COST-EFFECTIVENESS RATIO

Suppose costs and effects are measured in two independent samples. One sample receives a new treatment and the other sample receives standard care.

\[
C = \bar{C}_t - \bar{C}_s
\]

\[
E = \bar{E}_t - \bar{E}_s
\]

Incremental cost effectiveness ratio (ICER)

\[
\frac{C}{E}
\]

Characterizes tradeoff between costs and effects.

COST DATA FOR A COST-EFFECTIVENESS RATIO

Social perspective: Economic theory shows that by minimizing mean costs and differences in mean costs one can achieve social efficiency (Kaldor-Hicks).

Budgetary perspective: Mean costs are a better summary of budgetary impact than median costs or log of costs.
MULTIVARIATE ANALYSIS

- When cost data are available for each subject, model should be estimated in a multivariate model.

- Even if treatment is assigned in a randomized setting use multivariate analysis because:
  - 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.
  - Improves the power for tests of differences between groups (by explaining variation due to other causes).
  - Helps explain what is observed (e.g., coefficients for other variables should make sense economically).

- If treatment is not assigned in a randomized setting multivariate analysis is necessary to adjust for observable imbalances between treatment groups, but it may not be sufficient.

MULTIVARIATE TECHNIQUES USED FOR THE ANALYSIS OF COSTS

- Ordinary least squares (OLS) regression predicting costs.
- OLS regression predicting ln(Cost).
- Cox semiparametric regression.
- Generalized linear models (GLM).

SKewed COST DATA

- Common feature of cost data is right-skewness.

- Data tend to be skewed because:
  - Costs are positive values.
  - Most severe cases may require substantially more resources than less severe cases.
  - Certain events, which can be very expensive, occur only among a few patients.

- Results in small proportion of patients being responsible for a high proportion of health care costs.

- The problem with skewed cost data is not that high costs observations exist (some statistical approaches to the skewed data problem assume that these observations are mistakes in the data or provide very little information); it is that we cannot be confident that our data accurately reflects the frequency with which they occur or their magnitude.

DISTRIBUTION OF THE DATA
ADDRESSING SKEWED DATA

How one treats the highest cost patients in the analysis is important, given that they may:

- Be influential for some of the estimated parameters (e.g., means) but not for others (e.g., medians)
- Limit the power of parametric statistical tests
- Be predictable, based on the severity of the patient
- Be "influential observations" or "outliers"

LOG OF COSTS

A common transformation of costs

The goal is to approximate a normal distribution of the error terms so that OLS can be used without having to worry about the sensitivity of the results to skewness

ESTIMATING $C$ FROM OLS REGRESSION ON THE LOG OF COSTS

The coefficient on the treatment indicator from an OLS regression on the log of costs produces an estimate of the percentage difference in mean costs between treatment groups

For cost effectiveness analysis, the interest is in the difference in the arithmetic mean between treatment groups: $C$

It is possible to retransform the log difference into $C$ [Duan, 1983; Manning 1999]

ESTIMATING $C$ FROM OLS REGRESSION ON THE LOG OF COSTS WHEN HETEROSKEDASTICITY EXISTS

The problem of by-group heteroskedasticity:

- A particular concern when the variance of the residuals differs between treatment groups
- Almost always a problem for the estimate of treatment costs
- If heteroskedasticity is not dealt with correctly, the estimates of $C$ and its standard error from this model are likely to be highly inaccurate [Manning and Mullahy, 2001]

Addressing the problem of heteroskedasticity:

- See Manning (1998) for unbiased estimation of $C$ when by-group heteroskedasticity exists in the data
- See Ai and Norton (2000) for estimates of standard errors for $C$ when by-group heteroskedasticity exists
- This approach can be quite cumbersome
Generalized Linear Models (GLM)

- GLM provides an estimate of the exponential conditional mean function.
- Suggested use of the gamma distribution with a log link. (Blough, 2000)
- These models have the advantages of the log models, but C is estimated directly so it does not require any retransformation (e.g., none of problems addressed by Manning (2001) apply here.)

COX PROPORTIONAL HAZARD MODEL

- Because it makes no specific assumption about an error term distribution, the Cox proportional hazards specification is particularly attractive for modeling outcomes like costs whose probability distribution may be complex
- Assumes there is a proportional response to treatment
- See Harrell, Lipscomb, Etzioni, Dudley
- One should not use these models to control for censored cost data (Lipscomb, Hallstrom)

SUMMARY

- Use mean difference in costs between treatment groups estimated from a multivariate model as the numerator for a cost-effectiveness ratio
- It has yet to be clearly determined which model is best in which situations
  - Manning and Mullahy (2001) do have several helpful suggestions
- In our experience, the estimate of C and its standard error are not very sensitive to the model chosen
- Inference based on estimates of the % difference in means (estimated directly from log model) can differ from inferences based on estimates of C
  - For CEA, all inferences should be based on estimates of C
- Avoid the log of cost model unless you know that you are doing the retransformation correctly. Papers reporting results using the log model should be viewed with caution.
- Establish criteria for adopting a particular multivariate model for analyzing the data prior to unblinding the data (i.e., the fact that one model gives a more favorable result should not be a reason for its adoption)
- Given that no method will be without problems, it may be helpful to report the sensitivity of one’s results to different specifications of the multivariate model
REFERENCES

Measuring Treatment Costs


Alternative Multivariable Models


Non-parametric cost models (i.e. Cox)


Reasons not to use censoring


Generalized Linear Models