Sampling Uncertainty Appendices

ISPOR 18th Annual International Meeting

May 19, 2013

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

APPENDIX 1
Parametric Methods

---

iprogs.do

- Contains 6 **PROGRAMS** related to parametric estimates of sampling uncertainty, a (very basic) program for estimation of inputs for use in these programs, and a program that provides documentation for these programs
  - The command do iprogs simply loads these programs; it does not calculate anything
- Documentation program: iprogsdoc

iprogs.do (cont.)

- Programs for calculating sampling uncertainty
    - fielleri: Calculates Fieller's theorem CI for CER
    - nmbi: Calculates NMB, CI, and p-values for varying W
    - nmb1i: Calculates NMB, CI, and p-values for single W
    - accepti: Calculates % acceptable and p-values
    - accept1i: Calculates % acceptable and p-values for single W
    - ciboundi: Calculates t-statistics and p-values defining patterns 1, 2, and 3 for an experiment
- Program for calculating inputs
    - ipinputs: Calculates t-test means, SEs, and correlation for use with fielleri, etc. Usually want to use regression results, etc., instead

---

iprogs.do (cont.)

- Parameter values may be derived from any number of estimation methods, but these programs are particularly useful if we use multivariable regression to estimate point estimates and SEs.
- While these programs include a program for the derivation of the necessary parameters by use of t-tests (ipinputs), if t-tests are sufficient, consider using uprogs.do, which conducts the t-tests as well as assesses sampling uncertainty

---

ceagraphs.do

- Contains 3 **PROGRAMS** that graph the results of accepti (accgraph), nmbi (nmbgraph) and fielleri (fiellergraph)
    - The command do ceagraphs simply loads these programs; it does not calculate anything
- Documentation program: ceagraphsdoc
- Programs are meant to be run immediately after running calculation programs
    - e.g., accepti 1000 500 .1 .05 .1 498
        accgraph

Parametric Acceptability Curves

- Can derive parametric acceptability curves by estimating 1-tailed probability that NMB, calculated by use of values of WTP defined on X-axis, are greater than 0
- Stata commands for probability of acceptability for a willingness to pay of 20,000

  scalar w=20000

  scalar senmb = $((sec^2) + ((W^2)*(seq^2)) - (2*W*corrcq*sec*seq))^{0.5}$

  scalar acc20=1-ttail(DF,(((Q*W)-C)/senmb))

---

iprogsdoc: accepti

* PROGRAM: ACCEPTI

* CALCULATES WTP (W), % ACCEPTABLE, and P-VALUE

* COMMAND LINE: accepti [COST] [SEcost] [EFFECT] [SEeffect] [CORR] [DOF]

** The 6 arguments are all numbers
** `1' difference in costs
** `2' SE diff costs
** `3' difference in effects
** `4' SE diff effects
** `5' correlation of differences
** `6' degrees of freedom

---

iprogsdoc: accepti

Alternative command lines (e.g., for automated calculation):

```
ipinputs [cost] [effect] [group] [,if]
local a=r(meanc)
local b=r(sec)
local c=r(meanq)
local d=r(seq)
local e=r(rho)
local f=r(dof)
accepti `a' `b' `c' `d' `e' `f'
```

* Saved Results
* r(accmat)

Data Required for Acceptability Curve

- Needed data obtained from most statistical packages by:
  - Evaluating difference in costs (and obtaining C and s_C)
  - Evaluating difference in effects (and obtaining Q and s_E)
  - Estimating correlation between difference in costs and effects

---

accepti 88 103 .0408 .018 -.2440 494 .95

W	% Accept	P-value
.	.	.
-2828	0.02461	0.0492
-2806	0.02500	0.0500
-2370	0.03459	0.0692
.	.	.
1805	0.45049	0.9010
2531	0.54961	0.9008
2937	0.59910	0.8018
.	.	.
20000	0.96605	0.0679
28595	0.97500	0.0500
29241	0.97540	0.0492
.	.	.

Return List

```
.   return list

macros:
            r(cmd) : "accepti"

matrices:
            r(accmat) :  120 x 3
```

accgraph .95

accepti 88 103 .0408 .018 -.2440 494 .95
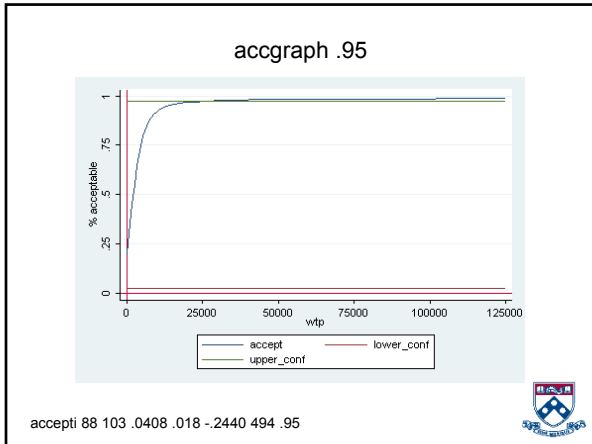
Saving Results of accepti

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

Parametric NMB Confidence Intervals

- Uses formula for a difference in two normally distributed continuous variables

$$\text{NMB CI} = \text{NMB} \pm t_{\alpha/2}\ \text{SE}_{\text{NMB}}$$

- Standard error for NMB equals:

$$\text{SE}^2_{\text{NMB}} = s_C^2 + \left(W^2\ s_Q^2\right) - 2\,W\,\rho\,s_C\,s_Q$$

$$\text{SE}_{\text{NMB}} = \left(s_C^2 + \left(W^2\ s_Q^2\right) - 2\,W\,\rho\,s_C\,s_Q\right)^{0.5}$$

- CI for NMB:

$$\text{CI}_{\text{NMB}} = ((Q\,W) - C) \pm t_{\alpha/2}(s_C^2 + W^2\ s_Q^2 - 2\,W\,\rho\,s_C\,s_Q)^{0.5}$$

iprogsdoc: nmbi

```
* PROGRAM:  NMBI

* CALCULATES NMB, CI, AND P-VALUE FOR VARYING
* WILLINGNESSES TO PAY

*   COMMAND LINE:  nmbi [COST] [SEcost] [EFFECT] [SEeffect]
    [CORR] [DOF] [CI]

* The 7 arguments are all numbers
** `1' difference in costs
** `2' SE diff costs
** `3' difference in effects
** `4' SE diff effects
** `5' correlation of differences
** `6' degrees of freedom
** `7' confidence interval, as decimal (e.g., 0.95 for a 95% interval)
```

iprogsdoc: nmbi (cont.)

Alternative command lines (e.g., for automated calculation):

```
    ipinputs [cost] [effect] [group] [,if]
    local a=r(meanc)
    local b=r(sec)
    local c=r(meanq)
    local d=r(seq)
    local e=r(rho)
    local f=r(dof)
    nmbi `a' `b' `c' `d' `e' `f' [CI]

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

nmbi 88 103 .0408 .018 -.2440 494 .95

W	NMB	95 % Lower limit	95 % Upper limit	P-value
.	.	.	.	.
-2828	-203	-406	-1	0.0492
-2860	-202	-405	0	0.0500
-2370	-185	-384	15	0.0692
.	.	.	.	.
1805	-14	-241	212	0.9010
2531	15	-225	256	0.9008
2937	32	-217	281	0.8018
.	.	.	.	.
20000	728	-54	1510	0.0679
28595	1079	0	2157	0.0500
29241	1105	4	2206	0.0492
.	.	.	.	.

nmbi Confidence Statements

```
Confidence Statements:
      For WTP <=28594.552, we can't be 95%
      confident that the 2 therapies differ in
      value;

      For WTP > 28594.552, we can be 95%
      confident that the therapy with the larger
      point estimate for effect represents good
      value compared with the alternative
```

Return List

. return list

scalars:
         r(CI) = 95

macros"
         r(cmd) : "nmbi"

matrices:
        r(nmbmat) : 120 x 5

nmbgraph
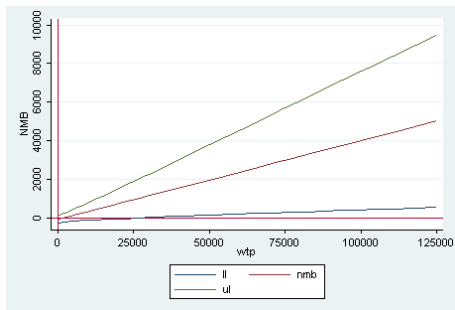
nmbi 88 103 .0408 .018 -.2440 494 .95

Saving the Results of nmbi

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

Fieller's Theorem Method

- Parametric method for calculating CI for CER based on assumption that difference in cost and effect follows a bivariate normal distribution
  - i.e., expression RQ - C is normally distributed with mean zero (where Q and C denote differences in mean effects and costs, respectively, and R = C/Q)
  - Standardizing statistic by its standard error and setting it equal to critical value from a normal distribution generates a quadratic equation in R
- Roots of quadratic equation give confidence limits

Fieller's Theorem Formula

Lower (clockwise) limit: $(M - [M^2 - NO]^{0.5}) / N$
Upper (counterclockwise) limit: $(M + [M^2 - NO]^{0.5}) / N$

Where:

$$M = CQ - t_{\alpha/2}^2 \, \rho \, s_c \, s_q$$

$$N = Q^2 - t_{\alpha/2}^2 \, s_q^2$$

$$O = C^2 - t_{\alpha/2}^2 \, s_c^2$$

C and Q denote difference in mean cost and effect; s_c and s_q denote estimated standard errors for difference in cost and effect; ρ equals estimated correlation between difference in cost and effect; and $t_{\alpha/2}$ is critical value from T distribution

iprogsdoc: fielleri

```
* PROGRAM:  FIELLERI

*CALCULATES FIELLER INTERVALS. Reports the specified interval
* as well as the last * defined interval; if the specified interval isn't
* defined, reports the limit of the last defined  interval

*   COMMAND LINE:  fielleri [COST] [SEcost] [EFFECT] [SEeffect]
    [CORR] [DOF] [CI]

* The 7 arguments are all numbers
** `1' difference in costs
** `2' SE diff costs
** `3' difference in effects
** `4' SE diff effects
** `5' correlation of differences
** `6' degrees of freedom
** `7' confidence interval, as decimal (e.g., 0.95 for a 95% interval)
```

---

iprogsdoc: fielleri (cont.)

```
Alternative command lines (e.g., for automated calculation):
   ipinputs [cost] [effect] [group] [,if]
   local a=r(meanc)
   local b=r(sec)
   local c=r(meanq)
   local d=r(seq)
   local e=r(rho)
   local f=r(dof)
   fielleri `a' `b' `c' `d' `e' `f' [CI]

* Saved Results
* r(R)
* r(CI)
* r(fll)
* r(ful)
* r(widestCI)
```

---

fielleri 88 103 .0408 .018 -.2440 494 .95

```
Cost-Effectiveness Analysis
      Point Estimate:                2157
      Quadrant:                   Upper right
      Fieller 95 % Confidence Interval
           Lower limit :             -2806
           Upper limit:              28595
 Confidence Statements:
      For WTP <=28595, we can't be 95% confident
      that the 2 therapies differ in value;

      For WTP > 28595, we can be 95% confident
      that the therapy with the larger point
      estimate for effect represents good value
      compared with the alternative
```

fielleri Saved Results

```
. return list

scalars:
                 r(R) =  2157
                r(CI) =  95
               r(fll) = -2806
               r(ful) =  28594.55
         r(widestCI) =  99.2645
         r(widestCL) = -10063


macros:
             r(cmd) :  "fielleri"
```
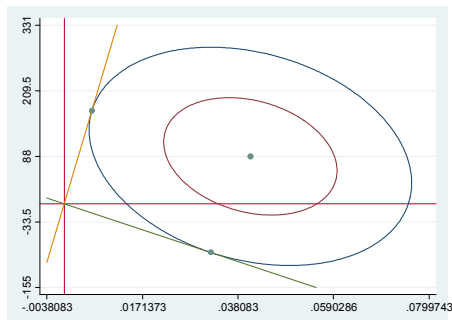
fiellergraph,ellipse

fielleri 88 103 .0408 .018 -.2440 494 .95

APPENDIX 2
Bootstrap ordering Hard Cases

Example #1 (isporex1.dta)

- Suppose we conducted an experiment and found that:
  - The difference in cost was 138 (SE, 102; p = 0.0005)
  - The difference in QALYs was 0.0625 (SE = 0.0179, p=0.18)
  - The correlation between the differences was -0.20
  - The degrees of freedom were 498
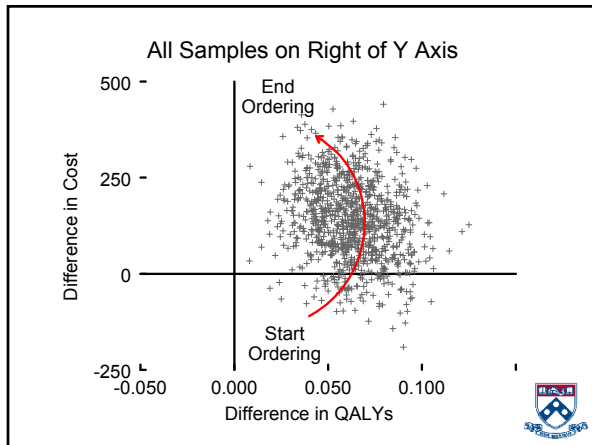- Point estimate CER: 2205
- Fieller's 95% CI: -891 to 7883

All Samples on Right of Y Axis

Resulting CI for CER
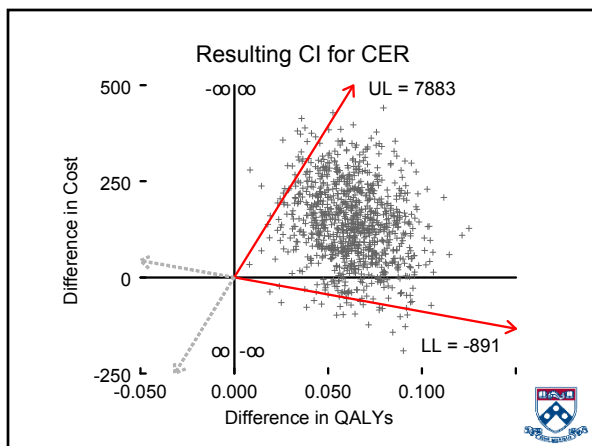
Correctly Ordering the Replicates

- gen cer=cost/qaly
- sort cer
- list cer if _n==26|_n==975

```
     +-------------+
     |        cer  |
     |-------------|
 26. | -891.0117 |
975. | 7882.697  |
     +-------------+
```

Bootstrap interval: -891 to 7883

Example 2. isporex2.dta: Ordering Monte Carlo Samples / Bootstrap Replicates

- IF samples / replicates fall on both sides of the y axis naïve ordering breaks down
- If samples fall in at most 3 quadrants, non-naïve ordering still gives dependably accurate CI
- Consider the following results:
  - The difference in cost was 2000 (SE, 200; p < 0.0000)
  - The difference in QALYs was 0.008 (SE = 0.008, p=0.32)
  - The correlation between the differences was 0.1
  - The degrees of freedom were 498
  - Point estimate ICER: 250,000
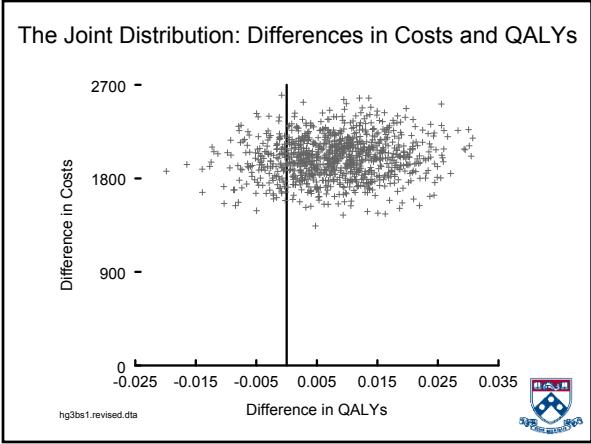 - Fieller interval -∞ to -251,572 and 83,507 to ∞

The Joint Distribution: Differences in Costs and QALYs

hg3bs1.revised.dta

Difference in Costs (y-axis): 0, 900, 1800, 2700
Difference in QALYs (x-axis): -0.025, -0.015, -0.005, 0.005, 0.015, 0.025, 0.035

Naïve Ordering, isporex2.dta

```
.  use isporex2
.  sort cer
.  list cer if _n==26|_n==1000

      +----------+
      |      cer |
      |----------|
 26.  | -2085492 |
975.  |  2456259 |
      +----------+
```

Defining the Lower Limit, Naïve Ordering

hg3bs1.dta

Defining the Upper Limit, Naïve Ordering

Naive ordering interval:
-2,085,492 to 2,493,136

hg3bs1.revised.dta

Slide 1: The Correct Limits

The Correct Limits

-281,202   81,684

② ①

End Ordering

Start Ordering

Non-naive interval:
81684 to infinity and
-infinity to -281202

Difference in Costs (y-axis): 0, 900, 1800, 2700
Difference in QALYs (x-axis): -0.025, -0.015, -0.005, 0.005, 0.015, 0.025, 0.035

hg3bs1.revised.dta

Slide 2: Correctly Ordering the Replicates

- gen quad=cd>=0&qd>=0
- replace quad=2 if cd>=0&qd<0
- gen cer=cd/qd
- sort quad cer
- list cer if _n==26|_n==975

```
      +--------------+
      |         cer  |
      |--------------|
 26.  |   81684.34   |
975.  |  -281202.4   |
      +--------------+
```

Bootstrap interval: -∞ to -281202 and 81684 to ∞

Slide 3: Example #3 (isporex3.dta) Replicates in 3 Quadrants

- Suppose we conducted an experiment and found that:
  - The difference in cost was 8500 (SE, 5000; p < 0.93)
  - The difference in QALYs was 0.5 (SE = 0.3, p=0.93)
  - The correlation between the differences -0.95
  - The degrees of freedom were 498
- Point estimate CER: 17,000
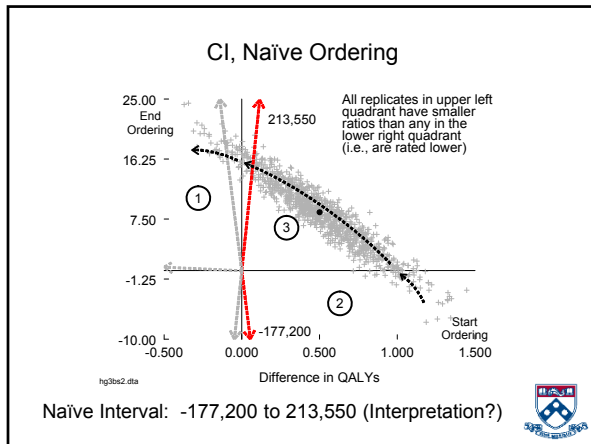- Fieller's 95% CI: -∞ to -198,933 and -1252 to ∞

CI, Naïve Ordering

All replicates in upper left quadrant have smaller ratios than any in the lower right quadrant (i.e., are rated lower)

213,550

-177,200

25.00 — End Ordering
16.25
7.50
-1.25
-10.00

-0.500 0.000 0.500 1.000 1.500

'Start Ordering

Difference in QALYs

hg3bs2.dta

Naïve Interval: -177,200 to 213,550 (Interpretation?)

Correct Ordering

-187,870

1025

25.00 — End Ordering
16.25
7.50
-1.25
-10.00

-0.500 0.000 0.500 1.000 1.500

'Start Ordering

Difference in QALYs

hg3bs1.dta

Fiellers Interval: -∞ to -198,933 and -1252 to ∞

Correctly Ordering the Replicates

- gen quad=cd<0&qd>=0
- replace quad=2 if cd>=0&qd>=0
- replace quad=3 if cd>=0&qd<0
- gen cer=cd/qd
- sort quad cer
- list cer if _n==26|_n==975

```
      +-------------+
      |       cer   |
      |-------------|
  26. | -1024.611   |
 975. | -187870.1   |
      +-------------+
```

Bootstrap interval: -∞ to -187870 and -1025 to ∞

Ordering Monte Carlo Samples / Bootstrap Replicates

- IF samples / replicates fall in all 4 quadrants, non-naïve ordering can break down (the notion of ordering may no longer have meaning)
  - Resulting intervals no longer need be dependably accurate
- The correct limits can both omit the same samples/replicates
- If defined, lines through the origin that each exclude 2.5% of the distribution will still define dependably accurate confidence interval

---

Replicates Falling in All 4 Quadrants

- There is no consensus in the literature about ordering replicates when an experiment's joint distribution of the difference in cost and effect has nonnegligible amounts of its density in all four quadrants of the cost-effectiveness plane
- By use of simulation we have found that the resulting limits are most like the Fieller's methods' results if the replicates are ranked counter clockwise lexicographically by quadrant and by ratio
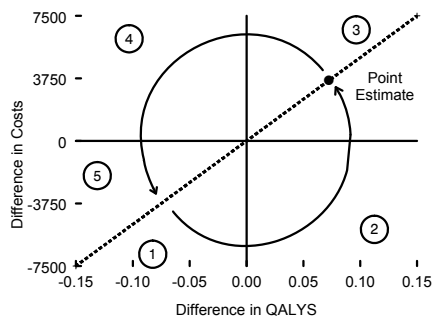

Ordering When Replicates Fall in All 4 Quadrants

Example #4 (isporex4.dta)

- Suppose we conducted an experiment and found that:
  - The difference in cost was 370 (SE, 200; p < 0.93)
  - The difference in QALYs was 0.09 (SE = 0.1, p=0.93)
  - The correlation between the differences 0.1
  - The degrees of freedam were 498
- Point estimate CER: 4111
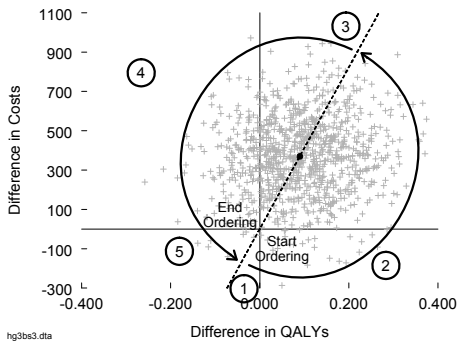- Fieller's 95% CI: -∞ to -1198 and -479 to ∞

Ordering for isporsex4.dta

hg3bs3.dta

Hardwired CI Code for this Experiment

```
gen cer=cd/qd
local pecer=370/.09
gen rquad=.
replace rquad=1 if cd<0&qd<0&cer>=`pecer'
replace rquad=2 if cd<0&qd>0
replace rquad=3 if cd>0&qd>0&cer>0
replace rquad=4 if cd>0&qd<0
replace rquad=5 if cd<0&qd<0&cer<`pecer'
sort rquad cer
list cer if _n==26|_n==975
```

Generalized CI Algorithm for Any Experiment

```
gen cer=cd/qd
local pecd=370
local peqd=.09
local pecer=370/.09
gen rquad=.
replace rquad=1 if cer/`pecer'>0&cd/`pecd'<0&qd/`peqd'<0&cer>=`pecer'
replace rquad=2 if cd/`pecd'<0&qd/`peqd'>0&`pecer'>=0
replace rquad=2 if cd/`pecd'>0&qd/`peqd'<0&`pecer'<0
replace rquad=3 if cer/`pecer'>0&cd/`pecd'>0&qd/`peqd'>0
replace rquad=4 if cd/`pecd'>0&qd/`peqd'<0&`pecer'>=0
replace rquad=4 if cd/`pecd'<0&qd/`peqd'>0&`pecer'<0
replace rquad=5 if cer/`pecer'>0&cd/`pecd'<0&qd/`peqd'<0&cer<`pecer'
sort rquad cer
list if _n==26|_n==975
```


Percentile Interval for isporex4.dta

Percentile Interval for isporsex4.dta

Percentile Interval for isporsex4.dta

Fiellers Interval: -∞ to -1198 and -479 to ∞
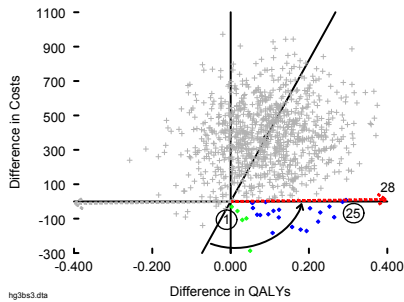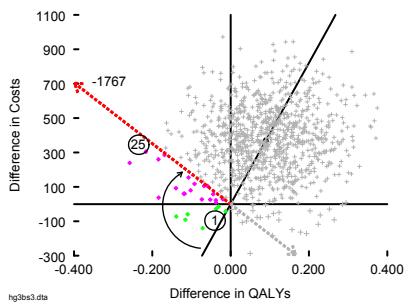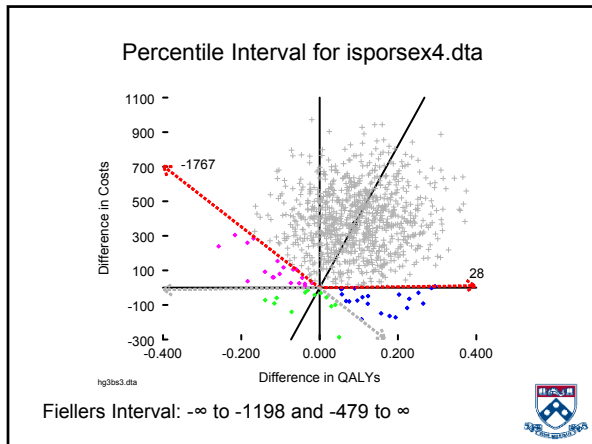
Example #5 (isporex5.dta)

- Suppose we conducted an experiment and found that:
  - The difference in cost was 300 (SE, 200; p < 0.13)
  - The difference in QALYs was 0.09 (SE = 0.1, p=0.37)
  - The correlation between the differences 0.1
  - The degrees of freedom were 498
- Point estimate CER: 3333
- Fieller's 95% CI: Undefined (WDI, 90.6%, -1064)
  - (i.e., no line through the origin that exludes 2.5% of the distribution

---
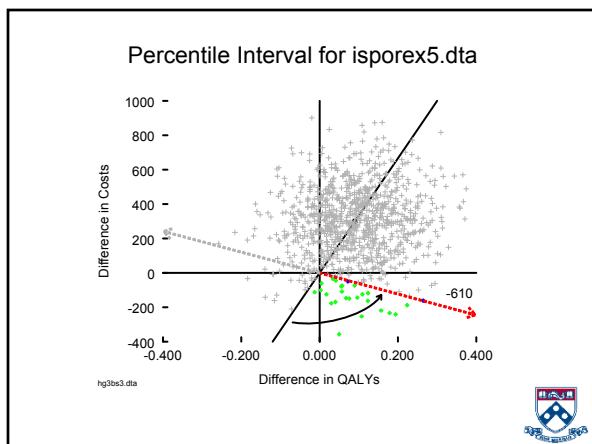
Percentile Interval for isporex5.dta

Percentile Interval for isporex5.dta

hg3bs3.dta

Percentile Limits Each Include ~4.9% of the Distribution

hg3bs3.dta

Fiellers Interval: Undefined; (90.2% ll = -659; ul = -1560)
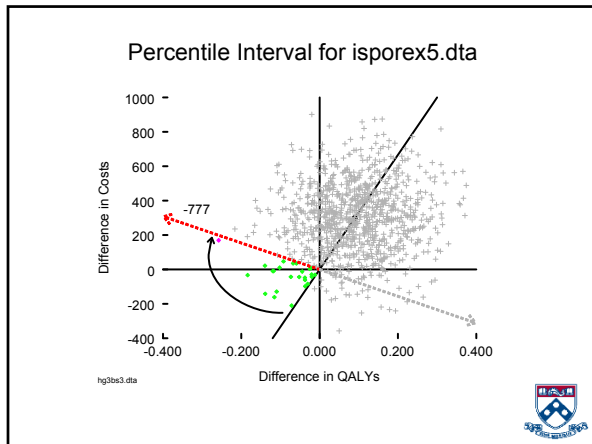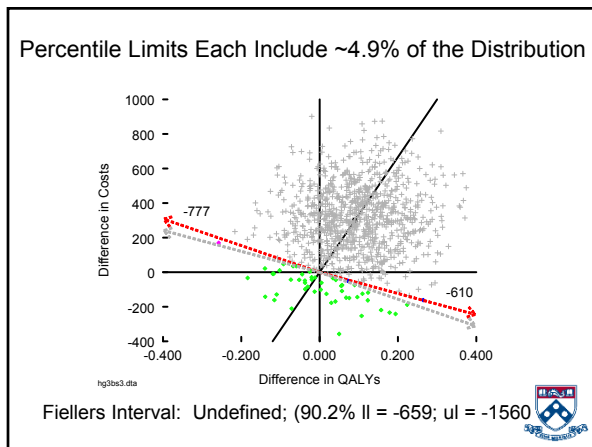
Test of Percentile Limits

- There is a simple test of whether the percentile method limits are likely to be dependably accurate
- Step 1: Use each of the confidence limits as W and calculate NMB
- Step 2: Determine if both limits exclude ~2.5% of the distribution
  - e.g., sum nmb if nmb>0
- Step 3: If the intervals each exclude substantially more than 2.5%, the percentile method is not likely to be dependably accurate