Sampling Uncertainty and Patient-Level Cost-Effectiveness Analysis (Part 1b)

Henry A. Glick, Ph.D.

EPI 550

April 22, 2020

Confidence Interval for Incremental Cost-Effectiveness Ratio

"Counting" Method 2: CI for ICER

- Can also determine fraction of results that fall on one side of W by identifying slopes of 0, 1, or 2 lines through the origin that exclude α/2 of distribution
  - Identify them by either counting/estimating distribution of results falling on each side of lines through origin
    - Slopes of lines that have 2.5% of distribution on 1 side and 97.5% on other define 95% CL for ICER
    - Slopes of these lines define values of W for which acceptability curve has heights of 2.5% and/or 97.5%
- Referred to as confidence interval for cost-effectiveness ratio

Construction of CI for Difference, OR, or RR

- Common algorithm
  - Develop distribution of difference (e.g., NMB), OR, or RR
    - e.g, create empiric distribution from bootstrap or assume a distribution such as normal or log normal
  - Order distribution from smallest to largest
  - Construct 95% CI by identifying 2.5th and 97.5th percentiles of rank-ordered distribution
    - Either by counting (nonparametric) or estimating density (parametric)
  - Values of outcome that bound these percentiles represent the 95% confidence limits
- Works well for differences, OR, or RR

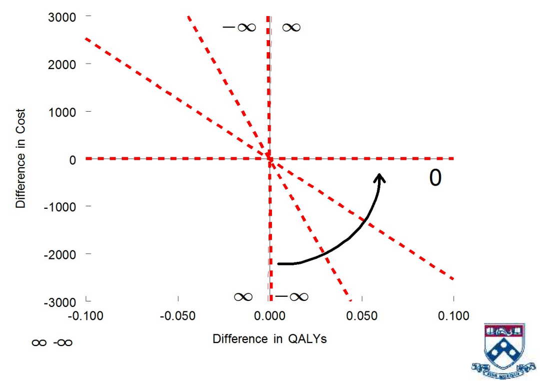
Construction of CI for ICER

- To use same algorithm for construction of CI for ICER:
  - Develop joint distribution of difference in C and Q and calculate ratios
  - Order ratios from smallest to largest (referred to as "naïve ordering")
  - For 95% CI, identify 2.5th and 97.5th percentiles of rank-ordered ratios
  - Values of ratios that bound 2.5th and 97.5th percentiles represent 95% confidence limits

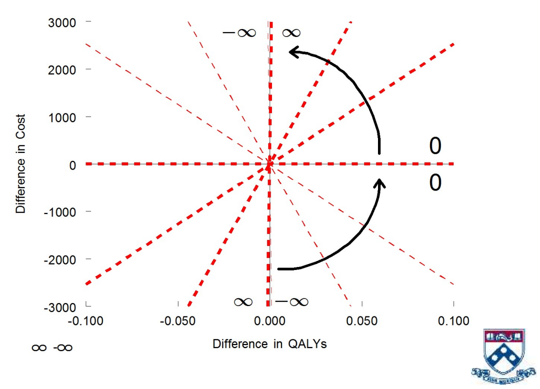
All of Density on One Side of Y-Axis

- Algorithm for constructing CI for ICER WORKS when all density/replicates are on ONE SIDE of Y-axis
- On CE plane, interval stretches COUNTER-CLOCKWISE from lower (CLOCKWISE) limit to upper (COUNTER-CLOCKWISE) limit

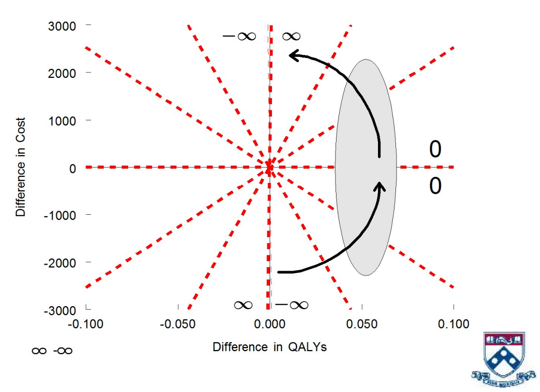
Order from -∞ to 0



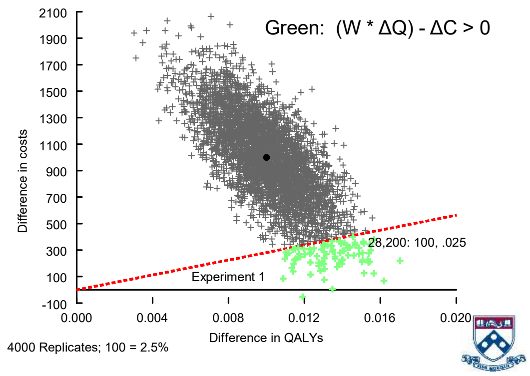
Continue from 0 to ∞



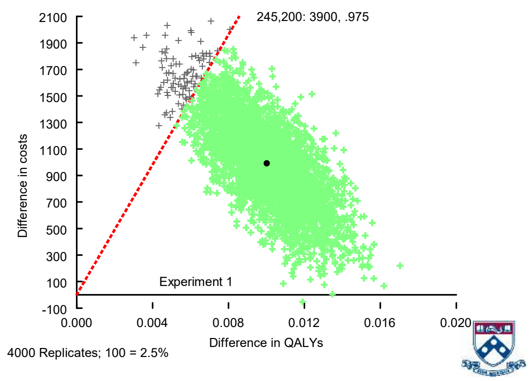
Makes Sense if Entire Density on One Side of X-Axis



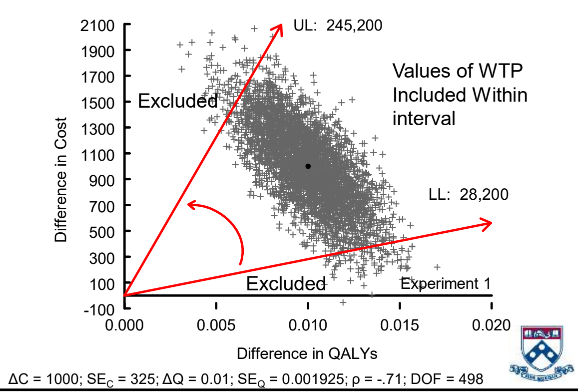
Lower 95% Confidence Limit (Slide 44, Part 1A)



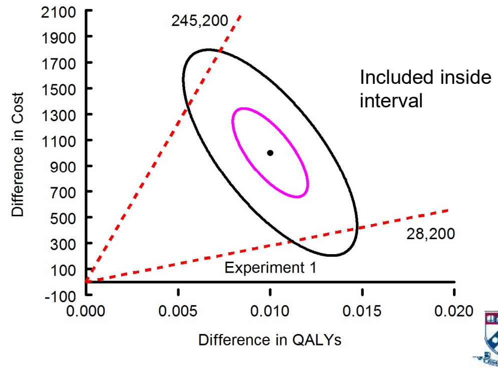
Upper 95% Confidence Limit (Slide 46, Part 1A)



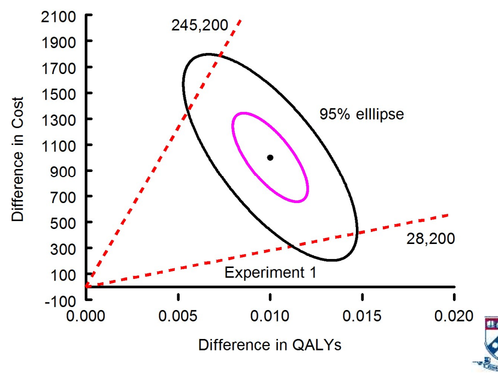
95% CI for ICER for Experiment 1



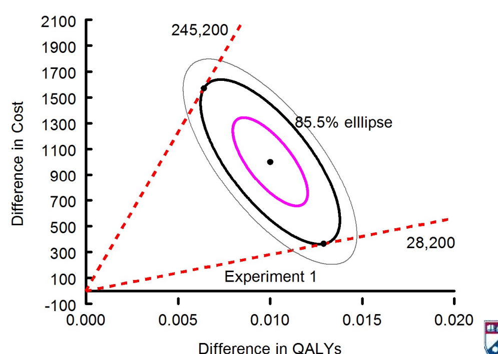
Parametric 95% CI for ICER for Experiment 1



95% CI for ICER Not Tangent to 95% Ellipse



95% CL Tangencies With 85.5% Confidence Ellipse



lower limit: 364.96, .0129; upper limit: 1570.38, .0064

Confidences Statements for CI for ICER

- Not confident of value if:
  - LL < W < UL
- Confident of value if:
  - LL < UL < W (confident of good value)
  - W < LL < UL (confident of bad value)

Confidences Statements for Current Experiment

- Can be confident of value when W not included in confidence interval
- ***When lower limit is a smaller number than upper limit***
  - Interval ranges between lower and upper limit
    28,200 to 245,200
  - Confident of value if WTP is either smaller than lower limit or greater than upper limit
    - Confident of bad value if WTP < 28,200
      - Because at least 97.5% of samples have ratios greater than 28,200
    - Confident of good value if WTP > 245,200
      - Because at least 97.5% of samples have ratios less than 245,200

"Common" Conclusions, CI for ICER

W	What is often said
<28,200	"95% confident Rx A not good value" (Rx B good value)
76,800	Can't be 95% confident value of Rxs differs
100,000	Can't be 95% confident value of Rxs differs
127,700	Can't be 95% confident value of Rxs differs
>245,200	"95% confident Rx A good value (Rx B not good value)

- Usually employ 2-tailed interpretation of CI for ICER

Confidence Interval for NMB

"Counting" Method 3: CI for NMB

- Can determine if W falls inside or outside distribution by constructing distribution of NMB for specified W and identifying whether 0 falls within interval
- As for any difference, construct interval by ordering distribution of NMB and identifying values of NMB that define the 2.5th and 97.5th percentiles
- In contrast with acceptability curve and CI for ICER, not (typically) defining lines through the origin of CE plane
  - But lines through origin have same meaning as for acceptability curves and CI for ICER

NMB Recap

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

- For a WTP of 50,000, NMB for experiment 1:

  (50,000 * .01) -1000 = -500
- Study result 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 is always defined
- Given not all decision making bodies have agreed upon maximum willingness to pay, routinely estimate net benefit over range of policy relevant values of willingness to pay

Net Benefit Graphically

- For a given W, can calculate value of NMB for every point on CE plane
- Formula: For replicate "i" NMB = W $\Delta Q_i - \Delta C_i$
- If W= 50,000, following points all fall on same NMB line (slope 50,000, intercept -500) and have same NMB value

ΔC	ΔQ	NMB
-500	0	(50,000 * 0) – (-500) = 500
49,500	1	(50,000 * 1) – (49,500) = 500
99,500	2	(50,000 * 2) – (99,500) = 500
149,500	3	(50,000 * 3) – (149,500) = 500

- Value of NMB for lines with 50,000 slope = -intercept
  - e.g., -(-500) = 500

Net Benefit Graphically (2)

- As with diagnostic test optimal operating slopes, NMB graphically defined on cost effectiveness plane using a family of lines
- Each line has a slope equal to W
- Each line represents a single value of NMB which equals −ΔC (i.e., − intercept, because when ΔQ=0, WΔQ drops out of equation
- 95% CI for NMB defined by identifying 2 NMB lines that each omit 2.5% of distribution

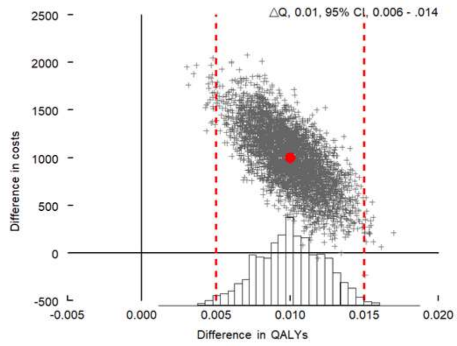
Net Benefit Graphically (3)

- As we just saw, following 4 points all fall on same NMB line when W=50,000
- If W= 100,000, the same 4 points all fall on different NMB lines (slope 100,000, varying intercepts) and have different values of NMB

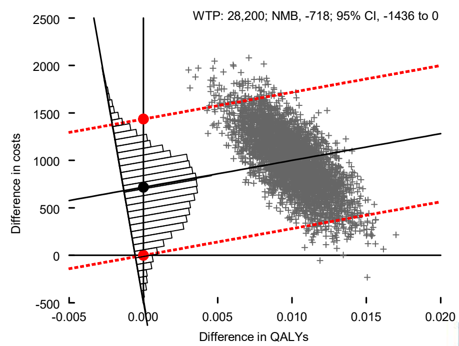
ΔC	ΔQ	NMB
-500	0	(100,000 * 0) – (-500) = 500
49,500	1	(100,000 * 1) – (49,500) = 50,500
99,500	2	(100,000 * 2) – (99,500) = 100,500
149,500	3	(100,000 * 3) – (149,500) = 150,500

- Value of NMB for lines with 100,000 slope = -intercept

Reminder: Constructing 95% CI for ΔQ

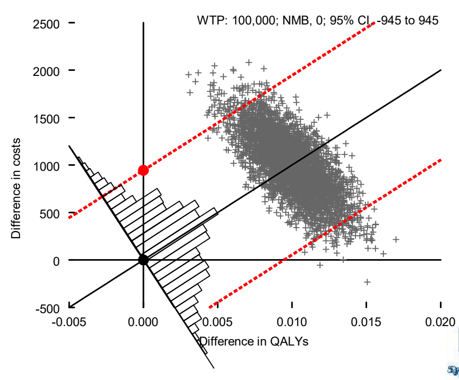


Constructing CI for NMB, WTP=28.2K

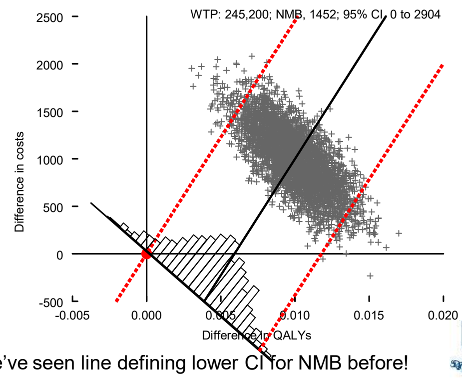


- We've seen line defining upper CI for NMB before!

Constructing CI for NMB, WTP=100K

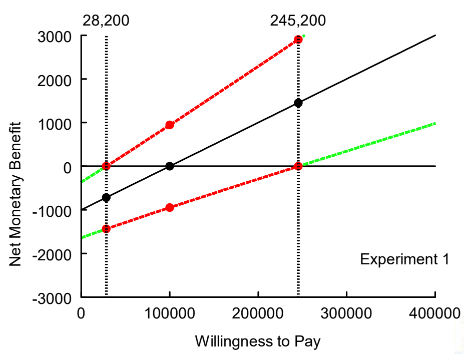


Constructing CI for NMB, WTP=245.2K

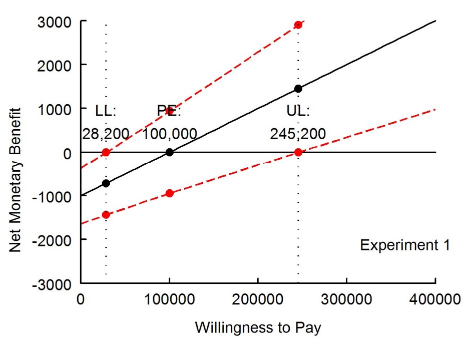


- We've seen line defining lower CI for NMB before!

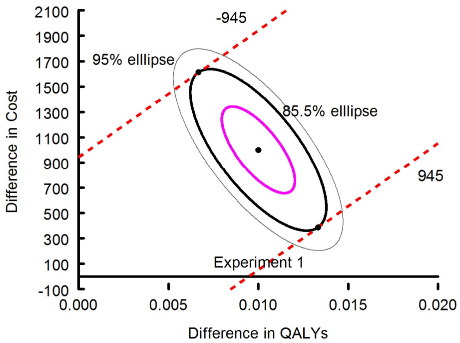
NMB Graph with CI



Additional Information About **CI for ICER** in NMB Graph



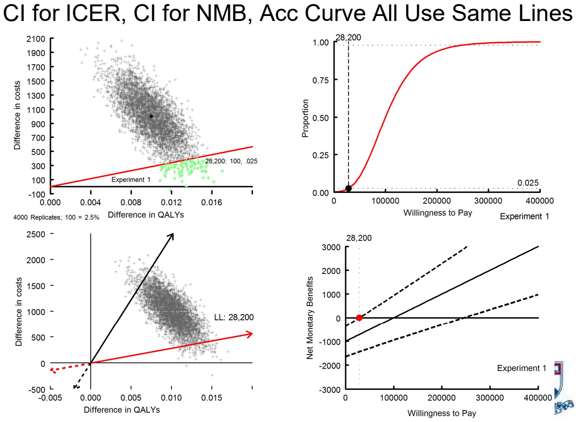
NMB 95% CL Also Tangent to 85.5% Ellipse



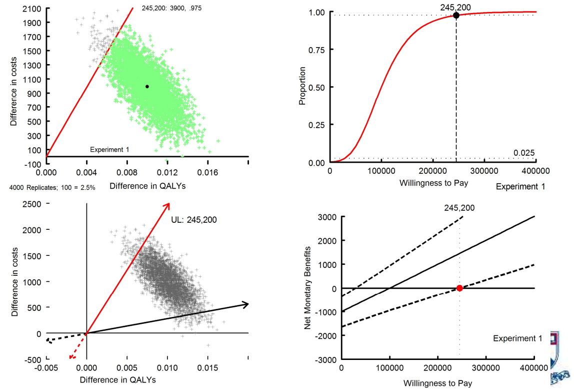
Confidences Statements for CI for NMB

- If both confidence limits negative, 95% confident therapy is bad value
  - i.e., for values of WTP $\leq$ 28,200
- If both confidence limits positive, 95% confident therapy is good value
  - i.e., for values of WTP $\geq$ 245,200
- If one confidence limit positive and one negative, cannot be 95% confident value of 2 therapies differs
  - i.e., for values of WTP > 28,200 and < 245,200

CI for ICER, CI for NMB, Acc Curve All Use Same Lines



CI for ICER, CI for NMB, Acc Curve All Use Same Lines



Similarities and Differences

- For magnitude estimation for a single value of W, NMB provides information that is NOT shared by acceptability curve or CI for ICER
  - i.e., generally isn't identifying lines through origin as are acceptability curve and CI for ICER
- For meta-question about ranges of W for which we can or can't be confident of value, NMB graph provides information that IS shared
  - Nonparametrically, identification of whether CI for NMB includes or excludes 0 relies on same lines through origin as acceptability curve and CI for ICER
  - Parametrically, CI for NMB and acceptability curve use transformation of Fieller's theorem equation for CI for ICER

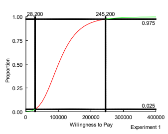
Acceptability & CI for ICER

- Acceptability curve plots confidence intervals for the cost-effectiveness ratio
  - e.g., the value of WTP where the height of the acceptability curve equals 0.025 and/or 0.975 represent the 95% confidence limits for the cost-effectiveness ratio
    - In current example, 95% CL = 28,200 and 245,200

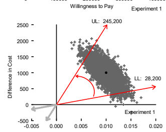
Acceptability & CI for NMB

- Acceptability curves also report values of WTP for which one of NMB confidence limits equals 0
  - e.g. if we calculate NMB using values of WTP where height of acceptability curve equals 0.025 and/or 0.975, one of 95% confidence limits for NMB will equal 0
  - If we calculate NMB using values of WTP where height of the acceptability curve equals 0.25 and/or 0.75, one of 50% confidence limits for NMB will equal 0

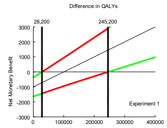
Review of Results for Experiment 1



Acceptability curve
Acceptability curve intersects 0.025 and 0.975 at 28,200 and 245,200



Confidence interval for ICER
CER CI: (28,200 to 245,200)

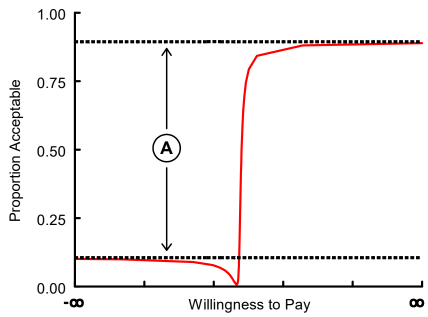


Confidence frontier for NMB
CI intersect decision threshold (0) at 28,200 and 245,200

“Pattern 1” Findings

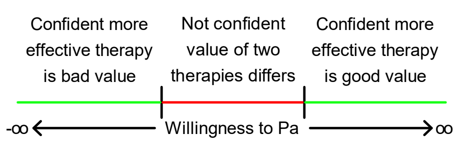
- Refer to findings like experiment 1’s as pattern 1 findings
- Occur when difference in effect is significant
- Know we are observing pattern 1 finding when:
  - Confidence interval for cost-effectiveness ratio excludes Y axis (i.e., LL < PE < UL)
  - Both NMB confidence limits curves intersect decision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

Region of Acceptability Related to Pattern 1



- For this curve, widest pattern 1 finding is 78.81% CI

3 Ranges of WTP for Pattern 1 Findings



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

Confidence vs Value of Information

- Requiring statistical significance (i.e., confidence) prior to the adoption of a new therapy that maximizes NMB runs counter to expected utility theory
  - Said to impose opportunity costs on patients

Quality of the Evidence

- Rejection of significance tests for cost-effectiveness ratios/NMB does not imply that decisions should be made using point estimates alone
  - Particularly if a decision can be made to collect more information
- "Value of information" represents difference in expected value of outcome given current decision and expected value of outcome that would result if we had perfect information (EVPI)
  - Determined based on probability decision is wrong and costs of wrong decision if it occurs

Per-Person Expected Value of Perfect Information

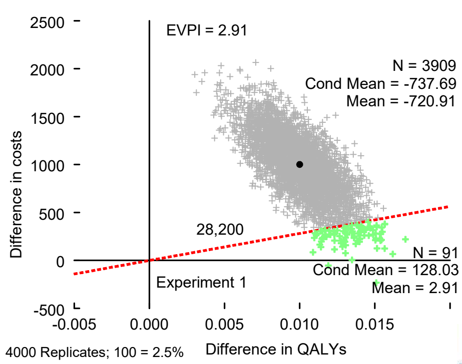
$$EVPI_{PP} = \min\left(V_j\ ,\ V_k\right)$$

where

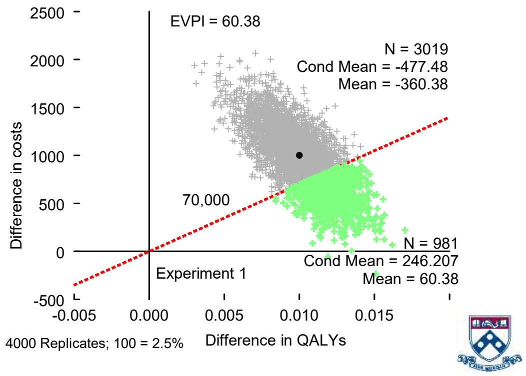
$$V_j = \forall_j\ NMB_j > 0:\ \frac{N_j}{N}\sum\frac{NMB_j}{N_j}$$

$$V_k = \forall_k\ NMB_k < 0:\ \text{abs}\left(\frac{N_k}{N}\sum\frac{NMB_k}{N_k}\right)$$

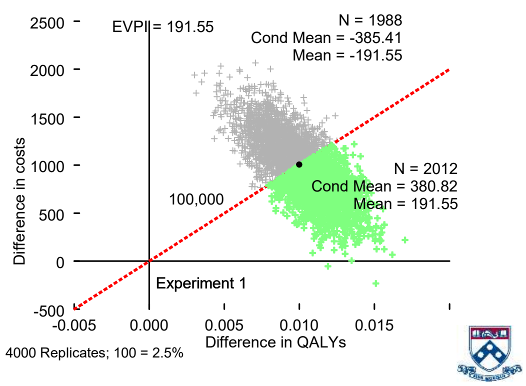
Calculating Per-Person EVPI, 28,200



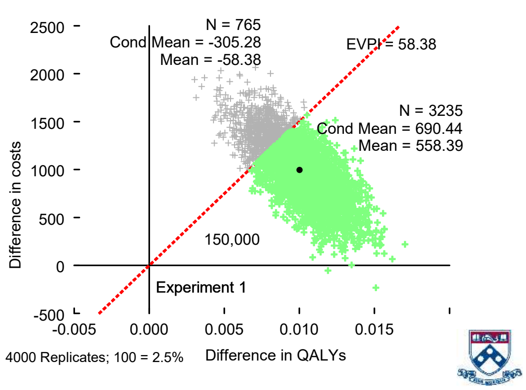
Calculating Per-Person EVPI, 70,000



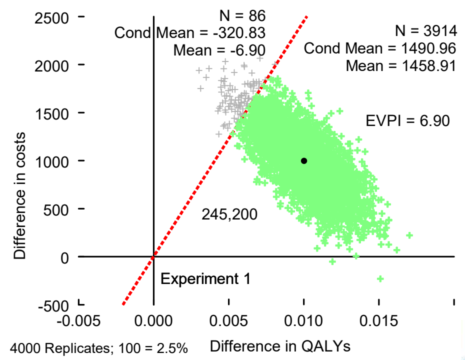
Calculating Per-Person EVPI, 100,000



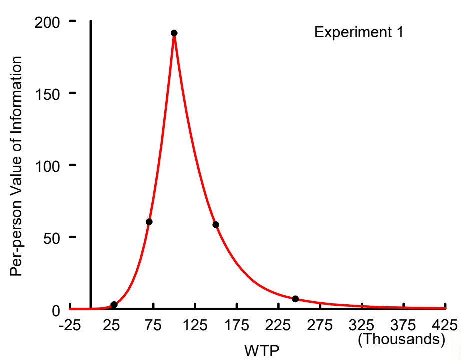
Calculating Per-Person EVPI, 150,000



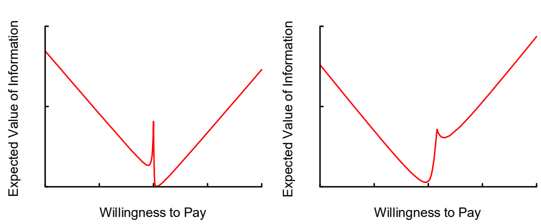
Calculating Per Person EVPI, 245,200



Per-Person EVPI Graph



Two Basic **Parametric** EVPI Curve Patterns



Per Person EVPI

- Can be large because either there is a lot of uncertainty or because cost of mistakes (i.e., W) is large
  - e.g., might already be very certain (e.g., 99.99% confident), but if cost of mistakes is extremely high might want even greater certainty
- Can be small because either there is a lot of certainty or because costs of mistakes are small
  - e.g., might be very uncertain (e.g., only 10% confident), but if the cost of mistakes is extremely low, might not need greater certainty

Total EVPI

- Total EVPI = N * $EVPI_{pp}$
  - where N = number of people for whom treatment is indicated
- Net EVPI = Total EVPI - Cost of gathering additional information
- Given additional research is unlikely to yield perfect information, net EVPI at best provides upper bound on how much additional research should be funded
  - Need to focus on value of expected change in information
- Can also be used to evaluate particular uncertainties for which research is needed: expected value of perfect information for a parameter (EVPPI)

Potential VOI Caveat

- "...value of information methods require consideration of the totality of the evidence base...."
- "...may not therefore be appropriate to simply base value of information estimates on the sampling variability from a single study where other studies exist."

Asserted EVPI Advantages

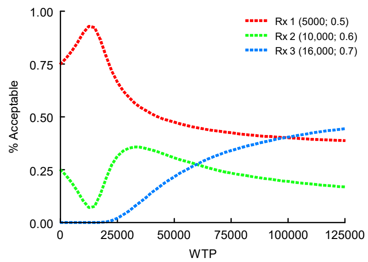
- Quantitative measure of when we have enough information to make a decision
- Avoids inference
- Avoids temptation to use 'need for evidence' to delay decision making
- Recognises information gathering is not costless
- Can distinguish value of different types of information which might guide study design

Rejection of Inference Applies to Everything

- Nothing different about economic decisions and other decisions
  - If we adopt an EVPI decision criterion – i.e., reject an inference criterion – for making economic decisions about therapies, should do same for other decisions
    - FDA should stop requiring significance for drug adoption decisions
- Economics (theoretical) vs medical (life and death) decision making?
- Significance testing may be transactionally efficient
  - Assuming there are costs of switching therapies, interpret significance tests as a mechanism for limiting switching and reducing these costs
    - Can build these (and other costs) into EVPI

Acceptability Curves When More Than 2 Therapies

Acceptability Curves When More than 2 Therapies



- When comparing more than 2 therapies, common to graph one curve per therapy with curves representing proportion of time therapy is best value (e.g., for Rx1: fraction that Rx1 > Rx2 AND Rx1 > Rx3)

Counting Methods

- Calculate MB for each Rx using WTP; count MBs where Rx 1's MB greater than both Rx 2's and Rx 3's
- Calculate NMB for 1 vs 2, 1 vs 3, and 2 vs 3 using WTP; count NMBs where Rx 1's NMB vs Rx 2 $\geq$ 0 AND Rx 1's NMB vs Rx 3 $\geq$ 0

Violation of Independence of Irrelevant Alternatives

- Best criterion violates Independence of irrelevant alternatives (IIA)
  - IIA a ubiquitous assumption in welfare economics / social choice theory
- IIA: Choice between alternatives x and y depends on preferences for x and y only (and is not affected by preferences for z)
  - e.g., if Rx 1 is chosen over Rx2 and Rx3, Rx1 must be both better than Rx2 and better than Rx3
- Focusing solely on fraction of time a therapy is best throws away information about the preference between 2 therapies (e.g., x and y) when a third therapy (e.g., z) is best

Fraction of Time Best

- Suppose making choice for 7 people between 3 mutually exclusive modes of travel. Choose single mode for all 7
  - buses (B), cabs (C), or walking trails (W)
- Suppose most preferred choices are as follows:

Obs	1	2	3	4	5	6	7
Pref	W	W	W	B	B	C	C

- If basing decision solely on first preferences, heights of “multi-way” acceptability curves would equal:
  - walk, 3/7; bus, 2/7; cab, 2/7
    - i.e., walking is “best”

Fraction of Time Better Value

- Suppose people who prefer cabs or buses least prefer walking; people who prefer walking least prefer cabs.
- Rank-ordered preferences would be:

Obs	1	2	3	4	5	6	7
1st	W	W	W	B	B	C	C
2nd	B	B	B	C	C	B	B
3rd	C	C	C	W	W	W	W

- B is preferred to both W (4/7) and C(5/7)
- C is preferred to W (4/7), but not B (2/7)
- W is least preferred (3/7 against both B and C)

Summary

- According to best rule, W is best and indifferent between B and C
- If instead consider complete set of preferences:
  - B preferred to both W and C
  - C preferred to W

What's the Alternative

- Return to use of multiple pairwise comparisons
  - Strong tradition in economic choice theory, e.g., basis of Arrow impossibility theorem
- Analog to "best" algorithm is to select therapy that in pair-wise comparison is better than all other therapies
  - ??? Significantly better ???

What's the Alternative (2)

- For each value of WTP plot lowest percentage acceptable against all other therapies
  - If B better than W 4/7 of time and better than C 5/7, height of B curve = 4/7
  - If C better than B 3/7 of time and better than W 4/7, height of C curve = 3/7
  - If W better than both B and C 3/7 of time, height of W curve = 3/7
- Best alternative has highest curve
  - i.e., select B because it is better than other 2 options at least 4/7 of the time
- Note, sum of heights of curves >1

Example For Single Value of W

- Assume 4 Rx, 1-4; WTP = 1900

Fraction of times NMB for Rx (row identifiers in column 1) exceeds NMB for other Rxs (column identifiers)

	Rx 1	Rx 2	Rx 3	Rx 4	BEST
Rx 1	--	**0.215**	0.5825	0.737	0.1875
Rx 2	**0.785**	--	0.872	0.914	0.7015
Rx 3	0.4175	**0.128**	--	0.7685	0.075
Rx 4	0.263	**0.086**	0.2315	--	0.036

Acccomp.dta, cost and q11 variables

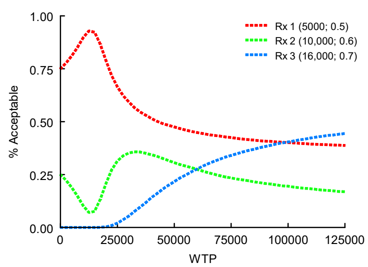
- Rx 2 better than Rx 1 78.5% of time, than Rx 3 87.2% of time, and Rx 4 91.4% of time
- Rx 2 better curve has height of 0.785 for W=1900

Simulation

	Rx1	Rx2	Rx3
Cost	5000 (5000)	10,000 (2500)	16000 (2500)
QALY	0.5 (0.3)	0.6 (0.2)	0.7 (0.2)

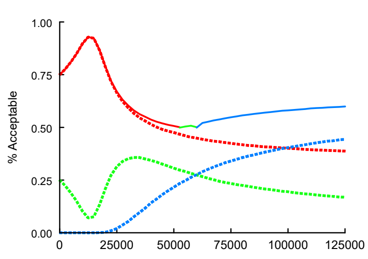
	C1	C2	C3	Q1	Q2	Q3
C1	1.0					
C2	-0.9	1.0				
C3	-0.9	0.9	1.0			
Q1	0.9	-0.9	-0.9	1.0		
Q2	-0.9	0.9	0.9	-0.9	1.0	
Q3	-0.9	0.9	0.9	-0.9	0.9	1.0

Multi-way Curve Simulation, Best Curves



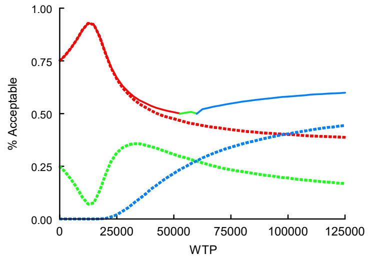
- Rx1 “best” for W between 0 and 97,500 (red dashed line)
- Rx2 never “best” (green dashed line
- Rx3 “best” for W greater than 97,500 (blue dashed line)

Multi-way Curve Simulation, Better Frontier



- While Rx2 never “best”, between 53K and 60K it is better (green solid line) than both Rx1and Rx3

Multi-way Curve Simulation, Better Frontier (2)



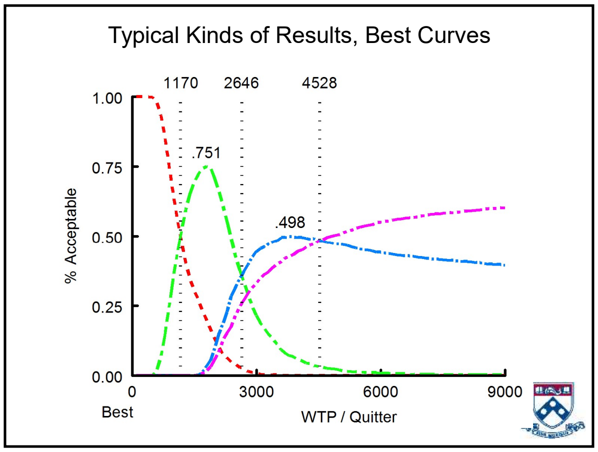
- While Rx1 “best” for W up to 97,500 (red dashed line), Rx3 (solid blue line) better than both Rx1 and Rx2 for W>60K

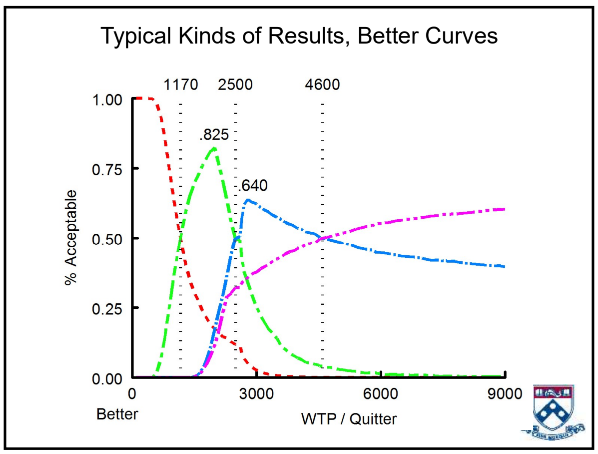
Counting Methods

- MB:
  - Calculate MB for each Rx using WTP
  - Count MBs where Rx 1’s MB greater than Rx 2’s
  - Count MB’s where Rx 1’s MB greater tha Rx3’s
  - Height of curve equals minimum of 2 fractions
- NMB:
  - Calculate NMB for Rx 1 vs 2, Rx 1 vs 3, and Rx 2 vs 3 using WTP
  - Count NMBs where Rx 1’s NMB greater than Rx 2’s
  - Count NMBs where Rx 1’s NMB greater than Rx 3’s
  - Height of curve equals minimum of 2 fractions

In Usual Practice...

- While example suggests differences can be dramatic, for typical kinds of results, 2 approaches probably have similar recommendations over wide ranges of W
- However:
  - Can observe differences around boundaries between therapies
  - Compared to “Best” algorithm, “Better” algorithm yields more appropriate measure of magnitude of probability therapy is better than alternative


Typical Kinds of Results, Best Curves
1170
2646
4528
.751
.498
1.00
0.75
0.50
0.25
0.00
% Acceptable
0
3000
6000
9000
Best
WTP / Quitter


Typical Kinds of Results, Better Curves
1170
2500
4600
.825
.640
1.00
0.75
0.50
0.25
0.00
% Acceptable
0
3000
6000
9000
Better
WTP / Quitter