Designing Economic Evaluations in Clinical Trials

ISPOR AUSTRALIA

December 2, 2020

Henry Glick
glickh@mail.med.upenn.edu

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 point estimates and confidence intervals of 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 rate		
0%	Dominates	Dom to 6350
7%	Dominates	Dom to 7000

Outline

- Steps in economic evaluation of clinical trials
- The gold standard and its tensions
- 3 strategic issues
  1) How naturalistic should study design be?
  2) How should costs (QALYs) be analyzed?
  3) How should we interpret results from multinational (multicenter) studies?

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
  – Compare 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. 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

#1a. 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

#1b. 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

#1c. 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 would require 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 #2. How Should Costs (QALYs) Be Analyzed?

Analysis of Costs (QALYs)?

- Cost data typically right skewed with long, heavy, right tails (QALYs typically left skewed)
  - Can also have extreme highliers, but statistical problems often due as much to heaviness of tails as they are 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 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

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 sample/arithmetic 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 Means Vs Geometric Means

$ 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)

Method 3, 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
- GLM 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 2, 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 #3. 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 represent 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., unit costs/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 Australia, use Australian price weights and conduct the analysis as if all participants were treated in Australia
  - Trial-wide clinical results and medical service use and price weights from one country, for example, Australia
- 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
† 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 tries to assess 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
- Method has potential added advantage of providing better estimates of the uncertainty surrounding pooled result than naive estimates of trial-wide result

Issue 3 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

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