UNIVARIATE AND MULTIVARIABLE ANALYSES OF ECONOMIC OUTCOMES

- Analysis plans for economic assessments should routinely include univariate and multivariable methods for analyzing the trial data
- Univariate analyses are used for the predictors of economic outcomes
  - Demographic characteristics
  - Clinical history
  - Length of stay, and other resource use before entry of study subjects into the trial
- Univariate and multivariable analyses should be used for the economic outcome data
  - Total costs
  - Hospital days
  - Quality-adjusted life years

OUTLINE

- Univariate Analysis
  - Statistical Tests
  - General Advice
- Multivariable Analysis
  - Multivariable Techniques
  - Diagnostic Tests
  - General Advice
- APPENDICES
  - Details on OLS and log OLS models
  - Technical Notes on Diagnostic Tests
  - Stata Programs
COST DATA

• 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, a relatively small number of patients
    → A minority of patients are responsible for a high proportion of health care costs

• Heavy tails vs. "outliers"
  - Distributions with long, heavy, right tails will have means that differ from the median
    * Median is not a better measure of the costs on average than is the mean

TYPICAL DISTRIBUTION OF COST DATA

<table>
<thead>
<tr>
<th>Sample</th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>23,019</td>
<td>16,052</td>
</tr>
<tr>
<td>&lt;75,000</td>
<td>20,430</td>
<td>15,960</td>
</tr>
</tbody>
</table>
UNIVARIATE ANALYSIS OF COSTS

- Report:
  - Arithmetic means and their difference
    * Economic analysis is based on differences in arithmetic mean costs (because n x mean = total), not median costs; thus means are the statistic of interest
  - Measures of variability and precision, such as:
    * Standard deviation
    * Quantiles such as 5%, 10%, 50%,... (particularly if data are skewed)
  - An indication of whether or not the difference in arithmetic means:
    * Occurred by chance
    * Is economically meaningfully

UNIVARIATE ANALYSIS: PARAMETRIC TESTS OF MEANS

- 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 the this assumption

RESPONSES TO VIOLATION OF ASSUMPTIONS (I)

- Adopt nonparametric tests of other characteristics of the distribution that are not as affected by the nonnormality of the distribution ("biostatistical" approach)
  - Wilcoxon rank-sum test for difference in medians
  - Kolmogorov-Smirnov test for difference in cumulative distribution function

RESPONSES TO VIOLATION OF ASSUMPTIONS (II)

- Transform costs so they approximate a normal distribution
  - Common transformations
    * Log (arbitrary additional transformations required if any observation equals 0)
    * Square root
  - Estimate and draw inferences about differences in transformed costs
  - Use these estimates and inferences to estimate and draw inferences about differences in untransformed costs
    * Estimation: Simple exponentiation of mean of log costs results in geometric mean (not arithmetic mean)
      - Need to apply smearing factor to correct
    * Inference: On the retransformed scale, inferences about the log of costs translate into inferences about differences in the geometric mean / the ratio of the treatment group means, rather than the arithmetic mean
TRANSFORMATION OF THE DATA?

- For economic analysis, the outcome of interest is the difference in untransformed costs (e.g., “Congress does not appropriate log dollars. First Bank will not cash a check for log dollars”)

- Thus, the results on the transformed scale must be retransformed to the original scale

- “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”

PRIMER ON THE LOG TRANSFORMATION OF COSTS

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group 0</th>
<th>Group 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>29</td>
</tr>
</tbody>
</table>

Arithmetic mean

<table>
<thead>
<tr>
<th></th>
<th>Group 0</th>
<th>Group 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>GM: ( \frac{1}{n} \sum_{i=1}^{n} y_i )</td>
<td>14.422496</td>
<td>7.5769849</td>
</tr>
</tbody>
</table>

Natural log

<table>
<thead>
<tr>
<th></th>
<th>Group 0</th>
<th>Group 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.3025851</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>2.7080502</td>
<td>2.7080502</td>
</tr>
<tr>
<td>3</td>
<td>2.9957323</td>
<td>3.3672958</td>
</tr>
</tbody>
</table>

Arithmetic mean

<table>
<thead>
<tr>
<th></th>
<th>Group 0</th>
<th>Group 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \exp^{\text{mean \log}} )</td>
<td>14.422496</td>
<td>7.5769849</td>
</tr>
</tbody>
</table>

- Observation: Simple exponentiation of the mean of the logs yields the geometric mean of costs, which in the presence of variability in costs (variance, skewness, kurtosis) is a biased estimate of the arithmetic mean
RETRANSFORMATION OF THE LOG OF COST

- “Smearing” retransformation eliminates this bias
- Duan's common smearing factor:
  \[ \Phi = \frac{1}{N} \sum_{i=1}^{N} e^{(\hat{z}_i - \bar{z}_i)} \]

where in univariate analysis, \( \hat{z}_i \) = the group mean

<table>
<thead>
<tr>
<th>Obs</th>
<th>ln</th>
<th>( z_i - \hat{z}_i )</th>
<th>( e^{(ln - \hat{ln})} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0, 1</td>
<td>2.302585</td>
<td>-0.366204</td>
<td>0.693361</td>
</tr>
<tr>
<td>0, 2</td>
<td>2.708050</td>
<td>0.039261</td>
<td>1.040042</td>
</tr>
<tr>
<td>0, 3</td>
<td>2.995732</td>
<td>0.326943</td>
<td>1.386723</td>
</tr>
<tr>
<td>Mean, 0</td>
<td>2.668789</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>1, 1</td>
<td>0.0</td>
<td>-2.025115</td>
<td>0.131979</td>
</tr>
<tr>
<td>1, 2</td>
<td>2.708050</td>
<td>0.682935</td>
<td>1.979679</td>
</tr>
<tr>
<td>1, 3</td>
<td>3.367296</td>
<td>1.342181</td>
<td>3.827380</td>
</tr>
<tr>
<td>Mean, 1</td>
<td>2.025115</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Smear</td>
<td>--</td>
<td>--</td>
<td>1.509861</td>
</tr>
</tbody>
</table>

COMMON SMEARING RETRANSFORMATION

- Retransformation formula
  \[ E(\bar{Y}_0) = \Phi \ e^{(\hat{z}_i)} \]
  \[ E(\bar{Y}_1) = \Phi \ e^{(\hat{z}_1)} \]

<table>
<thead>
<tr>
<th>Group</th>
<th>( \Phi )</th>
<th>( e^{(lnhat)} )</th>
<th>Predicted cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.509861</td>
<td>( \times 14.422496 )</td>
<td>21.775964</td>
</tr>
<tr>
<td>1</td>
<td>1.509861</td>
<td>( \times 7.5769849 )</td>
<td>11.440194</td>
</tr>
</tbody>
</table>

COMMON SMEARING RETRANSFORMATION

- Log transformations and normal assumptions:
  - If met, t-test of the log may be more efficient than t-test of cost
  - If not met there are no efficiency gains
  - In either case, retransformation translates differences in variance, skewness, and kurtosis into differences in means

- Why are the retransformed subgroup-specific means -- 21.78 and 11.44 -- so different from the untransformed subgroup means of 15?
  - Because the standard deviations of the subgroups’ logs are substantially different

\( SD_0 = 0.3482375; SD_1 = 1.784508 \)
SUBGROUP-SPECIFIC SMEARING FACTORS

- Manning has shown that in the face of heteroscedasticity, use of a common smearing factor in the retransformation of the predicted log of costs yields biased estimates of predicted costs.
- One obtains unbiased estimates by use of subgroup-specific smearing factors.
- Manning's subgroup-specific smearing factor:
  \[ \Phi_j = \frac{1}{N_j} \sum_{i=1}^{N_j} e^{(z_i - \bar{z}_j)} \]

<table>
<thead>
<tr>
<th>Observation</th>
<th>ln</th>
<th>exp(\hat{z}_i - \bar{z}_j)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0, 1</td>
<td>2.3025851</td>
<td>.6933613</td>
</tr>
<tr>
<td>0, 2</td>
<td>2.7080502</td>
<td>1.040042</td>
</tr>
<tr>
<td>0, 3</td>
<td>2.9957323</td>
<td>1.386723</td>
</tr>
<tr>
<td>Mean_0</td>
<td>2.668789</td>
<td><strong>1.040042</strong></td>
</tr>
<tr>
<td>1, 1</td>
<td>0.0</td>
<td>.1319786</td>
</tr>
<tr>
<td>1, 2</td>
<td>2.7080502</td>
<td>1.979679</td>
</tr>
<tr>
<td>1, 3</td>
<td>3.3672958</td>
<td>3.827379</td>
</tr>
<tr>
<td>Mean_1</td>
<td>2.025115</td>
<td><strong>1.979679</strong></td>
</tr>
</tbody>
</table>

SUBGROUP-SPECIFIC SMEARING RETRANFORMATION

- Retransformation formulas
  \[ E(\bar{Y}_0) = \Phi_0 e^{(\bar{z}_0)} \]
  \[ E(\bar{Y}_1) = \Phi_1 e^{(\bar{z}_1)} \]
- Retransformation

<table>
<thead>
<tr>
<th>Group</th>
<th>( \Phi_i )</th>
<th>e(\ln)</th>
<th>Predicted cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.040042</td>
<td>x</td>
<td>14.422496</td>
</tr>
<tr>
<td>1</td>
<td>1.979679</td>
<td>x</td>
<td>7.5769849</td>
</tr>
</tbody>
</table>

- All else equal, in the face of differences in variance (or skewness or kurtosis), use of subgroup-specific smearing factors yield unbiased estimates of subgroup means.
- Use of separate smearing factors eliminates efficiency gains from log transformation, because one cannot assume that p-value derived for the log of cost applies to the arithmetic mean of cost.
RESPONSES TO VIOLATION OF ASSUMPTIONS (II)

- Adopt tests of means that avoid parametric assumptions (most recent development)
  - Non-parametric bootstrap (Efron)
    * Estimates the distribution of the observed difference in arithmetic 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)

HOSPITAL COSTS, TIRILAZAD MESYLATE FOR SUBARACHNOID HEMORRHAGE *

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Vehicle</th>
<th>6 mg/kg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ($)</td>
<td>20,343</td>
<td>25,185</td>
</tr>
<tr>
<td>SD</td>
<td>22,488</td>
<td>22,619</td>
</tr>
<tr>
<td>Quantiles</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>
<tr>
<td>Geometric Mean</td>
<td>11,518</td>
<td>20,754</td>
</tr>
<tr>
<td>Mean of Log</td>
<td>9.351638</td>
<td>9.940484</td>
</tr>
</tbody>
</table>

Derived from Glick HA, Polsky D. Analytic approaches for the evaluation of costs. Hepatology. 1999;29:18S-22S.
HOSPITAL COSTS, TIRILAZAD MESYLATE FOR SUBARACHNOID HEMORRHAGE

- Statistical comparison of differences

<table>
<thead>
<tr>
<th>Test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-test of mean difference</td>
<td>0.15</td>
</tr>
<tr>
<td>Bootstrap (2,000 replicates)</td>
<td>0.07 (nonpar, 1-tailed)</td>
</tr>
<tr>
<td>Wilcoxon rank-sum:</td>
<td>0.0002</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov:</td>
<td>0.001</td>
</tr>
<tr>
<td>T-test, log of costs:</td>
<td>0.001</td>
</tr>
<tr>
<td>Breusch-Pagan test of</td>
<td>0.0000</td>
</tr>
<tr>
<td>heteroscedasticity</td>
<td></td>
</tr>
</tbody>
</table>

WHY WOULD DIFFERENT STATISTICAL TESTS LEAD TO DIFFERENT INFERENCES?

- The tests are evaluating differences in different statistics and may have different degrees of efficiency
  - T-test of untransformed costs indicates one cannot infer that the arithmetic means are different
  - Bootstrap leads to same (lack of) inference and does not make the normality assumption
  - 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
  - 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
- When distributions are skewed, means and medians can be measuring very different things
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 (Kaldor-Hicks)
  - Budgetary perspective: arithmetic mean costs are a better summary of budgetary impact than median costs or log of costs

- Cost-effectiveness ratios ($\Delta C/\Delta E$) and NMB ($[rc \Delta E] - \Delta C$) require an estimate of $\Delta C$

  Where:

  - $\Delta C = \bar{C}_i - \bar{C}_s$
  - $\Delta E = \bar{E}_i - \bar{E}_s$

TEST FOR DIFFERENCES IN MEANS

- 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

- Because of distributional problems related to evaluating the arithmetic mean, there has been a growing use of nonparametric tests such as Wilcoxon and KS tests
  - Problem: Their use divorces hypothesis testing from estimation
    * i.e., we want to 1) estimate the magnitude of the difference in arithmetic means and 2) test whether that difference was observed by chance
    * Use of tests of medians or distributions to address the second task does not help with the first task

- Tests of transformed variables such as the log or square root pose similar problems
MULTIVARIABLE ANALYSIS OF ECONOMIC OUTCOMES

● Multivariable estimates of costs generally thought to be better than univariate estimates

● Even if treatment is assigned in a randomized setting use of multivariable analysis may have added benefits because:
  - Improves the power for tests of differences between groups (by explaining variation due to other causes)
  - Facilitates subgroup analyses for cost-effectiveness (e.g., more and less severe; different countries/centers; etc.)
  - Variations in economic conditions and practice pattern differences by provider, center, or country may have a large influence on costs and the randomization may not account for all differences
  - Additional advantage: Helps explain what is observed (e.g., coefficients for other variables should make sense economically)

● If treatment is not randomly assigned, multivariable analysis is necessary to adjust for observable imbalances between treatment groups, but it may not be sufficient

MULTIVARIABLE TECHNIQUES USED FOR THE ANALYSIS OF COSTS

● More common techniques
  - Ordinary least squares regression predicting costs after randomization
  - Ordinary least squares regression predicting the log transformation of costs after randomization
  - Generalized Linear Models

● Other techniques
  - Cox semiparametric regression (Harrell et al., Statis. in Med. 1996;15:361-387.)
  - Generalized Gamma regression (Manning et al., NBER technical working paper 293)
  - Extended estimating equations (Basu and Rathouz, Biostatistics 2005)

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)
ORDINARY LEAST SQUARES REGRESSION (OLS)

\[ Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \epsilon \]

- 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
    * In large datasets with extreme observations
  - Can produce predictions with negative costs

- See Appendix 1 for details

LOG OF COSTS. ORDINARY LEAST SQUARES REGRESSION

\[ \ln(Y) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \epsilon \]

- Advantages
  - Widely known transformation for costs
  - Common in the literature
  - Reduces robustness problem
  - Improves efficiency

- Disadvantages
  - Retransformation problem can lead to bias
  - Coefficients not directly interpretable
  - Not easy to implement
  - Residual may not be normally distributed even after log transformation

- See Appendix 1 for details
GENERALIZED LINEAR MODELS (GLM)

- These models have the advantages of the log models, but (a direct transformation of) ΔC is estimated directly so it does not require any smearing correction.
- To build them, one must identify a "link function" and a "family" (based on the data).

STATA code:
```
glm y x, link(linkname) family (familyname)
```

SAS code:
```
proc genmod;
model y=x/ link=linkname dist=familyname;
run;
```

THE LINK FUNCTION

- Specifies the relationship between the covariates and the mean.
  - e.g. identity, log, power # (square root, etc.)
- GLMs are attractive because the link function directly characterizes how the mean on the raw untransformed scale is related to the predictors.
  - For example, if one uses a log link, one is assuming:
    \[ \ln(E(y/x)) = X\beta \]
    * GLM with a log link differs from log OLS in part because in log OLS, \( E(\ln(y)/x) = X\beta \) and
    \[ \ln(E(y/x)) \neq E(\ln(y)/x) \]
  - e.g. \( \ln(E(y/x))=X\beta \)

\[
\begin{array}{lcccc}
\text{Variable} & \text{Group 0} & \text{Group 1} \\
\hline
\text{Observations} & & & & \\
1 & 10 & 1 \\
2 & 15 & 15 \\
3 & 20 & 29 \\
\text{Arithmetic mean} & 15 & 15 \\
\hline
\text{ln(15)} & 2.7080502 & 2.7080502 \\
\text{Natural log} & & & & \\
1 & 2.3025851 & 0.0 \\
2 & 2.7080502 & 2.7080502 \\
3 & 2.9957323 & 3.3672958 \\
\text{Arithmetic mean} & 2.6687892 & 2.0251153 \\
\hline
\text{Regression Results} & & & & \\
\text{Variable} & \text{Coef.} & \text{S.E.} & \text{t} & \text{p-value} \\
\hline
\text{Log OLS} & & & & \\
\text{treat} & -0.6436739 & 1.049721 & -0.61 & 0.573 \\
\text{cons} & 2.668789 & 0.7422645 & 3.60 & 0.023 \\
\text{GLM, Log Link} & & & & \\
\text{treat} & -6.01e-12 & 0.5721986 & -0.00 & 1.000 \\
\text{cons} & 2.70805 & 0.4046055 & 6.69 & 0.000 \\
\end{array}
\]
THE LINK FUNCTION (II)

- Log link has been most commonly used in literature but may not necessarily be the best in all cases
- Little guidance in current literature for applied researcher on how to identify correct link function
  - Compare model performance of all permutations of candidate link and variance function
- Basu and Rathouz (2005) propose extended estimating equations (EEE) which estimate the link function and variance structure along with other components of the model based on the data

THE FAMILY (II)

- Modified Park test used to determine family
- Implement after GLM regression assuming a family & link e.g. glm cost treat $ivar, family(gamma) link(log)

(1) Predict value of y and log transform it
   predict yhat
   gen lnyhat=ln(yhat)

(2) Save raw scale residuals and square them
   gen res=cost-yhat
   gen r2=((res)^2)

(3) Regress ln(r2) on ln(yhat) and a constant using GLM with log link and gamma distribution
   glm r2 lnyhat , link(log) family(gamma) robust nolog

(4) Coefficient on ln(yhat) gives the family
   If λ=0  Gaussian NLLS
   If λ=1  Poisson
   If λ=2  Gamma
   If λ=3  Inverse Gaussian or Wald
   test lnyhat==0
   test lnyhat==1
   test lnyhat==2
   test lnyhat==3

See p. 39
GLM COMMENTS

- **Advantages**
  - No retransformation problems of log OLS models
    * Because the link function allows modeling of the log of mean costs [i.e. ln(E(y/x)=Xβ)] unlike the log OLS that models the mean of log costs [ i.e. E(ln(y)/x)=Xβ]
  - Gains in precision from estimator that matches data generating mechanism
  - Consistent even if not the correct family distribution
    * Choice of family only affects efficiency if link function and covariates are specified correctly

- **Disadvantages**
  - Can suffer substantial precision losses
    * If heavy-tailed (log) error term, i.e., log-scale residuals have high kurtosis (>3)
    * If variance function (i.e. family) is misspecified

RETRANSFORMATION

- GLM avoids the problem that simple exponentiation of the results of log OLS yields biased estimates of predicted costs
- It does not avoid the other complexity of nonlinear retransformations (also seen in log OLS models):
  - On the transformed scale, the effect of the treatment group is estimated holding all else equal; however, retransformation (to estimate costs) reintroduces the covariate imbalances
- Do not use the means of the covariates to avoid the reintroduction of covariate imbalance, because the mean of nonlinear retransformations does not equal the linear retransformation of the mean
- Rather, use the method of recycled predictions to create an identical covariate structure for the two groups by:
  - Coding everyone as if they were in treatment group 0 and predicting $\overline{\mu}_0$
  - Coding everyone as if they were in treatment group 1 and predicting $\overline{\mu}_1$
GLM (POISSON/LOG LINK) RESULT

```
replace treat=0
predict pois_0
replace treat=1
predict pois_1
gen pois_dif=pois_1-pois_0
replace treat=tmp
```

```
tabstat pois_1 pois_0 pois_dif   stats |     pois_1     pois_0   pois_dif
---------+------------------------------    mean |  10843.55  6825.096  4018.451
```

SPECIAL CASES

- A substantial proportion of observations have 0 costs
  - May pose problems to regression models
  - Commonly addressed by developing a “two-part” model in which
    the first part predicts the probability that the costs are zero or
    nonzero and the second part predicts the level of costs
    conditional on there being some costs
  * 1st part : Logit or probit model
  * 2nd part : log OLS or GLM model

- Censored costs
  - Results derived from analyzing only the completed cases or
    observed costs are often biased
  - Need to evaluate the “mechanism” that led to the missing data
    and adopt a method that gives unbiased results in the face of
    missingness
EXAMPLE OF ESTIMATES OF ΔC AND ITS 95% CI

<table>
<thead>
<tr>
<th>Estimation Method</th>
<th>ΔC</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>No covariates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS (untransformed costs)</td>
<td>4000</td>
<td>(2324 to 5706)</td>
</tr>
<tr>
<td>Log OLS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homoskedastic retransform</td>
<td>6684</td>
<td>(449 to 5381)</td>
</tr>
<tr>
<td>Heteroskedastic retransform</td>
<td>4000</td>
<td>(2324 to 5706)</td>
</tr>
<tr>
<td>GLM (gamma/log link)</td>
<td>4000</td>
<td>(2324 to 5706)</td>
</tr>
<tr>
<td>GLM (poisson/log link)</td>
<td>4000</td>
<td>(2324 to 5706)</td>
</tr>
<tr>
<td>Multivariate model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS (untransformed costs)</td>
<td>4027</td>
<td>(2362 to 5792)</td>
</tr>
<tr>
<td>Log OLS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homoskedastic retransform</td>
<td>6649</td>
<td>(4816 to 8479)</td>
</tr>
<tr>
<td>Heteroskedastic retransform</td>
<td>4268</td>
<td>(2566 to 6126)</td>
</tr>
<tr>
<td>GLM (gamma/log link)</td>
<td>4116</td>
<td>(2451 to 5984)</td>
</tr>
<tr>
<td>GLM (poisson/log link)</td>
<td>4018</td>
<td>(2351 to 5774)</td>
</tr>
</tbody>
</table>

WHICH ESTIMATE OF ΔC SHOULD ONE USE?

- While the ΔC estimates from the OLS and GLM models are quite similar, those from the log OLS models (particularly homoskedastic) are substantially different.
- Given that the different multivariable methods yield varying estimates of ΔC, how does one judge which result is a better estimate of ΔC?
- Series of diagnostic tests available to help compare performance of alternative multivariable models

DIAGNOSTIC TESTS

1. Skewness/Kurtosis
   - Tests for normality of residuals
2. Heteroskedasticity test (Breusch-Pagan test)
   - Tests whether residuals are homoskedastic
3. Modified Park test (GLM family test)
   - Used to determine the family distribution in GLM
4. Pregibon Link test
   - Checks linearity of response on scale of estimation
5. Modified Hosmer Lemeshow test
   - Checks for systematic bias in fit on raw scale
6. Pearson’s Correlation test
   - Checks for systematic bias in fit on raw scale
7. Copas test
   - Tests for overfitting and cross-validation

NOTE: Details on tests are provided in Appendix 2
BREUSCH-PAGAN TEST FOR HETEROSKEDASTICITY

- In our example, only the results from the log OLS with homoskedastic retransformation are substantively different

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test whether residuals are heteroskedastic</td>
<td>If reject null then residuals are heteroskedastic</td>
</tr>
<tr>
<td></td>
<td>If log-scale residuals are heteroskedastic, Log OLS will be biased if appropriate smearing correction not applied</td>
</tr>
</tbody>
</table>

- Test result:

<table>
<thead>
<tr>
<th>Test Statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breusch-Pagan test</td>
<td>32.3</td>
</tr>
</tbody>
</table>

- Log OLS residuals are heteroskedastic in treatment variable
- Hence, results from log OLS with homoskedastic retransformation are biased!

CURRENT STATE OF THE ART

- Common characteristics of the distribution of costs can cause problems for a large number of the available multivariable techniques
- It is most likely the case that no single model will always be most appropriate for estimating cost differences associated with medical therapies
- Criteria for comparing alternative models may include a series of diagnostic tests discussed here
  - Most tests detect problems but do not provide guidance on how to fix the problem
  - Test may or may not help you identify a single best model
  - Tests may help you clearly identify some models that perform very poorly and should be eliminated from consideration
  - Use the combination of all test results to make inferences
  - Use performance on tests to rank models

GENERAL ADVICE (I)

- Use mean difference in costs between treatment groups estimated from a multivariable model as the numerator for a cost-effectiveness ratio
- No model is best in every situation
- Inference based on estimates of the % difference in means (estimated directly from log model) can differ from inferences based on estimates of $\Delta C$
  - For CEA, all inferences should be based on estimates of $\Delta C$
GENERAL ADVICE (II)

- 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
  - When there is heteroskedasticity, the biases can be huge
  - Heteroskedasticity between treatment groups almost always exists
- Consider GLM models because they have the advantages of the log models without any transformation problems

GENERAL ADVICE (III)

- Establish criteria for adopting a particular multivariable 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 multivariable model

REFERENCES

Measuring Treatment Costs


Alternative Multivariable Models


Manning, WG. Mullahy J. Estimating log models: To transform or not to


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


/* Modified Park Test */

. gen r2 = ((cost-yhat)^2)
. gen lnyhat = ln(yhat)

. glm r2 lnyhat, link(log) family(gamma) robust nolog

Generalized linear models             No. of obs      =      200
Optimization   : ML: Newton-Raphson   Residual df     =      198
Scale parameter =  5.37055
Deviance         =  556.0966603       (1/df) Deviance = 2.808569
Pearson          =  1063.368955       (1/df) Pearson =  5.37055

Variance function: V(u) = u^2         [Gamma]
Link function    : g(u) = ln(u)       [Log]
Standard errors  : Sandwich

Log pseudo-likelihood = -3667.729811  AIC             =  36.6973
BIC                   =-492.9701783

|             Robust
| Robust
| Coef. | Std. Err. | z    | P>|z| | 95% Conf. Interval |
|----------|----------|------|------|-----------------|
| r2  | .8059514 | .6058605 | 1.33 | 0.183 | -.3815133   1.993416 |
| _cons | 10.04718 | 5.417169 | 1.85 | 0.064 | -.5702812   20.66463 |

/* Modified Park Test */

. test lnyhat==1

( 1) [r2]lnyhat = 1

    chi2(  1) =    0.10
    Prob > chi2 =    0.7488 → Poisson

. test lnyhat==2

( 1) [r2]lnyhat = 2

    chi2(  1) =    3.88
    Prob > chi2 =    0.0487 → Not Gamma

. test lnyhat==3

( 1) [r2]lnyhat = 3

    chi2(  1) =   13.11
    Prob > chi2 =    0.0003 → Not Inverse Gaussian

. }
****glm model (poisson/log)
. glm cost treat $ivar, family(poisson) link(log)

<table>
<thead>
<tr>
<th>Generalized linear models</th>
<th>No. of obs = 200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimization: ML: Newton-Raphson</td>
<td>Residual df = 193</td>
</tr>
<tr>
<td>Scale parameter = 1</td>
<td></td>
</tr>
<tr>
<td>Deviance = 700567.946</td>
<td>(1/df) Deviance = 3629.886</td>
</tr>
<tr>
<td>Pearson = 791555.8081</td>
<td>(1/df) Pearson = 4101.325</td>
</tr>
<tr>
<td>Variance function: V(u) = u</td>
<td>[Poisson]</td>
</tr>
<tr>
<td>Link function: g(u) = ln(u)</td>
<td>[Log]</td>
</tr>
<tr>
<td>Standard errors: OIM</td>
<td></td>
</tr>
<tr>
<td>Log likelihood = -351346.9719</td>
<td>AIC = 3513.54</td>
</tr>
<tr>
<td>BIC = 699545.3708</td>
<td></td>
</tr>
</tbody>
</table>

| cost | Coef. | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|------|-------|-----------|-------|------|----------------------|
|      | treat | .4629637  | .0015546 | 297.81 | 0.000 | .4599168   | .4660106 |
|      | age   | .0082989  | .0000756 | 109.72 | 0.000 | .0081507   | .0084472 |
| ejfract | -.0081781 | .0001135       | -72.07 | 0.000 | -.0084006  | -.0079557 |
| sex   | -.0721448| .0016935  | -42.60 | 0.000 | -.0754639  | -.0688256 |
| etiology | .2498528 | .0015617  | 159.99 | 0.000 | .2467919   | .2529137 |
| race  | .0462949 | .0023699  | 19.53  | 0.000 | .0416499   | .0509398 |
| _cons | 8.359824 | .005554  | 1505.18 | 0.000 | 8.348939   | 8.37071  |