Designing Economic Evaluations in Clinical Trials

Statistical Methods in Health Economic Evaluations

Henry Glick

Good Value for the Cost

- Economic data collected as primary or secondary endpoints in randomized trials are commonly used in evaluation of "value for the cost" of medical therapies
  - Short-term economic impacts directly observed
  - Longer term impacts potentially projected by use of decision analysis
  - Reported results: point estimates and confidence intervals for estimates of:
    - Both incremental costs and outcomes
    - Comparison of costs and outcomes
  - Impact of sensitivity analysis judged by its impact on both the point estimates and the confidence intervals of the ratios

Example

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
Hospitalization Cost		
-50%	Dominates	Dom to 5300
+50%	Dominates	Dom to 8400
Drug Cost		
-50%	Dominates	Dom to 4850
+50%	Dominates	Dom to 8750
Discount rage		
0%	Dominates	Dom to 6350
7%	Dominates	Dom to 7000

Outline

- Steps in economic evaluation
- The gold standard and its tensions
- 8 strategic issues
  1) What medical service use should we collect?
  2) How should we value medical service use
  3) At what level should medical service use be aggregated?
  4) How should we interpret results from multinational (multicenter) studies?
  5) What sized sample should we study?
  6) How naturalistic should study design be?
  7) How should costs (QALYs) be analyzed?
  8) How should we report sampling uncertainty?

Steps in Economic Evaluation

Step 1: Quantify the 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. report an incremental cost effectiveness ratio (ICER) or probability of acceptability

$$ICER = \frac{(Cost_A - Cost_B)}{(Effects_A - Effects_B)}$$

- Potential hypothesis: Cost per quality-adjusted life year saved significantly less than $60,000

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

Design Issues Not Unique To Trials

- A number of design issues apply equally to economic evaluations that are incorporated within clinical trials and to other economic evaluations:
  - Type of analysis conducted (e.g. cost-benefit, cost-effectiveness, or cost minimization)
  - Types of costs included (e.g. direct medical, direct nonmedical, productivity, and intangible)
  - Study perspective
- These issues well addressed in literature

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 larger than sample size required for clinical questions
- Length of follow-up needed for economic questions may be longer than follow-up needed for clinical questions

Trade-off

- These trials may be only source of information needed for important early decisions about adoption and diffusion of therapy

TRADE-OFF: Ideal vs best feasible

Issue #1. 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

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

During Trial Design, Document Expected Service Use

- 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
  - Easy for judgments to be flawed
- 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% ($p = 0.006$)

Other Types of Costs?

- Other types of costs that sometimes are documented within economic evaluations include:
  - Time costs: Lost due to illness or to treatment
  - Intangible costs
- Types of costs that should be included in an analysis depend on:
  - What is affected by illness and its treatment
  - What is of interest to decision makers
    - e.g., National Institute for Clinical Excellence (U.K.) and Australian Pharmaceutical Benefits Scheme have indicated lack of interest in time costs

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

Specific Recommendations

- Identify common patterns of medical service use in centers that will participate in trials
  - Speak with experts in multiple centers
  - Focus groups, etc.
- Design case report forms to collect important, common medical service use
- Collect the services independent of reason for their use
- Pilot test forms (if appropriate, in multiple centers)
- Consider collecting costs other than medical service use

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

Valuing 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

Price Weights from Which Centers / Countries

- Countries/centers from which price weights are collected might be ones that:
  - Enroll large numbers of participants
  - Have readily available price weights
  - Represent spectrum of economic conditions
  - Have regulators that require a submission
  - Sponsors wish to make economic claims for

Pricing Out All Recorded Services

- Eventually, need to identify price weights for all medical services recorded in case report form
- Because collecting price weights for all services may be expensive, we commonly:
  - Collect price weights for service use that:
    - Occurs most frequently in trial
    - Is considered likely to be affected by intervention
    - Has particularly high or low costs
- Presuming we are using a reliable method for imputing price weights (e.g. DRG weights), better to sample a smaller number of price weights in more countries/centers than to sample a larger number of price weights in fewer countries/centers

Center/Country-Specific vs Averaged Price Weights

- Once we have price weights from a number of countries/ centers, how should they be used to construct the cost outcome of the trial?
  - Ideal: Because relative prices can affect quantities of services provided, where ever feasible, multiply country-specific price weights times country-specific counts of medical services
  - For countries for which price weights aren't available:
    - Should use (averages of) price weights from similar countries
    - Shouldn't average price weights for all countries, independent of economic conditions

Issue #3. At What Level Should Medical Service Use Be Aggregated?

Level of Aggregation?

- If we count medical service use and multiply it times a set of price weight estimates, at what level of aggregation should services be recorded?
  - e.g., for inpatient care, should we count:
    - Hospitalizations?
    - Days in hospital?
    - Days in hospital stratified by location in hospital?
    - Days in hospital stratified by location plus individual services provided during hospitalization?

Factors Affecting Level of Aggregation

- Do we expect intervention to affect:
  - Number of hospitalizations that occur
  - Length of stay of a hospitalization when it occurs
  - Intensity of medical services utilized during stay
- In making decisions about level of aggregation, consider likely difference more or less aggregated information will have on study result as well as cost of collecting more or less aggregated data
- Resulting decisions affect price weight estimates required for calculation of cost

Hospital Care Decisions

- Types of hospital services that are counted often depend on setting in which therapies under investigation are expected to be used
  - For therapies used predominantly in hospital settings: common to sum individual costs of a hospital stay
    - e.g., days in hospital, stratified by intensity of care, laboratory evaluations, procedures, and medications
  - For therapies used predominantly in outpatient settings: common to collect information about hospital diagnoses and length of stay

Hospital Care Valuation

- Hospitalizations can be valued by use of aggregate measures of hospital cost, such as diagnosis-related group (DRG) payments or an estimate of cost per day times number of days in hospital
  - When using cost per day, might use a single cost estimate from a single center to value all hospitalizations at all centers
  - Alternatively might use diagnosis-specific price weight estimates from each center that participated in study
- Most studies adopt a strategy that falls somewhere between these extremes

Outpatient Care Decisions

- At most aggregate level, outpatient care can be recorded as number of visits
- Alternatively, diagnostic tests, procedures, and treatments can be recorded as well
- U.S. Medical Expenditure Panel Survey* reported direct payments for ER visits based on services performed:
  - Average expenditure: $560
  - Average if no special services provided: $302
  - Average if 1+ nonsurgical services provided: $637
  - Average if surgical procedure provided: $904

* Medical Expenditure Panel Survey Statistical Briefs. #111: Expenses for a Hospital Emergency Room Visit, 2003

Concomitant Medications

- Common to be very precise when costing investigational medications
- Greater problems posed by costing out concomitant medications
  - Number of agents / routes of administation / dosages / # of doses
- In many studies, 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 to represent cost for all members of class

Issue #4. How Should We Interpret Results From Multinational (Multicenter) Trials?

The Problem

- Long-standing concern that pooled (i.e., average) economic results from multinational trials may not be reflective of results that would be observed in individual countries that participated in trial
- Similar issues arise for any subgroup of interest in the trial (e.g., more and less severely ill patients)

Common Sources For Concern

- Transnational differences in:
  - Morbidity/mortality patterns
  - Practice patterns (i.e., medical service use)
  - Absolute and relative prices for this service use (i.e., price weights)
- Thus decision makers may find it difficult to draw conclusions about value for cost for therapies evaluated in multinational trials

Bad Solutions

- Use either:
  - Trial-wide clinical results, trial-wide medical service use, and price weights from one country
    - e.g., to tailor the results to the U.S., just use U.S. price weights, and conduct the analysis as if all participants were treated in the U.S.
  - Trial-wide clinical results, medical service use, and price weights from one country, for example the U.S.
- Both ignore influence clinical and economic outcomes may have on each other
  - Costs affect practice which affects outcomes AND practice affects outcomes which affect costs

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 the trial to address these issues -- have made their way into the 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
  - Nonsignificant p-value for test of homogeneity combined with belief that test had sufficient power to rule out economically meaningful differences in costs indicates can't reject that pooled economic result from trial applies to all of the countries in trial
  - Significant p-value indicates we should not use pooled estimate to represent result for individual countries
    - Method is 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 the trial-wide sample into subsets VS
  - Whether they are likely to have arisen due to systematic differences between countries
- Borrows information from the mean estimate to add precision to the country-specific estimates
- Methods have potential added advantage of providing better estimates of the uncertainty surrounding pooled result than naive estimates of trial-wide result

Issue 4 Summary

- Lots of weak methods used for applying results of multinational trials to individual countries
- Better methods include homogeneity testing and multilevel random effects modeling

Issue #5. What Sized Sample Should We Study?

What Sized Sample?

- Sample size for cost-effectiveness analysis typically calculated so experiment's result will have a specified likelihood that we an be confident that 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

Other Sample Size Traditions

- Sample size approach described here comes out of frequentist statistical tradition
- Other approaches include:
  - Bayesian (O'Hagan and Stevens)
  - Value of information (Koerkamp et al.)
  - Opportunity cost (Gafni et al.)
  - Decision model (Willan and O'Brien)

Sample Size Formula, Continuous Variable

- 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:

$$N = \frac{\text{Error Rates}}{\text{Difference}} \frac{2\,(z_{\alpha}+z_{\beta})^2 \;\; \text{Variance}}{\Delta^2}$$

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 = standard deviation for NMB; and Δ = expected difference in NMB

Why Does it Look So Much More Complex?

- Because 1) difference being assessed is difference in NMB (WΔQ – ΔC) and 2) standard deviation of NMB is a complicated formula
- Data needed to calculate sample size include:
  - Difference in cost (ΔC)
  - SD, difference in cost
  - *Difference in effect (ΔQ)*
  - *SD, difference in effect*
  - Z_{α} and Z_{β}
  - Correlation of difference in cost and effect (ρ)
  - Willingness to pay

Full Sample Size Formula, Cost-Effectiveness

- Assuming equal standard deviations and sample sizes per group:

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

Error Rates: $(z_{\alpha}+z_{\beta})^2$

Variance, NMB: $\left(sd_c^2 + (W\ sd_q)^2 - (2\ W\ \rho\ sd_c\ sd_q)\right)$

Difference: $(W\Delta Q - \Delta C)^2$

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

Where to Obtain the Necessary Data?

- When therapies are already in use: Expected differences in outcomes and standard deviations can be derived from feasibility studies or from records of patients
- Simple correlation between observed costs and effects may be an adequate proxy for the measure of correlation used for estimating sample size
- For novel therapies, information may need to be generated by assumption
  - e.g., sd from usual care will apply to new therapy, etc.

"Typical" Sample Size Table, W

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

* $\Delta C=25$; $\Delta Q=0.01$; $sd_c=2500$; $sd_q=.03$; $\rho=-.05$; $\alpha=.05$; $1-\beta=.8$

Sample Size Can Increase with Increasing W

	Sample Size Per Group	
WTP	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

* $\Delta C=-100$; $\Delta Q=0.01$; $sd_c=5000$; $sd_q=.15$; $\rho=-0.05$; $\alpha=.05$; $1-\beta=.8$

Sample Size Not Necessarily Monotonic With W

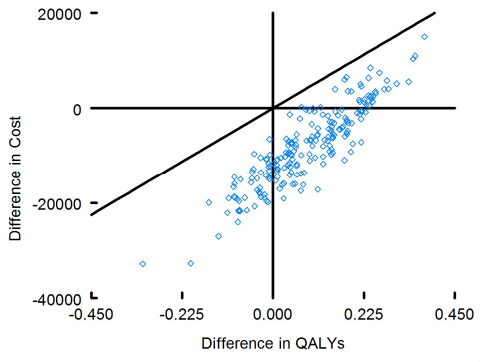
	Sample Size Per Group		
WTP	Exp 1	Exp 2	Exp 3 *
20,000	3466	387	178
30,000	1513	442	158
50,000	618	594	***151***
75,000	355	806	153
100,000	265	1011	156
150,000	200	1363	160

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

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

What Can We Conclude About ΔC, ΔQ, Value?



Willingness to Pay and Identification of an Appropriate Outcome Measure

- Sample size calculations require stipulation of willingness to pay to obtain a unit of outcome
- In many medical specialties, researchers use disease specific outcomes
- Yes, 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

Glick HA. Sample size and power for cost-effectiveness analysis (part 1). Pharmacoeconomics. 2011;29;189-98.

Glick HA. Sample size and power for cost-effectiveness analysis (part 2). The effect of maximum willingness to pay. Pharmacoeconomics. 2011;29:287-96.

Issue #6. How Naturalistic Should Study Design 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

#6a. Intention to Treat

- Economic questions relate to treatment decisions (e.g., whether to prescribe a therapy), not whether patient received prescribed drug 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

#6b. 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 was 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 changes in pattern of costs, discontinuation of these patients from economic study likely biases results
      - Continued follow-up reduces problems of non-ignorably missing data

#6c. 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
    - e.g., protocol may require substantial numbers 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)?"
- 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
  1) "Cadillac" costs may yield "Cadillac" outcomes
  2) If adjusting costs, would also need to adjust their effects on outcomes

Biases?

- Protocol-induced testing may bias testing cost to null
  - In truth, therapy might induce a difference in testing, but it can't be observed if protocol requires routine testing of all participants
- Protocol-induced testing may bias cost and outcome in an unknown direction
  - Trial's extra testing may lead to:
    - Detection and treatment of outcomes that wouldn't have been detected or treated in usual care
    - Earlier detection and treatment of problems when they are less severe and easier to treat
- Adjustment requires assumptions about what would or wouldn't have been detected in usual care

Specific Recommendations, Naturalism

- Use intention to treat sample for economic analysis
- Be aggressive in maintaining follow-up, including continuing to collect data on those who fail or switch therapy
- Use appropriate analytic methods to address missing data if and when they occur
- When possible, minimize effect of protocol on patient care

Issue #7. How Should Costs (QALYs) Be Analyzed?

Analysis of Costs (QALYs)?

- 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
- Statisticians' common reaction:
  - 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 are 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 transformed 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 (particularly if Ns are large)

Least Squares Regression Predicting Log of Cost

- Either OLS or identity/Gauss 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
  - Between group differences in variance/skewness/ kurtosis can cause a disconnect between magnitude and significance of differences in geometric means and differences in sample/arithmetic means
  - V/S/K differences can also affect percentage interpretation of coefficients
  - Retransformation problems (smearing estimators)

Sample Mean Vs Geometric Mean Disconnect

$ in 1000s				Sample Mean in 1000s	SD in 1000s	Geom Mean (1000s)
Obs 1	Obs 2	Obs 3	Obs 4			
S1: 50	55	65	70	**60**	9.1	**59**
S2: 30	45	75	90	60	27.4	55
S3: 10	35	85	110	**60**	45.6	**43**

- Sample means are all 60,000 and don't differ
  - OLS and GLM analyzing cost will indicate no difference in sample mean
- When variances differ, downwardly biased geometric means can differ
  - Log OLS will find a 16,000 difference in "means"
- What is commonly referred to as log OLS's "efficiency gain" can easily be quantification of bias (59 vs 43)

GLM Predicting Cost (Preferred)

- GLM with "appropriate" link and family
  - Log link / gamma family most typical in literature, but always using this combination is little different from always using OLS
- Advantages
  - Does not require (log) transformation and thus has no problems with retransformation
  - Relaxes normality and homoscedasticity assumptions
  - Consistent even if incorrect family is identified
  - Gains in precision due to having estimator that matches data generating function
  - Unaffected by differences in variance, skewness, or kurtosis

GLM Issues/Disadvantages

- Issues / Disadvantages
  - Can suffer substantial precision losses
  - Log link not necessarily 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
  - Link sometimes won't run with recommended family
  - Link sometimes won't run with any family
  - Model sometimes yields improbably large predictions
  - Estimation sometimes still requires 2-part models

Issue 7, GLM Summary

- Log/gamma not always preferred link/family
- Need to conduct diagnostic tests to identity appropriate link/family
- Establish criteria for choice of preferred link/family prior to unblinding data
  - Fact that one model gives a more favorable result should not be a reason for its adoption
- Report sensitivity of results to different link/family specifications

Bootstrapping the Analysis

- No matter what types of models one estimates, often bootstrap the entire analysis
  - Particularly important when estimating cost-effectiveness plane or when using a repeated measures design
    - e.g., estimating costs and QALY scores for 4 6-monthly periods, where reported SE is for per period differences, not for total difference
  - Also important for estimation of correlation of differences, which is used by all parametric methods for estimating sampling uncertainty for cost-effectiveness analysis

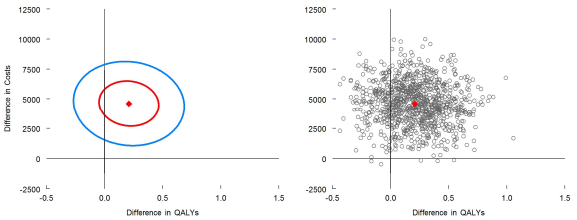
Issue #8. How Should We Report Sampling Uncertainty?

Two Most Frequently Published Uncertainty Graphs

- Cost-effectiveness plane
- Acceptability curve

- Other approaches not discussed today:
  - Net monetary benefit graph
  - Value of information graph

Cost-Effectiveness Plane



- Bivariate normal curves (Δc, SEc, Δq, SEq, ρ) (left)
- Bootstrap of patient level data (right) [Also can be derived from second order Monte Carlo simulation from decision analysis]

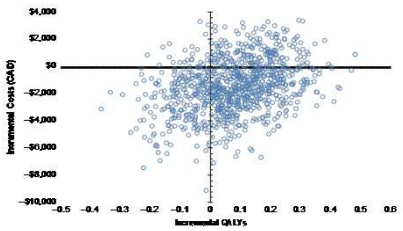
Information Derivable from Plane

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

Is CI for CER an Order Statistic?

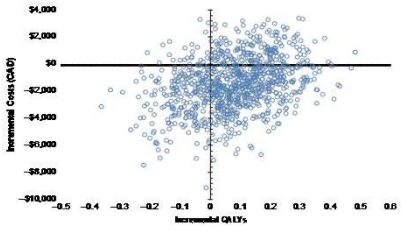
- Commonly CI for CER assumed to be an order statistic
  - Naïve ordering: order from lowest to highest ratio; identify ratios for the 2.5th and 97.5th ordered replicate
    - Works when **ALL** replicates on one side of Y axis
  - "Smart ordering": Order lexicographically (counter clockwise) first by 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 can be defined by lines through origin, but generally won'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

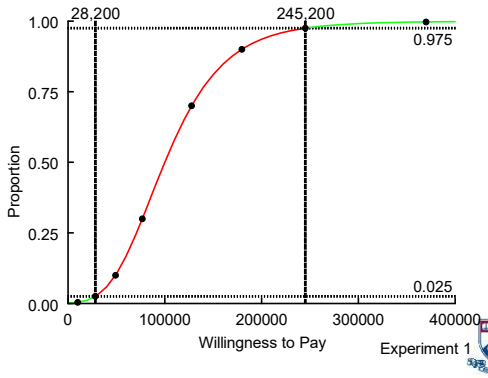
Cost-Effectiveness Plane



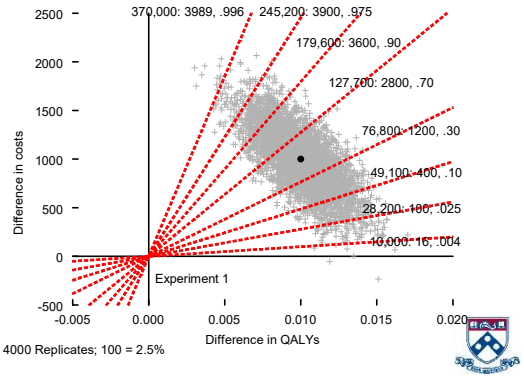
Reported cost difference: -1418, 95% CI -1540 to -1295
Reported QALY difference: 0.074, 95% CI, 0.066 to 0.082
Reported ICER -19511, 95% CI, -23815 to 2044

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

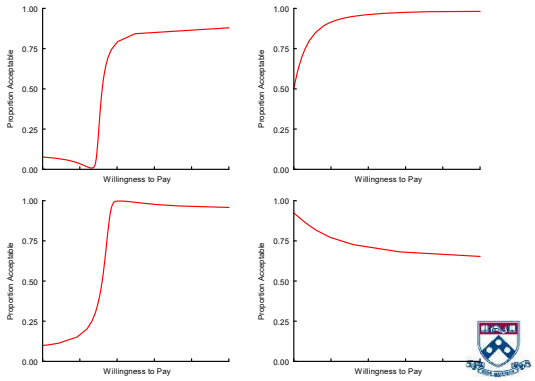
Acceptability Curve



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

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 they are analyzed appropriately, trial-based evaluations can 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