### Analysis of Costs Using Patient Level Data from Randomized Designs

Henry Glick, PhD and Jalpa Doshi, PhD  
University of Pennsylvania School of Medicine  
Leonard Davis Institute of Health Economics  
Astra Zeneca Methodology Seminar  
May 31, 2005

### Purpose of Seminar

- To familiarize audience with issues and common mistakes in the analysis of costs using patient-level data from randomized designs  
  - systematic review of 115 studies on randomized-trial based economic evaluations published in 2003  
- To illustrate the common mistakes and preferred analytic approaches for  
  - analysis of costs  
  - handling of incomplete cost data  
  - comparison of costs and effects with assessment of stochastic uncertainty

### Outline

<table>
<thead>
<tr>
<th>Background</th>
</tr>
</thead>
<tbody>
<tr>
<td>Literature Review: Objectives, Methods, and General Findings</td>
</tr>
<tr>
<td>Analysis of Costs</td>
</tr>
<tr>
<td>Handling of Incomplete Cost Data</td>
</tr>
<tr>
<td>Comparison of Costs and Effects &amp; Stochastic Uncertainty</td>
</tr>
<tr>
<td>Conclusion</td>
</tr>
</tbody>
</table>

### Background

- The number of randomized-trial based economic evaluations has increased considerably over the past few years  
  - 45 published in 1995 (Barber & Thompson 1997) vs.  
  - 115 published in 2003 (Doshi, Glick, and Polsky, 2005)  
- Serious issues with methodology and reporting had been identified in such studies (Barber and Thompson 1997)
Background

• In the past decade, the field has matured, methods have advanced, and consensus regarding appropriate statistical methods has emerged in several areas

• ISPOR RCT-CEA Taskforce (2005)
  – "GOOD RESEARCH PRACTICES FOR COST-EFFECTIVENESS ANALYSIS ALONGSIDE CLINICAL TRIALS: THE ISPOR RCT-CEA TASK FORCE REPORT"

• Use of good research practices will enhance credibility and usefulness of these studies to decision-makers

Literature Review: Objective

• Our objective was to assess the use of good research practices in published randomized trial-based economic evaluations

• Practice areas assessed were:
  (1) Analysis of costs
  (2) Handling of incomplete cost data
  (3) Comparison of costs and effects with assessment of stochastic uncertainty

Literature Review: Methods

• Medline search (Sep 2004) for all studies which included terms in the title, abstract, or MeSH headings related to
  – costs (e.g. "cost(s)", "economic evaluation(s)", or "health economic(s)")
  – clinical trials (e.g."trial(s)" or "randomized controlled trials")

• Search was limited to publications in English, involving human subjects, and published during 2003

• Exclusion criteria
  – Study was not a randomized trial
  – Study did not collect or analyze patient specific costs
  – Study applied clinical trial data in a decision analytic model

• 650 abstracts reviewed of all articles identified by Medline search
  – 162 full-text articles read
    • 115 studies met selection criteria for review

Literature Review: General Findings

• 83% of the articles listed the economic analysis as one of their primary objectives

• Majority of the trials were either conducted in the U.S. (23%), U.K. (23%), or were multinational (18%)

• 44% of the studies had drugs as one of their treatment comparators

• Trials addressed a variety of disease areas such as cardiovascular (22%), musculoskeletal (10%), cancer (10%), and mental health (9%) conditions
Analysis of Costs

Preferred Analytic Approaches & Findings on Common Mistakes Identified in Review

Cost Data 101

- Common feature of cost data is right-skewness (i.e., long, heavy, right tails)
- Data tend to be skewed because:
  - Can not have negative costs
  - Most severe cases may require substantially more services than less severe cases
  - Certain events, which can be very expensive, occur in a relatively small number of patients
    - A minority of patients are responsible for a high proportion of health care costs

Which Statistic Should be Used to Summarize Cost Data?

- What statistical formulation best characterizes the policy or decision problem of interest?
- For cost-effectiveness analysis: ∆C (arithmetic mean)
  - Social perspective: In economic theory, arithmetic mean costs and differences in arithmetic mean costs yield social efficiency
  - Budgetary perspective: arithmetic mean costs are a better summary of budgetary impact than median costs or log of costs (because n x mean = total)
- Cost-effectiveness ratios (∆C/∆E) and NMB ([rc ∆E] - ∆C) require an estimate of ∆C where:
  \[ \Delta C = C_t - C_s \]
  \[ \Delta E = E_t - E_s \]
Findings: Cost Statistic Reported Across Treatment Arms

- 82% used arithmetic mean costs only
- 12% used arithmetic mean and median costs
- 4% used median costs only
- 2% reported other methods

N=115

How Should Cost Data be Summarized?

- Arithmetic means and their difference
- Measures of variability and precision (e.g. std deviation)
- Quantiles such as 5%, 10%, 50%.....
- An indication of whether or not the difference in arithmetic means:
  - Occurred by chance (p-value or confidence intervals)
  - Is economically meaningful

Findings: Was Statistical Comparison of Treatment Arms Made?

- 82% of studies performed statistical comparison
- 18% did not perform statistical comparison

N=115
Findings: Was Statistical Comparison of Treatment Arms Made?

<table>
<thead>
<tr>
<th>% of Studies</th>
<th>N=115</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>82%</td>
</tr>
<tr>
<td>No</td>
<td>18%</td>
</tr>
</tbody>
</table>

1 out of 5 studies did not conduct a statistical test

Univariate Statistical Tests

- Usual starting point: T-tests and one way ANOVA
  - Used to test for differences in means in total costs, QALYS, etc.
  - Makes assumption that the costs are normally distributed
  - While the normality assumption is routinely violated for cost data, in large samples these tests have been shown to be robust to violations of this assumption
- Because of distributional problems related to evaluating the arithmetic mean, there has been a growing use of alternative tests

Common Mistakes (I)

- Adoption of nonparametric tests of other characteristics of the distribution that are not as affected by the nonnormality of the distribution
  - Wilcoxon rank-sum test for difference in medians
  - Kolmogorov-Smirnov test for difference in cumulative distribution function

Common Mistakes (II)

- Transformation of costs to approximate a normal distribution
  - Log transformation or square root transformation
    - For the log transformation, one is making estimates and inferences about the ratio of the treatment group means
  - For economic analysis, the outcome of interest is the difference in untransformed costs (e.g., "Congress does not appropriate log dollars")
    - Need to retransform log costs to original scale
    - Retransformation issues: Simple exponentiation of log costs results in geometric mean (not arithmetic mean). Need to apply appropriate smearing factors to obtain unbiased estimates
Common Mistakes (II)

- “There is a very real danger that the log scale results may provide a very misleading, incomplete, and biased estimate of the impact of covariates on the untransformed scale, which is usually the scale of ultimate interest” (Manning, 1998)

- “This issue of retransformation...is not unique to the case of a logged dependent variable. Any power transformation of y will raise this issue”

Preferred Statistical Test

- Adopt tests of arithmetic means that avoid parametric assumptions (most recent development)
  - Non-parametric Bootstrap
    - Estimates the distribution of the observed difference in mean costs
    - Yields a test of how likely it is that 0 is included in this distribution (by evaluating the probability that the observed difference in means is significantly different from 0)

Summary of Univariate Statistical Approach

- If arithmetic means are the most meaningful summary statistic of costs, one should test for significant differences in arithmetic mean costs
  - Parametric test of means
  - Non-parametric test of means
    - Bootstrap methods

- Do not use nonparametric tests such as Wilcoxon rank sum and KS tests because we want to estimate the magnitude of the difference in arithmetic means and test whether that difference was observed by chance

- Tests of transformed variables such as the log or square root pose similar problems

- Bootstrap and t-tests on untransformed costs easily provide confidence intervals (CI)
  - Non-parametric tests cannot provide CIs
  - Tests of log transformed costs can provide CIs but not easy to calculate

Findings: If YES, what type of statistical test was conducted?

- Non-parametric bootstrapping
- t-test on untransformed mean costs
- t-test on transformed mean costs
- Mann-Whitney test on mean costs
- Non-parametric test of median/distribution of costs
- Not clear

N=94
Findings: If YES, what type of statistical test was conducted?

<table>
<thead>
<tr>
<th></th>
<th>0%</th>
<th>20%</th>
<th>40%</th>
<th>60%</th>
<th>80%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-test on untransformed mean costs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8%</td>
<td>6%</td>
</tr>
<tr>
<td>t-test on transformed mean costs</td>
<td>9%</td>
<td>6%</td>
<td></td>
<td></td>
<td></td>
<td>46%</td>
</tr>
<tr>
<td>Mann-Whitney test on mean costs</td>
<td>4%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21%</td>
</tr>
<tr>
<td>Non-parametric test of median/distribution of costs</td>
<td>0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-parametric bootstrapping</td>
<td>6%</td>
<td>0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 out of 4 studies did not conduct an appropriate statistical test

Illustrative Examples

• To understand how different statistical tests can lead to different inferences

Illustrative Example 1

<table>
<thead>
<tr>
<th></th>
<th>Placebo</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ($CDM)</td>
<td>20,287</td>
<td>25,185</td>
</tr>
<tr>
<td>SD</td>
<td>22,542</td>
<td>22,619</td>
</tr>
<tr>
<td>Distribution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5%</td>
<td>4,506</td>
<td>10,490</td>
</tr>
<tr>
<td>25%</td>
<td>9,691</td>
<td>13,765</td>
</tr>
<tr>
<td>50%</td>
<td>13,773</td>
<td>18,834</td>
</tr>
<tr>
<td>75%</td>
<td>23,044</td>
<td>31,069</td>
</tr>
<tr>
<td>95%</td>
<td>53,728</td>
<td>51,771</td>
</tr>
</tbody>
</table>

Results from Different Statistical Tests Applied to Same Dataset

<table>
<thead>
<tr>
<th>Test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-test of mean difference</td>
<td>0.16</td>
</tr>
<tr>
<td>Non-parametric Bootstrap</td>
<td>0.09</td>
</tr>
<tr>
<td>T-test, log of cost difference</td>
<td>0.001</td>
</tr>
<tr>
<td>Wilcoxon rank-sum (Mann-Whitney)</td>
<td>0.0002</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Why Would Different Statistical Tests Lead To Different Inferences?

- The tests are evaluating differences in different statistics
  - T-test of untransformed costs indicates one cannot infer that the arithmetic means are different
  - Bootstrap leads to same inference as t-test and does not make the normality assumption
  - T-test of log costs indicates one can infer that the mean of the logs are different, and thus the geometric means of cost are different
  - Wilcoxon rank-sum test indicates one can infer that the medians are different
  - Kolmogorov-Smirnov test indicates one can infer that the distributions are different
- When distributions are skewed, means and medians can be measuring very different things

Illustrative Example 2

<table>
<thead>
<tr>
<th>Placebo</th>
<th>Treatment</th>
<th>Diff</th>
<th>P-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arithmetic Mean ($)</td>
<td>20000</td>
<td>20000</td>
<td>0.00</td>
</tr>
<tr>
<td>SD</td>
<td>16007</td>
<td>9599</td>
<td></td>
</tr>
<tr>
<td>Mean of Log Costs</td>
<td>9.6564</td>
<td>9.7957</td>
<td>0.13930</td>
</tr>
<tr>
<td>Geometric Mean</td>
<td>15620.69</td>
<td>17955.44</td>
<td>2334.75</td>
</tr>
</tbody>
</table>

* t-test and bootstrap resulted in similar p-values

Multivariate Analysis of Costs

- Even if treatment is assigned in a randomized setting use of multivariable analysis has benefits because:
  - Improves the power for tests of differences between groups (by explaining variation due to other causes)
  - Practice pattern differences by provider, center, or country may have a large influence on costs and the randomization may not account for all imbalances between groups
  - Variations in economic conditions often not controlled for in a randomized trial, therefore multivariable analysis takes on added importance
  - Additional advantage: Helps explain what is observed (e.g., coefficients for other variables should make sense economically)

Findings: Was Multivariate Adjustment of Incremental Costs Made?

- Yes | No
Multivariable Techniques for Analysis Of Costs

- Most common techniques
  - Ordinary least squares regression (OLS)
  - Ordinary least squares regression predicting the log transformation of costs (log OLS)
  - Generalized Linear Models (GLM)

- Different multivariable models make different assumptions
  - When assumptions are met, coefficient estimates will have many desirable properties
  - With cost analysis, assumptions are often violated, which may produce misleading or problematic coefficient estimates
    - Bias (consistency)
    - Efficiency (precision)

Comments On OLS

- Coefficient on treatment indicator produces an estimate of $\Delta C$

- Assumption of "error term is normally distributed" is generally violated with cost data

- Advantages
  - Easy
  - No retransformation problem (faced with log OLS)
  - Marginal/Incremental effects easy to calculate

- Disadvantages
  - Not robust in small to medium sized data sets and in large datasets with extreme observations
  - Can produce predictions with negative costs

Comments On Log OLS Model

- Coefficient on treatment indicator = % difference in mean costs between treatment groups

- Log OLS predicts log of costs, not costs
  - Retransformation is not trivial
  - Appropriate smearing correction is not applied (particularly in face of heteroskedasticity)

- Advantages
  - Widely known transformation for costs
  - Reduces robustness problem seen with OLS
  - Improves efficiency

- Disadvantages
  - Coefficients not directly interpretable
  - Retransformation problem could lead to bias
  - Not easy to implement

Comments on GLM Models

- These models have the advantages of the log models, but $\Delta C$ is estimated directly so it does not require any smearing correction

- To build them, one must identify the following:
  - "link" (e.g. identity, log, square root, etc.)
  - "family" (e.g. gaussian, gamma, poisson)

- Advantages
  - No retransformation problems of log OLS models
  - Consistent even if not the correct family distribution

- Disadvantages
  - Can suffer substantial precision losses if family is misspecified or heavy-tailed (log) error term
Multivariable Analysis

- The underlying distribution of costs should be carefully assessed to determine the most appropriate approach to conduct statistical inference on the costs between treatment groups
- No single model performs best in every situation
- May be helpful to report the sensitivity of one’s results to different multivariate approaches

Findings: If YES, what type of multivariate model was estimated?

- OLS
- log OLS without smearing retransformation
- log OLS with smearing retransformation
- GLM
- Other

Analysis of Costs: Summary

- While most studies presented arithmetic mean costs, more than 1 in 3 studies either did not conduct a statistical test or used an inappropriate test
- About 1 in 10 studies estimated incremental costs in a multivariate framework and most used OLS
  - Most studies did not report/assess distributional assumptions
  - Most studies did not test sensitivity of results to alternative multivariate techniques
  - No study used the GLM technique
# Handling of Incomplete Cost Data

- **Preferred Analytic Approaches**
- **Findings on Common Mistakes**
  Identified in Review

---

## Censored Data 101

- As economic data are increasingly collected alongside clinical trials attention has turned to the issue of censored cost data.
- Right censoring occurs whenever some individuals are not observed for the full duration of interest which results in information being incomplete for these patients.
  - Patients who are lost to follow-up, drop out of the study or are observed until the end of the study period without having reached the event of interest are said to be right censored.
- Note: Missing data can also be due to item-level missingness.
  - Multiple-imputation approach preferred method.

---

## Degree of Missingness

- No clear rule of thumb on what degree of missing data is problematic and requires adjustment.
- "Ignoring small amounts of missing data is acceptable if a reasonable case can be made that doing so is unlikely to bias treatment group comparisons" – ISPOR RCT CEA Taskforce.

## Mechanism of Missingness

- **Missing completely at random (MCAR)**
  - When outcome data are MCAR, the probability they are missing is unrelated to the value of the outcome or to any other variables that have been or could be measured.
- **Missing at random (MAR)**
  - When outcome data are MAR, the probability they are missing is unrelated to the value of the outcome after one has controlled for other variables in the analysis (MCAR is a special type of MAR).
- **Not missing at random (NMAR) or informatively (nonignorably) missing**
  - When outcome data are informatively missing, the probability they are missing is related to the value of the outcome even after one has controlled for other variables in the analysis.


Diagnosing Mechanism of Missingness

- No direct test of mechanism is available - in part because once participants have been censored, their costs are not observed

- Indirect tests include:
  - Comparison of demographic, clinical, and economic characteristics of participants who did and did not have censored cost data by the end of the study period
  - Comparison of the costs per study visit for participants who did and those who did not have censored cost data by the end of the study period
    - e.g., comparison of costs and QALYs at the 3 month visit between patients who were never censored and those who were censored at 6, 12, or 18 months

Common Mistakes

- Prevalent use of two “naive” estimators in the literature
  - Uncensored-cases estimator (Complete-case analysis)
  - Full-sample estimator (Average over all sample patients)

- Uncensored-cases estimator only uses the uncensored cases in the estimation of mean cost
  - Biased toward the costs of the patients with shorter survival times because larger survival times are more likely to be censored
  - Reduces power to test hypotheses

- Full-sample estimator uses all cases but does not differentiate between censored and uncensored observations
  - Always biased downward because the costs incurred after censoring times are not accounted for

Techniques to Handle Censored Costs

- Lin et al. 1997
- Carides et al. 2000
- Bang and Tsiatis 2000
- Lin 2000a & Lin 2000b
- Zhao and Tian 2001
- Jain and Strawderman 2002

- Relative advantages of some of these methods have been evaluated in
  - Raikou and McGuire 2004
  - O’Hagan and Stevens 2004

- These methods have been shown to perform better than “naive” methods

Findings: Was Cost Data Incomplete?

- 58% of studies indicate yes
- 23% indicate no
- 19% not reported

N=115
Findings: Was Cost Data Incomplete?

- Yes: 58%
- No: 23%
- Not Reported: 19% (N=115)

Findings: If YES, how was incomplete cost data handled?

- Complete case analysis: 48%
- Average costs over all patients: 10%
- Last observation carry forward: 6%
- Imputation using group-specific mean or median: 6%
- Imputation using regression: 7%
- Lin (1997) method: 1%
- Carides (2000) method: 1%
- Not clear: 19% (N=67)

Only 2 studies used a published statistical method for censored costs.

Findings: If YES, how was incomplete cost data handled?

Handling of Incomplete Cost Data:

- Summary
  - 1 in 5 studies did not even report whether cost data were complete or not.
  - Of those reporting, over three-quarters had incomplete cost data.
  - Most studies used "naive" methods to handle incomplete cost data and may have resulted in biased or inefficient estimates.
References: Analysis of Cost Data


References: Censored Cost Data

Findings: Economic Evaluation was Primary/Secondary Objective

- 83% of studies had economic evaluation as a primary objective.
- 17% of studies had economic evaluation as a secondary objective.

Findings: Country Setting

- 18% of studies were multinational.
- 23% of studies were conducted in the U.S.
- 23% of studies were conducted in the U.K.
- 8% of studies were conducted in Canada.
- 7% of studies were conducted in the Netherlands.
- 21% of studies were in other countries.

Findings: Treatment Comparators

- 44% of studies compared drugs.
- 10% of studies compared surgery.
- 4% of studies compared medical devices.
- 3% of studies compared diagnostic tests.
- 38% of studies compared other treatments.

Findings: Major Disease Area

- 22% of studies focused on cardiovascular conditions.
- 10% of studies focused on musculoskeletal conditions.
- 10% of studies focused on cancer.
- 9% of studies focused on mental health.
- 7% of studies focused on respiratory conditions.
- 42% of studies focused on other diseases.