Sample Size and Power for the Comparison of Cost and Effect

Statistical Considerations in Health Economic Evaluations

ISPOR 18th Annual International Meeting

May 19, 2013

Henry Glick and Jalpa Doshi
www.uphs.upenn.edu/dgimhsr

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 hypthesize that the point estimate for the cost-effectiveness ratio will be 20,000 per QALY
  - May want to design an experiment that provides 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

Basic Formula

- At the most basic level, sample size for cost-effectiveness is calculated using the same formula as the sample size for a difference in any continuous variable:

$$n = \frac{2\,(z_{\alpha}+z_{\beta})^2\;\; sd^2_{nmb}}{\Delta nmb^2}$$

where n = sample size/group; z_{α} and z_{β} = z-statistics for α (e.g., 1.96) and β (e.g., 0.84) errors; sd = standard deviation for cost (sd_c) and effect (sd_q)

Complexities

- Complexities arise because 1) difference being assessed is the difference in NMB (WΔQ – ΔC) and 2) standard deviation of NMB is a complicated formula
- Data needed to calculate sample size include:
  - Difference in cost
  - SD of cost
  - Difference in effect
  - SD of effect
  - Z_α and Z_β
  - Correlation of the difference in cost and effect
  - Willingness to pay

Sample Size / Power Formulas

- Sample Size

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

- Power

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

  - e.g., if z_β = -1.96 = 2.5% power; -0.84 = 20% power; 0 = 50% power; .84 = 80% power; 1.28 = 90%

Null Hypothesis, NMB

- Formula identifies a sample size that provides a 1−β% chance to have 1−α% confidence for rejection of null hypothesis that NMB (NMB = WQ − C) calculated by use of W equals 0
  - If assumptions about C, Q, sdc, sdq, and ρ are correct and if α=0.05 and β=0.2, then
    - In approximately 800 of 1000 repeated experiments, lower limit of 95% confidence interval for difference in NMB will be greater than 0
    - In approximately 200, 95% confidence intervals will either include 0 or have an upper limit less than 0

Null Hypothesis, CER and Acceptability

- Formula also identifies a sample size that provides a 1−β% chance to have 1−α% confidence for rejection of null hypothesis that cost-effectiveness ratio equals W (i.e., that 1−α% confidence interval for cost-effectiveness ratio excludes W)
- Or equivalently, identifies a sample size that provides a 1−β% chance for rejection of null hypothesis that at W, fraction of joint distribution of difference in cost and effect that is acceptable is greater than α/2% and less than 1−(α/2)%

Correlation of the Difference

- 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 positive correlation than when they are characterized by negative correlation

Effect of SD_q VS SD_c on Sample Size

- Commonly thought that sample size for cost-effectiveness driven more by the standard deviation for cost than it is by SD for effect
  - If not, why would we need a larger sample for the economic outcome than you do for the clinical outcome?
- However, if willingness to pay is substantially greater than the standard deviation for cost, percentage changes in QALY SD can have a substantially greater effect on sample size than will equivalent percentage changes in cost SD

Economic Vs Clinical Sample Sizes

- Sample size required to answer economic questions often larger than the sample size required to answer clinical questions
  - But it need not be
- ΔC and ΔQ are a joint outcome just as differences in nonfatal CVD events and all cause mortality are often combined into a joint outcome
- In the same way that we can have more power for the joint cardiovascular outcome than either individual outcome alone, we can have more power for cost-effectiveness than we do for costs or effects alone

Where to Obtain the Necessary Data?

- When therapies are already in use: Expected differences in outcomes and standard deviations can be derived from feasibility studies or from records of patients
- Simple correlation between observed costs and effects may be an adequate proxy for the measure of correlation used for estimating sample size
- For novel therapies, information may need to be generated by assumption
  - e.g., sd from usual care will apply to new therapy, etc.

Willingness to Pay and Identification of an Appropriate Outcome Measure

- Sample size calculations require stipulation of willingness to pay for a unit of outcome
- In many medical specialties, researchers use disease specific outcomes
- Can calculate a cost-effectiveness ratio for any outcome (e.g., cost/case detected; cost/abstinence day), to be informative, outcome must be one for which we have recognized benchmarks of cost-effectiveness
  - Argues against use of too disease-specific an outcome for economic assessment

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
  - Command do ssizeprg simply loads programs; it does not calculate anything
- “Doing” ssizeprg also loads a documentation program named ssizeprgdoc

3 Sample Size Programs

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

3 Power Programs

- cepow1i: Calculates power to detect NMB greater than 0 under assumption that sample size and standard deviations for cost and effect are common for both treatment groups
- cepow2i: Calculates power to detect NMB greater than 0 under assumption that sample size is same in both groups, but standard deviations for cost and effect differ
- cedpowi: Calculates power to detect NMB greater than 0 under assumption that sample size differs between 2 groups, but standard deviations for cost and effect are same

ssizeprg.do (cont.)

- All 6 programs report sample size and power for comparison of 2 arms in a trial (for multi-arm trials, programs report sample size and power for individual pair-wise comparisons)
- Sample size estimates from programs have been validated in simulations and yield results that match those derived from 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 expected difference in cost = 25; expected difference in QALYs = 0.05; expected SDs for cost and QALYs = 1000 and 0.195, respectively; expected correlation of difference = -0.1; maximum WTP = 75,000; and want a 2-tailed alpha = .05 and a 1-tailed beta = 0.8:
  - Point estimate = 25 / 0.5 = 500 / QALY
- Calculate necessary sample size:

cess1i 25 .05 1000 .195 -.1 75000 .05 .8

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 ***	247

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
```

Code for Looping Calculations

```
foreach wtp in 30000 50000 75000 100000 125000 {
cess1i 25 .05 1000 .195 -.1 `wtp' .05 .8
}
```

Fill In Following Table

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

WTP	Sample Size
30,000	263
50,000	252
75,000	247
100,000	245
125,000	244

Dropout

- These sample size estimates are appropriate if we expect no dropout from trial
- If we instead anticipate 10% dropout, we will want to divide these sample size estimates by 0.9

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 expected difference in cost = 25; expected difference in QALYs = 0.05; expected SDs for cost and QALYs = 1000 and 0.195, respectively; expected correlation of difference is -0.1; your maximum WTP is 75,000; want a 2-tailed alpha of .05; and current sample size plans are for 247 per group
- Calculate power of experiment

cepow1i 25 .05 1000 .195 -.1 75000 .05 247

cepow1i 25 .05 1000 .195 -.1 75000 .05 247

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(sampsize) =  247
       r(nmb) =  3725
        r(wdi) =  -26219
      r(power) =  .8009
```

Code for Looping Calculations

```
foreach ssize in 150 200 247 300 350 {
cepow1i 25 .05 1000 .195 -.1 75000 .05 `ssize'
}
```

Fill In Following Table

- Assuming that C=25; Q=0.05; SDc= 1000; SDq=0.195; correlation=-0.1; 2-tailed alpha=0.05; and sample size = 246/group, fill in following table

N/Group	Power
150	58.9
200	71.4
247	80.1
300	87.1
350	91.6

Power Tables

- If we anticipate 10% dropout, we will want to use "effective sample size" (e.g., 0.9 * 246) when we make our calculations

"Typical" Sample Size Table, W

	Sample Size Per Group
WTP	Exp 1 *
20,000	321
30,000	273
50,000	234
75,000	214
100,000	204
150,000	194

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

Sample Size Can Increase with Increasing W

	Sample Size Per Group	
WTP	Exp 1	Exp 2 *
20,000	321	36
30,000	273	42
50,000	234	68
75,000	214	92
100,000	204	108
150,000	194	127

* ΔC=-120; ΔQ=0.015; sd_c=1000; sd_q=.05; ρ=0.8; α=.05; 1-β=.8

Sample Size Not Necessarily Monotonic With W

	Sample Size Per Group		
WTP	Exp 1	Exp 2	Exp 3 *
20,000	321	36	178
30,000	273	42	158
50,000	234	68	***151***
75,000	214	92	154
100,000	204	108	156
150,000	194	127	160

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

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

Glick HA. Sample size and power for cost-effectiveness analysis (part 2). The effect of maximum willingness to pay. Pharmacoeconomics. 2011;29:287-96.