Economic Evaluation in Clinical Trials

Henry Glick
University of Pennsylvania
www.uphs.upenn.edu/dgimhsr

Cost-Effectiveness Analysis for Clinical Trials

Society for Clinical Trials
Montreal, Canada
05/15/16

Outline

- (Very) Brief introduction to economic evaluation
- (Very) Brief description of ideal economic evaluation in a clinical trial
- 7 issues in designing and analyzing economic evaluations in clinical trials
  - What Medical Service Use Should We Collect?
  - How Should We Value Medical Service Use?
  - How Naturalistic Should Study Be?
  - What Sized Sample Should We Study?
  - How Should We Analyze Cost (and QALY) Data?
  - How Should We Report Sampling Uncertainty for CEA?
  - How Should We Interpret Results From Multicenter (Multinational) Trials?

Brief Introduction to Economic Evaluation

- Types of Analyses
- Types of outcomes
- Perspective

Types of Analyses

Types of Analysis

- Types of analysis
  - Cost identification
  - Cost-effectiveness
  - Cost-benefit
  - Cost-utility
  - Net monetary benefit
- Generally distinguished by:
  - Outcomes included: e.g., costs only vs costs and effects
  - How outcomes are quantified: e.g., as money alone or as health and money

---

Cost-Identification / Cost-Minimization

- Estimates difference in costs between interventions, but not difference in outcomes
- Commonly conducted when no difference observed in effectiveness
- Introduction of sampling uncertainty undermines cost-identification analysis

IS FAILURE TO DETECT A DIFFERENCE SAME AS DEMONSTRATION OF EQUIVALENCE?

Cost-Effectiveness Analysis

- Estimates differences in costs and differences in outcomes between interventions
- Costs and outcomes are measured in different units
- Costs usually measured in money terms; outcomes in some other units

$$\frac{Costs_1 - Costs_2}{Effects_1 - Effects_2}$$

- Results meaningful in comparison with:
  - Predetermined threshold / cut-off for willingness to pay (e.g., $50,000 per QALY)
  - Other accepted and rejected interventions (e.g., league tables)

Cost-Benefit Analysis

- Estimates differences in costs and differences in benefits in same (usually monetary) units
- As with cost-effectiveness, requires a set of alternatives

Other Types of Analyses

- Cost-utility analysis
  - Form of cost-effectiveness analysis in which effectiveness expressed in terms of utility (e.g., quality-adjusted life years)
- Net monetary benefits
  - Multiply difference in effectiveness by threshold WTP and subtract costs ($W \Delta Q - \Delta C$)
  - Substitutes linear result for ratio
    - Avoids statistical problems that arise with ratios whose denominators can equal 0

Types of Outcomes

Types of Costs

- Direct: medical or nonmedical
- Time costs: Lost due to illness or to treatment
- Intangible costs
- Types of costs included in an analysis depend on:
  - What is affected by illness and its treatment
  - What is of interest to decision makers
    - e.g., a number of countries' decision makers have indicated they are not interested in time costs

---

What Effectiveness Measure?

- Can calculate a ratio for any outcome
  - Cost per toe nail fungus day averted
- For cost-effectiveness ratios to be an informative, must know willingness to pay for outcome
  - In many jurisdictions, quality-adjusted life year (QALY) is recommended outcome of cost-effectiveness analysis
  - In US, some resistance to this outcome, particularly from Congress

QALYS

- Economic outcome that combines preferences for both length of survival and quality into a single measure
- Help us decide how much to pay for therapies that:
  - Save fully functional lives/life years

    VS
  - Save less than fully functional lives/life years
    - e.g., heart failure drug that extends survival, but extra time spent in NYHA class III

    VS
  - Don't save lives/life years but improve function
    - e.g., heart failure patients spend most of their remaining years in class I instead of class III

QALY Scores

- QALY or preference scores generally range between 0 (death) and 1 (perfect health)
  - E.g., health state with a preference score of 0.8 indicates that year in that state is worth 0.8 of year with perfect health
  - There can be states worse than death with preference scores less than 0

Prescored Health State Classification Instruments

- Dominant approach for QALY measurement uses prescored health state classification instruments (indirect utility assessment)
- Participants' report their functional status across a variety of domains
- Preference scores derived from scoring rules that usually have been developed from samples from general public

EQ-5D, HUI2, HUI3 and SF-6D

- EQ-5D, HUI2, HUI3, and SF-6D are 4 most commonly used prescored preference assessment instruments
- All share features of ease of use
  - e.g., high completion rates and ability to be filled out in 5 minutes or less
- All have been used to assess preferences for a wide variety of diseases

Superiority?

- Widespread direct comparison of instruments doesn't provide answer about which instrument to use
  - Evaluation of correlations between instruments' preference scores find good correlation
  - Evaluation of correlations between instruments' scores and convergent validity criteria find good correlation
  - Evaluation of instruments' responsiveness find good responsiveness
- Most studies have concluded:
  - The instruments differ in their scores
  - Little evidence that one instrument superior to others

Study Perspective

Study Perspective

- Economic studies should adopt 1 or more "perspectives"
  - Societal
  - Payer (often insurer)
  - Provider
  - Patient
- Perspective helps identify services that should be included in analysis and how they should be costed out
  - e.g., patient out-of-pocket expenses may be excluded from insurer perspective
  - Not all payments may represent costs from societal perspective

Good Value for the Cost

- Economic data collected as secondary (or primary) endpoints in randomized trials commonly used in evaluation of" value for the cost"
  - Short-term economic impacts directly observed
    - Within-trial analysis
  - Longer term impacts potentially projected by use of decision analysis
    - Long term projection
  - Reported results: point estimates and confidence intervals for estimates of:
    - Incremental costs and outcomes
    - Comparison of costs and effects

Sample Results Table

Analysis	Point Estimate	95% CI
Incremental Cost	-713	-2123 to 783
Incremental QALYs	0.13	0.07 to 0.18
Cost-Effectiveness Analysis		
Principal Analysis	Dominates	Dom to 6650
Survival Benefit		
-33%	Dominates	Dom to 9050
+33%	Dominates	Dom to 5800
Drug Cost		
-50%	Dominates	Dom to 4850
+50%	Dominates	Dom to 8750
Discount rate		
0%	Dominates	Dom to 6350
7%	Dominates	Dom to 7000

Steps in Economic Evaluation

Steps in Economic Evaluation

Step 1: Quantify costs of care

Step 2: Quantify outcomes

Step 3: Assess whether and by how much average costs and outcomes differ among treatment groups

Step 4: Compare magnitude of difference in costs and outcomes and evaluate "value for costs"

- – e.g. by reporting a cost-effectiveness ratio, net monetary benefit, or probability that ratio is acceptable
- – Potential hypothesis: Cost per quality-adjusted life year saved significantly less than $75,000

Step 5: Perform sensitivity analysis

Ideal Economic Evaluation Within a Trial

- Conducted in naturalistic settings
  - – Compares therapy with other commonly used therapies
  - – Studies therapy as it would be used in usual care
- Well powered for:
  - – Average effects
  - – Subgroup effects
- Designed with an adequate length of follow-up
  - – Allows assessment of full impact of therapy
- Timely
  - – Can inform important decisions in adoption and dissemination of therapy

Ideal Economic Evaluation Within a Trial (II)

- Measure all costs of all participants prior to randomization and for duration of follow-up
  - Costs after randomization—cost outcome
  - Costs prior to randomization—potential predictor
- Independent of reasons for costs
- Most feasible when:
  - Easy to identify when services are provided
  - Service/cost data already being collected
  - Ready access to data

Difficulties Achieving an Ideal Evaluation

- Settings often controlled
- Comparator isn't always most commonly used therapy or currently most cost-effective
- Investigators haven't always fully learned how to use new therapy under study
- Sample size required to answer economic questions may be greater than sample size required for clinical questions
- Ideal length of follow-up needed to answer economic questions may be longer than follow-up needed to answer clinical questions

TRADE-OFF: Ideal vs best feasible

Issue #1. What Medical Service Use Should We Collect?

What Medical Service Use Should We Collect?

- Real/perceived problem: Don't have sufficient resources to track all medical service use
- Availability of administrative data may reduce costs of tracking all medical service use

What if Administrative Data are Unavailable?

- Measure services that make up a large portion of difference in treatment between patients randomized to different therapies under study
  - Provides an estimate of cost impact of therapy
- Measure services that make up a large portion of total bill
  - Minimizing unmeasured services reduces likelihood that differences among them will lead to biased estimates
  - Provides a measure of overall variability

Measure as Much as Possible

- Best approach: measure as many services as possible
  - No a priori guidelines about how much data are enough
  - Little to no data on incremental value of specific items in economic case report form
- While accounting for expense of collecting particular data items

Document Likely Service Use During Trial Design

- Can improve decisions by documenting types of services used by patients who are similar to those who will be enrolled in trial
  - Review medical charts or administrative data sets
  - Survey patients and experts about kinds of care received
  - Have patients keep logs of their health care resource use
- Guard against possibility that new therapy will induce medical service use that differs from current medical service use

Limit Data to Disease-Related Services?

- Little if any evidence about accuracy, reliability, or validity judgments about relatedness
- Investigators routinely attribute AEs to intervention, even when participants received vehicle/placebo
- Medical practice often multifactorial: modifying disease in one body system may affect disease in another body system
  - In Studies of Left Ventricular Dysfunction, hospitalizations "for heart failure" (and death) reduced by 30% ($p<0.0001$)
  - Hospitalizations for noncardiovascular reasons reduced 14% ($p = 0.006$)

General Recommendations

- General Strategy: Identify a set of medical services for collection, and assess them any time they are used, independent of reason for use
- Decision to collect service use independent of reason for use does not preclude ADDITIONAL analyses testing whether designated "disease-related" costs differ

Issue #2. How Should We Value Medical Service Use?

How Should We Value Medical Service Use?

- Availability of billing data may simplify valuation
- If billing data aren't available, common strategy is to measure service use in trial and identify price weights (unit costs) to value this use

---

Common Sources of US Price Weights

- Hospital care
  - Hospital bills adjusted by Federal cost-to-charge ratios
  - DRG payments
  - National inpatient sample
    - Calculator or dataset
  - Other administrative databases that include patient-level clinical and cost information
- Physician services
  - Medicare fee schedule
  - Other administrative databases

Common Sources (2)

- Laboratory tests
  - Clinical Diagnostic Laboratory Fee Schedule
- Durable equipment
  - Medicare Durable Good Fee Schedule
- Pharmaceuticals
  - Federal Supply Schedule
  - Adjusted AWP
  - National Average Drug Acquisition Cost (NADAC)
  - National Average Retail Prices (NARP)

Concomitant Medications

- Common to be very precise when costing study medications
- Greater problems posed by costing out concomitant medications
  - Number of agents / routes of administration / dosages / # of doses
- To facilitate use of data, some investigators simplify process:
  - Categorize drugs into classes
  - Identify 1 or 2 representatives of class (including route / dosage / # of doses)
  - Cost out representative drugs and use their cost as cost for all members of class

Issue #3. How Naturalistic Should Study Be?

How Naturalistic?

- Primary purpose of cost-effectiveness analysis:

    Inform real-world decision-makers about how to respond to real-world health care needs

- Greater naturalism, in terms of participants, analysis based on intention to treat, and limitation of loss to follow-up, implies greater likelihood that data developed within trial will speak directly to decision question

---

#3a. Intention to Treat

- Economic questions relate to treatment decisions (e.g., whether to prescribe a therapy), not whether patient received drug prescribed nor whether, once they started prescribed drug, they were switched to other drugs
    - Implication: costs and effects associated with these later decisions should be attributed to initial treatment decision
- Thus, trial-based cost-effectiveness analyses should adopt an intention-to-treat design

---

#3b. Loss to Follow-up

- Trials should be designed to minimize occurrence of missing data
    - Study designs should include plans to aggressively pursue participants and data throughout trial
    - Strategies may include:
        1) intensive outreach to reschedule assessment, followed by
        2) telephone assessment, followed by
        3) interview of a proxy who had been identified and consented at time of randomization

Loss to Follow-up (2)

- Investigators should also ensure that:
  - Follow-up continues until end of study period
  - Data collection isn't discontinued simply because a participant reaches a clinical or treatment stage such as failure to respond (as often happens in antibiotic, cancer chemotherapy, and psychiatric drug trials)
    - Given that failure often is associated with a change in pattern of costs, discontinuation of these patients from economic study likely biases results

#3c. Protocol-Induced Costs and Effects

- Common concerns:
  - Standardization of care in clinical trial protocols often means that care delivered in trials differs from usual care
    - Protocol may require substantial number of investigations and diagnostic tests that would not be performed under normal clinical practice
  - Protocols often prescribe aggressive documentation and treatment of potential adverse effects that differ from usual care
- Omit these costs???

Omission of Protocol-Induced Costs?

- Criterion for including costs should NOT be "Would services have been provided in usual care"
- Should be: "Could services have affected care / outcomes (and thus costs and effects)"
- No problem omitting services that cannot affect care / services
  - e.g., Cost of genetic samples that will not be analyzed until after follow-up is completed
- More problematic to omit services that can change treatment and affect outcome
  - "Cadillac" costs may yield "Cadillac" outcomes
  - Would have to adjust both costs and their effects on outcomes

Biases?

- Protocol-induced testing may bias testing cost to null
  - There might be a difference in this testing in usual care, but it can't be observed if everyone is routinely tested
- Protocol-induced testing may bias cost and outcome in an unknown direction
  - Trial's extra testing may lead to:
    - Avoidance of outcomes that would have occurred had there been no extra detection and treatment
    - Early detection and treatment of outcomes when they are less severe and easier to treat
    - Detection and treatment of outcomes that wouldn't have been detected and treated in usual care

---

Issue #4. What Sized Sample Should We Study?

What Sized Sample?

- Goal of sample size and power calculation for cost-effectiveness analysis is to identify likelihood that an experiment will allow us to be confident that a therapy is good or bad value when we adopt a particular willingness to pay
  - e.g., We may:
    - Expect a point estimate for cost-effectiveness ratio of 20,000 per QALY
    - Be willing to pay at most 75,000 per QALY
    - Want an experiment that provides an 80% chance (i.e., power) to be 95% confident (alpha) that therapy is good value

Sample Size Formula

- At most basic level, sample size for cost-effectiveness is calculated using same formula as used for sample size for a difference in any continuous variable:

$$n = \frac{2\,(z_\alpha + z_\beta)^2\,\,sd_{nmb}^2}{\Delta nmb^2}$$

where n = sample size/group; z_α and z_β = z-statistics for α (e.g., 1.96) and β (e.g., 0.84) errors; sd_{nmb} = standard deviation for NMB; and Δnmb = expected difference in NMB

Sample Size Formula (2)

- Complexities arise because 1) difference being assessed is difference in NMB ($W\Delta Q - \Delta C$) and 2) standard deviation of NMB is a complicated formula
- Data needed to calculate sample size include:
  - Difference in cost
  - SD, difference in cost
  - *Difference in effect*
  - *SD, difference in effect*
  - Z_α and Z_β
  - Correlation of difference in cost and effect
  - Willingness to pay

Full Formula

$$n = \frac{2\left(z_\alpha + z_\beta\right)^2\left(sd_c^2 + \left(W\,sd_q\right)^2 - \left(2\,W\,\rho\,sd_c\,sd_q\right)\right)}{\left(W\Delta Q - \Delta C\right)^2}$$

Correlation of Difference

- When increasing effects are associated with decreasing costs, a therapy is characterized by a negative (win/win) correlation between difference in cost and effect
  – e.g., asthma care
- When increasing effects are associated with increasing costs, a therapy is characterized by a positive (win/lose) correlation between difference in cost and effect
  – e.g., life-saving care
- All else equal, fewer patients need to be enrolled when therapies are characterized by a positive correlation than when therapies are characterized by negative correlation

---

Effect of SD_q VS SD_c on Sample Size

- Commonly thought that sample size for cost-effectiveness driven more by standard deviation for cost than it is by SD for effect
  – If not, why would we need a larger sample for economic outcome than we do for clinical outcome?
- However, if willingness to pay is substantially greater than standard deviation for cost, percentage changes in QALY SD can have a substantially greater effect on sample size than will equivalent percentage changes in cost SD

"Typical" Sample Size Table, W

WTP	Sample Size Per Group
	Exp 1 *
20,000	3466
30,000	1513
50,000	618
75,000	355
100,000	265
150,000	200

 * ΔC=25; ΔQ=0.01; sd_c=2500; sd_q=.03; ρ=-.05; α=.05; 1-β=.8

Sample Size Can Increase with Increasing W

WTP	Sample Size Per Group	
	Exp 1	Exp 2 *
20,000	3466	387
30,000	1513	442
50,000	618	594
75,000	355	806
100,000	265	1011
150,000	200	1363

* ΔC=-100; ΔQ=0.01; sd_c=5000; sd_q=.15; ρ=-0.05; α=.05; 1-β=.8
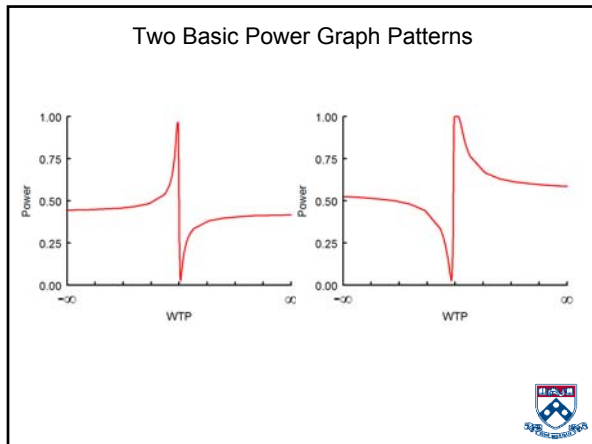
Sample Size Not Necessarily Monotonic With W

WTP	Sample Size Per Group		
	Exp 1	Exp 2	Exp 3 *
20,000	3466	387	178
30,000	1513	442	158
50,000	618	594	*151*
75,000	355	806	153
100,000	265	1011	156
150,000	200	1363	160

* ΔC=-120; ΔQ=0.015; sd_c=1000; sd_q=.05; ρ=0.0; α=.05; 1-β=.8

Six Power Patterns Associated with W

Two Basic Power Graph Patterns
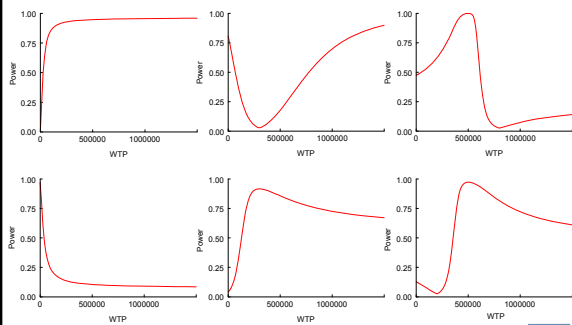
Economic Vs Clinical Sample Sizes

- Sample size required to answer economic questions often larger than sample size required to answer clinical questions
  - But it need not be
- ΔC and ΔQ are a joint outcome just as differences in nonfatal CVD events and all cause mortality are often combined into a joint outcome
- In same way that we can have more power for joint cardiovascular outcome than either individual outcome alone, we can have more power for cost-effectiveness than we do for costs or effects alone

Willingness to Pay and Identification of an Appropriate Outcome Measure

- Sample size calculations require stipulation of willingness to pay for a unit of outcome
- In many medical specialties, researchers use disease specific outcomes
- Can calculate a cost-effectiveness ratio for any outcome (e.g., cost/case detected; cost/abstinence day), but to be informative, outcome must be one for which we have recognized benchmarks of cost-effectiveness
  - Argues against use of too disease-specific an outcome for economic assessment

Issue #5. How Should Costs (QALYs) Be Analyzed?

How Should Costs (QALYs) Be Analyzed?

- Cost data typically right skewed with long, heavy, right tails
  - Can also have extreme highliers, but statistical problems often due as much to heaviness of tails as it is to highliers
- Common reactions of statisticians:
  - Adopt nonparametric tests of other characteristics of distribution that are not as affected by nonnormality of distribution ("biostatistical" approach)
  - Transform data to approximate normal distribution ("classic econometric" approach)

Policy Relevant Parameter for CEA

- In welfare economics, projects cost-beneficial if winners from any policy gain enough to be able to compensate losers and still be better off themselves
- Decision makers interested in total program cost/budget
- What we should be estimating comes out of theory, not statistical convenience
  - Policy relevant parameter should allow us to determine how much losers lose, or cost, and how much winners win, or benefit

Parameters of interest are estimates of difference in per-person population mean cost and mean effect (e.g., QALYs)

Common Multivariable Techniques Used for Analysis of Cost

- Common Techniques
  - Ordinary least squares regression predicting costs after randomization (OLS/glm with identity link and gauss family)
  - Ordinary least squares regression predicting the log transformation of costs after randomization (log OLS/identity/gauss glm predicting log cost)
  - Generalized Linear Models (GLM)
- Other Techniques:
  - Generalized Gamma regression (Manning et al.)
  - Extended estimating equations (Basu and Rathouz)

Least Squares Regression Predicting Cost

- Either OLS (SAS, proc reg; Stata, regress) or GLM with identity link and gauss family (SAS, proc glm; Stata, glm)
- Advantages
  - Easy to perform
  - No transformation problem
  - Marginal/incremental effects easy to calculate
- Disadvantages
  - Not robust
  - Can produce predictions with negative costs
- Some researchers believe disadvantages primarily theoretical
  - Claim few if any differences observed in actual practice

Least Squares Regression Predicting Log of Cost

- Either OLS or GLM predicting log of cost
- Advantages
  - Easy to perform
- Disadvantages
  - Estimation and inference directly related to log of cost / geometric mean of untransformed cost, not to arithmetic/sample mean of untransformed cost
  - In presence of differences in variance/skewness/ kurtosis, magnitude and significance of differences in geometric means can be unrelated to magnitude and significance of differences in arithmetic means
  - V/S/K differences affect percentage interpretation
  - Retransformation problems (smearing estimators)

(Log/Gamma) GLM Predicting Cost

- GLM with "appropriate" link and family
- Advantages
  - Relaxes normality and homoscedasticity assumptions
  - Consistent even if incorrect family is identified
  - Gains in precision from estimator that matches data generating function
  - Unaffected by differences in V/S/K
  - No problems with retransformation

(Log/Gamma) GLM Predicting Cost (2)

- Issues / Disadvantages
  - Can suffer substantial precision losses
  - Log link not always appropriate / best fitting
    - No agreed upon algorithm for selecting best link
      - Manning, combination of Pregibon link test, Pearson Correlation test, modified Hosmer and Lemeshow test; Hardin and Hilbe, AIC / BIC
      - Different tests recommend different links
  - Sometimes link doesn't run with recommended family
  - Sometimes link won't run with any family
  - Sometimes yield improbably large predictions
  - Still can require 2-part models

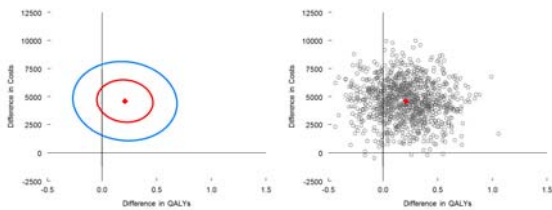### Issue #6. How Should We Report Sampling Uncertainty?

Two Most Frequently Published Uncertainty Graphs

- Cost-effectiveness plane
- Acceptability curve

Cost-Effectiveness Plane

- Bivariate normal curves (Δc, SEc, Δq, SEq, ρ) (left)
- Bootstrap of patient level data (right)

Information Derivable from Plane

- Cost-effectiveness plane provides information about point estimates, confidence intervals and p-values for:
  - Difference in effect
    - If $\leq 2.5\%$ of replicates on one or the other sides of Y axis, two-tailed $p<0.05$
  - Difference in cost
    - If $\leq 2.5\%$ of replicates on one or the other sides of X axis, two-tailed $p<0.05$
  - Cost-effectiveness analysis
    - Lines through origin that each exclude $\alpha/2$ of distribution represent $1-\alpha$ CL for CER
    - If line through origin with slope equal to WTP, falls outside interval, can be confident of value

Is CI for CER an Order Statistic?

- Commonly CI for CER assumed is an order statistic
  - Naïve ordering: order from lowest to highest ratio; identify ratios for the 2.5th and 97.5th ordered replicate
    - Works when all replicates on one side of Y axis
  - "Smart ordering": Order lexicographically (counter clockwise) first by quadrant and second by ratios within quadrant
    - Generally works when replicates on both sides of Y axis but in no more than 3 quadrants
- Ordering generally fails when replicates fall in all 4 quadrants
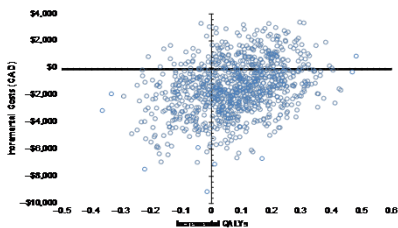 - Possible that CI for CER to be defined by lines through origin, but in most cases it can't be defined

---

Cost-Effectiveness Plane
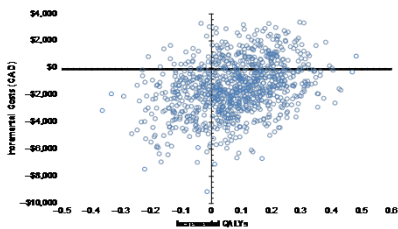
Brown ST, et al. Cost-effectiveness of insulin glargine versus sitagliptin in insulin-naïve patients w/ T2DM. Clin Therapuetics.2014; 36: 1576-87
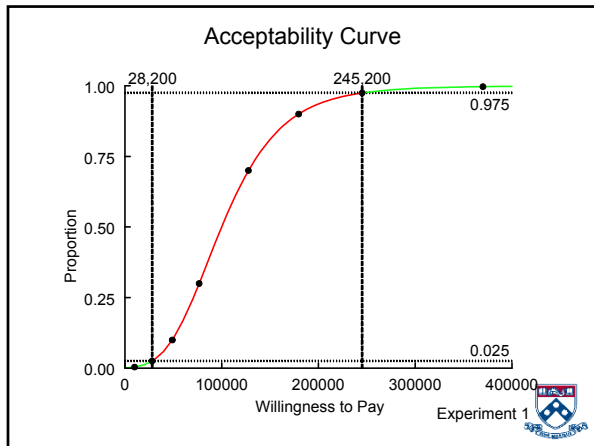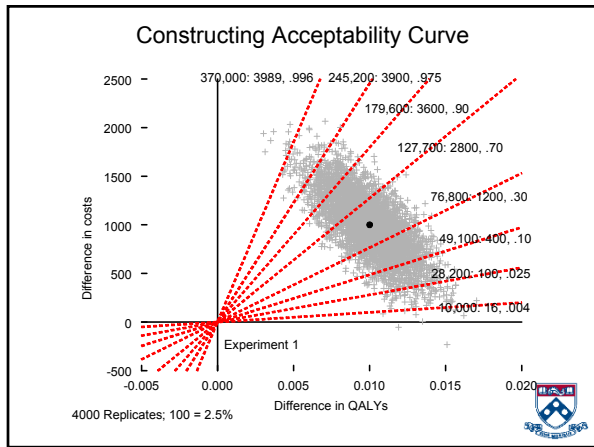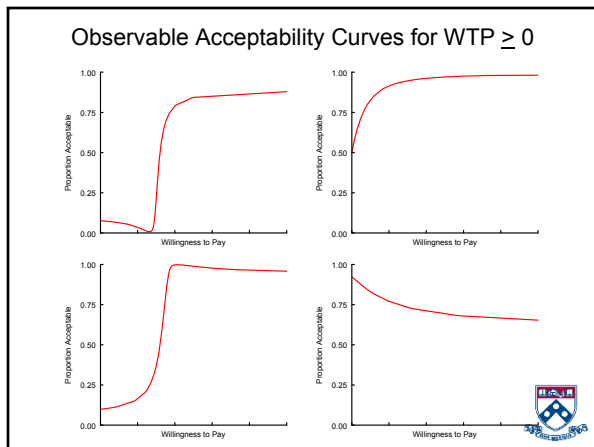

Cost-Effectiveness Plane

Reported cost difference:	-1418, 95% CI -1540 to -1295
Reported QALY difference:	0.074, 95% CI, 0.066 to 0.082
Reported ICER	-19511, 95% CI, -23815 to 2044

Brown ST, et al. Cost-effectiveness of insulin glargine versus sitagliptin in insulin-naïve patients w/ T2DM. Clin Therapuetics.2014; 36: 1576-87

Acceptability Curve

Constructing Acceptability Curve

Observable Acceptability Curves for WTP ≥ 0

"Common" Conclusions from Acceptability Curves

W	What is often said
28,200	"97.5% chance Rx A not good value"
76,800	"70% chance Rx A not good value"
100,000	"50% chance either therapy good value"
127,700	"70% chance Rx A good value"
245,200	"97.5% chance Rx A good value"

- Common to adopt 1-tailed interpretation of acceptability curve
- Ignores fact that 50% – not 0% – represents no information

Issue #7. How Should We Interpret Results From Multicenter (Multinational) Trials?

How Should We Interpret Results From Multicenter (Multinational) Trials?

- Problem:
  - There has been growing concern that pooled (i.e., average) economic results from multicenter (multinational) trials may not be reflective of results that would be observed in individual centers (countries) that participated in trial
  - Similar issues arise for any subgroup of interest in trial (e.g., more and less severely ill patients)

Common Sources of Concern

- Differences in morbidity/mortality patterns; practice patterns (i.e., medical service use); and absolute and relative prices for this service use (i.e., price weights)
- Decision makers may find it difficult to draw conclusions about value of therapies that were evaluated in multicenter (multinational) trials

Bad Solutions

- Use trial-wide clinical results, trial-wide medical service use, and price weights from one center (country)
  - e.g., to tailor results to U.S., just use U.S. price weights, and conduct analysis as if all participants were treated in U.S.
- Use trial-wide clinical results and use costs derived from subset of patients treated in country
- Ignore fact that clinical and economic outcomes may influence one another (cost affects practice which affects outcome; practice affects outcome which affects cost)

Impact of Price Weights vs Other Variation

Country	Trial-Wide Effects Price weight	Country-Specific Costs	Country-Specific Costs and Effects†
1	46,818	5921	11,450
2	57,636	91,906	60,358
3	53,891	90,487	244,133
4	69,145	93,326	181,259
5	65,800	**	**
Overall	45,892	45,892	45,892

* Willke RJ, et al. Health Economics. 1998;7:481-93
† Country-specific resource use ℗ Country-specific price weights
** New therapy dominates

Two Analytic Approaches To Transferability

- Two approaches -- which rely principally on data from trial to address these issues -- have made their way into literature
  - Hypothesis tests of homogeneity (Cook et al.)
  - Multi-level random-effects model shrinkage estimators

Drummond M, Barbieri M, Cook J, Glick HA, Lis J, Malik F, Reed S, Rutten F, Sculpher M, Severens J. Transferability of Economic Evaluations Across Jurisdictions: ISPOR Good Practices ResearchTask Force Report. Value in Health. 2009;12:409-18.

Hypothesis Tests Of Homogeneity

- Evaluate homogeneity of results from different countries
  - If no evidence of heterogeneity (i.e., a nonsignificant p-value for test of homogeneity), and test considered powerful enough to rule out economically meaningful differences in costs, can't reject that pooled economic result from trial applies to all of countries that participated in trial
  - If evidence of heterogeneity, should not use pooled estimate to represent result for individual countries
    - Method less clear about result that should be used instead

Estimation

- Multi-level random-effects model shrinkage estimation assesses whether:
  - Observed differences between countries are likely to have arisen simply because we have divided trial-wide sample into subsets VS
  - Whether they are likely to have arisen due to systematic differences between countries
- Borrows information from mean estimate to add precision to country-specific estimates
- Methods have potential added advantage of providing better estimates of uncertainty surrounding pooled result than naive estimates of trial-wide result

Summary

- Clinical trials may provide best opportunity for developing information about a medical therapy's value for cost early in its product life
- When appropriate types of data are collected and when data are analyzed appropriately, trial-based evaluations may provide data about uncertainties related to assessment of value for cost of new therapies that may be used by policy makers, drug manufacturers, health care providers and patients when therapy is first introduced in market

Glick HA, Doshi JA, Sonnad SS, Polsky D.
Economic Evaluation in Clinical Trials
Oxford: Oxford University Press, 2015