Designing Economic Evaluations in Clinical Trials

Statistical Methods in Health Economic Evaluations

ISPOR 8th Asia-Pacific Conference

September 9, 2018

Jalpa Doshi, Chee-Jen Chang, and Henry Glick

Good Value for the Cost

- Economic data collected as primary or secondary endpoints in randomized trials are commonly used in the evaluation of the 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:
    - Incremental costs and outcomes
    - Comparison of costs and effects
  - 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
- 4 strategic issues
  - What medical service use should we collect?
  - How should we value medical service use
  - How should we interpret results from multicenter studies?
  - What is the appropriate sample size?

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

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

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
  - 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?

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

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:
    - Use (averages of) price weights from similar countries
    - Should NOT average them and use results for all "like" services measured in trial

Issue #3. How Should We Interpret Results From Multicenter (Multinational) 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 and use costs derived from the subset of patients treated in the country
- 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		Country-Specific Costs and Effects
	Price weight	Country-Specific Costs	
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
H 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 believe 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 or 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
  - These methods have the potential added advantage of providing better estimates of the uncertainty surrounding the pooled result than naive estimates of the trial-wide result

Issue #4. 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:

$$N = \frac{2\,(z_{\alpha}+z_{\beta})^2 \;\; sd^2}{\Delta^2}$$

Error Rates Variance

Difference

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?

- 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 Sample Size Formula, Cost-Effectiveness

Error Rates

Variance, NMB

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

Difference

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

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

* 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

	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

* 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

	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

Willingness to Pay and Identification of an Appropriate Outcome Measure

- Sample size calculations require us to stipulate what we are willing to pay to obtain a unit of outcome
- In many medical specialties, researchers use disease specific outcomes
- While we can calculate a cost-effectiveness ratio for any outcome we want (e.g., cost/case detected or cost/additional abstinence day), to be convincing that a new, more costly and more effective therapy is good value, the 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.

Summary

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