

## Designing Economic Evaluations in Clinical Trials

### Statistical Considerations in Health Economic Evaluations

ISPOR 18th Annual International Meeting

May 19, 2013

Henry Glick and Jalpa Doshi  
[www.uphs.upenn.edu/dgimhsr](http://www.uphs.upenn.edu/dgimhsr)



---

---

---

---

---

---

---

---

### Good Value for Cost

- Economic data collected as secondary (or primary??) endpoint in randomized trials commonly used in evaluation of value for 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



---

---

---

---

---

---

---

---

### 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
- Gold standard and its tensions
- 5 Strategic issues



---

---

---

---

---

---

---

---

### 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



---

---

---

---

---

---

---

---

### Design Issues Not Unique To Trials

- A number of design issues apply equally to economic evaluations in clinical trials and to other economic evaluations:
  - Type of analysis that will be conducted
  - Types of costs that will be included
  - Study perspective
- 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 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

- 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



---

---

---

---

---

---

---

---

### 5 Strategic Issues

- What medical service use should we collect?
- At What Level Should Medical Service Use Be Aggregated?
- Sources of Price Weights?
- How naturalistic should study design be?
- How should we interpret results from multicenter / multinational trials?



---

---

---

---

---

---

---

---

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

- Real/perceived problem
  - Don't have sufficient resources to track all medical service use
  - Don't always expect to affect all medical service use



---

---

---

---

---

---

---

---

### Limited Data Collection Resources

- Availability of administrative data may reduce costs of tracking all medical service use
- 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
  - E.g., collecting 6700 blood gas tests that accounted for 1.8% of lab costs vs 420 cardiac studies that represented 4.3%



---

---

---

---

---

---

---

---

### 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 of such judgments
- 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% ( $p = 0.006$ )



---

---

---

---

---

---

---

---

### Blinded Vs Unblinded Studies

- Potential biases more of a problem in unblinded studies, but need not "balance out" in double-blinded studies



---

---

---

---

---

---

---

---

### 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 they have little 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 their use
- Decision to collect service use independent of its reason does not preclude ADDITIONAL analyses testing whether designated "disease-related" costs differ



---

---

---

---

---

---

---

---

### Specific Recommendations, Which Services

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



---

---

---

---

---

---

---

---

### Issue #2: At What Level Should Medical Service Use Be Aggregated?

- 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 the 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 study medications
- Greater problems posed by costing out concomitant medications
  - Number of agents / routes of administration / 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 #3. Sources of Price Weights?

- Assuming we have collected information about medical service use, we commonly translate this use into a cost by identifying price weights and multiplying medical service use by price weights for those services



---

---

---

---

---

---

---

---

### National Tariffs

- Wherever available, national tariffs are a commonly used and well accepted source of price weights for costing out medical service use
  - Diagnosis-related group (DRG) payments in Australia or United States
  - Health resource group (HRG) payments in Great Britain



---

---

---

---

---

---

---

---

### Advantages of National Tariffs

- Usually provide price weights for most if not all of services that are measured in study
- Are inexpensive to obtain
- Within individual countries, they are usually developed by use of a common methodology
- Make it difficult for investigators to pick and choose among price weights to make an intervention look more or less favorable than it should
- Fact that tariffs represent what is spent by governments may also be considered an advantage, particularly by governmental decision-making bodies



---

---

---

---

---

---

---

---

### Center-Specific Price Weights

- Center-specific price weights also a commonly used source of price weights



---

---

---

---

---

---

---

---

### Advantages of Center-Specific Price Weights

- If price weights come from centers in which study was conducted, provide a more accurate estimate of cost that was actually incurred within study
- Cost and medical service use can have important interactions that can lead to biased results
  - Because efficient producers use greater amounts of relatively less costly services and smaller amounts of relatively more costly services, use of a single set of price weights for all providers tends to overstate cost
  - Principle holds for studies performed in a single center, in multiple centers in a single country, or in multiple centers in multiple countries



---

---

---

---

---

---

---

---

### Which Price Weights?

- Most appropriate source of price weights depends in part on whether question we are asking is more national or center specific, but question being asked cannot be our only consideration
  - e.g., if our goal is to make national resource allocation decisions, we might conclude that use of a set of national price weights would yield data appropriate for making such decisions
  - But, representativeness of resulting cost estimates depends on representativeness of both medical service use and price weights
  - Once one factor is unrepresentative, not clear that ensuring that other factor is representative necessarily yields best estimates



---

---

---

---

---

---

---

---

### Sources of Price Weights

- Sources of price weights differ by country and by medical service
- For inpatient services, many countries now use patient classification systems like DRGs or HRGs
  - Some provide measures of both relative cost of a hospital stay – a relative value or a relative weight – and reimbursement/cost for a stay
  - Others provide measures of relative cost, and we must then independently identify cost per relative weight
- For studies that enroll participants in developing world, The WHO has developed estimates of price weights for inpatient care in at least 49 countries



---

---

---

---

---

---

---

---

### Fee Schedules

- UK:
  - [http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH\\_123459](http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_123459)
- US
  - [www.cms.hhs.gov/FeeScheduleGenInfo/](http://www.cms.hhs.gov/FeeScheduleGenInfo/)
  - [www.cms.hhs.gov/PropMedicareFeeSvcPmtGen/](http://www.cms.hhs.gov/PropMedicareFeeSvcPmtGen/)
  - <http://www.pbm.va.gov/DrugPharmaceuticalPrices.aspx>
- Australia (Round 14 cost reports)
  - [http://www.health.gov.au/internet/main/publishing.nsf/Content/ADF42B9AC16D4017CA257864000FBD0E/\\$File/R14CWNatEst\\_v6x.xls](http://www.health.gov.au/internet/main/publishing.nsf/Content/ADF42B9AC16D4017CA257864000FBD0E/$File/R14CWNatEst_v6x.xls)
- Fee Schedules' Web Addresses Change Routinely



---

---

---

---

---

---

---

---

### Center/Country-Specific vs Averaged Price Weights

- Once we have a number of different sets of price weights (e.g., weights from multiple countries that participated in trial), how should they be used to construct cost outcome of trial?



---

---

---

---

---

---

---

---

### Center/Country-Specific vs Averaged Price Weights (2)

- Ideal: Because relative prices can affect quantities of services provided, whenever feasible, multiply country-specific price weights times times country-specific counts of medical services
- For centers/countries for which price weights aren't available:
  - Use (averages of) price weights from similar centers/countries
  - e.g., in a trial that enrolls patients in Western and Eastern Europe and Latin America, we might average price weights from other Western European countries to value service use in Germany, but wouldn't want to use this average for Eastern Europe or Latin America



---

---

---

---

---

---

---

---

Center/Country-Specific vs Averaged Price Weights (3)

- Corollary: If we have a set of price weights for each center/country that participated in trial, we should not average them and use this average for all services measured in trial
  - Most common reasons suggested for such a strategy are:
    - Reducing variability in price weights reduces variability in estimated costs
    - Average set of price weights may be more representative



---

---

---

---

---

---

---

---

Center/Country-Specific vs Averaged Price Weights (4)

- However:
  - Empirically, use of a single set of price weights need not reduce variance
  - If substitution effects are strong, strategy may introduce bias in estimates of cost differences
  - Why is it more “representative” to use a set of price weights that no one faces?



---

---

---

---

---

---

---

---

Issue # 4. How Naturalistic Should Study Design Be?

- 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



---

---

---

---

---

---

---

---

#### #4a. 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



---

---

---

---

---

---

---

---

#### #4b. 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



---

---

---

---

---

---

---

---

#### #4c. 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 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)?”
- 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 need 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 testing in usual care, but it can't be observed if everyone routinely receives a test
- Protocol induced testing may bias treatment 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 #5. How Should We Interpret Results From Multicenter (Multinational) Trials?

- Problem:
  - There has been growing 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 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); and absolute and relative prices for this service use (i.e., price weights)
- Thus decision makers may find it difficult to draw conclusions about value of therapies that were evaluated in multinational trials



---

---

---

---

---

---

---

---



### Bad Solutions

- Use trial-wide clinical results, trial-wide medical service use, and price weights from one 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		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

\* Wilke 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 Research Task 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 they are analyzed appropriately, these evaluations can provide data about uncertainties related to assessment of value of new therapies that may be used by policy makers, drug manufacturers, health care providers and patients when therapy is first introduced in market



---

---

---

---

---

---

---

---