JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

SENSITIVITY ANALYSIS AND INFERENCE

Lecture 12

Points of Discussion

1. Univariate Sensitivity Analyses
2. Tornado Diagrams (many univariate)
3. Bivariate/multivariate sensitivity analyses
4. How to think about confidence intervals
5. How to calculate/interpret confidence intervals with primary data (bootstrapping)
6. How to calculate/interpret confidence intervals with simulations (probabilistic SA)
7. How to combine 5 and 6
8. Cost-effectiveness acceptability curves
9. How to transform back to something that looks more like a net benefit measure

Sensitivity analysis

- Simple definition
  - Vary a ‘critical component’ of the calculation by a meaningful amount
- Objective
  - To determine whether results are insensitive to substantial but plausible variation in a parameter
  - To judge the robustness of conclusions
  - To indicate where more detailed studies to obtain better data to pin down the ICER are needed

What are critical components?

- Prices and quantities
- Sensitivity, specificity, and prevalence
- Functional relationships
  - Are risks always the same?
    - Multiple exposures
    - Different personal characteristics
- Health related quality of life measures
- Discount rate
- Measures of costs and effectiveness may be differentially sensitive to changes

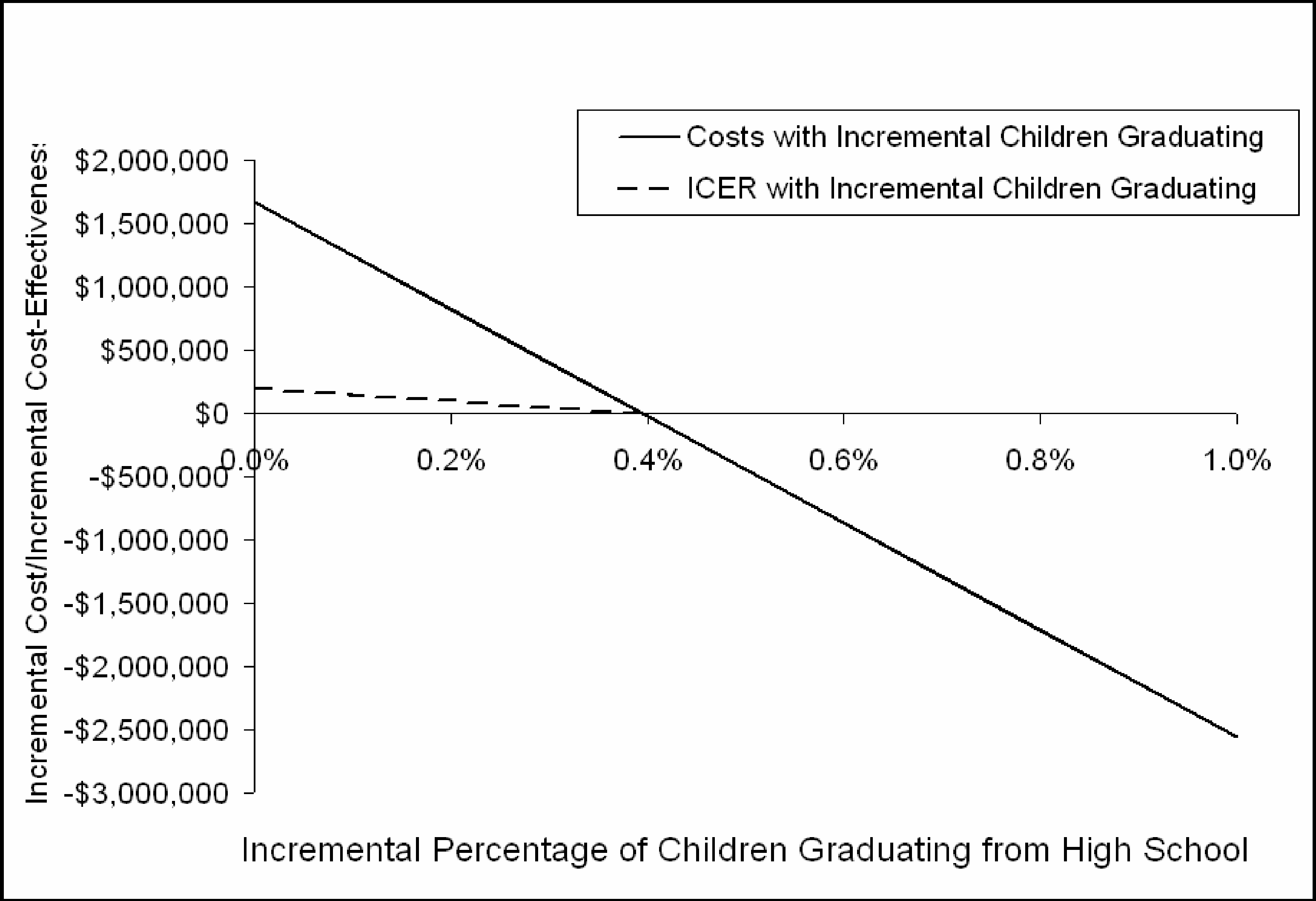
Univariate sensitivity analysis

- Vary one parameter at a time over a range
  - Best estimate, low value, high value
  - 95% confidence interval
  - “Extreme values” in plausible range
- What to do if it is difficult to find a value
  - Perform threshold analyses
    - Assess (usually reasoning-based) whether value is likely to be above or below threshold of interest
  - Can suggest a need for further research to determine range of parameter

Threshold analysis

- Solve equation for what value of a given parameter would make the incremental cost effectiveness ratio equal to a chosen value
  - $50,000/QALY
  - $100,000/QALY
  - The same ICER as was found for another intervention in a previous analysis
- Can be relatively simple algebra in some cases
- May be more intricate if looking at one parameter in a complex decision tree
  - Search over multiple values

Threshold Sensitivity Analysis



Pros and cons of univariate sensitivity analyses

- Pros
  - Can indicate whether anything alters the conclusion about the ICER's relationship with the "acceptable" range
- Cons
  - In the real world more than one thing varies at a time
  - No sense of the likelihood of being at different places in the range
  - Ignoring correlation between variation in multiple parameters can overstate uncertainty

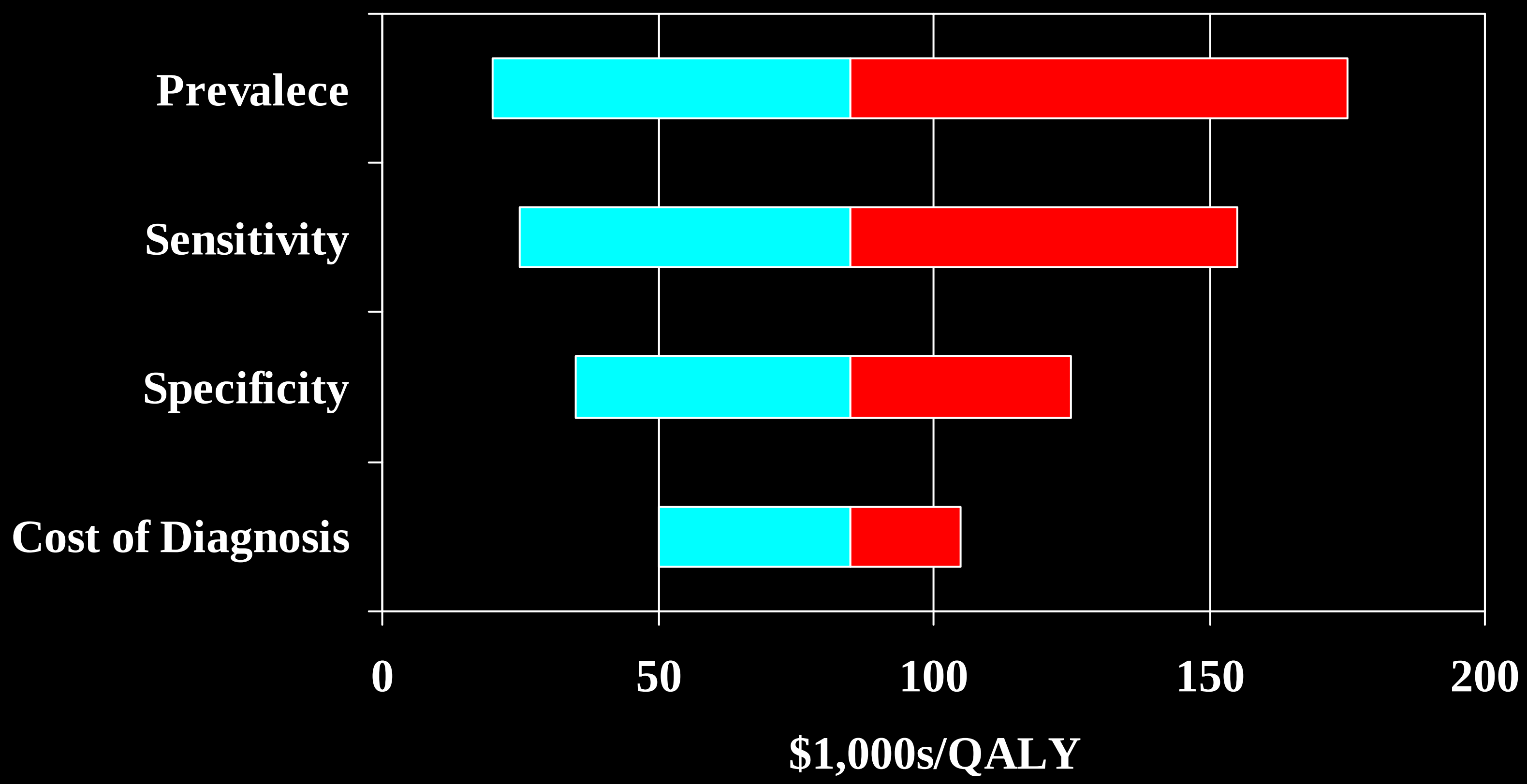
What can we do with univariate sensitivity analyses?

- Single univariate sensitivity analysis
  - Not much
- Multiple univariate sensitivity analyses
  - Tornado diagram
  - Provides a graphical way to show the varying effects of different parameters

Tornado Diagram Methods

1. Calculate the point estimate of the ICER
2. Calculate the most favorable ICERs (lowest $/QALY) when varying each parameter of interest
3. Calculate the least favorable ICERs (highest $/QALY) when varying each parameter of interest
4. Calculate the range for each parameter
5. Order parameters from widest range to most narrow range
6. Draw graph “centered” at point estimate

Tornado Diagram Example



Reporting on a Tornado Diagram

- Give ranges for parameters
- Give ranges for ICERs
- Tell which parameters were tested
  - Sometimes do not show all on tornado diagram
- Interpret the variable for which the uncertainty about the value has the biggest effect on the ICER
  - Suggests need for more information

Why might we need more than univariate sensitivity analysis?

- The ICER depends on multiple parameters
  - Ratio of two numbers
    - Numbers are functions of other parameters
    - Parameters may be be correlated
- Interaction of certain factors may imply that total effect is larger or smaller than the sum of the parts

What to do with multiple sources of uncertainty?

- Bivariate or more generally multivariate sensitivity analysis
  - How to choose which variables and how to choose how to have them vary together?
- Combination of extreme values
  - Only useful if demonstrates insensitivity to changes
  - Unlikely that all extreme parameters will occur together
- Look for set of values that provides best and worst case
  - All most of least favorable rather than all highest or lowest
- Simulation
  - Return to this one later

Confidence intervals for ICERs

- Confidence interval of ratio is not simple because…
  - Even if absolute and incremental are normally distributed…
  - Even if absolute and incremental effectiveness are normally distributed…
  - The ratio of two normally distributed variables is not normally distributed

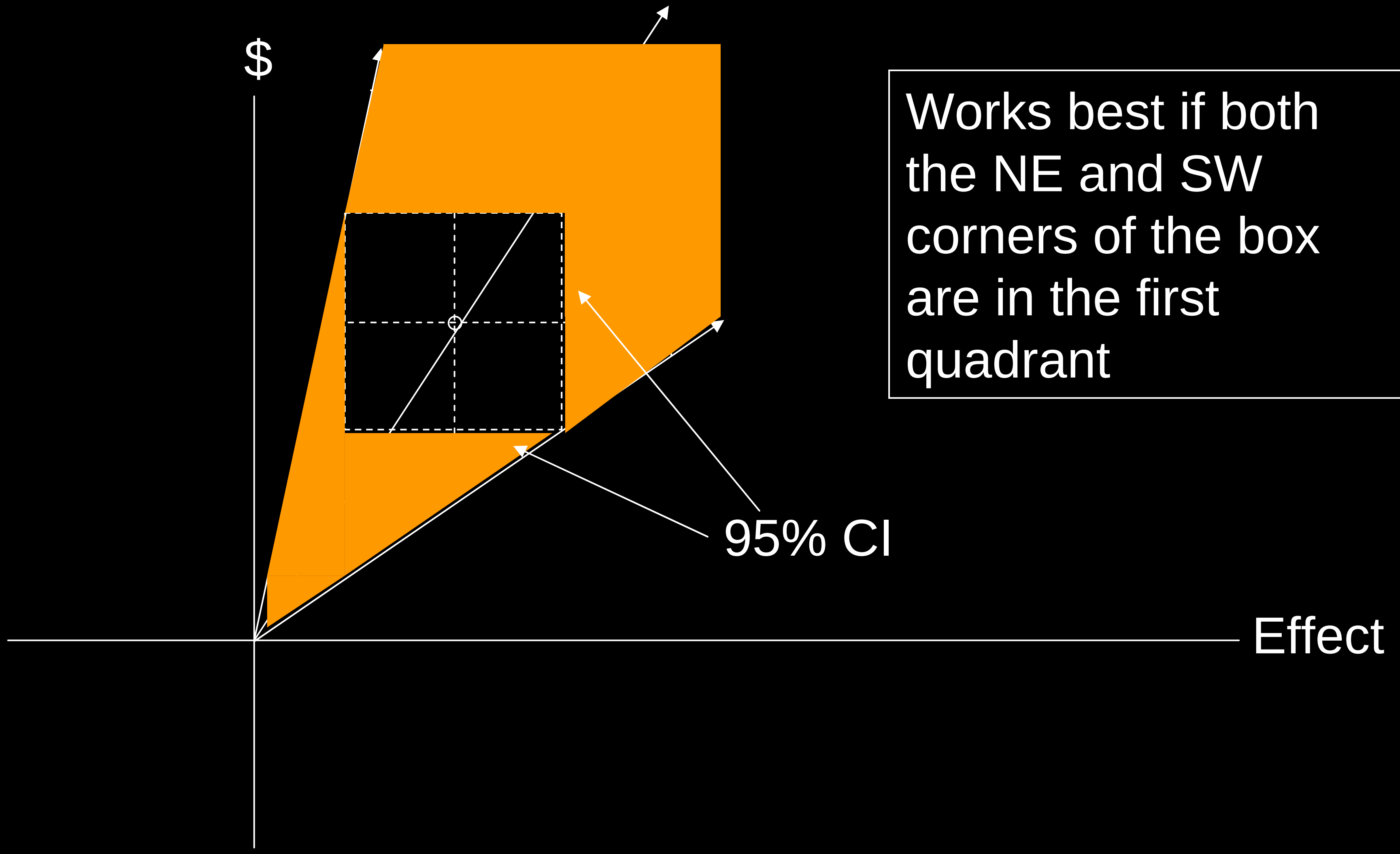
Complication #1 – Zero Denominator

- If the 95% effectiveness confidence interval includes zero, cannot calculate a 95% confidence interval for the ICER
  - ICER goes to infinity as effect goes to zero
  - If the CI for the effect includes zero, the CI for the ICER is discontinuous
    - Sign of ICER changes when cross zero effectiveness
      - Both negative and positive infinity

Early Efforts at CIs – Confidence Box Concept

- Calculate separate confidence intervals for the numerator and denominator of the incremental cost-effectiveness ratio
- Draw the “box” formed by the minimum and maximum values in the confidence intervals of each measure
- Use the ratios implied by the “northwest” and “southeast” corners

Confidence Box - Graphically



Works best if both the NE and SW corners of the box are in the first quadrant

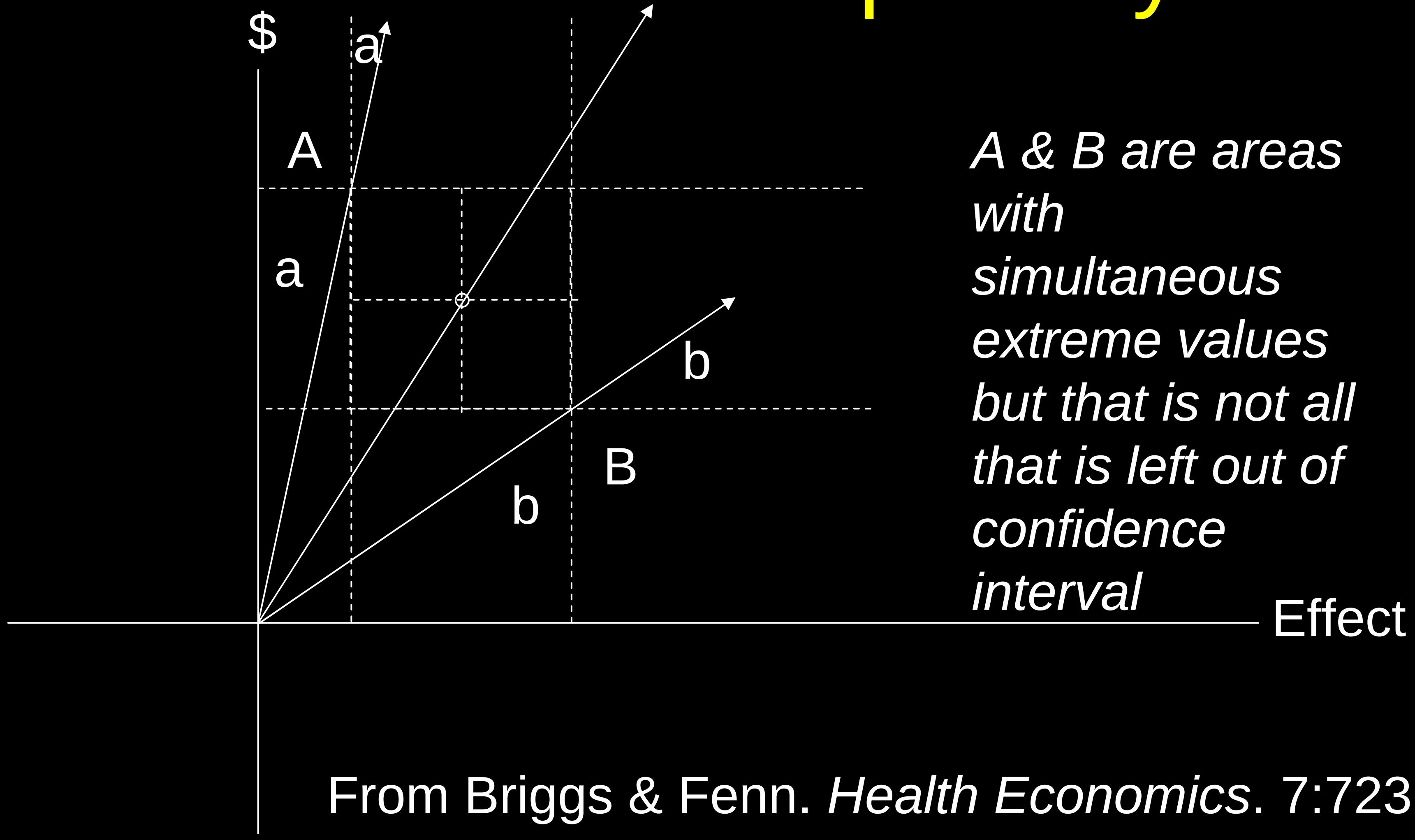
Confidence Box - Implications

- Assumes costs and effects are independent
  - Each confidence interval is calculated with no consideration of the confidence interval for the other measure
- Potential uncertainty about…
  - Zero effectiveness in the denominator
  - Dominance

Confidence Box – What is the Real Interval?

- Recall that CIs reflect Type I error
  - Proportion of times that the true parameter lies outside the estimated interval in repeated sampling
- Chance that the true parameter is within the interval is $(1-\alpha)$
  - Based on probability, the chance of two events simultaneously is $(1- \alpha)^2$
    - The confidence interval formed by the confidence box is approximately a 90% confidence interval (0.95*0.95)
    - So, what (if anything) do we know about the wedge formed by the rays from the origin through the corners?

More Confidence Box Problems - Graphically



A & B are areas with simultaneous extreme values but that is not all that is left out of confidence interval

From Briggs & Fenn. *Health Economics*. 7:723.

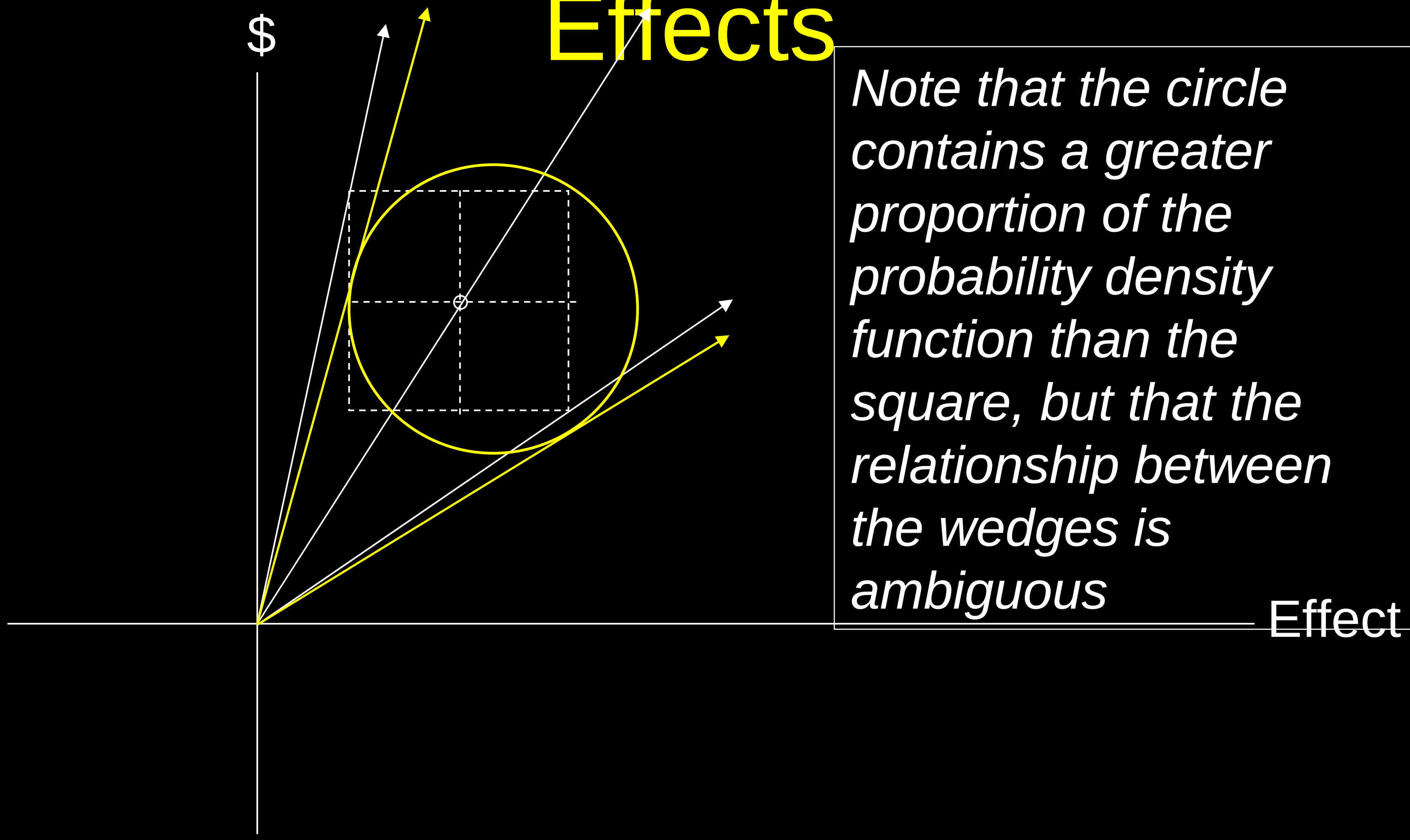
Confidence Ellipse - Concept

- Consider the fact that the costs and effects may not be independent
  - More effective treatment may cost more
  - Individuals who live longer have an opportunity to accrue greater costs over time
- We can use an ellipse to capture the confidence interval rather than a box

Confidence Ellipse - Implication

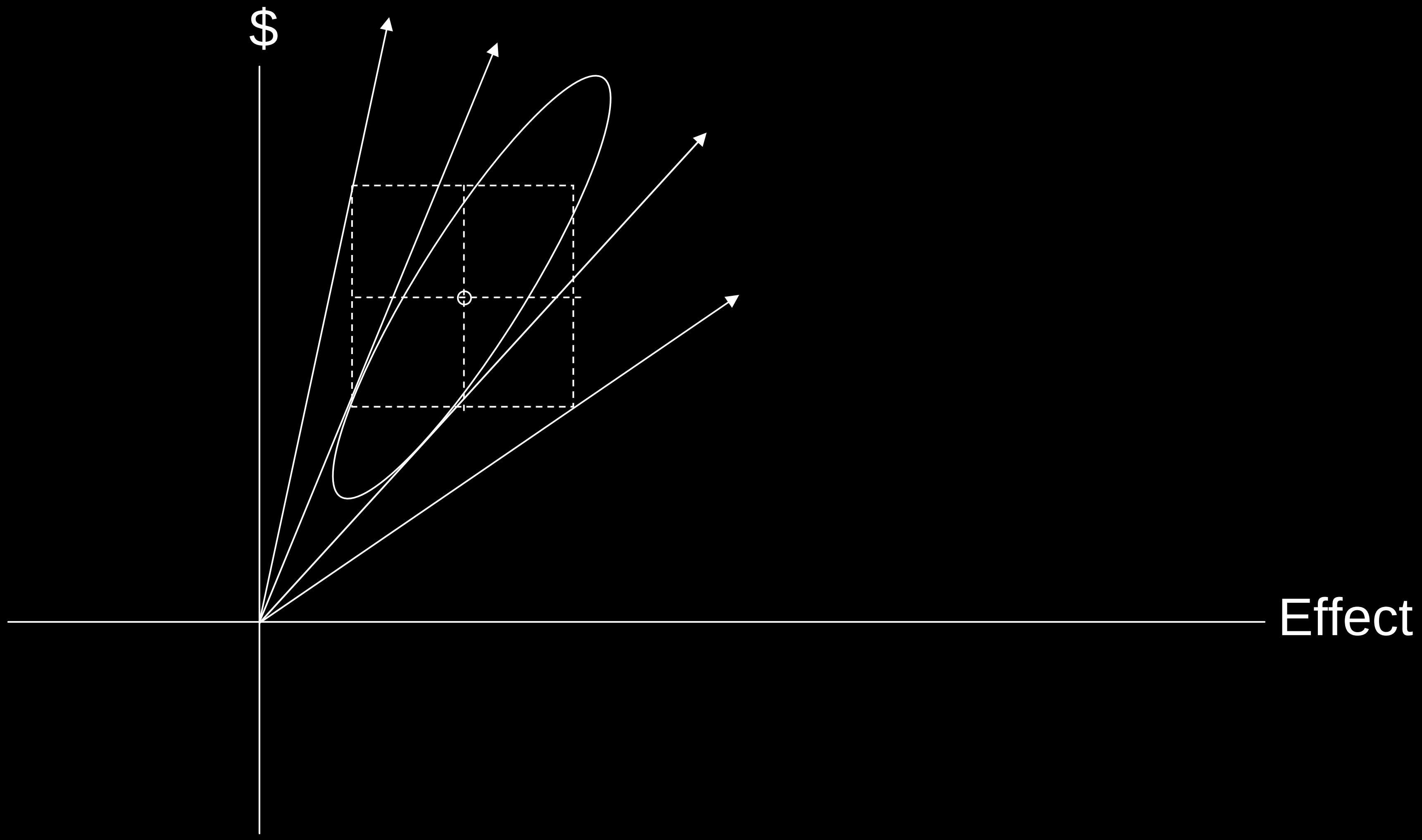
- If costs and effectiveness are positively correlated the CI will be smaller
- In contrast, the CI will be wider if costs and effectiveness are negatively correlated
- Thus, correlation can substantially influence the interpretation of the results

Confidence Ellipse – Independent Costs and Effects

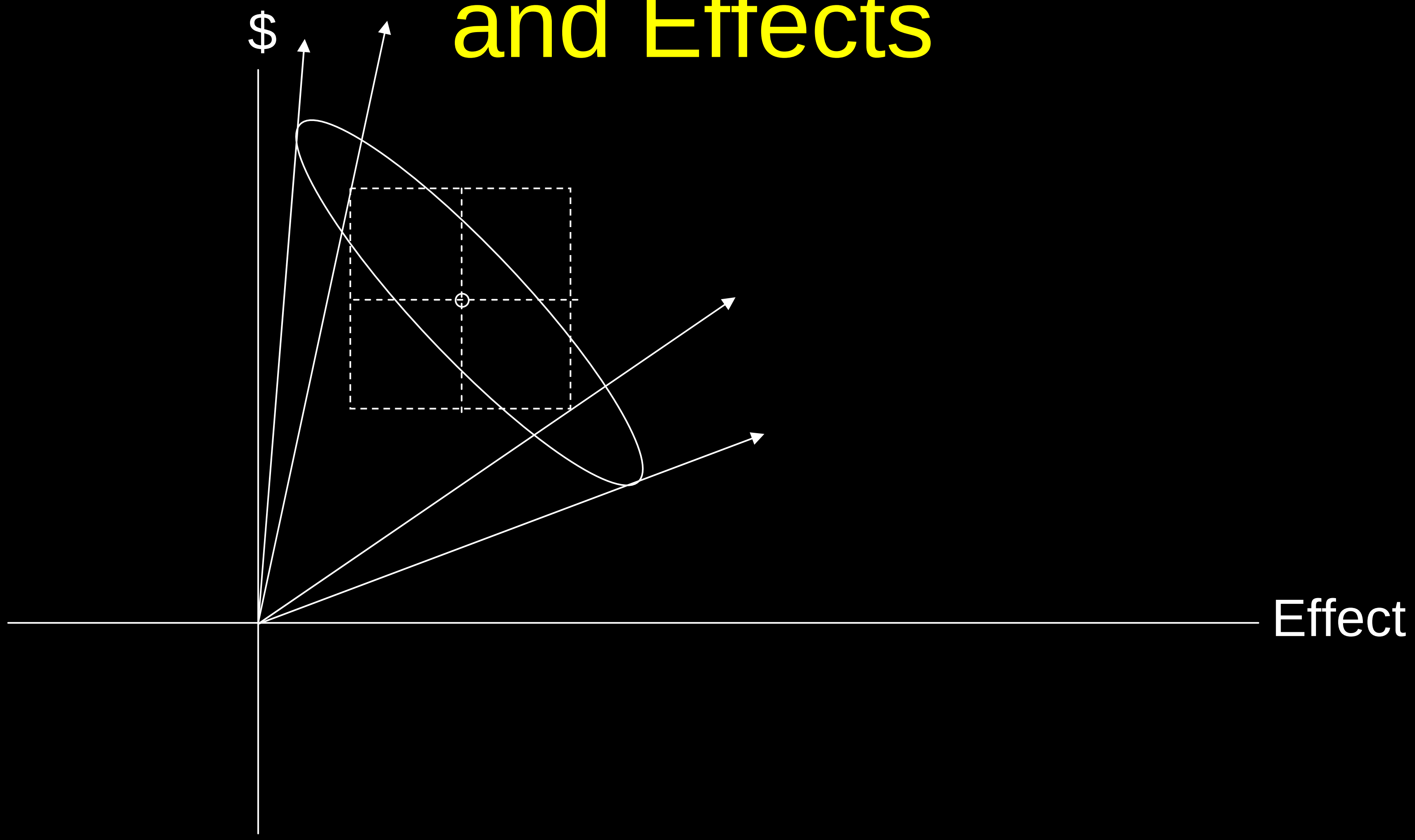


Note that the circle contains a greater proportion of the probability density function than the square, but that the relationship between the wedges is ambiguous

Confidence Ellipse - Positively Correlated Costs and Effects



Confidence Ellipse – Negatively Correlated Costs and Effects



Confidence Ellipse – Comparison with Confidence Boxes

- To calculate ellipse use the joint probability density
- Relationship between ellipses and boxes
  - The ellipse contains more of the probability density function than the box
  - But, the “wedges” formed by the rays tangential to the box and ellipse contains more of the probability density than either shape alone

Confidence Interval Objective

- Find a way to determine the wedge that contains 95% of the probability density function
  - Issue remains that the confidence interval may still include zero effectiveness
- Alternatives
  - Mathematical expression
  - Bootstrapping
  - Cost effectiveness acceptability curves

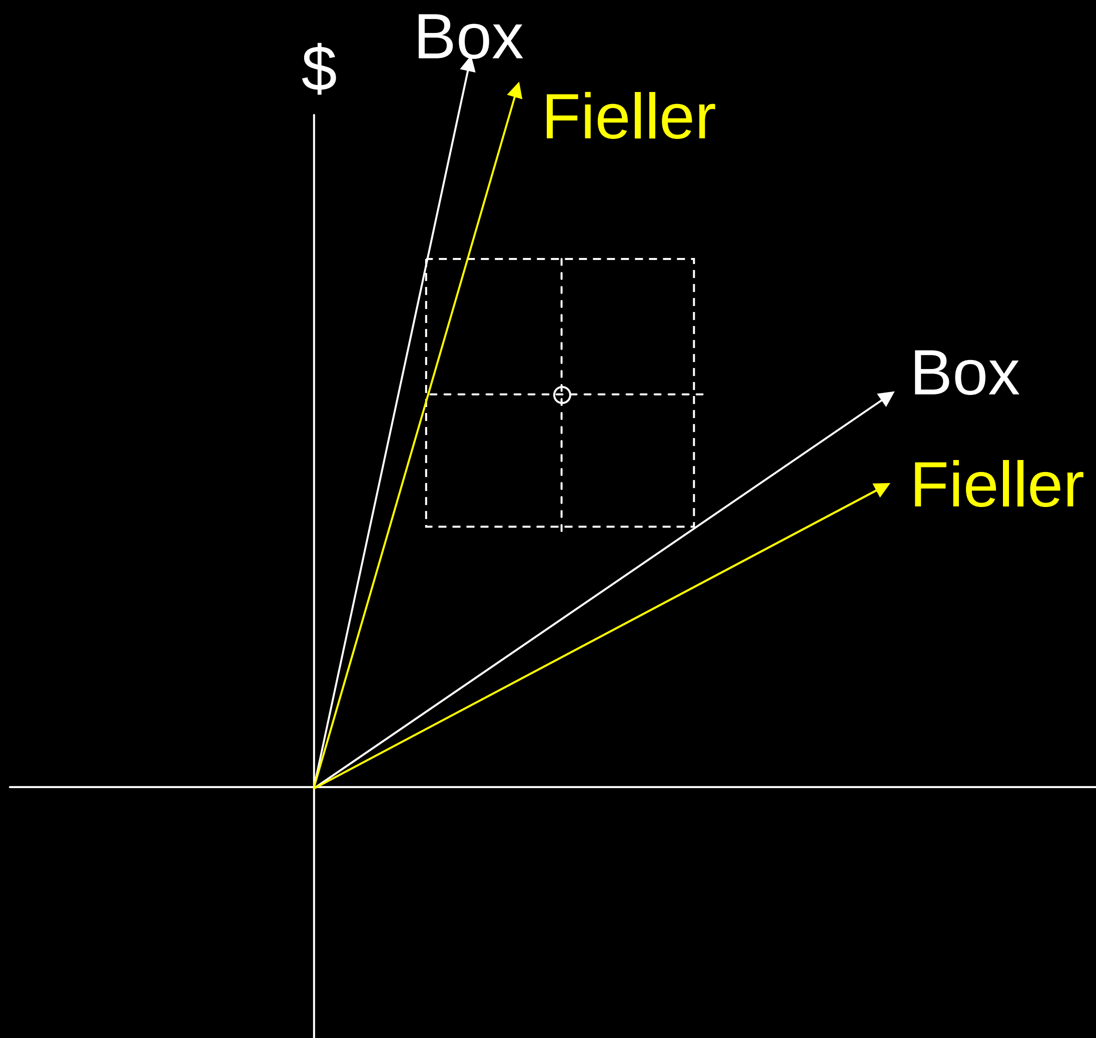
Fieller - Concept

- Distribution of ratio
- Assume that the costs and effects follow a joint normal distribution

$$\frac{R\Delta E - \Delta C}{\sqrt{R^2 \operatorname{var}(\Delta E) + \operatorname{var}(\Delta C) - 2R\operatorname{cov}(\Delta E, \Delta C)}} \sim N(0,1)$$

- Calculate exact CI such that the wedge in a graph on the CE plane will be exactly 95%
- Results are conditional on correctness of parametric assumptions

Fieller - Graphical



Relationship between Fieller wedge and box wedge is ambiguous. Primarily know that Fieller wedge is exact CI.

Fieller -Mathematically

- Find roots of quadratic equation

$$R^2\left[\Delta E^2 - z_{\alpha/2}^2 \operatorname{var}(\Delta E)\right] - 2R\left[\Delta E \cdot \Delta C - z_{\alpha/2}^2 \operatorname{cov}(\Delta E, \Delta C)\right] + \left[\Delta C^2 - z_{\alpha/2}^2 \operatorname{var}(\Delta C)\right] = 0$$

Roots = [-b ± sqrt(b^2 – 4ac)] / 2a

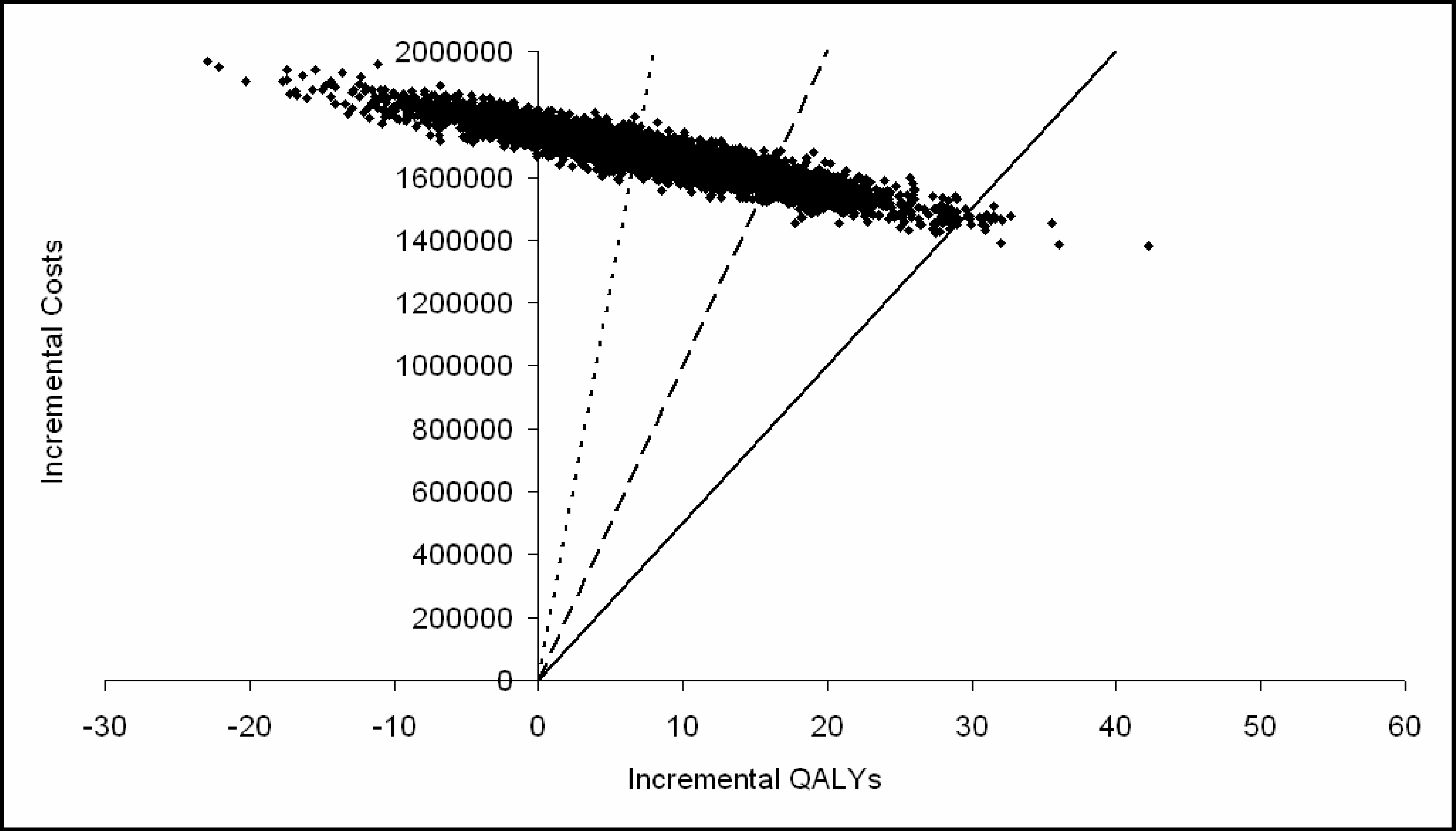
Bootstrapping - Concept

- Empirical sample distribution
- Multiple draws from a data set
- Calculate statistics based on each draw
- Draw conclusions based on those results

Bootstrapping – Basic Method

- Draw sample of same size as original sample with replacement from data
- Obtain confidence limits by taking observations at 2.5 percentile and 97.5 percentile
- Distinct from probabilistic sensitivity analysis because drawing observations from data set rather than random points in a distribution

Bootstrapping-type Example



Bootstrapping Interpretation

- Note that cannot necessarily interpret CI directly
- May be better to ask questions like:
  - What proportion of draws lead to ICERs below $50,000 per QALY?
  - What proportion of draws lead to ICERs below $100,000 per QALY
  - What proportion of draws indicate that the intervention is dominated?

Probabilistic sensitivity analyses through simulation

- Perhaps sensitivity analyses are not useful
  - Not based on the likelihood of values occurring
- Probabilistic sensitivity analyses
  - Use information on priors about each parameter
    - A somewhat Bayesian approach
    - How likely is it that the ICER is in an acceptable range
  - Can indicate usefulness of additional

What does a probabilistic sensitivity analysis really do?

- Define a distribution of a parameter
- Draw a random value from the distribution
- Calculate the ICER (usually just two alternatives) with the value drawn
- Repeat the process a number of times
- Report the mean & standard deviation of the distribution, or percentiles of the distribution, or what proportion of the results are below a threshold value for the ICER

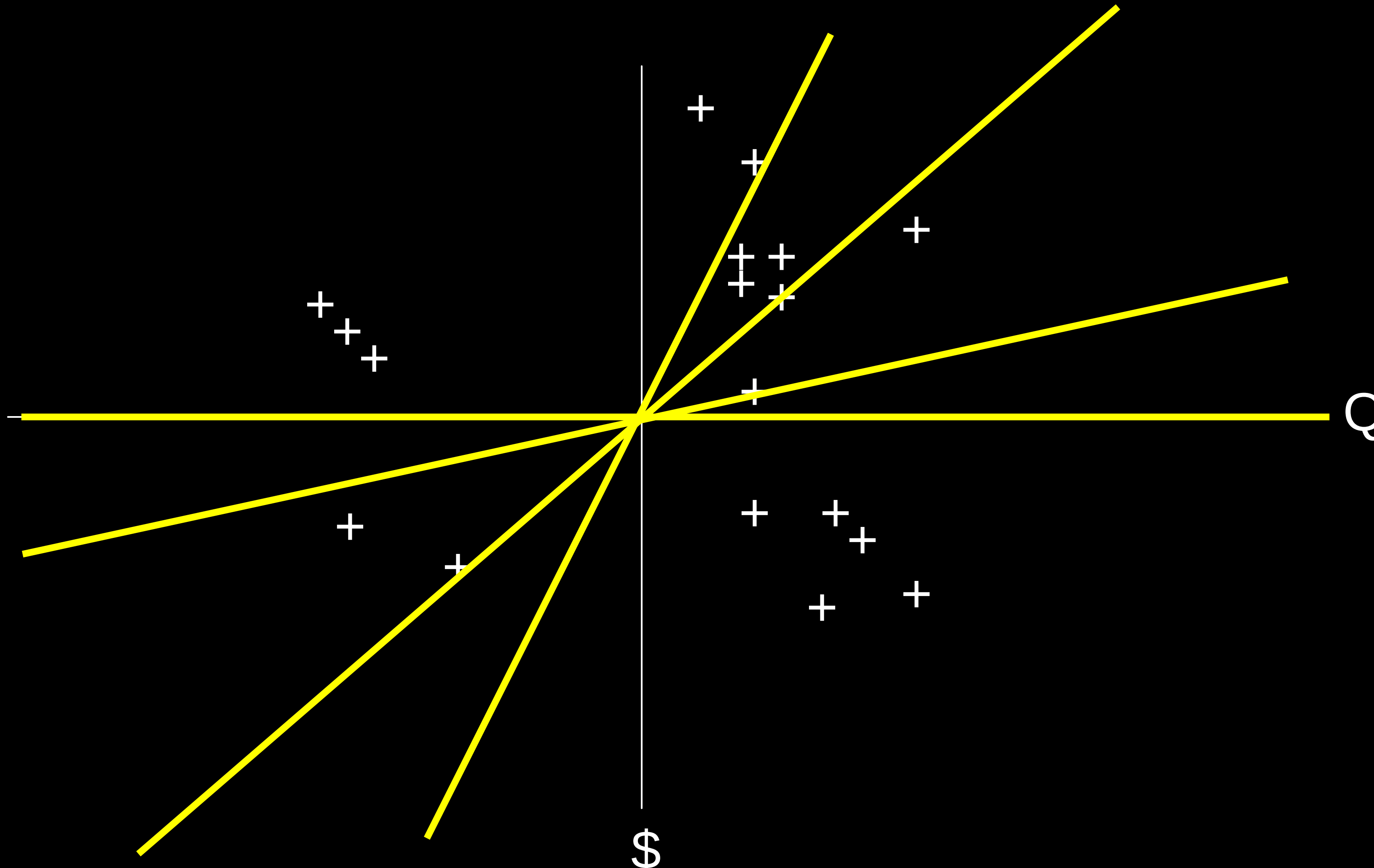
Cost-Effectiveness Acceptability Curves

- Crossover between CEA and CBA
- CEAC’s can be used for two purposes
  - Summarize (for every possible value of health effect) the evidence in favor of the intervention leading to a positive net benefit
  - Construct a confidence interval

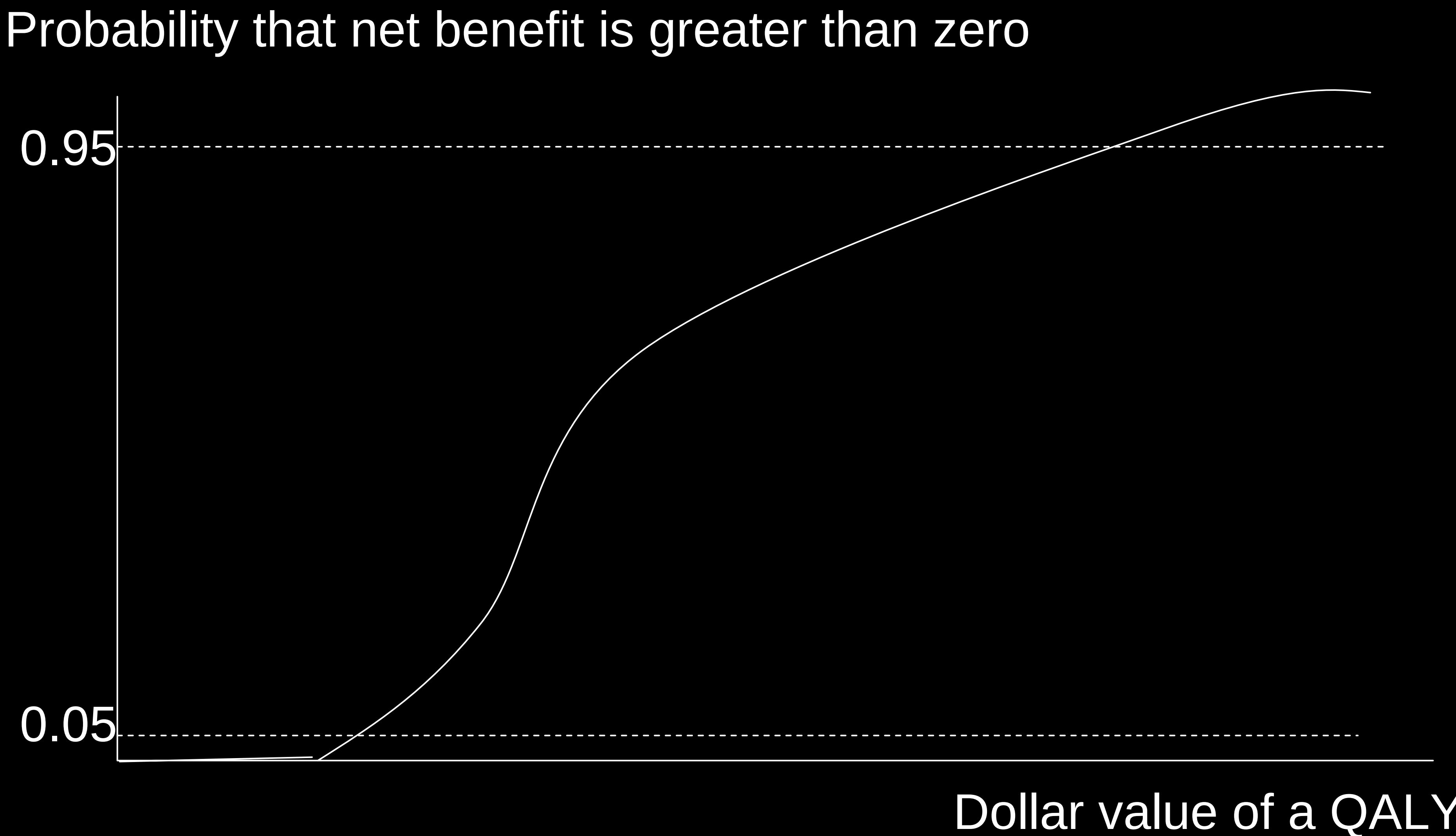
CEAC - Approaches

- Parametric
  - Use joint normality assumption from Fieller's theorem and looks at the proportion of the joint density that is cost-effective
- Non-parametric
  - Use bootstrapping or probabilistic sensitivity analysis and determine what proportion of ICER estimates lie below any value of health effect

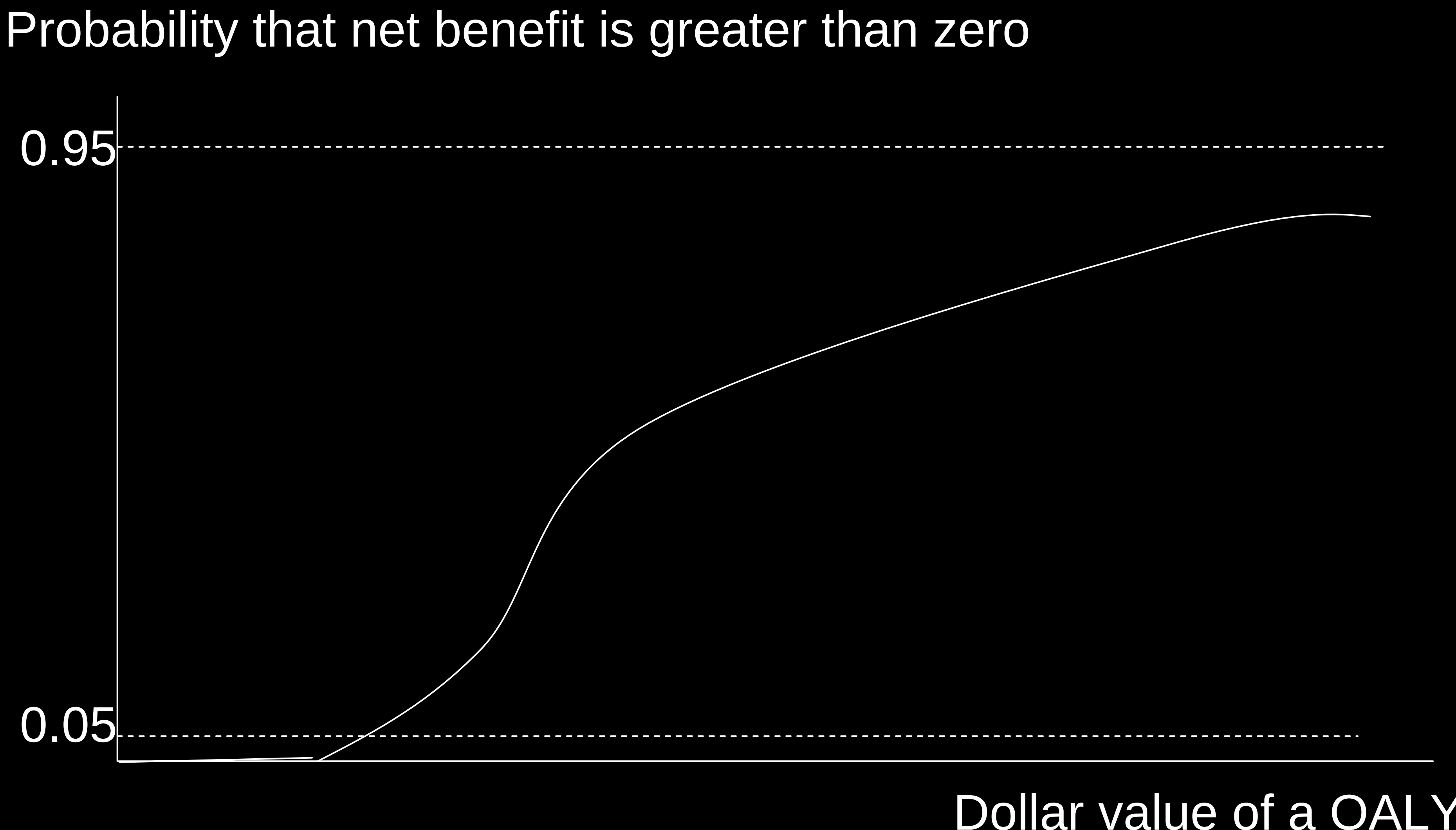
CEAC Graphical Example (1)



CEAC Graphical Example (2)



CEAC Graphical Example (2)



Lack of upper end of CEAC

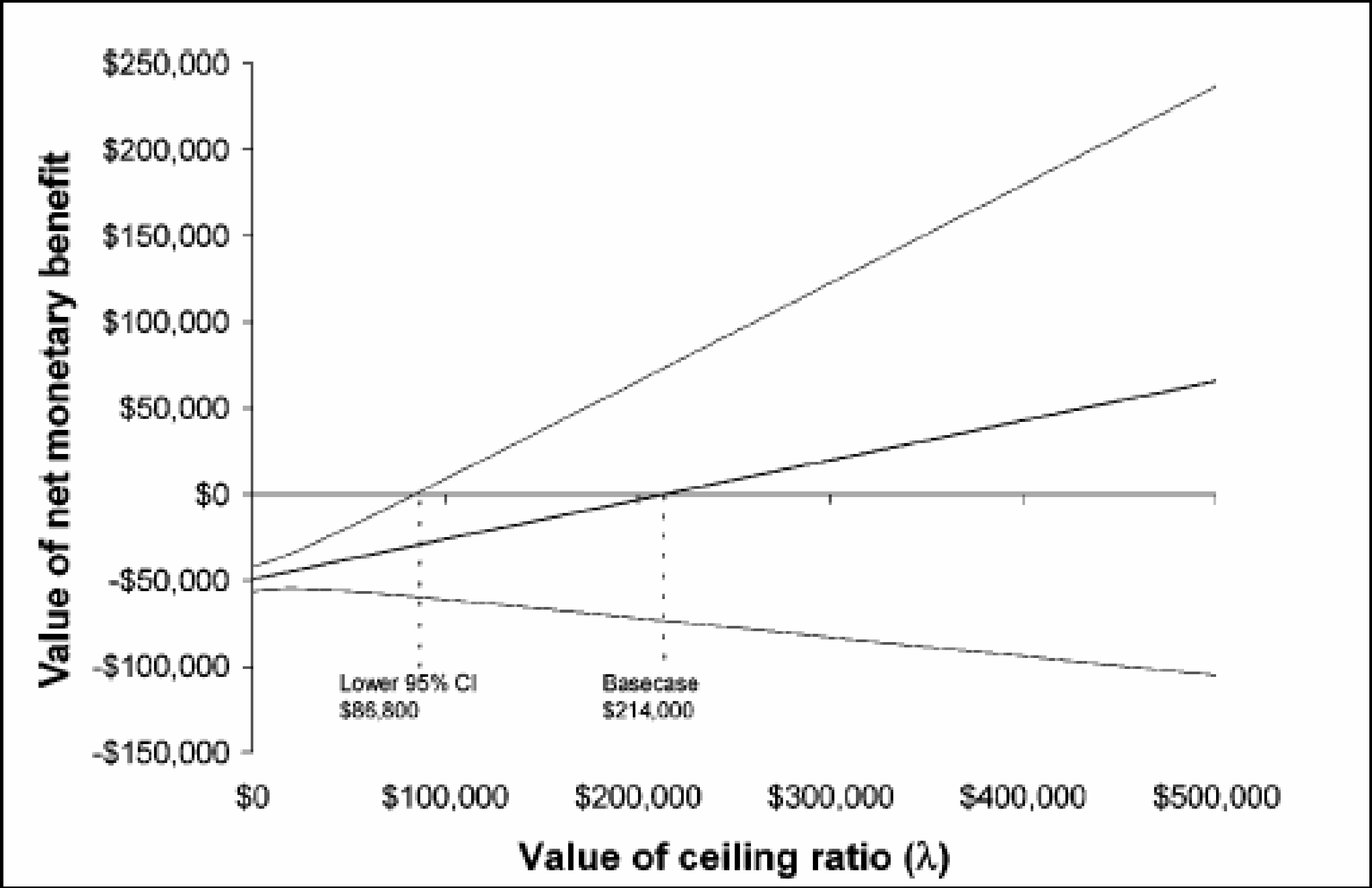
- Frick et al. Journal of Urban Health example from several slides back is perfect example
  - The distribution is so unfavorable for the intervention in question that there is never a 97.5% chance that the intervention has a positive net benefit
  - Alternatively, may have so many cost saving alternatives that never have only 2.5% with a positive net benefit

Compare & Contrast Techniques Generally

- If parametric and non-parametric results are similar, use parametric results
  - Replicability
  - Generally more powerful
- The acceptability curve tends toward a Bayesian approach, i.e. the probability that the intervention is cost-effective
- Some authors have compared theoretical implications of techniques
- Some authors have compared empirical results with techniques

Back to Net Benefits

- While some had problems with CBA because of the need to place a dollar value on life, it is easier to work with confidence intervals
  - $NMB = \lambda \Delta E - \Delta C$
  - λ = dollar value of a QALY
  - $var(NMB) = \lambda^2 var(\Delta E) + var(\Delta C) - 2\lambda\, cov(\Delta E, \Delta C)$



Briggs AH, et al. Thinking outside the box: recent advances in the analysis and presentation of uncertainty in cost-effectiveness studies. Annu Rev Public Health. 2002;23:377-401. . Reprinted, with permission, from the *Annual Review of Public Health,* Volume 23 ©2002 by Annual Reviews www.annualreviews.org"

Polsky et al.'s Monte Carlo Experiment

- Repeated samples from data
  - Box method using a different assumption about how wide to make the individual cost/effect confidence intervals
  - Taylor series method
    - An alternative mathematical approximation
  - Fieller
  - Non-parametric bootstrap

Polsky et al.'s objectives

- Determine which method came closest to the target miscoverage level of 5%
- Determine whether or not the miscoverage was symmetric
  - 2.5% of time entire CI is greater than true value and 2.5% of time entire CI is less than true value

Polsky et al. Results

- Fieller was most consistently near 5%
- Bootstrap was also fairly consistent
- Box and Taylor methods were less consistent as the correlation between costs and effectiveness increased in the positive range
- Taylor series method had asymmetric missed coverage