BOOTSTRAPPING CONFIDENCE INTERVALS FOR COST-EFFECTIVENESS RATIOS

Henry Glick
5.05

NONPARAMETRIC BOOTSTRAP METHOD

Step 1: Develop a set of bootstrap estimates of differences in costs and effectiveness

- Simultaneously resample costs and effects from the study sample and compute cost-effectiveness ratios in each of the multiple samples
    * Draw a sample with replacement from the original treatment arms in the study and using the resulting data to compute a bootstrap replication of the cost-effectiveness ratio
    * Repeat this sampling and calculation of the ratio (by convention, between 1000 and 2000 times for confidence intervals)

DEFINE A BOOTSTRAP CONFIDENCE INTERVAL

Step 2: Use the set of bootstrap estimates to define a confidence interval

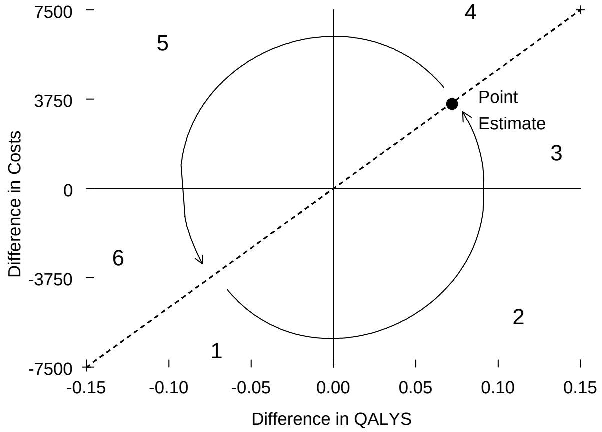
- Most common method in the literature: percentile method
    * In cases where the confidence interval for effects includes 0, potentially not dependably accurate
- Alternative dependably accurate method: "Acceptability" method
    * A nonparametric equivalent of Fieller's theorem intervals

BOOTSTRAP PERCENTILE METHOD

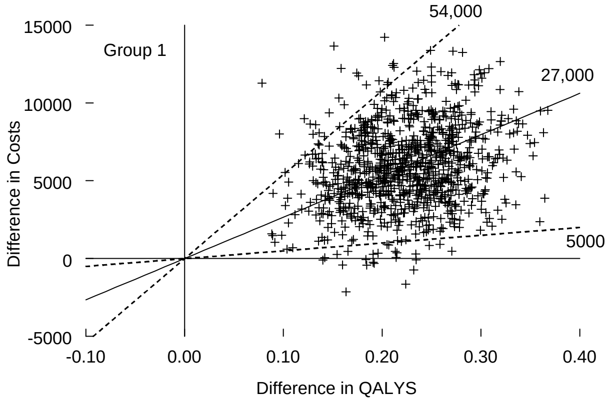
Step 2: Use the set of bootstrap estimates to define a confidence interval

- Order the repeated estimates of the ratio counter-clockwise lexicographically by quadrant and by ratio
    - Begins with the point-estimate for the cost-effectiveness ratio that falls in the quadrant opposite that of the point estimate for the cost-effectiveness ratio (e.g., if the point estimate is 40,000 and falls in the northeast quadrant of the cost-effectiveness plane, ordering begins at 40,000 in the southwest quadrant)
    - It continues counter clockwise through each of the quadrants of the cost-effectiveness plane until it again reaches the point estimate (e.g., 40,000) in the quadrant opposite that of the point estimate.
- Identify a 95% confidence interval from this rank-ordered distribution
    - When one has made 1000 repeated estimates, use the 26th and 975th ranked cost-effectiveness ratios to define the confidence interval
    - Do not use bootstrap methods that assume one can estimate the ratio’s standard error

PERCENTILE METHOD ORDERING

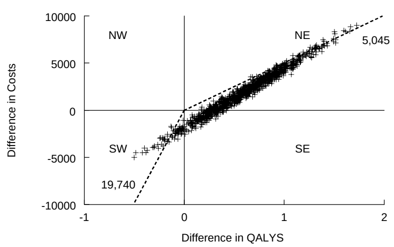


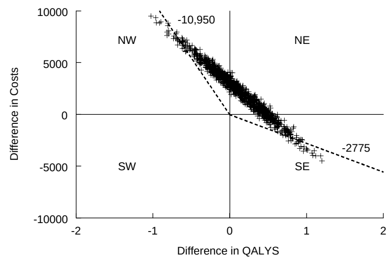
BOOTSTRAP CONFIDENCE INTERVALS, ONE OR TWO QUADRANTS

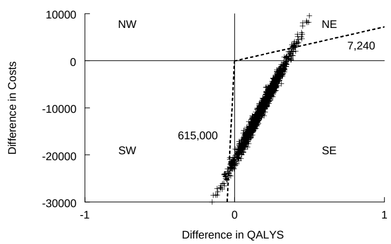


- 25 replicates fall below the 5000 limit; 25 fall above the 54,000 limit
- The percentile confidence interval will be dependably accurate when the difference in effect is statistically significant. In this case, the preponderance of replicates will fall either in a single quadrant of the cost-effectiveness plane (if the cost difference is also significant) or (if the cost difference is not significant) in one of two pairs of quadrants

REPLICATES FALLING IN 3 QUADRANTS



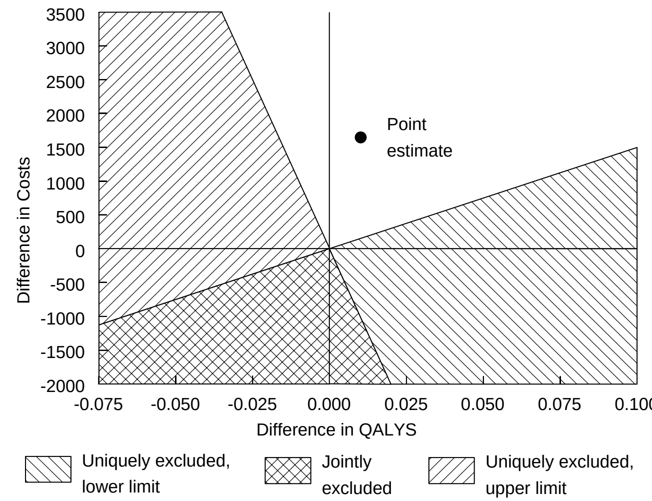




- When replicates fall in 3 quadrants (and are "far from the origin"), the rsulting interval will match the Fieller's theorem interval (and share in its dependable accuracy)

REPLICATES FALLING IN 4 QUADRANTS

- In the examples above, when each limit excludes 2.5% of the distribution, the 2 limits together -- when interpreted as lines through the origin -- exclude approximately 5% of the joint distribution
- When nontrivial densities of the joint distribution fall in all four quadrants, one must give up one of these relationships



- Do dependably accurate CI exclude α of the distribution or does each CL exclude α/2?

PERCENTILE METHOD AND LOSS OF DEPENDABLE ACCURACY

- The bootstrap percentile method loses its dependable accuracy when there are nontrivial densities of the joint distribution of the difference in costs and effects in all four quadrants

  - e.g.

Cost*	QALY*	CER	Coverage, 80% CI	Avg % excluded
2080	0.052	40,000	0.787	0.103
1580	0.0395	40,000	0.801	0.111
1236	0.0309	40,000	0.793	0.121
928	0.0232	40,000	0.779	0.136
544	0.0136	40,000	0.756	0.165
5	0.000125	40,000	0.736	0.179

* Standard error, difference in costs, 894.42719; Standard error, difference in QALYS, 0.02012461; Correlation, difference in costs and QALYS, 0.25

INTUITION BEHIND COVERAGE FAILURE

- When there are nontrivial densities of the joint distribution of the difference in costs and effects in all four quadrants:

  - The loss in dependable accuracy is due to the fact that each of the limits -- when interpreted as lines through the origin -- are excluding more than $\alpha/2$ of the distribution

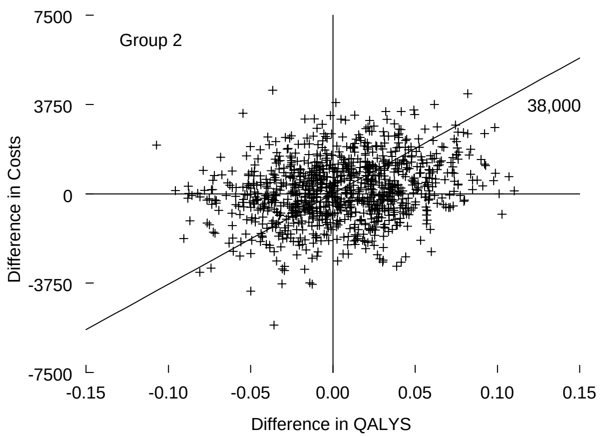
    * Because, by construction, the wedge used to define bootstrap percentile limits always excludes α of the replicates

  - Given that the same replicates need to be excluded by both the lower and upper limits, a single ordering is impossible

    * How can the same replicate be ordered both at the bottom and top of the distribution?

- To maintain dependable accuracy, one needs to identify lines through the origin that each exclude approximately $\alpha/2$ of the distribution

BOOTSTRAP CONFIDENCE INTERVALS



- Neither the percentile method, the acceptability method, nor Fieller's method can construct a dependably accurate 95% CI for this distribution
  - Because there is no line through the origin that excludes α/2 of the distribution
  - Thus, CI are undefined for these data

BOOTSTRAP ACCEPTABILITY METHOD

Step 2: Use the set of bootstrap estimates to define a confidence interval

- Identify lines through the origin that exclude P/2% of the bootstrap replicates
- Calculate the fraction of replicates that are excluded by each potential limit
- The limits are the replicates that exclude (approximately) α/2 of the replicates
- Discontinuities in the density of the bootstrap replicates can make precise identification of the bootstrap acceptability limits difficult
  - It is possible that no ratio excludes exactly α/2% of the replicates
  - Alternatively, it is possible that a number of ratios with “similar” magnitudes all exclude α/2% of the replicates
- The algorithm in appendix 1 indicates which result is the upper limit and which is the lower

DEPENDABLE ACCURACY, ACCEPTABILITY METHOD

- The bootstrap acceptability method maintains its dependable accuracy when there are nontrivial densities of the joint distribution of the difference in costs and effects in all four quadrants:

Cost	QALY	CER	Coverage, 80% CI
2080	0.052	40,000	0.790
1580	0.0395	40,000	0.809
1236	0.0309	40,000	0.813
928	0.0232	40,000	0.814
544	0.0136	40,000	0.802
5	0.000125	40,000	0.809

MIXED BOOTSTRAP ALGORITHM

- Because the percentile method is easier to evaluate, but potentially less dependably accurate, than is the acceptability method, one may want to pursue the following:
- First use the percentile method
- Assess the fraction of the replicates that are excluded by the resulting limits
  - If each excludes approximately α/2 of the replicates, use the results of the percentile method
- If the limits from the percentile method exclude more than α/2 of the replicates, implement the bootstrap acceptability algorithm

COMPARISON OF FIELLER, PERCENTILE BOOTSTRAP AND BOOTSTRAP ACCEPTABILITY CI (in '000) *

		Bootstrap	
CI	Fieller	Percentile	Acceptability
10%	32 to 46	32 to 46	32 to 47
20%	26 to 52	26 to 55	26 to 58
30%	21 to 69	21 to 66	20 to 71
40%	16 to 92	17 to 84	16 to 92
50%	10 to 146	12 to 115	11 to 138
60%	2 to 583	6 to 215	5 to 574
65%	-4 to -512	4 to 456	0 to -387
70%	-13 to -127	-2 to -1,539	-7 to -89
73%	-27 to -59	-5 to -336	-11 to -57
80%	undef	-18 to -89	undef

* Cost difference, 1150; SE cost difference, 1319.271; effect difference, .03; SE effect difference, .0337529; correlation of difference, .2436; dof, 498. Fieller and acceptability limits become undefined for CI > 73.5%

APPENDIX 1. Identifying the Upper and Lower Bootstrap Acceptability Limits

- Once one has identified a pair of limits (i.e., two lines through the origin that each exclude $\alpha/2$ of the distribution), one must identify which is the lower limit and which is the upper. The following algorithm can be used for this identification:

    - If one limit is less than the point estimate and one is greater than the point estimate, then the smaller limit is the lower limit and the larger limit is the upper limit

    - If both limits are greater than the point estimate or both limits are less than the point estimate, then the larger limit is the lower limit and the smaller limit is the upper limit