

Sampling Uncertainty and Patient-Level Cost-Effectiveness Analysis

Statistical Considerations in Health Economic Evaluations

ISPOR 18th Annual International Meeting

May 19, 2013

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



Good Value for Cost

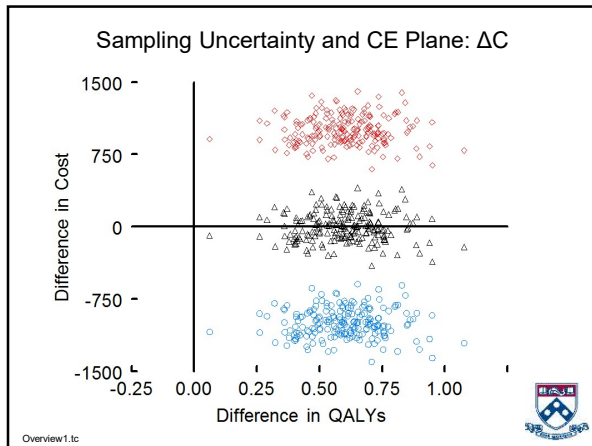
- Common goal of economic analysis: identify when we can be confident that a therapy is good value compared to another
- Threat to confidence arises because economic result observed in an experiment may not truly reflect result in population
 - Single sample drawn from a population
- Referred to as sampling (or stochastic) uncertainty
- Methods for estimating sampling uncertainty for economic outcomes have much in common with methods used for clinical findings

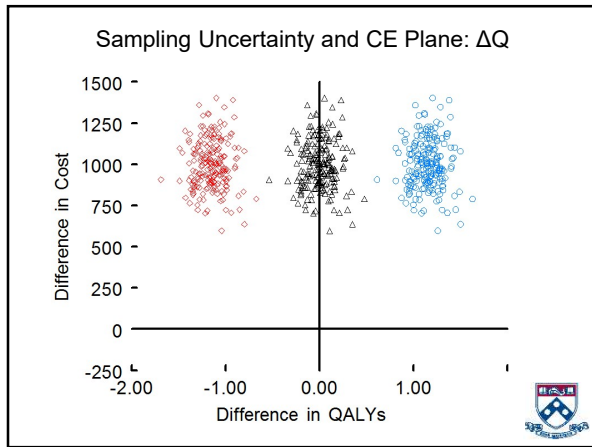


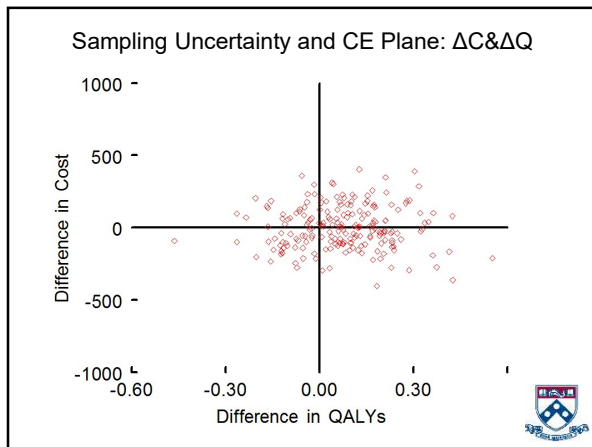
Outline

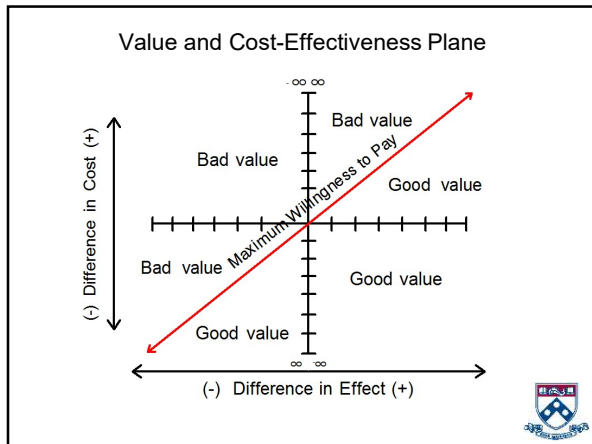
- Describe methods for identifying when we can and cannot be confident about a therapy's value
 - Acceptability curves
 - CI for NMB
 - CI for CER
- Goal is to demonstrate quantification and interpretation of sampling

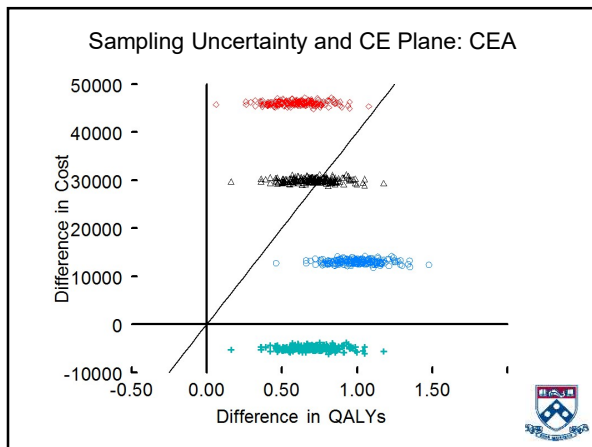


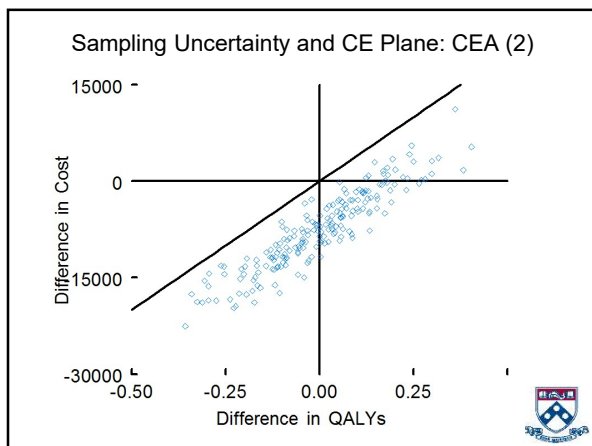


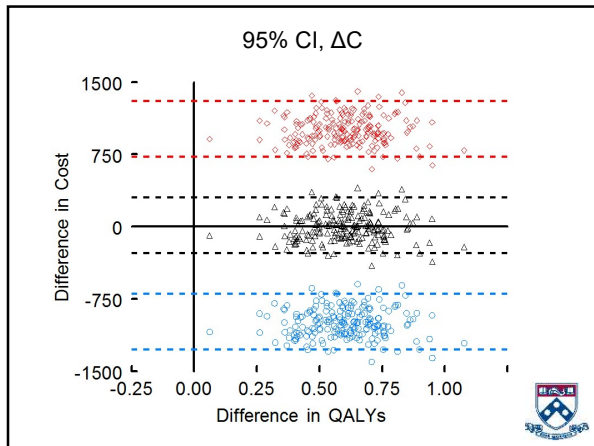


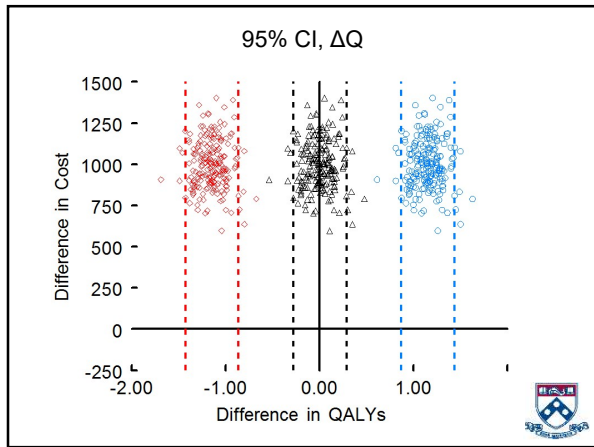


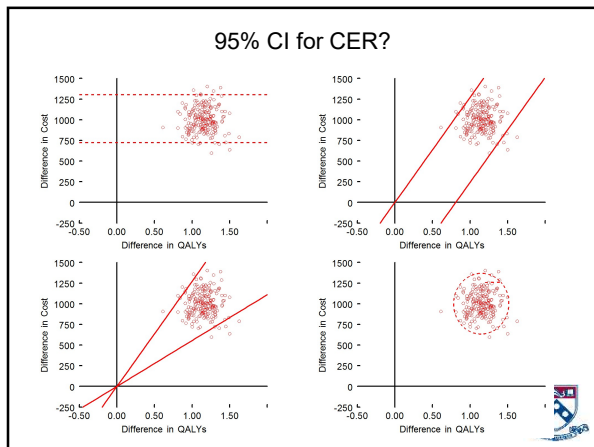










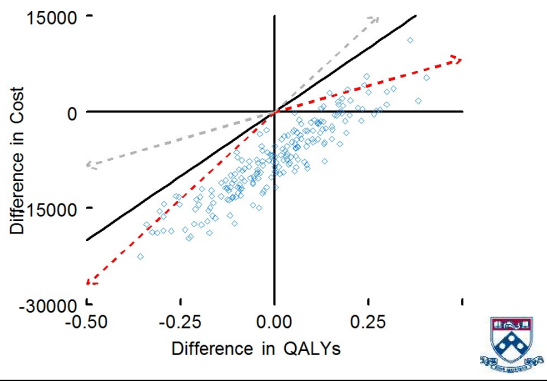


CI Issues;

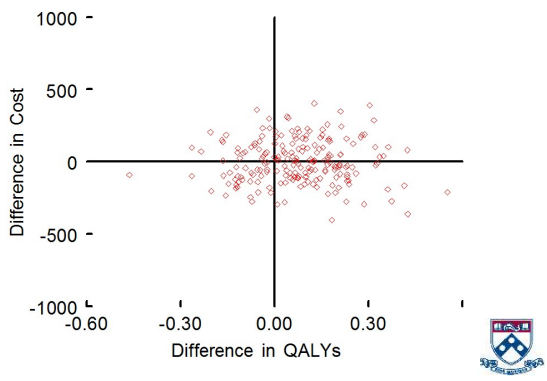
- # of methods available?
- What is the threshold, maximum willingness to pay?
 - Differs across jurisdictions
 - Differs within jurisdictions
- Should we be 95% confident?



95% CI for CER, ΔC and ΔQ Not Significant



95% CI for CER ???



Nonparametric Methods

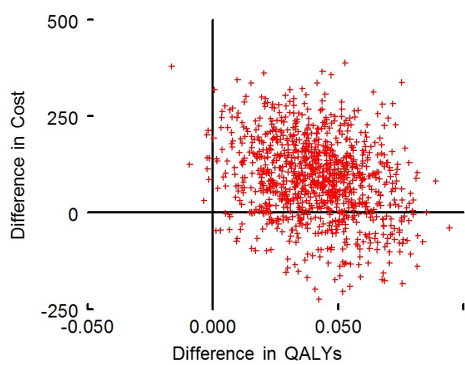


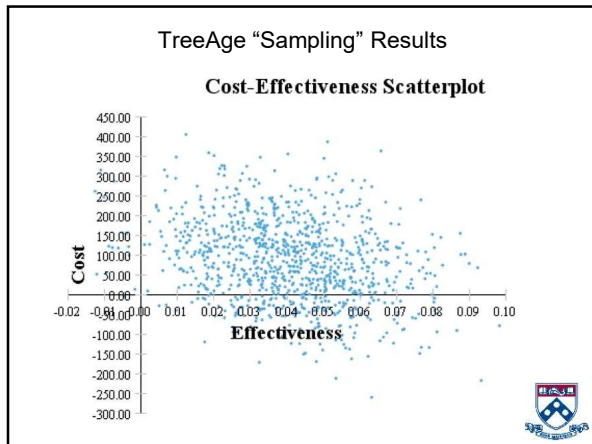
Example #1

- Subsample (N=1000) of bootstrapped rchapter5.dta regression results from Jalpa's lecture
 - Cost, power 0.65, poisson
 - ΔC , mean = 88.09, SE = 103
 - QALYs, power 1.65, poisson
 - ΔQ , mean = 0.0408, SE = 0.018
 - Correlation of difference, -0.2523
 - Incremental cost-effectiveness ratio: 2159

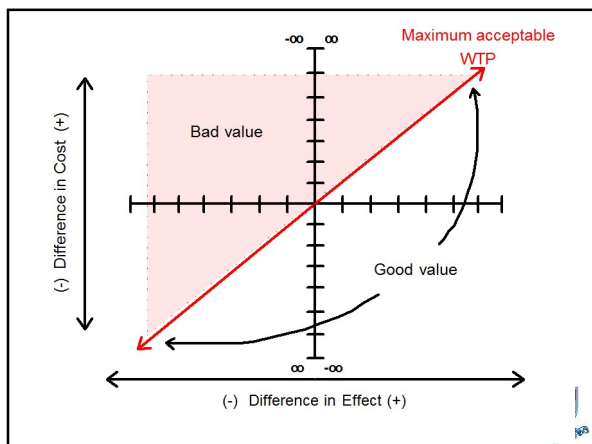


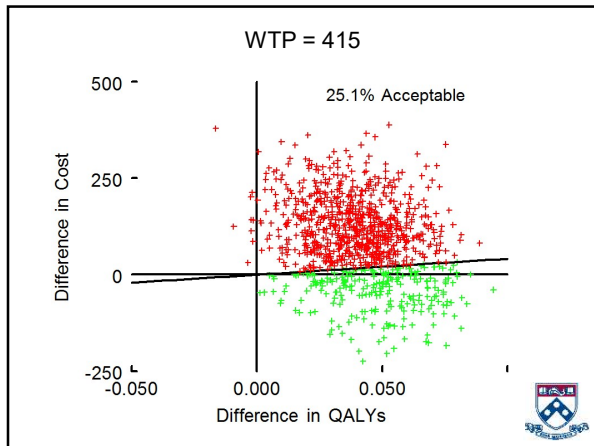
Bootstrapped Data

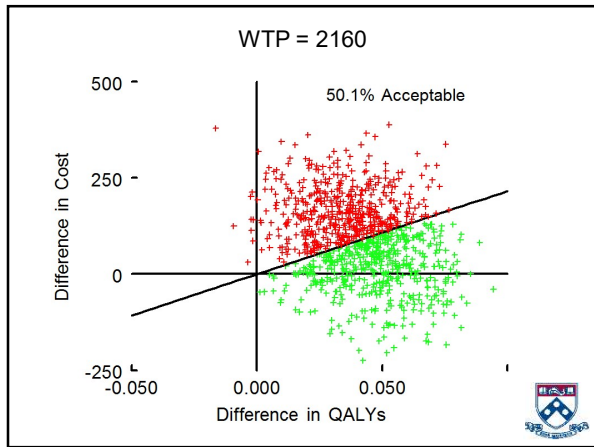


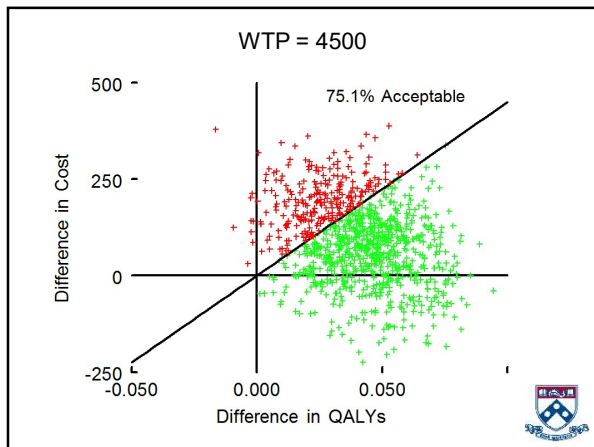


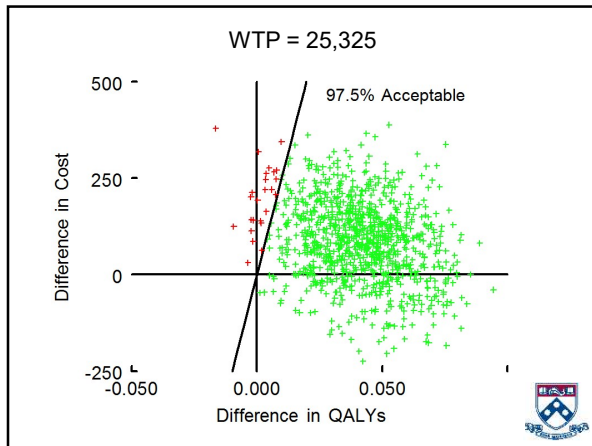
- ### Acceptability Curve
- Acceptability criterion defined on cost-effectiveness plane as a line through origin with slope equal WTP
 - Proportion of distribution of difference in cost and effect falling below and to right of line is "acceptable" (i.e., has positive NMB)
 - Proportion acceptable for one therapy = 1-proportion acceptable for alternative therapy
 - In pairwise comparison, no additional information from plotting 2 lines, one for each therapy
 - Proportion falling above and to left of line is "unacceptable"
 - Proportion unacceptable for one therapy = 1-proportion unacceptable for alternative therapy

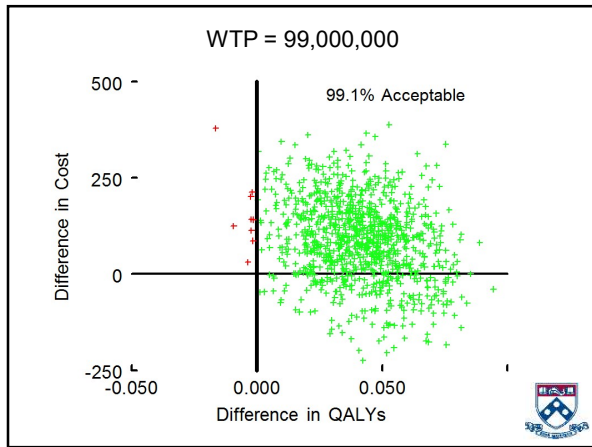


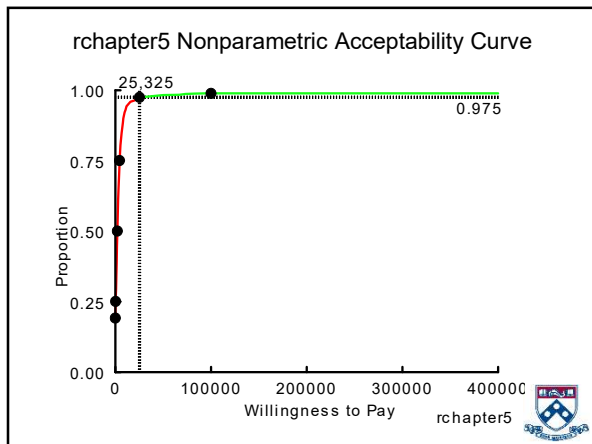












2-tailed Confidences Statements for Acceptability Curve

- If curve has a height that is ≤ 0.025 th, 95% confident therapy is bad value
 - i.e., for current study, for most negative values of WTP (data not shown)
- If curve has a height that is ≥ 0.975 th, 95% confident therapy is good value
 - i.e., for values of WTP $\geq 25,325$
- If curve falls between 0.025 and 0.975, cannot be 95% confident that value of 2 therapies differs
 - i.e., for values of WTP > 0 and $< 25,325$



Stata Programs

- Provide 4 Stata .do files that contain programs for calculating and plotting analyses of sampling uncertainty
- 2 calculate and plot nonparametric measures of sampling uncertainty (focused on today)
 - bsceaprogs
 - bsceagraphs
- 2 calculate and plot parametric measures of sampling uncertainty
 - iprogs
 - ceagraphs
- Running .do files (e.g., do bsceaprogs) loads programs; it does not calculate anything



bsceaprogs.do

- Contains 5 programs related to sampling uncertainty for datasets that represent bootstrap replicates or repeated samples from second order Monte-Carlo analysis PLUS a help file (bsceaprogsdoc)
 - bsaccept1: Calculates % acceptable and p-value for a user-specified value of WTP
 - bsaccept: Calculates % acceptable and p-values for program-determined values of WTP
 - bsnmb1: Calculates NMB point estimate, CI, and p-value for a user-specified value of WTP
 - bsnmb: Calculates NMB point estimates, CI, and p-values for program-determined values of WTP
 - bscicer: Calculates CI for CER

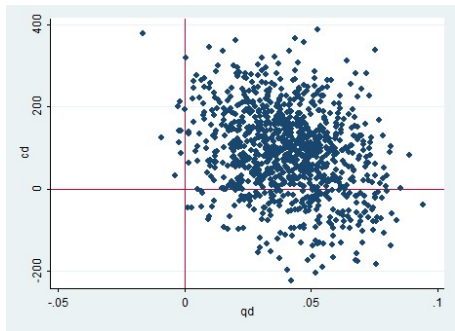


bsceagraphs.do

- Contains 3 programs that graph results of bsaccept, bsnmb, and bscicer PLUS a help file (bsceagraphsdoc)
 - bsaccgraph (works with bsaccept): Draws acceptability curve
 - bsnmbgraph (works with bsnmb): Draws NMB graph
 - bscicergraph (works with bscicer): Graphs bootstrap cloud as well as upper and lower limits of its CI (if defined)



twoway scatter pglmpqcd pglmpqpd,xline(0) yline(0)



bsceaprogdoc: bsaccept

- * PROGRAM: BSACCEPT
- * USES BOOTSTRAPPED DATA TO DEFINE
- * NONPARAMETRIC ACCEPTABILITY CURVE
- * COMMAND LINE: bsaccept [COST] [EFFECT]
- * The 2 arguments are both names of variables
- ** [COST] = Name of difference in cost variable
- ** [EFFECT] = Name of difference in effect variable
- * Saved Results
- * r(accmat)



Run bsaccept for power/poisson Cost and QALYs

```
. quietly do bsceaprogs  
. use dataforslides  
. bscicer pglmppcd pglmppqd
```



bsaccept pglmppcd pglmppqd

W	% Accept	P-value
.	.	.
-3374	0.02200	0.0440
-2803	0.02700	0.0540
-2350	0.03900	0.0780
.	.	.
1805	0.44700	0.8940
2158	0.50100	0.9980
2533	0.54600	0.9080
.	.	.
20000	0.97400	0.0720
29723	0.97600	0.0480
30000	0.97600	0.0480
.	.	.



Return List

```
. return list  
macros:  
  r(cmd) : "bsaccept"  
matrices:  
  r(accmat) : 122 x 3
```



Return List (cont.)

- To view r(accmat): matrix list r(accmat)
- To access data in r(accmat):
 - First, create a new matrix: matrix [name]=r(accmat)
 - Second, transform new matrix into a dataset (svmat [matrix name])
 - Results in 3 variables named [name]1, [name]2, and [name]3, where
 - [name]1 = wtp
 - [name]2 = % acceptable
 - [name]3 = p-value



Saving Results of bsaccept

```
preserve
drop _all
matrix accmat=r(accmat)
svmat accmat
ren accmat1 wtp
ren accmat2 accept
ren accmat3 pval
save [FILENAME.DTA], replace
restore
```



bsceagraphsdoc: bsaccgraph

- * PROGRAM: bsaccgraph
- * This program draws the acceptability curve. It is meant to be run directly after running the bsaccept program (or soon enough after that the r(accmat) return matrix is still resident in memory.
- * As currently written, the program draws the curve for values of wtp between 0 and 125,000. To change the upper bound wtp in the graph, open the program file and revise the statement wtp<125000.
- * The default setting draws the acceptability curve alone.
- * Optionally, you can add horizontal confidence lines by specifying your desired confidence level (e.g., for 2-tailed 95% confidence, 0.95).



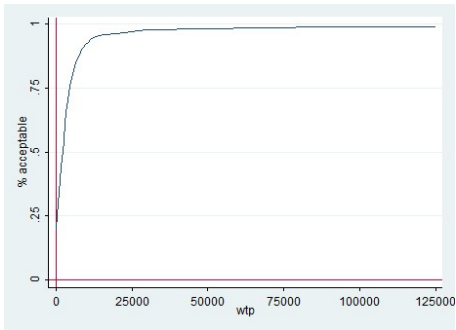
bsceagraphsdoc: bsaccgraph (cont.)

- * Command Line: bsaccgraph
- * For optional horizontal lines (e.g., for 95% 2-tailed confidence):
- * confidence):
* bsaccgraph 0.95
- *
- * Saved Results
- * r(accmat)
- * r(cmd)
- *

quietly do bsceagraphs



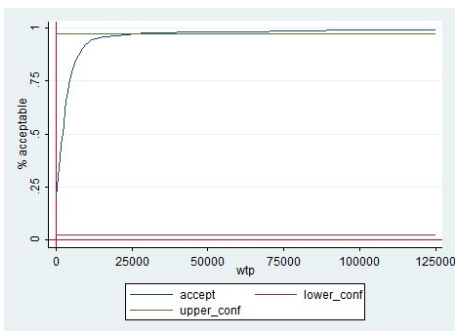
bsaccgraph



bsaccept pglmppcd pglmppqd



bsaccgraph .95



bsaccept pglmppcd pglmppqd



Saving the Graph

graph export [filename.extension],replace

e.g. graph export rc5acc.png,replace

Commonly used extensions include .png, .wmf, .pdf, .ps, and .tif



Net Monetary Benefit

- Composite measure (part cost-effectiveness, part cost benefit analysis), usually expressed in dollar terms, derived by rearranging cost-effectiveness decision rule:

$$W^* > \Delta C / \Delta Q$$

where W^* = maximum acceptable cost-effectiveness ratio (e.g., 50,000 per QALY)

- NMB routinely (but not necessarily) expressed on cost scale, known as net monetary benefits (NMB)

$$(W \times \Delta Q) - \Delta C$$

- Particularly important for statistical evaluation of cost-effectiveness analysis (e.g., sample size; direct statistical testing by use of patient-level data; etc.)



Expected NMB

$$NMB = (W^* \Delta Q) - \Delta C$$

- For a WTP of 50,000, NMB for rchapter5:
 $(50,000 * .0408) - 88 = 1952$
- Study result is a difference in means of net benefits, not a ratio of means, and is always defined (i.e., no odd statistical properties like ratio) and continuous
- Unlike cost-effectiveness ratio, standard error of net benefits always defined
- Given not all decision making bodies have agreed upon maximum willingness to pay, routinely estimate net benefit over a range of policy relevant values of willingness to pay

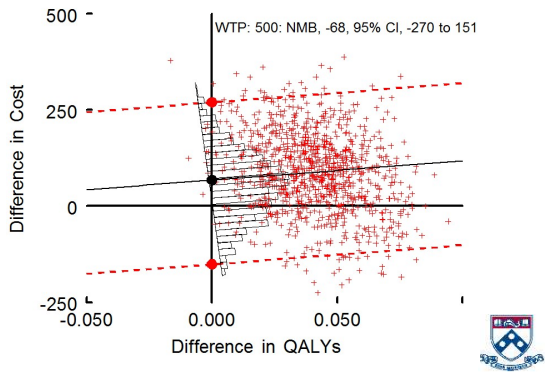


Net Benefit Graphically

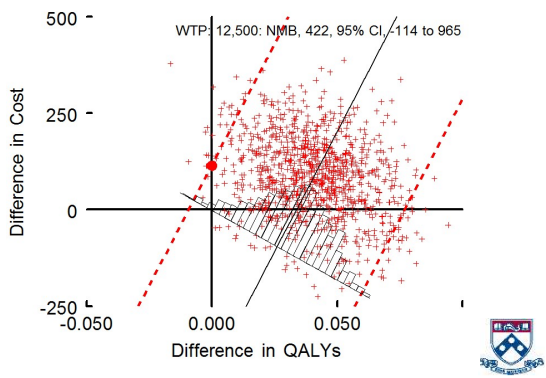
- Defined on cost effectiveness plane using a family of lines
- Slope of all lines equals W
- Each line represents a single value of NMB and equals $-intercept$
 - Because when $\Delta Q=0$, $W\Delta Q$ drops out of equation and left with $-\Delta C$
- 95% CI for NMB defined by identifying 2 NMB lines that each omit 2.5% of distribution of difference in cost and effect

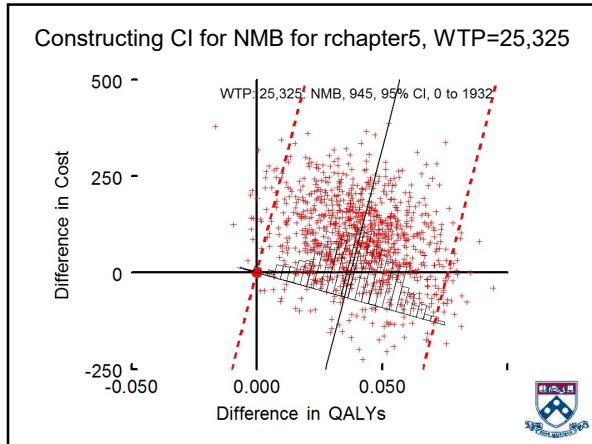


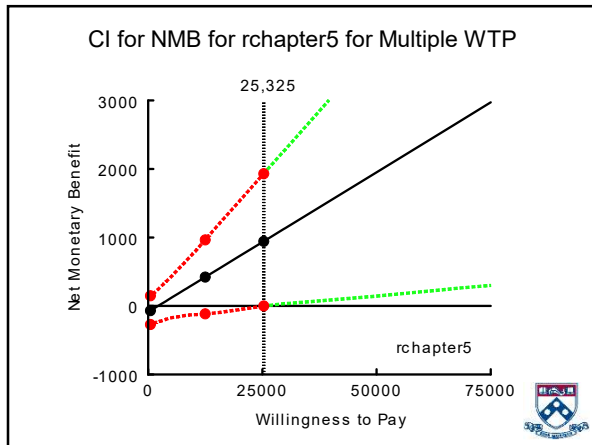
Constructing CI for NMB for rchapter5, WTP=500



Constructing CI for NMB for rchapter5, WTP=12,500







- Confidence Statements for CI for NMB
- If both confidence limits are negative, confident therapy is bad value
 - i.e., for current study, for most negative values of WTP (data not shown)
 - If both confidence limits are positive, confident therapy is good value
 - i.e., for values of $WTP \geq 25,325$
 - If one confidence limit is positive and one is negative, can't be confident that value of therapies differs
 - i.e., for values of $WTP > 0$ and $< 25,325$

```

bsceaprogdoc: bsnmb

* PROGRAM: BSNMB
* USES BOOTSTRAPPED DATA TO DEFINE POINT
* ESTIMATES AND CI FOR NMB GRAPH

* COMMAND LINE: bsnmb [COST] [EFFECT] [CI]

* The 2 arguments are both names of variables
** `1' Name of difference in cost variable
** `2' Name of difference in effect variable
** `3' confidence interval, as decimal (e.g., 0.95 for 95%)

* Saved Results
* r(CI)
* r(nmbmat)

```



bsnmb pglmpcd pglmpqd .95

W	NMB	95 % Lower limit	95 % Upper limit	P-value
.
-3374	-226	-423	-12	0.0440
-2803	-202	-394	2	0.0540
-2350	-184	-372	20	0.0780
.
1805	-13	-238	221	0.8940
2158	-0	-230	243	0.9980
2533	14	-222	267	0.9080
.
20000	728	-51	1531	0.0720
29723	1125	35	2262	0.0480
30000	1136	36	2282	0.0480
.



```

Return List

. return list

scalars:
    r(CI) = 95

macros:
    r(cmd) : "bsnmb"

matrices:
    r(nmbmat) : 122 x 5

```



Viewing and Accessing r(nmbmat)

- Same as viewing and accessing r(accmat)
 - See prior slides for r(accmat)



Saving Results of bsnmb

```
preserve
drop _all
matrix nmbmat=r(nmbmat)
svmat nmbmat
ren nmbmat1 wtp
ren nmbmat2 nmb
ren nmbmat3 ll
ren nmbmat4 ul
ren nmbmat5 p
save [FILENAME.DTA], replace
restore
```



bsceagraphsdoc: bsnmbgraph

- * PROGRAM: bsnmbgraph
- * This program draws the nmbgraph. It is meant to be
- * run directly after running the bsnmb program (or soon
- * enough after that the r(nmbmat) return matrix is still
- * resident in memory.
- * As currently written, the program draws the curve for
- * values of wtp between 0 and 125,000. To change the
- * upper bound wtp in the graph, open the program file and
- * revise the statement wtp<125000.
- *

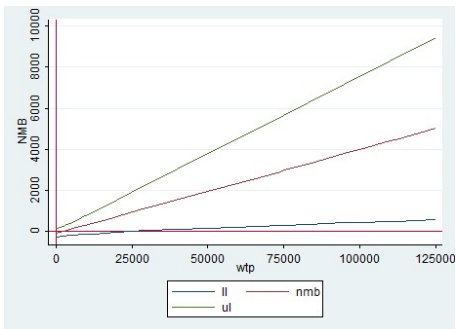


bsceagraphsdoc: bsnmbgraph (cont.)

- * Command Line: bsnmbgraph
- *
- * Saved Results
- * r(Cl)
- * r(nmbmat)
- * r(cmd)
- *



bsnmbgraph



bsnmb pglmpcpd pglmpcpd .95



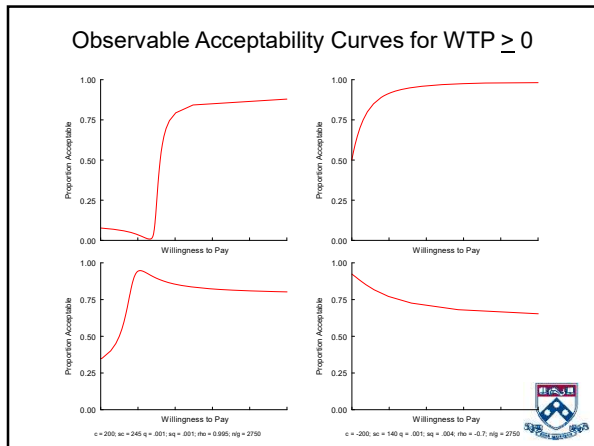
Saving the Graph

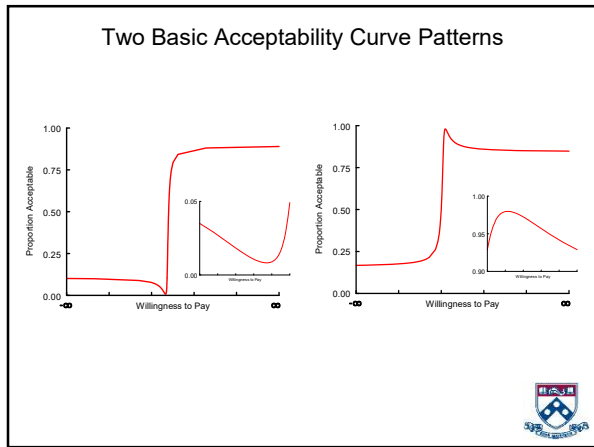
graph export [filename.extension],replace

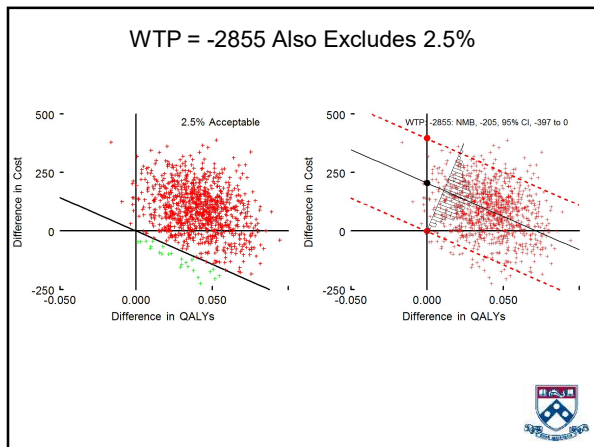
e.g. graph export rc5nmb.png,replace

Commonly used extensions include .png, .wmf, .pdf, .ps,
and .tif

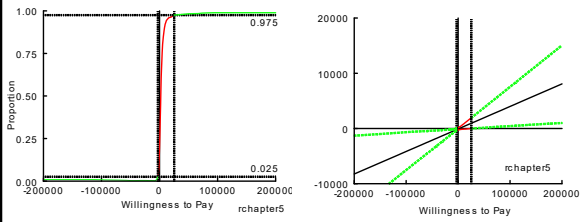








Extended Acceptability Curve and NMB Graph



- For both graphs, confident of bad value for WTP < -2855; confident of good value for WTP > 25,325; and not confident of value between -2855 and 25,325

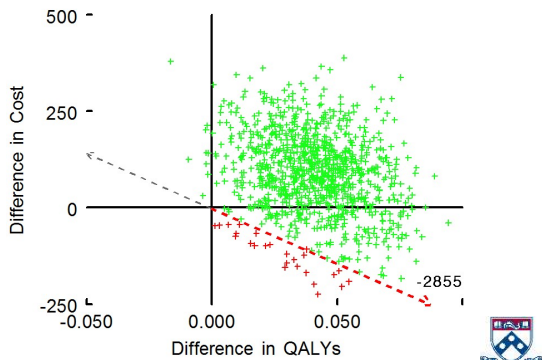


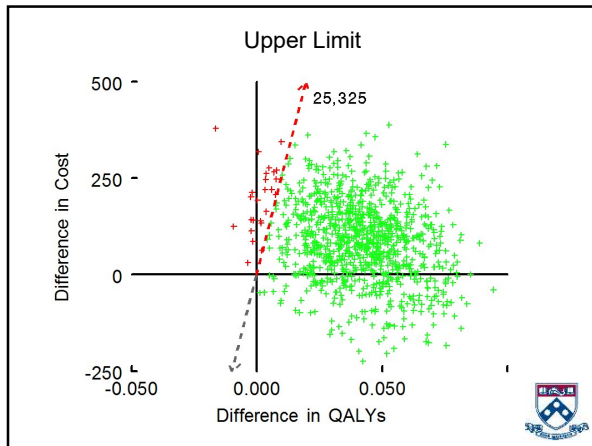
Confidence Intervals for Cost-Effectiveness Ratios

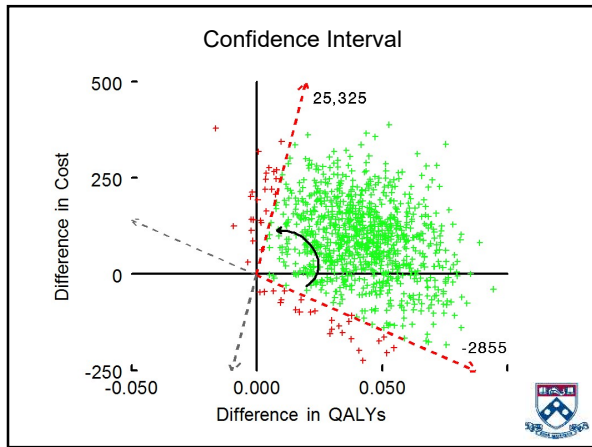
- Common suggestion for constructing CI:
 - Order ratios from smallest to largest
 - Identify ICER of 2.5th percentile (e.g., 26th ordered observation out of 1000) and 97.5th percentile (e.g., 975th observation out of 1000)
- Technically, not an order statistic (although in many cases equivalent to one)
- Technically, lines through origin that exclude $\alpha/2$ of joint distribution of difference in cost and effect
- Independent of whether lower limit is a larger or smaller number than upper limit, on cost-effectiveness plane, interval stretches counter-clockwise from lower (clockwise) limit to upper (counter-clockwise) limit

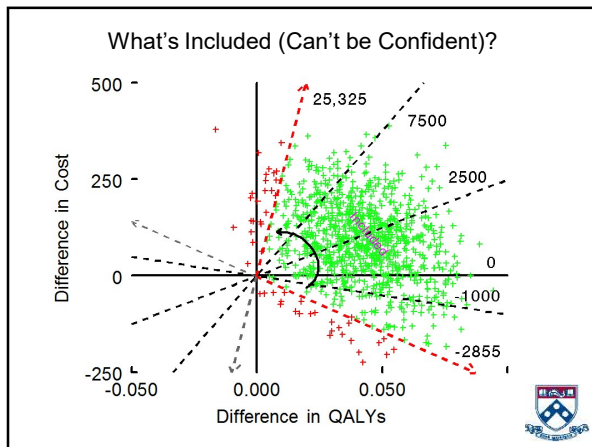


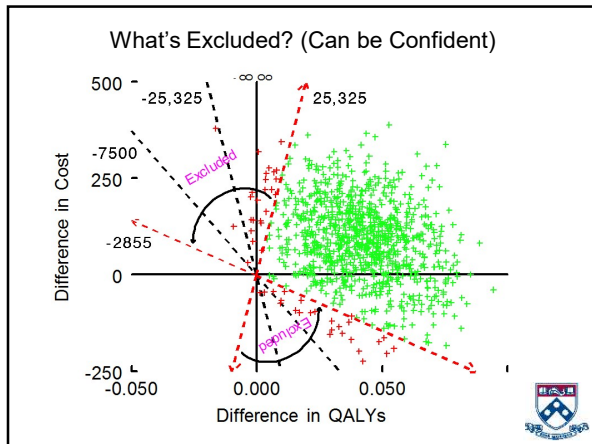
Lower Limit, CI for CER











- ### Confidence Statements for CI for CER
- When (as in current experiment) lower limit is a smaller number than upper limit:
 - If lower limit is greater than WTP, confident therapy is bad value
 - i.e., for current study, for most values < -2855
 - If upper limit is less than WTP, confident therapy is good value
 - i.e., for values of WTP $\geq 25,325$
 - If WTP is greater than lower limit and less than upper limit, can't be confident that value of therapies differ
 - i.e., for values of WTP > -2855 and < 25,325

bsceaprogdoc: bscicer


```

* PROGRAM: BSCICER
* USES BOOTSTRAPPED DATA TO DEFINE
* NONPARAMETRIC PERCENTILE AND ACCEPTABILITY
* METHOD CI FOR CER
* COMMAND LINE: bscicer [COST] [EFFECT] [CI]
* The 2 arguments are both names of variables; the 3rd is a
* number
** `1' Name of difference in cost variable
** `2' Name of difference in effect variable
** `3' confidence interval, as decimal (e.g., 0.95 for 95%)
  
```

```

bscicer pglmpcd pglmpqd .95
Bootstrap percentile 95 % Confidence
Interval
Lower limit (quadrant):    -2852  ( 4 )
Upper limit (quadrant):    25323  ( 1 )
Density omitted by:
  Lower limit:              2.5 %
  Upper limit:              2.5 %
Fraction of density
uniquely excluded:         5 %
Fraction of density excluded,
wedge interpretation:      5 % (cont.)


```



```

bscicer (cont.)
Bootstrap acceptability 95 % Confidence
Interval
Lower limit:                -2852
Upper limit:                25323
Density omitted by:
  Lower limit:              2.5 %
  Upper limit:              2.5 %
Fraction of density
uniquely excluded:         4.6 %
Data for Immediate Form programs
Difference in costs:        88.085314
SE, difference in costs:    103.00336
Difference in effects:      .04081035
SE, difference in effects:  .01801509
Correlation of differences: -.2523440


```



```

Return List
. return list
scalars:
  r(bsp1l) = -2851.925396
  r(bspul) = 25322.526603
  r(bsall) = -2852
  r(bsaul) = 25323
  r(ci) = .95
macros:
  r(cmd) : "bscicer"
  r(cost) : "pglmpcd"
  r(effect) : "pglmpqd"

```



“Acceptability” vs “Percentile” CI

- Acceptability CI for CER
 - Defined by identifying lines through origin that each exclude $\alpha/2$ of joint distribution of difference in cost and effect
 - Can be shown to be dependably accurate
- Percentile CI for CER
 - Defined by use of non-naïve ordering of replicates (orders lexicographically by quadrant and by ratio)
 - Naïve ordering (most negative to most positive) fails in accuracy when replicates fall on both sides of y axis
 - Non-naïve ordering can fail in accuracy when replicates fall in 3 or all 4 quadrants of CE plane



bsceagraphsdoc: bscicergraph

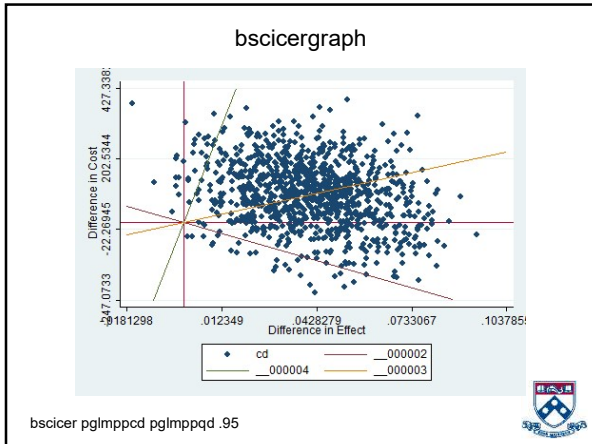
- * PROGRAM: bscicergraph
- * This program graphs the bootstrap cloud as well as the upper and lower limits of its confidence interval on the cost-effectiveness plane. It is meant to be run directly after running the bscicer program (or soon enough after that the return list is still defined in memory.
- * Command Line: bscicergraph



bsceagraphsdoc: bscicergraph (cont.)

- * Saved Results
- * r(bsppl) (percentile lower limit)
- * r(bspul) (percentile upper limit)
- * r(bsall) (acceptability lower limit)
- * r(bsaul) (acceptability upper limit)
- * r(CI)
- * r(cost)
- * r(effect)
- * r(cmd)
- *



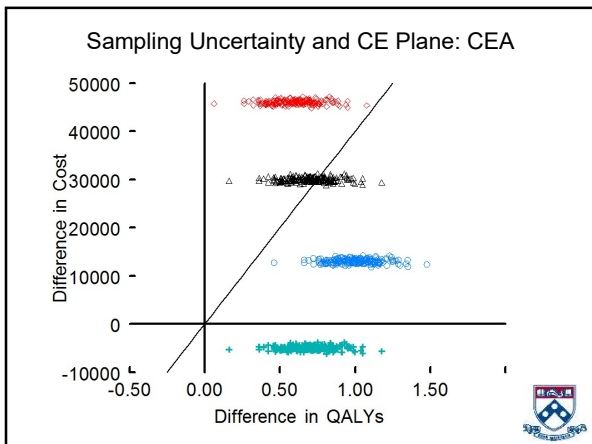


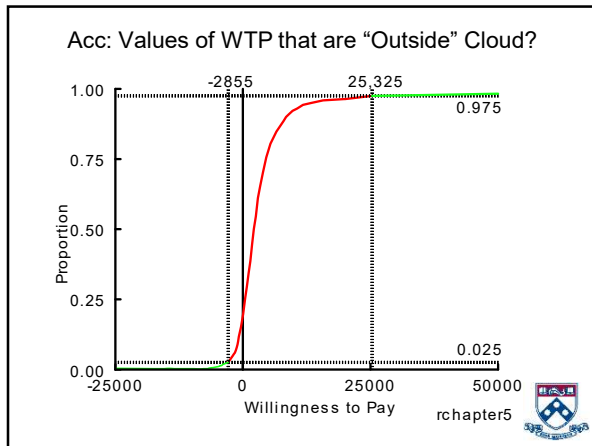
Saving the Graph

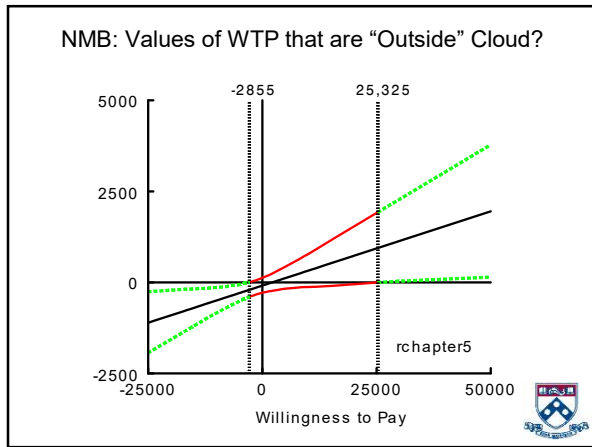
graph export [filename.extension],replace

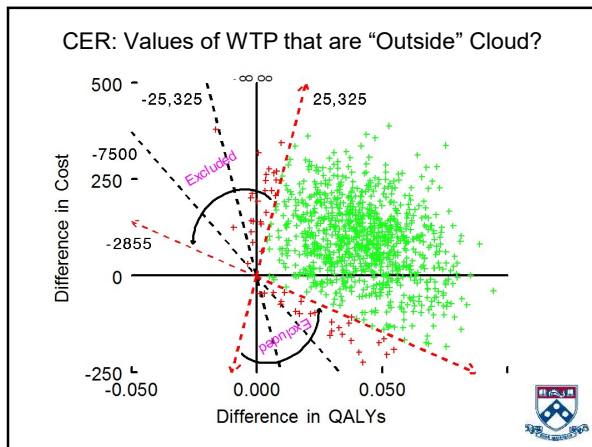
e.g. graph export rc5cicer.png,replace

Commonly used extensions include .png, .wmf, .pdf, .ps, and .tif







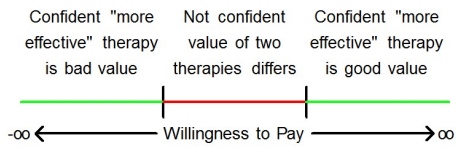


"Pattern 1" Findings

- Refer to findings like those in rchapter5 experiment as pattern 1 findings
- Occur when difference in effect is significant
- Know pattern 1 finding being observed when:
 - Confidence interval for cost-effectiveness ratio excludes Y axis (i.e., $LL < PE < UL$)
 - Both NMB confidence limits curves intersect x-axis (0) once
 - Acceptability curve intersects horizontal lines drawn at both 0.025 and 0.975



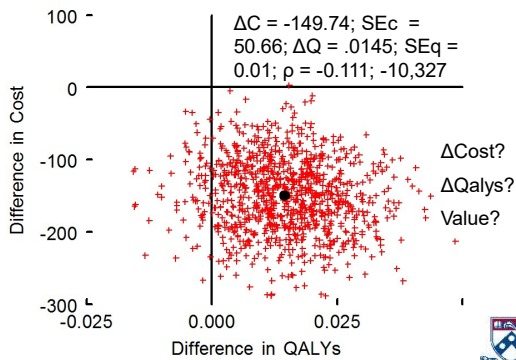
Pattern 1 Findings (2)

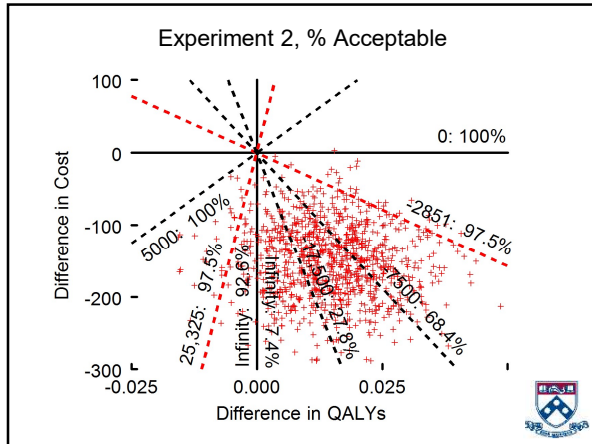


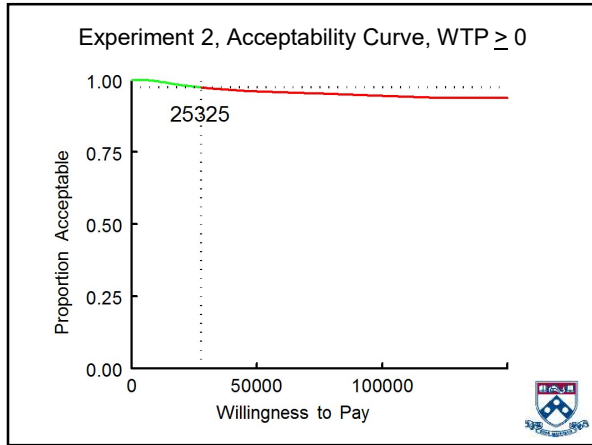
* In cases where boundaries between regions occur at negative values of willingness to pay, may not always observe all 3 regions on an acceptability curve or NMB plot

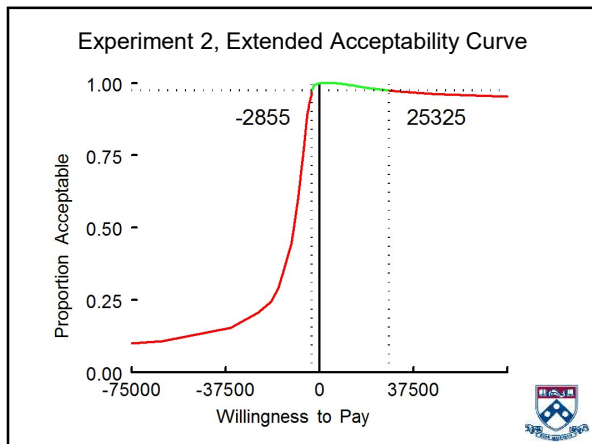


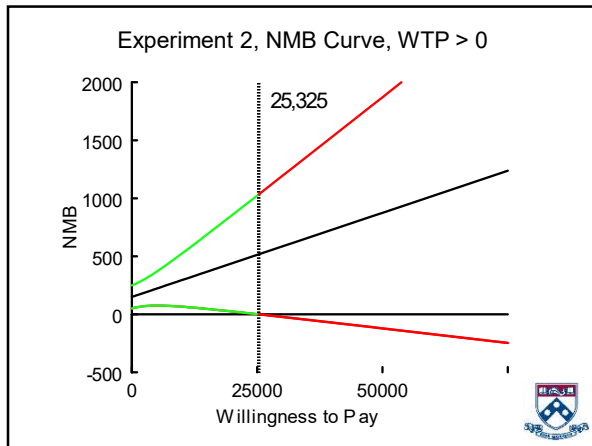
Experiment 2: View Data

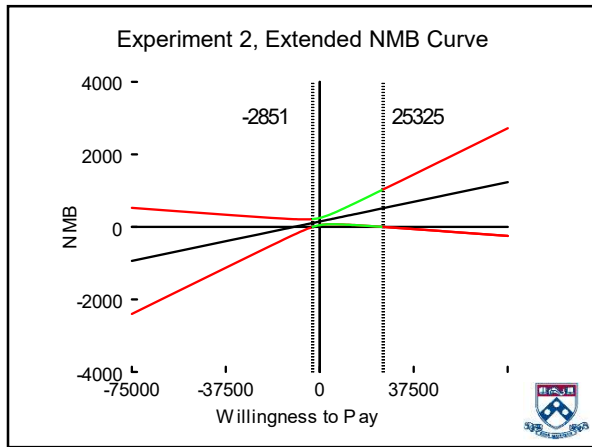


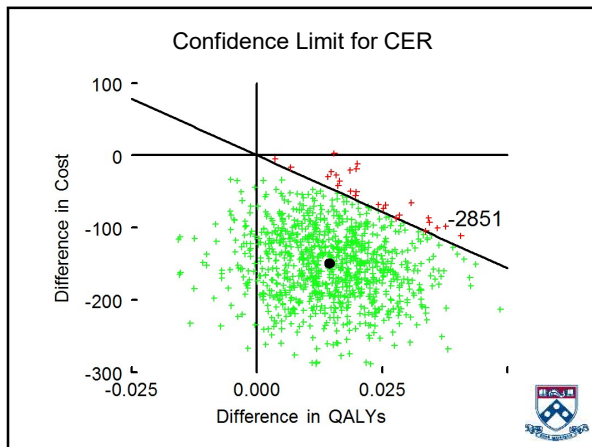


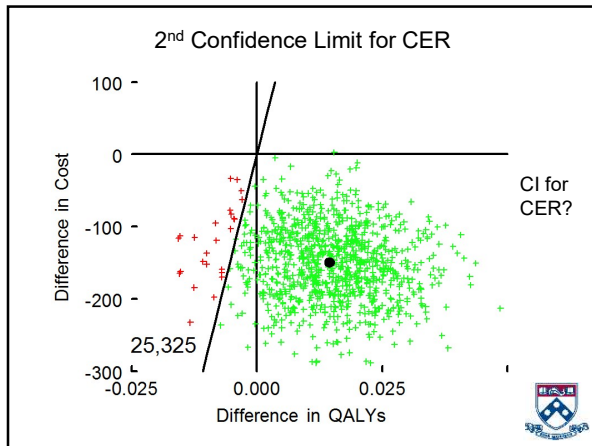


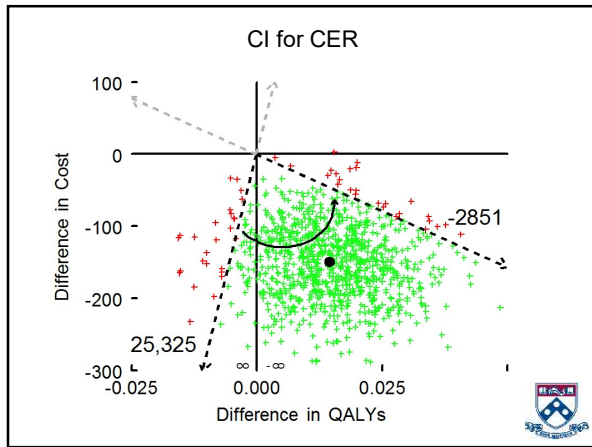


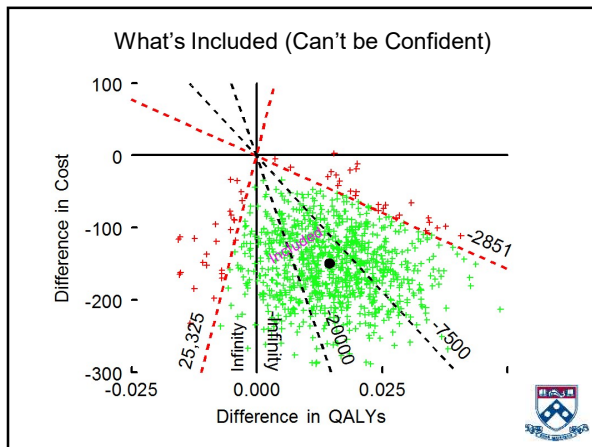


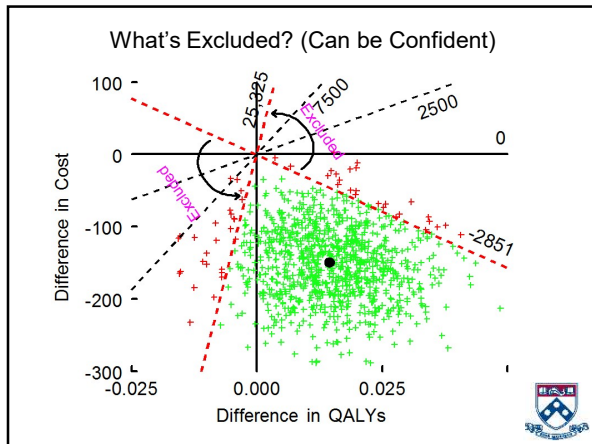












Confidence Statements for CI for CER

- When lower limit is a larger number than upper limit
Interval ranges between $-\infty$ and upper limit and between lower limit and ∞
 - If WTP greater than upper limit and less than lower limit, can be confident that one of the therapies is good value
 - i.e., for current study, for values values of WTP ≥ -2851 and $\leq 25,325$
 - If WTP less than upper limit or greater than lower limit, can't be confident that value of therapies differ
 - i.e., for current study, for values of WTP < -2851 and $> 25,325$

When the Lower Limit is Larger than Upper Limit

- One of limits indicates that one therapy may be delivering more health at increased or decreased cost
- The other limit indicates alternative therapy may be delivering more health at increased or decreased cost
- Q is not statistically significant at the α level represented by the interval
- The interval thus includes y axis

When Lower Limit is "Larger" than Upper Limit (2)

- Point estimate is either larger than both limits or smaller than both limits, but meets expectations for point estimate and limit when both are on same side of Y axis
 - If point estimate and lower limit are on same side of Y axis, point estimate is larger than lower limit
 - Upper limit, which is on opposite side of y axis, \leq lower limit
 - If point estimate and upper limit are on same side of Y axis, upper limit larger than point estimate
 - Lower limit, which is on opposite side of y axis, \geq upper limit



Common Mistakes, CI for CER (See TreeAge)

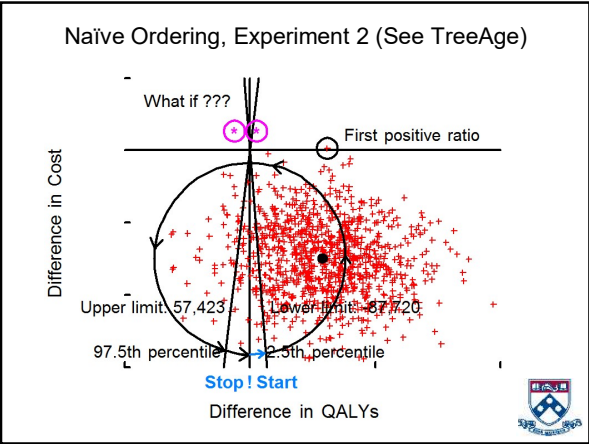
- CER equals ratio of mean differences in cost and effect
- Ratio of mean differences does not equal mean of ratios
 - Can't use result of Stata sum command for ratios to derive point estimate
- "SD" generated by summing ratios is not a good measure of SE of ratio (which can be undefined)
- When all replicates on one side of y axis (e.g., all on right or all on left), ordering ratios and identifying 2.5th and 97.5th percentiles of replicates yields a dependably accurate CI for CER
 - Equivalent to identifying lines through origin that exclude 2.5% of distribution

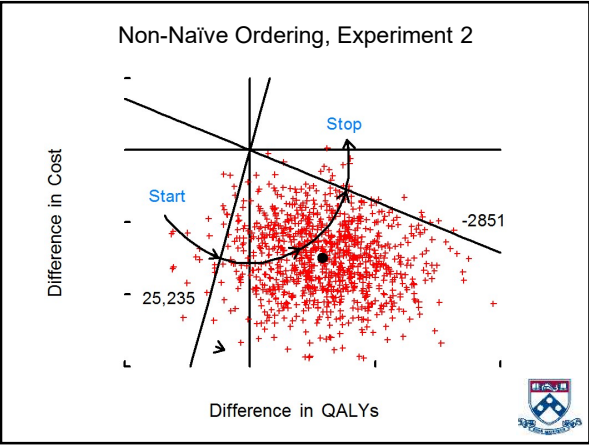


Common Mistakes (2)

- When replicates fall on both sides of y axis and are in at most 3 quadrants, cannot simply order ratios from lowest to highest
 - In most, but not all cases, must instead order lexicographically counter clockwise by quadrant and by magnitude of ratios within each quadrant
 - Can also yield dependably accurate confidence interval
- When replicates fall in all 4 quadrants, ordering can fail
 - Identifying lines through origin that exclude 2.5% of distribution guarantees dependably accurate CI, while CI based on ordering does not

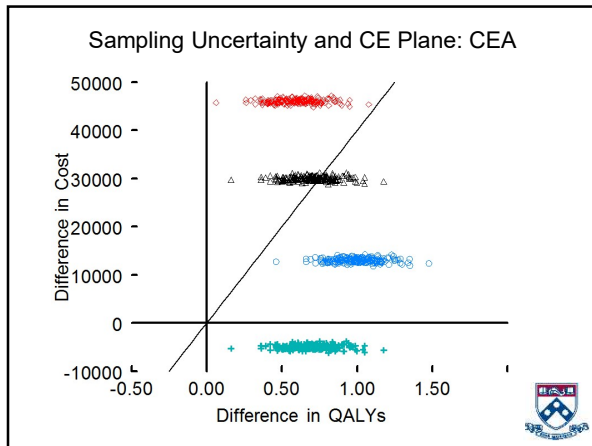


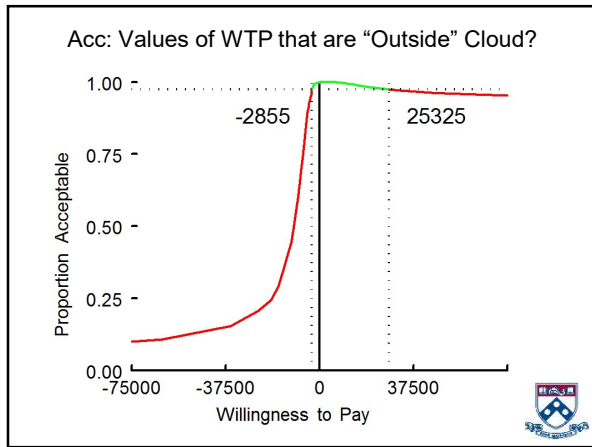


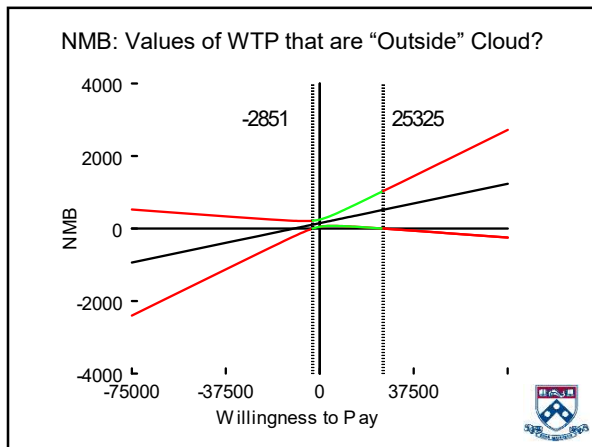


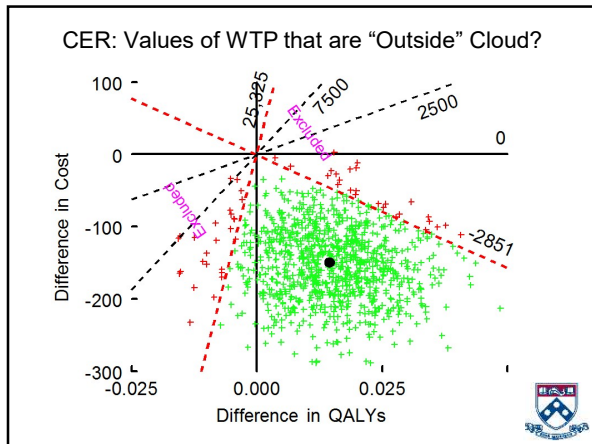
TreeAge/Stata Naïve Ordering, ICER Stats Report, and Correct ICER, SE, 95% CI

Statistic	Stata, Naive	TreeAge, Naive	Correct
ICER:	-13808	-15090	-10,327
"SD" (SE)	75466	2,373,023	--
Lower limit	-87,720	-90,162	25,325
Upper limit	57,423	63,616	-2851

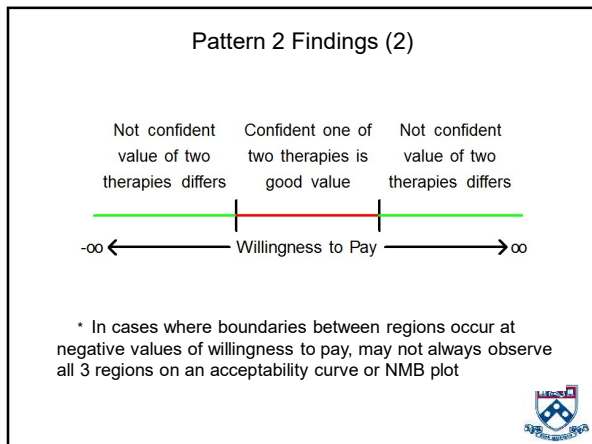


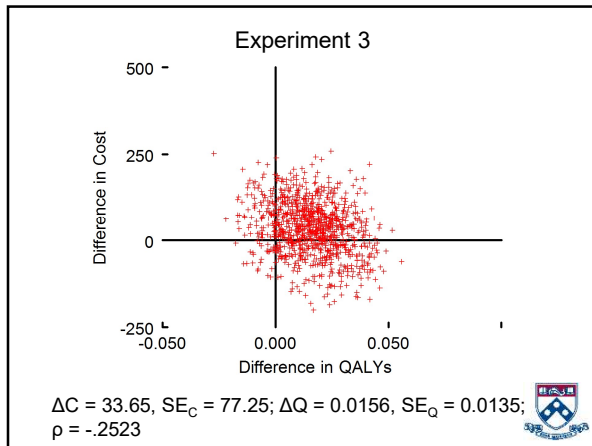


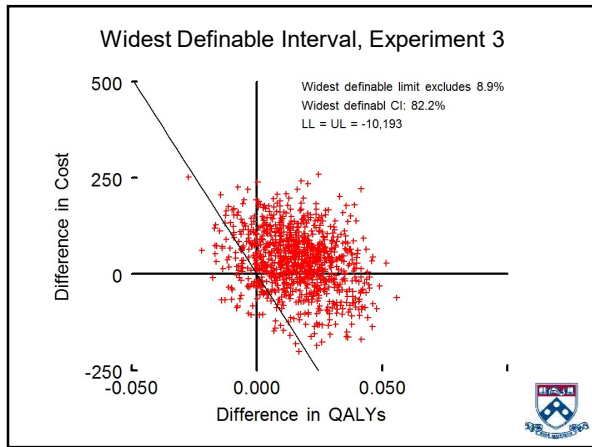


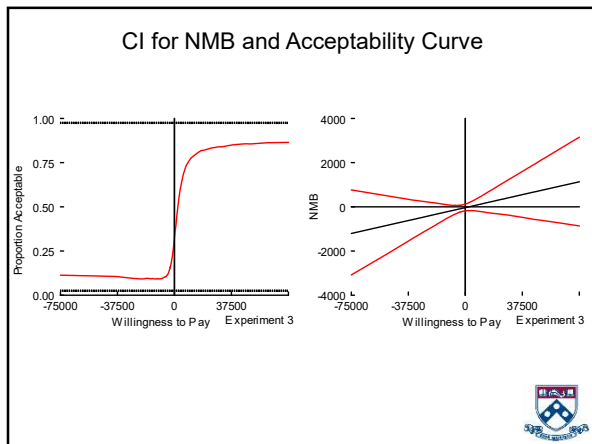


- ### Pattern 2 Findings
- Refer to findings like those in experiment 2 as pattern 2 findings
 - 1 of 2 patterns that occur only when difference in effect is not significant
 - Know pattern 2 is observed when:
 - Confidence interval for ICER includes Y axis (i.e., $LL > UL > PE$ OR $PE > LL > UL$)
 - One NMB confidence limit curve intersects x-axis (0) twice; other limit never intersects x-axis
 - Acceptability curve intersects a horizontal line drawn at either 0.025 and 0.975 on Y axis twice and never intersects other line (e.g., intersects 0.975 twice and never intersects 0.025)







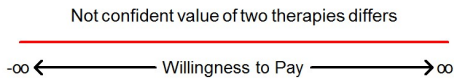


Pattern 3 Findings

- Refer to findings like those in experiment 3 as pattern 3 findings
- 1 of 2 patterns that occur only when difference in effect is not significant
- Know pattern 3 is observed when:
 - Confidence interval for ICER is undefined
 - Neither NMB confidence limit curve intersects x-axis (0)
 - Acceptability curve never intersects horizontal lines drawn at either 0.025 or 0.975 on Y axis

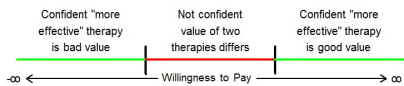


Pattern 3 Findings (2)

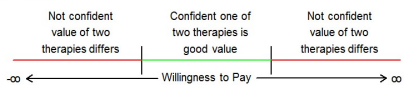


3 Patterns

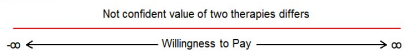
PATTERN 1



PATTERN 2



PATTERN 3



Conclusions (1)

- For any given willingness to pay, an experiment **ALWAYS** allows us to draw one of three conclusions:
 - Can be confident therapy is good value compared to alternative
 - Can be confident alternative is good value compared to therapy
 - Cannot be confident value of 2 therapies differs



Conclusions (2)

- If goal is to identify which of 3 conclusions holds for a given willingness to pay, confidence intervals for cost-effectiveness ratios, confidence intervals for NMB, and acceptability curves **ALWAYS** provide the same answer
 - e.g., if fraction acceptable at our WTP falls between 0.025 and 0.975:
 - CI for NMB calculated by use of our WTP includes 0, and
 - WTP is included within the CI for CER

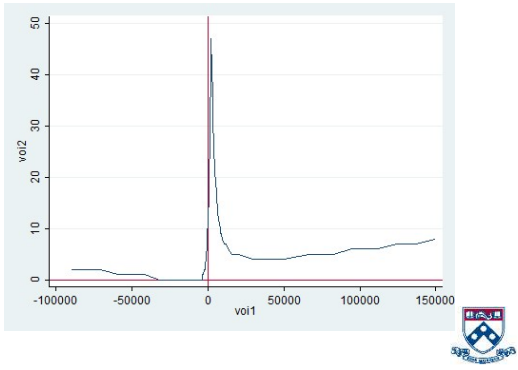


Conclusions (3)

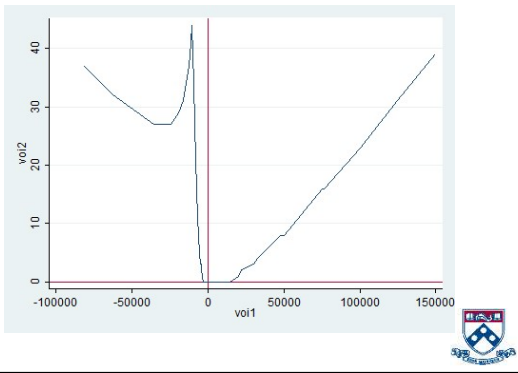
- CI for CER provide decision makers with concise information (i.e., 0, 1, or 2 numbers) that allows them to determine – based on own WTP -- if they can be confident about a therapy's value
- Acceptability curves allow decision makers to assess alternate levels of confidence if alternate levels are of interest



Per-Person Value of Information, RC5



Per-Person Value of Information, Experiment 2



Per-Person Value of Information, Experiment 3

