Computer Workshop: Sample Size and Power for the Comparison of Cost and Effect

Applications of Statistical Considerations in Health Economic Evaluations

ISPOR 16th Annual International Meeting

May 2011

Henry Glick and Jalpa Doshi

www.uphs.upenn.edu/dgimhsr

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

Goal of Sample Size and Power Calculation

- Sample size and power calculation allow us to conduct experiments with an expected likelihood that at the conclusion of the experiment will allow us to be confident in the resulting comparison of 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:

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

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 NMB ($NMB = WQ - C$) 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 $1- (\alpha/2)\%$).
Similarities With Clinical Sample Size Formulas

Error Rate  NMB Variance

\[ n = \frac{2 (z_{\alpha} + z_{\beta})^2 (sd_c^2 + (W^2 sd_q^2) - (2 W \rho sd_c \cdot sd_q))}{\Delta NMB^2} \]

\[ n = \frac{2 (z_{\alpha} + z_{\beta})^2 (sd_q^2)}{\Delta Q^2} \]

Differences in Formulas

\[ \text{Var}_{\text{NMB}} = sd_c^2 + (W^2 sd_q^2) - (2 W \rho sd_c \cdot sd_q) \]

- 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

- When increasing effects are associated with decreasing costs, a therapy is characterized by a negative (win/win) correlation between the difference in cost and effect
  - e.g., asthma care
- When increasing effects are associated with increasing costs, a therapy is characterized by a positive (win/lose) correlation between the difference in cost and effect
  - e.g., life-saving care
- All else equal, fewer patients need to be enrolled when therapies are characterized by a positive correlation than when therapies 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 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

- Power is calculated by use of the following formula:

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

- 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 (V9) code: `power = norm(zbeta)`
  - E.g., -1.96 = 2.5% power; -0.84 = 20% power; 0 = 50% power; .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, $sd_c=5000; sd_q=.2; \rho=-.1; \alpha=.05; \beta=.8$

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>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>20,000</td>
<td>321</td>
<td></td>
</tr>
<tr>
<td>30,000</td>
<td>273</td>
<td></td>
</tr>
<tr>
<td>50,000</td>
<td>234</td>
<td></td>
</tr>
<tr>
<td>75,000</td>
<td>214</td>
<td></td>
</tr>
<tr>
<td>100,000</td>
<td>204</td>
<td></td>
</tr>
<tr>
<td>150,000</td>
<td>194</td>
<td></td>
</tr>
</tbody>
</table>

* $\Delta C=-120; \Delta Q=0.015; sdc=1000; sdq=.05; \rho=-.8; \alpha=.05; 1-\beta=.8$

### 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>321</td>
<td>36</td>
</tr>
<tr>
<td>30,000</td>
<td>273</td>
<td>42</td>
</tr>
<tr>
<td>50,000</td>
<td>234</td>
<td>68</td>
</tr>
<tr>
<td>75,000</td>
<td>214</td>
<td>92</td>
</tr>
<tr>
<td>100,000</td>
<td>204</td>
<td>108</td>
</tr>
<tr>
<td>150,000</td>
<td>194</td>
<td>127</td>
</tr>
</tbody>
</table>

* $\Delta C=-120; \Delta Q=0.015; sdc=1000; sdq=.05; \rho=0.8; \alpha=.05; 1-\beta=.8$
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>321</td>
<td>36</td>
<td>178</td>
</tr>
<tr>
<td>30,000</td>
<td>273</td>
<td>42</td>
<td>158</td>
</tr>
<tr>
<td>50,000</td>
<td>234</td>
<td>68</td>
<td>151</td>
</tr>
<tr>
<td>75,000</td>
<td>214</td>
<td>92</td>
<td>154</td>
</tr>
<tr>
<td>100,000</td>
<td>204</td>
<td>108</td>
<td>156</td>
</tr>
<tr>
<td>150,000</td>
<td>194</td>
<td>127</td>
<td>160</td>
</tr>
</tbody>
</table>

* ΔC=-120; ΔQ=0.015; sd_c=1000; sd_q=.05; ρ=0.0; α=.05; 1-β=.8

Power Tables

- When sample size per group is fixed, we commonly calculate the power for multiple values of WTP

<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>

ΔC=250; ΔQ=.1; sd_c=10,000; sd_q=.3; ρ=-.25; α=.05
Sample size per group=300
Power Tables

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

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
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 that the 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
ssizeprg.do (cont.)

- 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

ssizeprgdoc: cess1i

* PROGRAM: CESS1I

* cess1i is used to estimate sample size when we assume that
  * the 2 treatment groups have a common sample size and
  * common standard deviations for cost and effect (SDs, not SEs for the difference.

* COMMAND LINE: cess1i [diffc] [diffe] [sdc] [sde] [corr] [wtp] [alpha] [beta]

* The 8 arguments are all numbers
  ** `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' Maximum willingness to pay
  ** `7' Two-tailed alpha level (e.g., 0.05)
  ** `8' One-tailed beta level (e.g., 0.80)
ssizeprgdoc: cess1i (cont.)

• Saved results (scalars)
  * r(diffc)
  * r(diffq)
  * r(sd_c)
  * r(sd_e)
  * r(rho)
  * r(wtp)
  * r(alpha)
  * r(beta)
  * r(nmb)
  * r(wdi)
  * r(sampsize)

Implementing cess1i

• Suppose the expected difference in cost is 25; the expected difference in QALYs is 0.05; the expected SDs for cost and QALYs are 1000 and 0.195, respectively; the expected correlation of the difference is -0.1; your maximum WTP is 75,000; and you want a 2-tailed alpha of .05 and a 1-tailed beta of 0.8:
  – Point estimate = 25 / 0.5 = 500 / QALY
  – Calculate the necessary sample size:

      cess1i 25 .05 1000 .195 -.1 75000 .05 .8
### SAMPLE SIZE CALCULATION (Common SD Costs and Effects)

**Assumptions**

- **Difference in costs:** 25
- **Difference in effects:** .05
- **Standard deviation, costs:** 1000
- **Standard deviation, effects:** .195
- **Correlation, difference in costs and effects:** -.1
- **Willingness to pay:** 75000
- **Two-tailed alpha level:** .05
- **One-tailed beta level:** .8
- **Expected NMB:** 3725
- **Widest definable interval:** -26219

### *** SAMPLE SIZE PER GROUP ***

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>cess1i</td>
<td>247</td>
</tr>
</tbody>
</table>

### Saved Results, cess1i

- **. return list**

- **Scalars:**
  - `r(diffc) = 25`
  - `r(diffq) = .05`
  - `r(sd_c) = 1000`
  - `(sd_e) = .195`
  - `r(rho) = -.1`
  - `r(wtp) = 75000`
  - `r(alpha) = .05`
  - `r(beta) = .8`
  - `r(nmb) = 3725`
  - `r(wdi) = -26219`
  - `r(sampsize) = 247`
Fill In the Following Table

- Assuming that $C=25$; $Q=0.05$; $SD_c=1000$; $SD_q=0.195$; the correlation=$-0.1$; the 2-tailed alpha=$0.05$; and the 1-tailed beta=$0.8$, fill in the following table:

<table>
<thead>
<tr>
<th>WTP</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>30,000</td>
<td></td>
</tr>
<tr>
<td>50,000</td>
<td></td>
</tr>
<tr>
<td>75,000</td>
<td></td>
</tr>
<tr>
<td>100,000</td>
<td></td>
</tr>
<tr>
<td>125,000</td>
<td></td>
</tr>
</tbody>
</table>

Code for Looping Calculations

```bash
foreach wtp in 30000 50000 75000 100000 125000 {
  cess1i 25 .05 1000 .195 -.1 `wtp' .05 .8
}
```
Fill In the Following Table

- Assuming that \(C=25\); \(Q=0.05\); \(SDc= 1000\); \(SDq=0.195\); the correlation=-0.1; the 2-tailed alpha=0.05; and the 1-tailed beta=0.8, fill in the following table

<table>
<thead>
<tr>
<th>WTP</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>30,000</td>
<td>263</td>
</tr>
<tr>
<td>50,000</td>
<td>252</td>
</tr>
<tr>
<td>75,000</td>
<td>247</td>
</tr>
<tr>
<td>100,000</td>
<td>245</td>
</tr>
<tr>
<td>125,000</td>
<td>244</td>
</tr>
</tbody>
</table>

ssizeprgdoc: cepow1i

- PROGRAM: CEPOW1i
- cepow1i is used to assess power when we assume that
  - the 2 treatment groups have a common sample size and
  - common standard deviations for costs and effects (SDs, not SEs for the difference in cost and effect).
- COMMAND LINE: cepow1i [diffc] [diffe] [sdc] [sde] [corr] [wtp] [alpha]
  - [sampsize]
  - The 8 arguments are all numbers
  - '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' Willingness to pay
  - '7' Two-tailed level (e.g., 0.05)
  - '8' Sample size per group
ssizeprgdoc: cepow1i

- Saved results (scalars)
  * r(diffc)
  * r(diffq)
  * r(sd_c)
  * r(sd_e)
  * r(rho)
  * r(wtp)
  * r(alpha)
  * r(sampsize)
  * r(nmb)
  * r(wdi)
  * r(power)

Implementing cepow1i

- Suppose the expected difference in cost is 25; the expected difference in QALYs is 0.05; the expected SDs for cost and QALYs are 1000 and 0.195, respectively; the expected correlation of the difference is -0.1; your maximum WTP is 75,000; you want a 2-tailed alpha of .05; and the current sample size plans are for 246 per group
- Calculate the power of this experiment

    cepow1i 25 .05 1000 .195 -.1 75000 .05 246
POWER CALCULATION (Common SD Costs and Effects)

Assumptions

Difference in costs: 25
Difference in effects: .05
Standard deviation, costs: 1000
Standard deviation, effects: .195
Correlation, difference in costs and effects: -.1
Willingness to pay: 75000
Two-tailed alpha level: .05
Sample size per group: 247
Expected NMB: 3725
Widest definable interval: -26219

*** POWER TO DETECT DIFFERENCE *** .8009

Saved Results, cpow1i

. return list

scalars:

r(diffc) = 25
r(diffq) = .05
r(sd_c) = 1000
r(sd_e) = .195
r(rho) = -.1
r(wtp) = 75000
r(alpha) = .05
r(sampsiz) = 247
r(nmb) = 3725
r(wdi) = -26219
r(power) = .8009
Fill In the Following Table

- Assuming that $C=25$; $Q=0.05$; $SDc=1000$; $SDq=0.195$; the correlation=-0.1; the WTP=75,000; and the 2-tailed alpha=0.05, fill in the following table:

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td></td>
</tr>
<tr>
<td>246</td>
<td></td>
</tr>
<tr>
<td>300</td>
<td></td>
</tr>
<tr>
<td>350</td>
<td></td>
</tr>
</tbody>
</table>

Code for Looping Calculations

```bash
foreach ssize in 150 200 246 300 350 {
  cepow1i 25 .05 1000 .195 -.1 75000 .05 `ssize'
}
```
Fill In the Following Table

- Assuming that $C=25; Q=0.05; SD_c=1000; SD_q=0.195; \text{the correlation}=-0.1; \text{the WTP}=75,000; \text{and the 2-tailed alpha}=0.05$, fill in the following table:

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td>0.589</td>
</tr>
<tr>
<td>200</td>
<td>0.714</td>
</tr>
<tr>
<td>247</td>
<td>0.801</td>
</tr>
<tr>
<td>300</td>
<td>0.871</td>
</tr>
<tr>
<td>350</td>
<td>0.916</td>
</tr>
</tbody>
</table>

Sample Size Often More Sensitive to $SD_q$ than to $SD_c$

\[
\frac{2(z_{\alpha/2}+z_{\beta})^2}{\Delta \text{NMB}^2} \left( sd_c^2 + (W^2 \cdot sd_q^2) - (W \rho (2 \cdot sd_c^2)^{0.5} (2 \cdot sd_q^2)^{0.5}) \right)
\]

- 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
**Fill In the Following Table**

- Assuming that $C=25; \ Q=0.05$; unless otherwise specified, $SDc= 1000$ and $SDq=0.195$; the correlation=$-0.1$; the $wtp=75,000$; the 2-tailed alpha=$0.05$; the 1-tailed beta=$0.8$, fill in the following table:

<table>
<thead>
<tr>
<th>$SDc$</th>
<th>Sample Size/G</th>
<th>$SDq$</th>
<th>Sample Size/G</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1000</strong></td>
<td><strong>0.195</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1500</td>
<td>0.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3000</td>
<td>0.60</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Code for Looping Calculations**

For $SDc$:

```bash
foreach sdc in 500 1000 1500 2000 3000 {
cess1i 25 .05 `sdc' .195 -.1 75000 .05 .8
}
```

For $SDq$

```bash
foreach sdq in .1 .195 .3 .4 .6 {
cess1i 25 .05 1000 `sdq' -.1 75000 .05 .8
}
```
Fill In the Following Table

- Assuming that C=25; Q=0.05; unless otherwise specified, SDc= 1000 and SDq=0.195; the correlation=-0.1; the wtp=75,000; the 2-tailed alpha=0.05; the 1-tailed beta=0.8, fill in the following table:

<table>
<thead>
<tr>
<th>SDc</th>
<th>Sample Size/G</th>
<th>SDq</th>
<th>Sample Size/G</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>244</td>
<td>0.10</td>
<td>67</td>
</tr>
<tr>
<td>1000</td>
<td><strong>247</strong></td>
<td><strong>0.195</strong></td>
<td><strong>247</strong></td>
</tr>
<tr>
<td>1500</td>
<td>250</td>
<td>0.30</td>
<td>579</td>
</tr>
<tr>
<td>2000</td>
<td>254</td>
<td>0.40</td>
<td>1027</td>
</tr>
<tr>
<td>3000</td>
<td>263</td>
<td>0.60</td>
<td>2303</td>
</tr>
</tbody>
</table>

Fill In the Following Table

- Fill in the following table if we had instead paired an SDc=15,000 and SDq=0.195:

<table>
<thead>
<tr>
<th>SDc</th>
<th>Sample Size/G</th>
<th>SDq</th>
<th>Sample Size/G</th>
</tr>
</thead>
<tbody>
<tr>
<td>7500</td>
<td>331</td>
<td>0.10</td>
<td>344</td>
</tr>
<tr>
<td><strong>15,000</strong></td>
<td><strong>547</strong></td>
<td><strong>0.195</strong></td>
<td><strong>547</strong></td>
</tr>
<tr>
<td>22,500</td>
<td>890</td>
<td>0.30</td>
<td>904</td>
</tr>
<tr>
<td>30,000</td>
<td>1360</td>
<td>0.40</td>
<td>1375</td>
</tr>
<tr>
<td>45,000</td>
<td>2682</td>
<td>0.60</td>
<td>2699</td>
</tr>
</tbody>
</table>
Effects of Variation in W

- For the parameters we’ve been looking at there are values common to all experiments for which sample size reaches a minimum and power reaches a maximum
  - For example, for the sd’s or z’s this value equals 0, for C or Q these values equal $\pm \infty$
- Similarly, for most of the parameters there are values common to all experiments for which sample size approaches a maximum
  - For example, for the sd’s this value equals $\infty$, for the z’s these values equal $\pm \infty$
Effects of Variation in W (2)

• When \( W = C/Q \), power reaches a minimum and sample size approaches 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:

\[
W_{\text{maxpow}} = \frac{\left(1 - \rho^2\right) \frac{C}{Q} sd_c^2}{\left(1 - \rho^2\right) sd_q^2 - \left(sd_c^2 + \frac{C}{Q} sd_q^2\right)} - \left(\frac{C}{Q} sd_q^2 - \left(2 \frac{C}{Q} sd_c^2 \rho\right)\right)
\]

Two Underlying Patterns of Power for W

• When we plot power patterns for \(-\infty \leq W \leq \infty\), 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

- The pattern we observe depends on where the 0 WTP falls

Truncation at 0 Creates the 6 Patterns

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

- The pattern we observe depends on where the 0 WTP falls
Summary

• Goal of sample size and power calculation for cost-effectiveness analysis is to identify the likelihood that an experiment will allow us to be confident that a therapy is good or bad value when we adopt a particular willingness to pay
• Sample size and power depend on the difference in cost and effect, the SD of cost and effect, the correlation of the difference, our willingness to pay, and our target confidence level

Summary (2)

• When we estimate sample size or power, we often do so for varying levels of WTP
  – Sample size is undefined / power reaches a minimum when wtp equals the point estimate for the cost-effectiveness ratio (NMB=0)
• When WTP is substantially greater than the SD for cost, changes in the SD for effect generally have greater impact on sample size than do changes in the SD for cost
• So long as W>0, positive correlations decrease sample size / increase power
Additional Programs

ssizeprgdoc: cess2i

* PROGRAM: CESS2I

* cess2i is used to assess sample size when we assume that
* the 2 treatment groups have common sample sizes, but
* different standard deviations for costs and effects (SDs not,
* SEs for the difference in costs and effects)

* COMMAND LINE: cess2i [diffc] [diffe] [sdc0] [sdc1] [sde0] [sde1] [corr] [wtp]
[alpha] [beta]

* The 10 arguments are all numbers
  * '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' Willingness to pay
  * '9' Two-tailed alpha level (e.g., 0.05)
  * '10' One-tailed beta level (e.g., 0.80)
ssizeprgdoc: cess2i (cont.)

* Saved results (scalars)
  * r(diffc)
  * r(diffq)
  * r(sd_c0)
  * r(sd_c1)
  * r(sd_e0)
  * r(sd_e1)
  * r(rho)
  * r(wtp)
  * r(alpha)
  * r(beta)
  * r(nmb)
  * r(sampsize)

Implementing cess2i

- Suppose the expected difference in cost is 25; the expected difference in QALYs is 0.05; the expected SDs for cost are 800 and 1200; the expected SDs for QALYs are 0.19 and 0.20; the expected correlation of the difference is -0.1; your maximum WTP is 75,000; and you want a 2-tailed alpha of .05 and a 1-tailed beta of 0.8:
SAMPLE SIZE CALCULATION (Different SD, Costs and Effects)

Assumptions

Difference in costs: 25
Difference in effects: .05
Standard deviation, costs, group 0: 800
Standard deviation, costs, group 1: 1200
Standard deviation, effects, group 0: .19
Standard deviation, effects, group 1: .2
Correlation, difference in costs and effects: -.1
Willingness to pay: 75000
Two-tailed alpha level: .05
One-tailed beta level: .8
Expected NMB: 3725

*** SAMPLE SIZE PER GROUP ***

247

Saved Results, cess2i

. return list

scalars:

r(diffc) = .25
r(diffq) = .05
r(sd_c0) = 800
r(sd_c1) = 1200
r(sd_e0) = .19
r(sd_e1) = .2
r(rho) = -.1
r(wtp) = 75000
r(alpha) = .05
r(beta) = .8
r(nmb) = 3725
r(sampsize) = 247
• PROGRAM: CEPOW2i
  * cepow2i is used to assess power when we assume that
  * the 2 treatment groups have common sample size but
  * different standard deviations for cost and effect
  * (SDs, not SEs for the difference in cost and effect)
  * COMMAND LINE: cepow2i [diffc] [diffq] [sd_c0] [sd_c1] [sd_e0] [sd_e1] [corr] [wtp]
  * alpha] [sampsize]
  * The 10 arguments are all numbers
  * '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' Willingness to pay
  * '9' Two-tailed level (e.g., 0.05)
  * '10' Sample size per group

• Saved results (scalars)
  * r(diffc)
  * r(diffq)
  * r(sd_c0)
  * r(sd_c1)
  * r(sd_e0)
  * r(sd_e1)
  * r(rho)
  * r(wtp)
  * r(alpha)
  * r(sampsize)
  * r(nmb)
  * r(power)
Implementing cepow2i

• Suppose the expected difference in cost is 25; the expected difference in QALYs is 0.05; the expected SDs for cost are 800 and 1200; the expected SDs for QALYs are 0.19 and 0.20; the expected correlation of the difference is -0.1; your maximum WTP is 75,000; you want a 2-tailed alpha of .05; and the current sample size plans are for 246 per group

cepow2i 25 .05 800 1200 .19 .20 -.1 75000 .05 246

POWER CALCULATION (Different SD, Costs and Effects)
Assumptions
Difference in costs: 25
Difference in effects: .05
Standard deviation, costs, group 0: 800
Standard deviation, costs, group 1: 1200
Standard deviation, effects, group 0: .19
Standard deviation, effects, group 1: .2
Correlation, difference in costs and effects: -.1
Willingness to pay: 75000
Two-tailed alpha level: .05
Sample Size: 247
Expected NMB: 3725
*** POWER TO DETECT DIFFERENCE *** .8005
Saved Results, cepow2i

. return list

 scalars:
   r(diffc) = 25
   r(diffq) = .05
   r(sd_c0) = 800
   r(sd_c1) = 1200
   r(sd_e0) = .19
   r(sd_e1) = .2
   r(rho) = -.1
   r(wtp) = 75000
   r(alpha) = .05
   r(sampsize) = 247
   r(nmb) = 3725
   r(power) = .8005

ssizeprgdoc: cedssi

* PROGRAM: CEDSSI

* cedssi is used to estimate sample size when we assume that
  * the 2 treatment groups have a different sample sizes but
  * common standard deviations for cost and effect (SDs, not
    SEs for the difference in cost and effect).

* COMMAND LINE: cedssi [diffc] [diffe] [sdc] [sde] [corr] [wtp] [alpha] [beta]
  [n2/n1 ratio]

* The 9 arguments are all numbers
  * '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
  * '7' Maximum willingness to pay
  * '7' Two-tailed alpha level (e.g., 0.05)
  * '6' One-tailed beta level (e.g., 0.80)
  * '9' ratio of n2 / n1 (e.g., if n2 is twice n1, r=2
ssizeprgdoc: cedssi (cont.)

* Saved results (scalars)
  * r(diffc)
  * r(diffq)
  * r(sd_c)
  * r(sd_e)
  * r(rho)
  * r(wtp)
  * r(alpha)
  * r(beta)
  * r(nmb)
  * r(wdi)
  * r(sampsize1)
  * r(sampsize1)

Implementing cedssi

  • Suppose the expected difference in cost is 25; the expected difference in QALYs is 0.05; the expected SD for cost is 1000; the expected SD for QALYs is 0.195; the expected correlation of the difference is -0.1; your maximum WTP is 75,000; you want a 2-tailed alpha of .05 and a 1-tailed beta of 0.8; and you want to enroll 2 participants in group 2 for every 1 participant in group 1:
SAMPLE SIZE CALCULATION (Different Ns, but commons sds)

Assumptions

Difference in costs: 25
Difference in effects: .05
Standard deviation, costs: 1000
Standard deviation, effects: .195
Correlation, difference in costs and effects: -.1
Willingness to pay: 75000
Two-tailed alpha level: .05
One-tailed beta level: .8
Expected NMB: 3725
Widest definable interval: -26219

*** SAMPLE SIZE PER GROUP ***

Group 1: 185
Group 2: 370

Saved Results, cedssi (cont.)

. return list

Scalars:

r(diffc) = 25
r(diffq) = .05
r(sd_c) = 1000
r(sd_e) = .195
r(rho) = -.1
r(wtp) = 75000
r(alpha) = .05
r(beta) = .8
r(nmb) = 3725
r(wdi) = -26219
r(sampsize1) = 185
r(sampsize2) = 370
ssizeprgdoc: cedpowi

- PROGRAM: CEDPOWI

* cedpowi is used to assess power when we assume that
* the 2 treatment groups have different sample sizes but
* common standard deviations for cost and effect (SDs, not SEs
* for the difference in costs and effects)

* COMMAND LINE: cedpowi [diffc] [diffe] [sdc] [sde] [corr] [wtp] [alpha]
  [sampsize1] [sampsize2]

* The 9 arguments are all numbers
* '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' Maximum willingness to pay
* '7' Two-tailed alpha level (e.g., 0.05)
* '8' Sample size for group 1
* '9' Sample size for group 2

ssizeprgdoc: cedpowi

- Saved results (scalars)

* r(diffc)
* r(diffq)
* r(sd_c)
* r(sd_e)
* r(rho)
* r(wtp)
* r(alpha)
* r(sampsize1)
* r(sampsize2)
* r(nmb)
* r(wdi)
* r(power)
Implementing cedpowi

- Suppose the expected difference in cost is 25; the expected difference in QALYs is 0.05; the expected SD for cost is 1000; the expected SD for QALYs is 0.195; the expected correlation of the difference is -0.1; your maximum WTP is 75,000; you want a 2-tailed alpha of .05; and the current sample size plans are 185 for group 1 and 370 for group 2:

```plaintext
cedpowi 25 .05 1000 .195 -.1 75000 .05 185 370
```

POWER CALCULATION (Different SD, Costs and Effects)

Assumptions
- Difference in costs: 25
- Difference in effects: 0.05
- Standard deviation, costs: 1000
- Standard deviation, effects: 0.19
- Correlation, difference in costs and effects: -0.1
- Willingness to pay: 75000
- Two-tailed alpha level: 0.05
- Sample size group 1: 185
- Sample size group 2: 370
- Expected NMB: 3725
- Widest definable interval: -26219

*** POWER TO DETECT DIFFERENCE *** 0.804
Saved Results, cepowi

. return list

Scalars:

  r(diffc) = 25
  r(diffq) = .05
  r(sd_c) = 1000
  r(sd_e) = .195
  r(rho) = -.1
  r(wtp) = 75000
  r(alpha) = .05
  r(sampsize1) = 185
  r(sampsize2) = 470
  r(nmb) = 3725
  r(wdi) = -26219
  r(power) = .799