Analysis of Costs Using Patient Level Data from Randomized Trials

Jalpa Doshi, PhD, Henry Glick, PhD, Daniel Polsky, PhD

University of Pennsylvania School of Medicine
Leonard Davis Institute of Health Economics

Society for Clinical Trials
27th Annual Meeting
May 24, 2006

Outline

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

Background

- The number of randomized-trial based economic evaluations has increased considerably over the past few years
- However, serious issues with methodology and reporting have 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) Comparison of costs and effects with assessment of sampling uncertainty
  - (3) Handling of incomplete cost data

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
- 115 studies met selection criteria for review

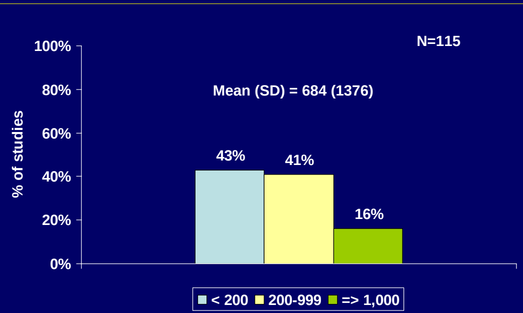
Literature Review: Findings

General Study Information

General Findings

- Studies covered a variety of clinical areas such as cardiovascular disease, musculoskeletal conditions, cancer, and psychiatry
- 85% of the studies were published in general medical, surgical, or subspecialty clinical journals and the remainder were in methods or policy journals
- The trials in which these economic analyses were performed were conducted in either the United States (24%), the UK (24%), multinationally (21%), or in other countries (31%)

Sample Size



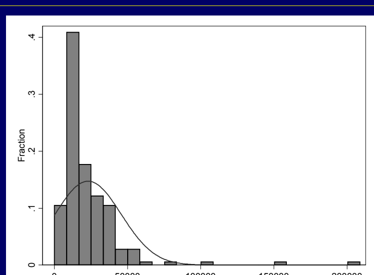
Analysis of Costs

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)
- Cost data tend to be skewed because:
  - Can not have negative costs
  - More severe cases require substantially more services than less severe cases (“long heavy right tails”)
  - Catastrophes can yield small numbers of patients with astronomical costs (“Outliers”)

Typical Distribution Of Cost Data



Sample	Mean	Median
All	23,019	16,052
<75,000	20,430	15,960

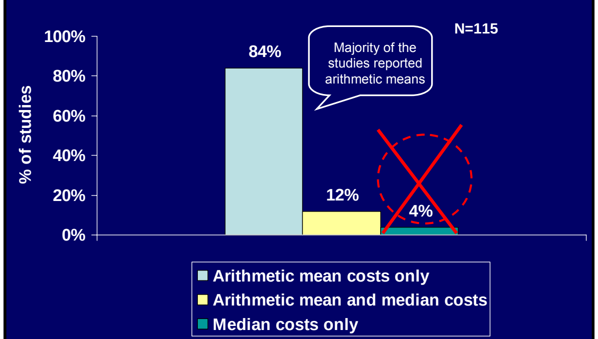
Which Statistic Should be Used to Summarize Cost Data?

- What statistical formulation best characterizes the policy or decision problem of interest?
- For cost-effectiveness analysis: ΔC (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 (ΔC/ΔE) require an estimate of ΔC where:

  ΔC= Ct - Cs

  ΔE= Et - Es

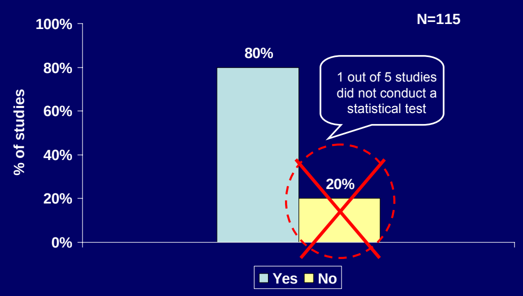
Findings: Cost Statistic Reported Across Treatment Arms



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% (median),…75%.....
- An indication of whether or not the difference in arithmetic means occurred by chance

Findings: Was Statistical Comparison of Treatment Arms Made?



Univariate Statistical Tests

- Usual starting point for comparing arithmetic means: T-tests and one way ANOVA
  - 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): Non-parametric tests

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

Problems With Non-parametric Tests

- Tests don't provide an estimate of the difference in arithmetic mean cost
- Tests don't yield inferences about the difference in arithmetic means
- BOTTOM LINE: tests may tell you some measure of the cost distributions differs between the treatment groups, but they DON'T tell you whether the parameter of interest – the arithmetic mean -- is different

Common Mistakes (II): Log Transformation

- Attempt to make the distribution more normal by taking the log transformation of cost
  - Estimate and draw inferences about the difference in log cost with the goal of applying these estimates and inferences to the arithmetic mean of cost
- If distribution of log cost is normal, t-test of log cost may be more efficient than t-test of nonnormally distributed cost

Problems With Log Transformations

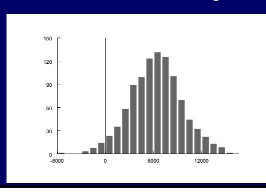
- Transformation does not always yield normal distribution
- For the log transformation, one is making estimates and inferences about the ratio of the treatment group means or differences in geometric means
- For economic analysis, the outcome of interest is the difference in untransformed costs (e.g., "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

Problems With Log Transformations

- "There is a very real danger that the log scale results may provide a very misleading, incomplete, and biased estimate…………………………..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"
  (Manning, 1998)

Better Statistical Approach: Non-Parametric Bootstrap

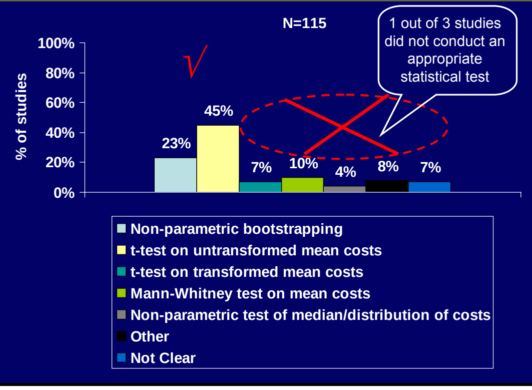
- Provides direct test of arithmetic mean of cost that avoids parametric assumptions
  - Estimates distribution of the observed difference in arithmetic mean cost using N bootstrap replicates
  - Yields a test of how likely it is that 0 is included in this distribution (by evaluating the probability 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 arithmetic means
    - T-test on untransformed costs
  - Non-parametric test of arithmetic means
    - Bootstrap methods

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



Illustrative Example 1

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

Results from Different Statistical Tests Applied to Same Dataset

Test	P-value
T-test of mean difference	0.16
Non-parametric Bootstrap	0.09
T-test, log of cost difference	0.001
Wilcoxon rank-sum (Mann-Whitney)	0.0002
Kolmogorov-Smirnov	0.001

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

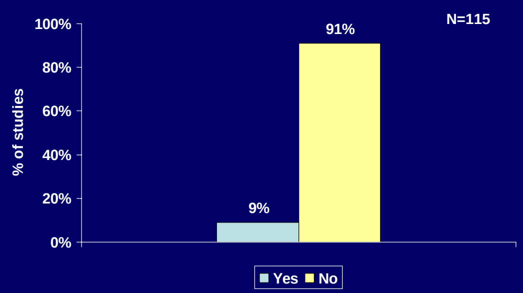
Illustrative Example 2

	Mean	SD	Skewness	Kurtosis	Diff.	P-value
Costs ($)						
Group 0	20000	10263	1.60	8.06	0	1.00
Group 1	20000	3123	0.58	3.24		
Log Costs						
Group 0	9.78	0.49	-0.02	2.77	0.107	0.0000
Group 1	9.89	0.15	0.16	2.86		

Multivariable Analysis of Costs

- Even if treatment is randomly assigned, multivariable analysis may be superior to univariate analysis because:
  - Improves the power for tests of differences
  - Facilitates subgroup analyses for cost-effectiveness (e.g., more and less severe; different countries; etc.)
  - Accounts for potential variations in economic conditions and practice pattern (by provider, center, or country) that may influence costs and that may not be accounted for by randomization
  - Helps explain what is observed (e.g., coefficients for other variables should make sense economically)

Findings: Was Multivariate Adjustment of Incremental Costs Made?



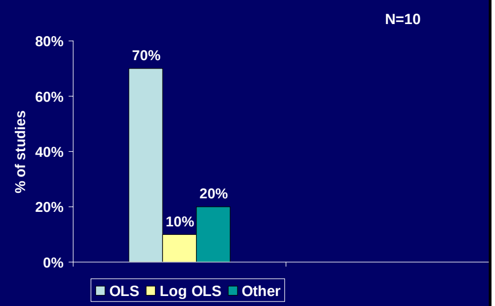
Multivariable Techniques Used for the Analysis Of Costs

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

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

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



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

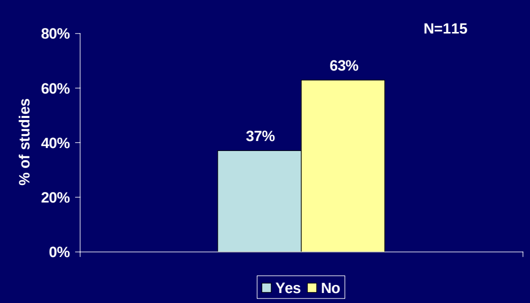
Comparison of Costs and Effects and Assessing Sampling Uncertainty

Preferred Analytic Approaches
&
Findings on Common Mistakes
Identified in Review

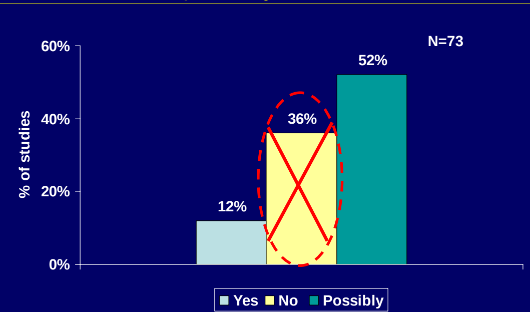
Joint Comparison Of Costs And Effects

- The incremental cost-effectiveness ratio (ICER) is a useful decision tool to help determine whether the new therapy offers good value to the alternative
- Lack of joint comparison justified
  - YES: One therapy is unambiguously dominant over its alternative (i.e. significantly more effective and significantly less costly)
    - Joint comparison may not be necessary
  - NO: Tradeoff between costs and effects
    - Joint comparison necessary
  - POSSIBLY: No significant difference in effect
    - Need for a joint comparison still remains under most circumstances

Findings: Was Comparison of Costs and Effects Made?



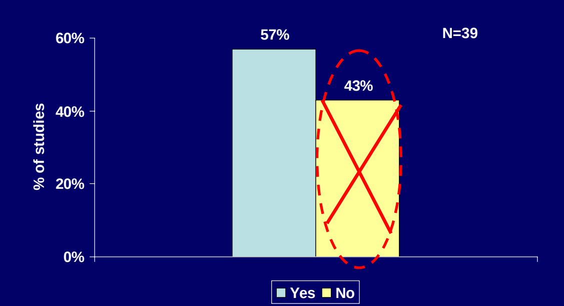
Findings: If NOT, was the lack of joint comparison justified?



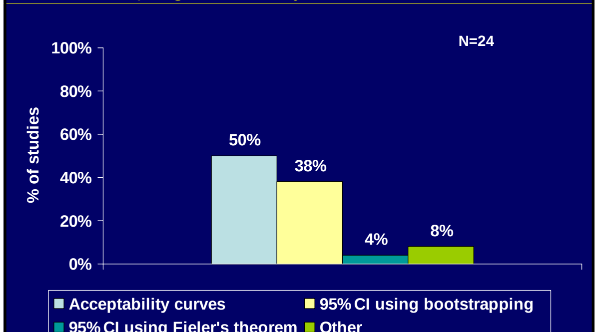
Sampling Uncertainty

- Because cost-effectiveness ratios estimated from trial data are the result of samples drawn from the population, one should report the uncertainty in this outcome that derives from such sampling.
  - Confidence intervals for cost-effectiveness ratios
  - Confidence intervals for net monetary benefit curves
  - Cost-effectiveness acceptability curves

Findings: If joint comparison conducted, was sampling uncertainty measured?



Findings: If YES, how was sampling uncertainty measured?



* Other includes studies such as those that calculated 95% CI but did not specify how these were estimated (based on t-statistic, bootstrapping, etc.) or studies that calculated 95% CI for ICER based on 95% CI values for only the numerator (i.e. costs)

Joint Comparison and Uncertainty: Summary

- Only 37% of the studies conducted a joint comparison of costs and effects
- Depending on the strictness of the criteria, 23% to 56% of the 115 studies should have estimated costs and effects jointly, but failed to do so
- Among the studies that compared costs and effects, only half reported sampling uncertainty

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
- Incomplete cost data can also be due to item-level missingness
  - Multiple-imputation approach preferred method

Degree and Mechanism 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
- Need to diagnose mechanism of missingness
  - Missing completely at random (MCAR)
  - Missing at random (MAR)
  - Not missing at random (NMAR) or informatively (nonignorably) missing

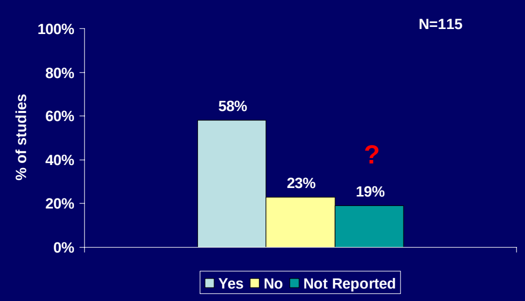
Common Mistakes

- 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

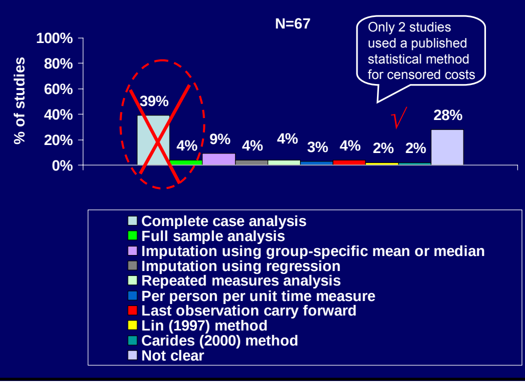
Techniques to Handle Censored Costs

- 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?



Handling of Incomplete Cost Data: Summary

- 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

Conclusion

- Our review finds a substantial number of clinical trial-based economic studies using statistical methods of poor quality
- Efforts are needed from different stakeholders to ensure that future clinical trial-based cost-effectiveness analyses address these issues to enhance the validity of their findings and ensure their usefulness in health-care decision making

Comments or Questions?