Analysis of Costs Using Patient Level Data from Randomized Designs

Contributed Workshop (W10)

Jalpa Doshi, PhD, Henry Glick, PhD, Daniel Polsky, PhD
University of Pennsylvania School of Medicine
Leonard Davis Institute of Health Economics
Funding Source: NIAAA & NIDA
ISPOR 10th Annual International Meeting
May 16, 2005

Purpose of Workshop

• To familiarize participants 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

• Background
• Literature Review Objectives and Methods
• General Findings
• Analysis of Costs
• Handling of Incomplete Cost Data
• Comparison of Costs and Effects with Assessment of Stochastic Uncertainty

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)
• However, serious issues with methodology and reporting have been identified in such studies (Barber and Thompson 1997)

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, or

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

Literature Review: Findings

General Study Information

Findings: Economic Evaluation was Primary/Secondary Objective

N=115

Findings: Country Setting

N=115

Findings: Treatment Comparators

N=115

Findings: Major Disease Area

N=115
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: \( \Delta C \) (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 \( n \times \text{mean} = \text{total} \))
• Cost-effectiveness ratios (\( \Delta C/\Delta E \)) and NMB ([rc \( \Delta E \) - \( \Delta C \)) require an estimate of \( \Delta C \) where:
  \( \Delta C = C_t - C_s \)
  \( \Delta E = E_t - E_s \)

Findings: Cost Statistic Reported Across Treatment Arms

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
  – Is economically meaningful
Findings: Was Statistical Comparison of Treatment Arms Made?

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

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

<table>
<thead>
<tr>
<th>Non-parametric bootstrapping</th>
<th>T-test on untransformed mean costs</th>
<th>T-test on transformed mean costs</th>
<th>Mann-Whitney test on mean costs</th>
<th>Non-parametric test of median/distribution of costs</th>
<th>Not clear</th>
</tr>
</thead>
<tbody>
<tr>
<td>21%</td>
<td>46%</td>
<td>6%</td>
<td>8%</td>
<td>9%</td>
<td>0%</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>

Illustrative Example 1

<table>
<thead>
<tr>
<th></th>
<th>Placebo</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ($)</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>

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></th>
<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>
<td>1.00</td>
</tr>
<tr>
<td>SD</td>
<td>16007</td>
<td>9599</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean of Log Costs</td>
<td>9.6564</td>
<td>9.7957</td>
<td>0.13930</td>
<td>0.008</td>
</tr>
<tr>
<td>Geometric Mean</td>
<td>15620.69</td>
<td>17965.44</td>
<td>2334.75</td>
<td>0.007</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: 91%
- No: 9%

Multivariable Techniques Used for the 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)

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

- OLS: 70%
- log OLS without retransformation: 10%
- log OLS with smearing retransformation: 10%
- GLM: 0%
- Other: 20%

Multivariable Analysis

- 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)
  - 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
  - May be helpful to report the sensitivity of one's results to different multivariate approaches

Analysis of Costs: Summary

- Most studies presented arithmetic mean costs; however, 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
  - No studies tested 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 the accommodation of censoring is becoming increasingly important within this context
  - Only recently the 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

- Two 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 “naïve” methods.

Findings: Was Cost Data Incomplete?

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

- 1 in 5 studies did not even report whether cost data were complete or not
- Of those reporting, almost three-quarters had incomplete cost data
- Most studies used “naïve” methods to handle incomplete cost data and may have resulted in biased or inefficient estimates

References: Analysis of Cost Data


References: Censored Cost Data