximately 200, the 95% confidence intervals will either include 0 or have an upper limit that is less than 0.

Null Hypothesis, CER and Acceptability

- The formula also identifies a sample size that provides a \( 1-\beta\% \) chance to have \( 1-\alpha\% \) confidence for the rejection of the null hypothesis that the cost-effectiveness ratio equals W (i.e., that the \( 1-\alpha\% \) confidence interval for the cost-effectiveness ratio excludes W).

  - Or equivalently, it identifies a sample size that provides a \( 1-\beta\% \) chance for the rejection of the null hypothesis that at W the fraction of the joint distribution of the difference in cost and effect that is acceptable is greater than \( \alpha/2\% \) and less than \( 1- (\alpha/2)\% \) (i.e., that the acceptability curve lies above \( 1- (\alpha/2)\% \)).
W and the Point Estimate

- When W is greater than the expected point estimate, the resulting sample size and power are for experiments that allow us to be confident that the therapy is good value
  - Because the confidence statements from these trials will be that the point estimate is less than willingness to pay
- When W is less than the expected point estimate, the resulting sample size and power are for experiments that allow us to be confident that the therapy is bad value
  - Because the confidence statements from these trials will be that the point estimate is greater than willingness to pay

Similarities With Clinical Sample Size Formulas

Error Rate NMB Variance

\[
\begin{align*}
n &= \frac{2 (z_{\alpha} + z_{\beta})^2 \left( \text{Var}_{\text{NMB}} \right)}{\Delta \text{NMB}^2}
\end{align*}
\]

\[
\begin{align*}
n &= \frac{2 (z_{\alpha} + z_{\beta})^2 \left( \text{Var}_{\text{Q}} \right)}{\Delta Q^2}
\end{align*}
\]

Differences in Formulas

\[
\text{Var}_{\text{NMB}} = \text{sd}_{i}^2 + (W^2 \text{sd}_{i}^2) - (2 W \rho \text{sd}_{i} \text{sd}_{j})
\]

- Variance of NMB more complicated than variance for usual continuous clinical differences
  - Includes \(\rho\), the correlation of the difference between cost and effect
  - Includes W, the decision threshold we are trying to rule out
Correlation
• 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 a positive correlation than when they are characterized by negative correlation

Ability to Shift W
• W is to cost-effectiveness analysis as 1 is to OR and RR
  – It is the decision threshold we are trying to rule out if we are to have confidence about value
• While we rarely consider comparing OR and RR to a decision threshold other than 1 (noninferiority trials may be the exception), we often choose W because in many countries there is no clear consensus on what its value is
• Moving W “nearer to” or “further away from” the expected point estimate reduces or increases the power we have to be confident of value
  – Caution: “Nearer” and “further away” are not measured on the real number line

Power Formula, Common SDs
• Assuming equal standard deviations for cost and effect and equal sample sizes, the power formula is:

\[
z^* = \sqrt{\frac{n \cdot (W \Delta Q - \Delta C)^2}{\left( \frac{\text{sd}_Q^2 + (W \text{sd}_Q^2)}{2} \right) - \left( 2 W \rho \text{sd}_Q \text{sd}_C \right)}} \cdot z_{\beta}
\]
• Unlike sample size equation where result = N, result of formula is \( z_{\beta} \), not power
• To estimate power, use the normal distribution table to identify the fraction of the tail that is to the left of \( z_{\beta} \)
  – Stata code: power = norm(zbeta)
  – E.g., -1.96 = 2.5% power; -0.84 = 20% power; 0 = 50% power; 0.84 = 80% power; 1.28 = 90%
WTP and the Point Estimate

- When WTP is greater than the expected point estimate, the resulting sample size and power are for experiments that allow us to be confident that the therapy is good value
  - Because the confidence statements from these trials will be that the point estimate is less than willingness to pay
- When WTP is less than the expected point estimate, the resulting sample size and power are for experiments that allow us to be confident that the therapy is bad value
  - Because the confidence statements from these trials will be that the point estimate is greater than willingness to pay

Sample Size Tables, SD

- We commonly construct sample size tables for different values of $\Delta C$, $\Delta Q$, the standard deviations for $C$ and $Q$, and $W$

<table>
<thead>
<tr>
<th>$SD_c$</th>
<th>N/Group</th>
<th>$SD_q$</th>
<th>N/Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>2500</td>
<td>306</td>
<td>0.1</td>
<td>114</td>
</tr>
<tr>
<td>5000</td>
<td>340</td>
<td>0.2</td>
<td>340</td>
</tr>
<tr>
<td>7500</td>
<td>389</td>
<td>0.3</td>
<td>710</td>
</tr>
<tr>
<td>10,000</td>
<td>455</td>
<td>0.4</td>
<td>1224</td>
</tr>
<tr>
<td>15,000</td>
<td>634</td>
<td>0.6</td>
<td>2685</td>
</tr>
</tbody>
</table>

$\Delta C=250; \Delta Q=0.05$; unless otherwise specified, $sd_c=5000$; $sd_q=2; \rho=-1; \alpha=0.05; \beta=0.8$

Dropout

- These sample size estimates are appropriate if we expect no dropout from the trial
- If we instead anticipate 10% dropout, we will want to divide these sample size estimates by 0.9
### “Typical” Sample Size Table, W

<table>
<thead>
<tr>
<th>WTP</th>
<th>Exp 1 *</th>
</tr>
</thead>
<tbody>
<tr>
<td>20,000</td>
<td>693</td>
</tr>
<tr>
<td>30,000</td>
<td>462</td>
</tr>
<tr>
<td>50,000</td>
<td>337</td>
</tr>
<tr>
<td>75,000</td>
<td>295</td>
</tr>
<tr>
<td>100,000</td>
<td>279</td>
</tr>
<tr>
<td>150,000</td>
<td>266</td>
</tr>
</tbody>
</table>

* ΔC=-10; ΔQ=0.05; sd_c=5000; sd_q=.20; ρ=-.1; α=.05; 1-β=.8

### Sample Size Can Increase with Increasing W

<table>
<thead>
<tr>
<th>WTP</th>
<th>Exp 1</th>
<th>Exp 2 *</th>
</tr>
</thead>
<tbody>
<tr>
<td>20,000</td>
<td>693</td>
<td>560</td>
</tr>
<tr>
<td>30,000</td>
<td>462</td>
<td>831</td>
</tr>
<tr>
<td>50,000</td>
<td>337</td>
<td>1527</td>
</tr>
<tr>
<td>75,000</td>
<td>295</td>
<td>2573</td>
</tr>
<tr>
<td>100,000</td>
<td>279</td>
<td>3717</td>
</tr>
<tr>
<td>150,000</td>
<td>266</td>
<td>6059</td>
</tr>
</tbody>
</table>

* ΔC=-250; ΔQ=0.001; sd_c=1000; sd_q=.05; ρ=-.3; α=.05; 1-β=.8

### Sample Size Not Necessarily Monotonic With W

<table>
<thead>
<tr>
<th>WTP</th>
<th>Exp 1</th>
<th>Exp 2</th>
<th>Exp 3 *</th>
</tr>
</thead>
<tbody>
<tr>
<td>20,000</td>
<td>693</td>
<td>560</td>
<td>3383</td>
</tr>
<tr>
<td>30,000</td>
<td>462</td>
<td>831</td>
<td>3205</td>
</tr>
<tr>
<td>50,000</td>
<td>337</td>
<td>1527</td>
<td>3167</td>
</tr>
<tr>
<td>75,000</td>
<td>295</td>
<td>2573</td>
<td>3185</td>
</tr>
<tr>
<td>100,000</td>
<td>279</td>
<td>3717</td>
<td>3204</td>
</tr>
<tr>
<td>150,000</td>
<td>266</td>
<td>6059</td>
<td>3229</td>
</tr>
</tbody>
</table>

* ΔC=-40; ΔQ=0.04; sd_c=5500; sd_q=.58; ρ=1; α=.05; 1-β=.8
Power Tables

- When sample size per group is fixed, we commonly calculate the power for multiple values of WTP (or of other parameters used to identify a sample size)

<table>
<thead>
<tr>
<th>WTP</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>20,000</td>
<td>.383</td>
</tr>
<tr>
<td>30,000</td>
<td>.61</td>
</tr>
<tr>
<td>50,000</td>
<td>.829</td>
</tr>
<tr>
<td>75,000</td>
<td>.912</td>
</tr>
<tr>
<td>100,000</td>
<td>.941</td>
</tr>
</tbody>
</table>

$\Delta C = 250$; $\Delta Q = 1$; $sd_{C} = 10,000$; $sd_{Q} = .3$; $\rho = -.25$; $\alpha = .05$;
Sample size per group = 300

- If we anticipate 10% dropout, we will want to use the effective sample size (e.g., 0.9 * 300) when we make our calculations

Patterns of Power: SD, Z, $\rho$, $\Delta C$ (Q)
Sample Size Often More Sensitive to SD_q than to SD_c

\[ \frac{2 (z_{1-\alpha}, z_{1-\beta})^2 (sd_c^2 + (W^2 sd_q^2) - (W p (2 sd_c)^2 + 2 sd_c^2))^2}{\text{NMB}} \]

- The sample size formula is generally symmetric for the SDs of cost and effect except for the following:
- Changes in the square of the QALY SD are weighted by the square of WTP; changes in the square of the cost SD are unweighted
- When WTP is substantially greater than SD for cost, percentage changes in the QALY SD will have a substantially greater effect on sample size than will equivalent percentage changes in cost SD

Effects of Variation in W

- For the parameters we’ve been looking at there are values common to all experiments for which power reaches a maximum and sample size reaches a minimum
  - For example, for the sd’s or z’s this value equals 0, for C or Q these values equal +\infty
- Similarly, for most of the parameters there are values common to all experiments for which power reaches a minimum and sample size approaches a maximum
  - For example, for the sd’s this value equals +\infty, for the z’s these values equal +\infty

Effects of Variation in W (2)

- When W = C/Q, power reaches a minimum and sample size reaches a maximum
- However, the value of W at which power equals a maximum and sample size reaches a minimum is a function of the assumed values for C, Q, sd_c, sd_q, and \( \rho \) (but is independent of z\( \alpha \) and z\( \beta \)).
- The equation that defines the W at which sample size reaches a minimum equals:

\[ W_{\text{min power}} = \left( \frac{1 - \rho^2}{1 - \rho^2} \right) \cdot \left( \frac{sd_c^2}{sd_c^2 + \left( \frac{C}{Q} sd_q^2 \right)} \right) \cdot \left( \frac{sd_q^2}{sd_q^2 + \left( \frac{C}{Q} sd_c^2 \right)} \right) \cdot \left( \frac{2 C sd_c sd_q}{2 C sd_c sd_q + \rho} \right) \]
Two Underlying Patterns of Power for $W$

- When we plot power patterns for $-\infty < W \leq \infty$, patterns differentiated by whether the value of $W$ where power reaches a maximum is greater than or less than the value of $W$ where power reaches a maximum.
- When we plot power patterns for $W > 0$.

Six Power Patterns Associated with $W$

- For each of the two underlying power patterns, we've added 3 different potential 0 values of willingness to pay.

Truncation at 0 Creates the 6 Patterns

- The pattern we observe depends on where the 0 WTP falls.
Power Patterns and W

• As can be seen, power need not reach a maximum nor even be increasing as W approaches $\infty$.
• Similarly, power need not be monotonically increasing or decreasing and it can be multimodal.

Confidence about $\Delta C$ and $\Delta Q$

• Before we calculated confidence intervals for cost-effectiveness ratios, it was generally accepted that if neither $\Delta C$ nor $\Delta Q$ were significant, there was no reason to assess cost-effectiveness.
• But $\Delta C$ and $\Delta Q$ can be considered a joint outcome just as differences in nonfatal CVD events and all cause mortality are often combined into a joint outcome.
  – A trial can be “clinically” successful when the differences in the individual outcomes are not significant but the difference in the joint outcome is.
• Similarly, a trial can be “economically” successful when neither the difference in cost nor the difference in effect is significant, but the joint outcome of the difference in costs and effects is.

Economic Vs Clinical Sample Sizes

• While the sample size required to answer economic questions tends to be larger than the sample size required to answer clinical questions, it need not be for all experiments.
• As previously mentioned, the sample size needed to answer the economic question is more likely to be smaller than that needed to answer the clinical question when:
  – The correlation of cost and effect is positive.
  – We have more power for the joint outcome of difference in cost and effect than we do for either outcome alone.
Sample Size Programs: ssizeprg.do

- quietly do ssizeprg
- ssizeprg.do is a text file that contains 6 “immediate form” PROGRAMS that estimate 2-sample sample sizes and power to detect NMB differences that are greater than 0
  - The command do ssizeprg simply loads these programs; it does not calculate anything
- “Doing” ssizeprg also loads a documentation program named ssizeprgdoc

http://www.uphs.upenn.edu/dgimhsr/stat-samps.htm

3 Sample Size Programs

- cess1i: Calculates sample size under the assumption that the sample size and the standard deviations for cost and effect are common for the 2 treatment groups
- cess2i: Calculates sample size under the assumption that the sample size is the same in the 2 groups, but the standard deviations for cost and effect differ
- cddssi: Calculates sample size under the assumption that the sample size differs between the 2 groups, but the standard deviations for cost and effect are the same

3 Power Programs

- cepow1i: Calculates power to detect NMB greater than 0 under the assumption the that sample size and the standard deviations for cost and effect are common for the 2 treatment groups
- cepow2i: Calculates power to detect NMB greater than 0 under the assumption that the sample size is the same in the 2 groups, but the standard deviations for cost and effect differ
- cedpowi: Calculates power to detect NMB greater than 0 under the assumption that the sample size differs between the 2 groups, but the standard deviations for cost and effect are the same
• All 6 programs report sample size and power for the comparison of 2 arms in a trial (for multi-arm trials, the programs report sample size and power for individual pair-wise comparisons)
• The sample size estimates from these programs have been validated in simulations and yield results that match those derived from the NHB formula in: Willan AR. Analysis, sample size, and power for estimating incremental net health benefit from clinical trial data. Control Clin Trials 2001;22:228-237