Evaluating Patient Level Costs

Statistical Considerations in Health Economic Evaluations

ISPOR 18th Annual International Meeting

May 19, 2013

Jalpa Doshi and Henry Glick

www.uphs.upenn.edu/dgimhsr

Outline

- Policy relevant parameter for CEA
- Cost data 101
- Univariate analysis
- Multivariable analysis

Policy Relevant Parameter for CEA (I)

- Policy relevant parameter: differences in the arithmetic, or sample, mean
  - In welfare economics, a project is cost-beneficial if the winners from any policy gain enough to be able to compensate the losers and still be better off themselves
    - Thus, we need a parameter that allows us to determine how much the losers lose, or cost, and how much the winners win, or benefit
  - From a budgetary perspective, decision makers can use the arithmetic mean to determine how much they will spend on a program

Policy Relevant Parameter for CEA (II)

- Other summary statistics such as median cost may be useful in describing the data, but do not provide information about the difference in cost that will be incurred or the cost saved by treating patients with one therapy versus another
  - They thus are not associated with social efficiency
- Lack of symmetry of cost distribution does not change fact that we are interested in the arithmetic mean
- Evaluating some other difference, be it in medians or geometric means, simply because the cost distribution satisfies the assumptions of the tests for these statistics, may be tempting, but does not answer the question we are asking

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

Sample Dataset

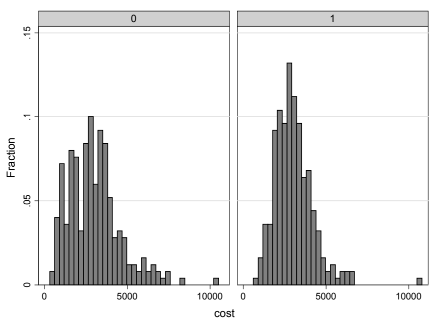
```
. clear
. set more off
. use rchapter5
. sum
Variable| Obs      Mean  Std. Dev.       Min       Max
--------+---------------------------------------------
      id |  500     250.5  144.4818          1       500
   treat |  500        .5  .5005008          0         1
    cost |  500    3027.5  1389.921        315     10499
    qaly |  500  .5941654  .2121149     .04798    .95119
  dissev |  500   .347486  .1124773       .025      .729
--------+---------------------------------------------
    race |  500      .506  .5004647          0         1
  blcost |  500  1634.859  770.5504   111.0891  4926.931
  blqaly |  500  .7857801   .145283   .4895464         1
    male |  500      .484  .5002444          0         1
```

* Data taken from Glick HA, Doshi JA, Sonnad SS, Polsky D. Chapter 5 in "Economic Evaluation in Clinical Trials", 2007.

Sample Dataset

```
. describe

Contains data from D:\henry\HGClass\rchapter5.dta
  obs:           500
 vars:             9                        18 Apr 2008 14:25
 size:        16,500 (99.9% of memory free)
-----------------------------------------------------------
              storage  display    value
variable name   type   format     label      variable label
-----------------------------------------------------------
id              int    %9.0g                 Patient ID
treat           byte   %9.0g                 Treatment group
cost            int    %9.0g                 Total cost
qaly            float  %9.0g                 QALYs
dissev          float  %9.0g                 Disease severity
race            float  %9.0g                 Race
blcost          float  %9.0g                 Baseline cost
blqaly          float  %9.0g                 Baseline QALY
male            float  %9.0g
-----------------------------------------------------------
Sorted by:  id
```

Typical Distributions Of Cost Data (I)



Inspect the Cost Data (I)

```
. summary cost if treat==0,detail

                         Total cost
-------------------------------------------------------------
      Percentiles      Smallest
 1%          622            315
 5%          899            589
10%         1093            622       Obs                 250
25%         1819            640       Sum of Wgt.         250

50%       2825.5                      Mean               3015
                        Largest       Std. Dev.      1582.802
75%         3752           7361
90%         4952           7540       Variance        2505262
95%         6103           8232       Skewness        1.03501
99%         7540          10483       Kurtosis       4.910192
```

Inspect the Cost Data (II)

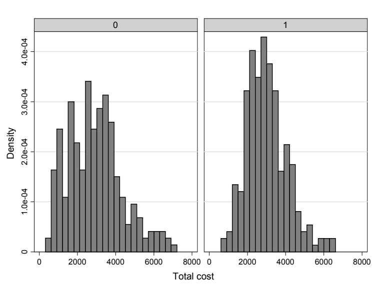
```
. summary cost if treat==1,detail

                        Total cost
-------------------------------------------------------------
      Percentiles      Smallest
 1%         1093            681
 5%         1426            899
10%         1832           1093       Obs                 250
25%         2226           1170       Sum of Wgt.         250

50%       2900.5                      Mean               3040
                        Largest       Std. Dev.      1168.737
75%         3604           6296
90%         4404           6470       Variance        1365946
95%         5085           6520       Skewness       1.525386
99%         6470          10499       Kurtosis       9.234913
```

Typical Distribution Of Cost Data (II)

- Heavy tails vs. "outliers“
  - Distributions with long, heavy, right tails will have means that differ from the median
    - Median is a biased estimate of the sample mean

Problem Not Related Solely to “Outliers”

- Distribution when 5 observations with cost > 7200 are eliminated



Mean, SD When 5 Observations with Cost > 7200 are Eliminated

	Full Sample		Trimmed (3*SD) *	
	Group 0	Group 1	Group 0	Group 1
Mean	3015	3040	2927	3010
Median	2826	2901	2816	2885

* p = 0.003 and 0.0001 for nonnormality of groups 0 and 1, respectively

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 participants into the trial
- Univariate and multivariable analyses should be used for the economic outcome data
  - Total costs, hospital days, quality-adjusted life years

Univariate Analysis of Costs

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%,...
  - An indication of whether or not the difference in arithmetic means
    - Occurred by chance
    - Is economically meaningful

Univariate Analysis: Parametric Tests Of Raw Means

- Usual starting point: T-tests and one way ANOVA
  - Used to test for differences in arithmetic means in total costs, QALYS, etc.
  - Makes assumption that the costs are normally distributed
  - Normality assumption is routinely violated for cost data, but t-tests have been shown to be robust to violations of this assumption when:
    - Samples moderately large
    - Samples are of similar size and skewness
    - Skewness is not too extreme

Steps in Performing a T-test

- Evaluate whether or not the outcome is normally distributed
  - Stata command: sktest (joint test of skewness and kurtosis)

    sktest cost if treat==0

    sktest cost if treat==1
- Evaluate whether or not the standard deviations of costs for the treatment groups are similar
  - Stata command: sdtest

    sdtest cost, by(treat)
- Perform the t-test and interpret it in relationship to the prior two tests
  - Stata command: ttest

    ttest cost, by(treat) unequal

Results of Tests of Normality and Equivalence of S.D. of Costs

Test	P-value	Conclusion
Normality		
sktest, group 0	0.0	Failed
sktest, group 1	0.0	Failed
Equality of standard deviations		
sdtest	0.0	Failed

T-test for Cost

```
. ttest cost,by(treat) unequal

Two-sample t test with unequal variances
---------------------------------------------------------------
   Group | Obs    Mean  Std. Err  Std. Dev  [95% Conf. Interval]
---------+-----------------------------------------------------
       0 | 250    3015  100.1052  1582.802  2817.839    3212.161
       1 | 250    3040  73.91742  1168.737  2894.417    3185.583
---------+-----------------------------------------------------
combined | 500  3027.5  62.15917  1389.921  2905.374    3149.626
---------+-----------------------------------------------------
    diff |         -25  124.4381           -269.5399    219.5399
---------------------------------------------------------------
    diff = mean(0) - mean(1)                        t =  -0.2009
Ho: diff = 0       Satterthwaite's degrees of freedom =  458.304

    Ha: diff < 0          Ha: diff != 0           Ha: diff > 0
 Pr(T < t) = 0.4204  Pr(|T| > |t|) = 0.8409   Pr(T > t) = 0.5796
```

Responses To Violation Of Normality Assumption

- Adopt nonparametric tests of other characteristics of the distribution that are not as affected by the nonnormality of the distribution (“biostatistical” approach)
- Transform the data so they approximate a normal distribution (“classic econometric” approach)
- Adopt tests of arithmetic means that avoid parametric assumptions (most recent development)
- OBSERVATION: If we abandon statistical testing of the arithmetic mean because distributional assumptions of the t-test are violated, does not imply that we are not interested in differences in the arithmetic mean

Response 1: Non-parametric Tests of Other Characteristics of the Distribution

- Rationale: Can analyze the characteristics that are not as affected by the nonnormality of the distribution
  - Wilcoxon rank-sum test
  - Kolmogorov-Smirnov test

Potential Problem with Testing Other Characteristics of the Distribution

- Tests indicate that some measure of the cost distribution differs between the treatment groups, such as its shape or location, but not necessarily that the arithmetic means differ
- The resulting p-values need not be applicable to the arithmetic mean
- While we might decide to compare cost by use of tests like the Mann-Whitney U test, the numerator and denominator of the cost-effectiveness ratio should never be represented as a difference in median cost or effect

Response 2: Transform the Data

- 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

Estimates and Inferences Not Necessarily Applicable to Arithmetic Mean

- Goal is to 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, which is a biased estimate of the arithmetic mean
    - Need to apply smearing factor
  - Inference: On the retransformed scale, inferences about the log of costs translate into inferences about differences in the geometric mean rather than the arithmetic mean

Primer on the Log Transformation of Cost

Raw Cost	Group 2	Group 3
Obs: 1	15	35
2	45	45
3	75	55
Arithmetic mean	45	45
Log of arithmetic mean	3.806662	3.806662
Geometric mean $\sqrt[n]{\prod Y}$	36.993	44.247
Log Cost		
Obs: 1	2.70805	3.555348
2	3.806662	3.806662
3	4.317488	4.007333
Arithmetic mean of logs	3.610734	3.789781
$\text{Exp}^{(\text{mean ln})}$	36.993	44.247

Data taken from Glick HA, Doshi JA, Sonnad SS, Polsky D. chapter 5 in Economic Evaluation in Clinical Trials, 2007.

Primer On The Log Transformation Of Costs

- 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
  - All else equal, the greater the variance, the skewness, or kurtosis, the greater the downward bias of the exponentiated mean of the logs
  - e.g., $(25 * 30 * 35)^{0.333} = 29.7196$
    $(10 * 30 * 50)^{0.333} = 24.6621$
    $(5 * 30 * 55)^{0.333} = 20.2062$
    $(1 * 30 * 59)^{0.333} = 12.0664$
- “Smearing” factor attempts to eliminate bias from simple exponentiation of the mean of the logs

Retransformation Of The Log Of Cost (I)

- Duan's common smearing factor:

$$\Phi = \frac{1}{N}\sum_{i=1}^{N} e^{(Z_i - \hat{Z}_i)}$$

where in univariate analysis, $\hat{Z}_i$ = the group mean

- Common smearing factor equals the mean of the exponentiation of the log residuals
- Most appropriate when treatment group variances are equivalent

Retransformation Of The Log Of Cost (II)

Group	Observ	ln	$z_i - \hat{z}_i$	$e^{(z_i - \hat{z}_i)}$
2	1	2.708050	-0.9026834	0.4054801
2	2	3.806663	0.1959289	1.216440
2	3	4.317488	0.7067545	2.027401
Mean, 2	--	3.610734	--	--
3	1	3.555348	-0.2344332	0.7910191
3	2	3.806663	0.0168812	1.017025
3	3	4.007333	0.2175519	1.24303
Mean, 3	--	3.789781	--	--
Smear				**1.116732** Φ

Common Smearing Retransformation (I)

- Retransformation formula

$$E(\bar{Y}_2) = \Phi\, e^{(\bar{Z}_2)}$$

$$E(\bar{Y}_3) = \Phi\, e^{(\bar{Z}_3)}$$

- Retransformation

Group	Φ	$e^{(\ln)}$	Predicted cost
2	1.116732	36.993	41.3
3	1.116732	44.247	49.4

Common Smearing Retransformation (II)

- Why are the retransformed subgroup-specific means -- 41.3 and 49.4 -- so different from the untransformed subgroup means of 45?
- Because the standard deviations of the subgroups' logs are substantially different

$$SD_2 = 0.8224;\ SD_3 = 0.2265$$

- The larger standard deviation for group 2 implies that compared with the arithmetic mean, its geometric mean has greater downward bias than does the geometric mean for group 3
- Thus, multiplication of the 2 groups' geometric means by a common smearing factor cannot give accurate estimates for both groups' arithmetic means

Log Transformations and Normal Assumptions

- 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

Subgroup-specific Smearing Factors (I)

- Manning has shown that in the face of heteroscedasticity – i.e., differences in variance -- use of a common smearing factor in the retransformation of the predicted log of costs yields biased estimates of predicted costs
- We obtain unbiased estimates by use of subgroup-specific smearing factors
- Manning's subgroup-specific smearing factor:

$$\Phi_j = \frac{1}{N_j} \sum_{i=1}^{N_j} e^{(Z_{ij} - \hat{Z}_j)}$$

Subgroup-specific Smearing Factors (II)

Group	Observ	ln	$z_i - \hat{z}_i$	$e^{(z_i - \hat{z}_i)}$	
2	1	2.708050	-0.9026834	0.4054801	
2	2	3.806663	0.1959289	1.216440	
2	3	4.317488	0.7067545	2.027401	
Mean, 2	--	3.610734	--	**1.21644**	Φ_2
3	1	3.555348	-0.2344332	0.7910191	
3	2	3.806663	0.0168812	1.017025	
3	3	4.007333	0.2175519	1.24303	
Mean, 3	--	3.789781	--	**1.0170245**	Φ_3

Subgroup-specific Smearing Retransformation (I)

- Retransformation formulas

$$E(\bar{Y}_2) = \Phi_2 \, e^{(\bar{Z}_2)}$$

$$E(\bar{Y}_3) = \Phi_3 \, e^{(\bar{Z}_3)}$$

- Retransformation

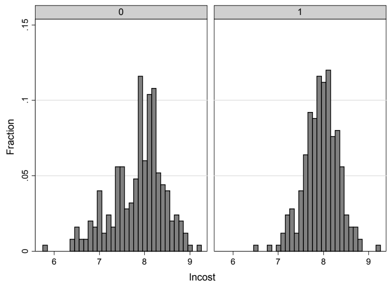
Group	Φ_i	$e^{(\ln)}$	Predicted cost
2	1.21644	36.993	45.00
3	1.0170245	44.247	45.00

Subgroup-specific Smearing Retransformation (II)

- 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 we cannot assume that p-value derived for the log of cost applies to the arithmetic mean of cost

Potential Problems with Testing Transformation of the Data (I)

• Log transformation doesn't always result in normality



P- value for normality = 0.002 and p=0.01 for the two groups

Potential Problems with Testing Transformation of the Data (II)

- When we use a t-test to evaluate log cost, the resulting p-value has direct applicability to the difference in the log of cost
- It generally also applies to the difference in the geometric mean of cost (i.e., we see similar p-values for the difference in the log and the difference in the geometric mean)
- The p-value for the log may or may not be directly applicable to the difference in arithmetic mean of untransformed cost

Potential Problems with Testing Transformation of the Data (III)

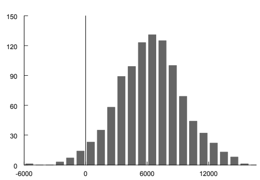
- Whether the p-value for the log is applicable to the difference in the arithmetic mean of untransformed cost depends on whether the two distributions of the log are normal and whether they have equal variance and thus standard deviation
  - If log cost is normally distributed and if the variances are equal, inferences about the difference in log cost are generally applicable to the difference in arithmetic mean cost
  - If log cost is normally distributed and if the variances are unequal, inferences about the difference in log cost generally will not be applicable to the difference in arithmetic mean cost

Potential Problems with Testing Transformation of the Data (IV)

- 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”

Response 3: Tests of Means that Avoid Parametric Assumptions

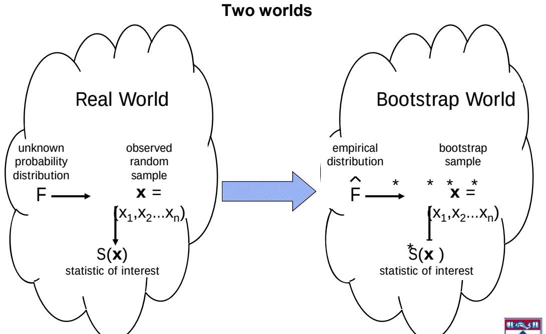
- Bootstrap estimates the distribution of the observed difference in arithmetic mean costs



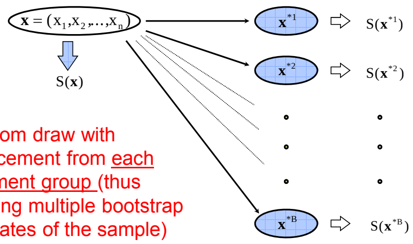
- Yields a test of how likely it is that 0 is included in this distribution (by evaluating the probability that the observed difference in means is significantly different from 0)

Bootstrap Simulation

Two worlds

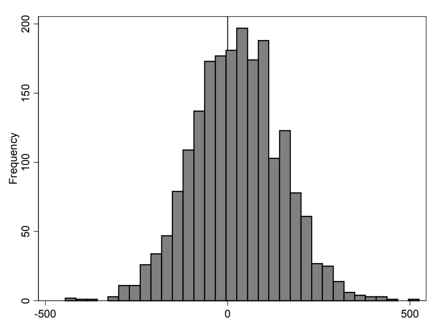


Bootstrap Simulation



- Random draw with replacement from each treatment group (thus creating multiple bootstrap replicates of the sample)
- Calculate the difference in the mean between the two treatment groups for each bootstrap replicate

Histogram of Bootstrap Results



Data Example: Distribution of Costs, Chapter 5

	Group 0	Group 1
Arith Mean	3015	3040
Std. Dev.	1582.802	1168.737
Quantiles		
5%	899	1426
25%	1819	2226
50%	2825.5	2900.5
75%	3752	3604
95%	6103	5085
Skewness	1.03501	1.525386
Kurtosis	4.910192	9.234913
Geom Mean	2600.571	2835.971
Mean ln	7.8634864	7.9501397
SD ln	.57602998	.37871479
Obs	250	250

Univariate Analysis with STATA

- Provide a log file with full set of commands for all types of (appropriate & inappropriate) univariate tests in STATA
- Provide documentation for bootstrap when we perform multivariable analysis of cost below
- In the next slide, we summarize the results of the univariate tests using STATA

Results from Univariate Analysis of cost

	Plac	Act	Diff	P-val	-- 95% CI --
Mean cost:	3015	3040	25	0.8409	-220 to 270
Median cost:	2826	2901	75	0.3722	
Kolm-Smirn:				0.0017	
Log cost					
Common SD:	2901	3164	263	0.0475	
Heterosk:				0.0000	
Bootstrap					
Nonparamet:				0.8050	-210 to 265
Parametric:				0.8371	-214 to 264

Why Do Different Statistical Tests Lead To Different Inferences?

- The tests are evaluating differences in different statistics
  - T-test of untransformed costs indicates we cannot infer that the arithmetic means are different
  - Wilcoxon rank-sum test also leads to the same inference, but its p-value relates more to the probability that the medians differ
  - Kolmogorov-Smirnov test indicates we can infer that the distributions are different
  - T-test of log costs indicates we can infer that the mean of the logs are different, and thus the geometric means of cost are different
  - Bootstrap leads to same (lack of) inference as t-test and does not make the normality assumption

Univariate Analysis: Summary/Conclusion (I)

- Cost-effectiveness ratios (ΔC /ΔE) and NMB ([WTP ΔE] - ΔC) require an estimate of ΔC and ΔE, the differences in arithmetic means
- If arithmetic means are the most meaningful summary statistic of costs, we should test for significant differences in arithmetic mean costs
  - Parametric test of means
  - Non-parametric test of means (e.g., bootstrap methods)

Univariate Analysis: Summary/Conclusion (II)

- 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

Multivariate Analysis of Costs

Multivariable Analysis Of Economic Outcomes (I)

- Even if treatment is assigned in a randomized setting use of multivariable analysis may have added benefits:
  - 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/less severe; different countries/centers)
  - 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
  - Added advantage: Helps explain what is observed (e.g., coefficients for other variables should make sense economically)

Multivariable Analysis Of Economic Outcomes (II)

- 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 Cost

- Common Techniques
  - Ordinary least squares regression predicting costs after randomization (OLS)
  - Ordinary least squares regression predicting the log transformation of costs after randomization (log OLS)
  - Generalized Linear Models (GLM)
- Other Techniques:
  - Generalized Gamma regression (Manning et al. Journal of Health Economics 2005)
  - 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)

Multivariate Analysis with STATA: Outline

- Estimate of adjusted mean difference in costs
  - Start with everyone's "old" favorite: OLS
  - Check the fit of the gauss family used in OLS
    - Revise family if necessary
  - Start with everyone's "new" favorite: GLM gamma/log
  - Check the fit of the gamma family
    - Revise family if necessary
  - Tune the link
- P-values and confidence intervals for the adjusted mean difference in costs using bootstrapping
  - Parametric tests
  - Non-parametric tests

Ordinary Least Squares (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 set
    - In large datasets with extreme observations
  - Can produce predictions with negative costs

```
regress cost treat dissev blcost blqaly race
```

					F(6,493)	34.09
Source	SS	Df	MS		Prob>F	0.0000
Model	2473e+5	5	416e+5		R-squared	0.2565
Resid	7167e+5	494	145e+4		Adj R-sq	0.2490
Total	9640e+5	499	193e+4		Root MSE	1204.5

Cost	Coef	Std Err	T	P>|t|	[95% Conf Int]	
treat	**21.993**	**107.77**	**0.20**	**0.838**	**-189.74**	**233.74**
dissev	4033.41	516.34	7.81	0.000	3018.92	5047.91
blcost	.3945	.0758	5.20	0.000	0.2455	0.5435
blqaly	-773.30	371.98	-2.08	0.038	-1504.16	-42.45
race	-768.02	118.75	-6.47	0.000	-1001.35	-534.69
_cons	1966.32	366.11	5.37	0.000	1247.00	2685.64

eeict1.dta

Predicted Cost

- Coefficient from OLS (21.99) equals predicted cost difference
- Alternatively, can use mean values for the other explanatory variables and calculate the difference in the predictions for treat = 0 and one for treat = 1:
  - **Control**: 1966.32+(.347*4033.41)+(1634.86*.3945)-((.786*773.30) + (.506*768.02)) = **3014.43**
  - **Treatment**: 1966.32+(.347*4033.41)+(1634.86*.3945) - ((.786*773.30)+(.506*768.02))+21.99 = **3036.42**

3036.42 - 3014.43 = **21.99**

Don't Take Means of Individual Predictions

- Don't predict cost for each individual and take means:

```
predict olscost
tab treat, sum(olscost)
```

Treatment group	Summary of Fitted values Mean	Std. Dev	Freq
0	3015	746.32806	250
1	3040	660.17754	250
Total	3027.5	703.97565	500

3040 - 3015 = 25 ≠ 21.99

- This method re-introduces the covariate imbalance that OLS was meant to eliminate

Method of Recycled Predictions

- Alternative method of using the mean values for the explanatory variables is to use the method of recycled predictions
  - i.e., alternative method for obtaining $\beta_i \bar{X}_i$
- To recycle predictions, code everyone as if they were in treatment group 0 and make a prediction; then code everyone as if they were in treatment group 1 and make a second prediction

```
gen temp=treat
regress cost temp dissev blcost blqaly race
replace temp=0
predict olscost0
replace temp=1
predict olscost1
```

Results of Recycled Predictions

```
sum olscost0 olscost1

Variable | Obs      Mean Std. Dev.     Min      Max
---------+---------------------------------------
olscost0 | 500  3016.503  703.866 1184.116 5527.065
olscost1 | 500  3038.497  703.866 1206.109 5549.059
```

3038.497 - 3016.503 = 21.99 *

- Recycled predictions are simply another way to use the sample means for the covariates but at the same time make patient-level predictions

* Differences between this method and multiplication of sample-wide means times the coefficients due to rounding

Generalized Linear Models (GLM)

- OLS can be run as a generalized linear model
- Rerunning as a GLM facilitates comparison of model fit to the fit of other model specifications
- GLM model has the advantages of the log model, but
  - Doesn't require normality or homoscedasticity,
  - Evaluates a transformation of the difference in arithmetic mean cost, not a transformation of individual patient level costs
  - Doesn't raise problems related to retransformation from the scale of estimation to the raw scale
- To run a GLM, must identify a "link function" and a "family" (based on the data)

The Link Function

- Link function directly characterizes how the linear combination of the predictors is related to the prediction on the original scale
- Examples of links include:
  - Identity Link: $\hat{Y}_j = \sum_i \beta_i X_{ij}$
  - Log link: $\hat{Y}_j = \exp^{(\sum_i \beta_i X_{ij})}$
- Availability of alternative links relaxes linearity assumption
  - $E(y/x) = \Sigma\beta_i X_i$ (OLS)
  - $E(\ln(y)/x)=\Sigma\beta_i X_i$ (log OLS)

Which link is used by OLS?

Family

- Specifies distribution that reflects mean-variance relationship
- Currently, families for continuous data available in Stata include:
  - Gaussian (constant variance)
  - Poisson (variance proportional to mean)
  - Gamma (variance proportional to square of mean)
  - Inverse gaussian (variance proportional to cube of mean)
- Availability of the poisson, gamma, and inverse Gausian families relaxes assumption of constant variance

Which family is used by OLS?

Rerun OLS as GLM With Identity Link and Gauss Family

```
replace temp=treat

glm cost temp dissev blc blq race,link(identity)
family(gauss)
```

General syntax:
glm [depvar] [indepvars] [if xxx],link(xxx) family(xxx)

rchapter5.dta

```
glm cost temp dissev blcost blqaly race,
      link(identity) family(gauss)
```

Variance function: V(u) = 1 [Gaussian]
Link function: g(u) = u [Identity]
Log likelihood = -4253 AIC 17.037
BIC 7.17e+08

cost	Coef	Std Err	z	P>|z|	95% CI	
temp	**21.99324**	**107.7662**	**0.20**	**0.838**	**-189.2247**	**233.2112**
dissev	4033.414	516.3404	7.81	0.000	3021.406	5045.423
blcost	.3944632	.0758403	5.20	0.000	.2458189	.5431076
blqaly	-773.301	371.9785	-2.08	0.038	-1502366	-44.23705
race	-768.020	118.7549	-6.47	0.000	-1000.775	-535.2645
_cons	1966.319	366.1061	5.37	0.000	1248.765	2683.874

eeict1.dta

Predicted Cost

- As with OLS, coefficient from GLM, identify link, gauss family (21.99) equals predicted cost difference
- As with OLS, can use mean values for the other explanatory variables and calculate one difference in the predictions for treat = 0 and another for treat = 1
- As with OLS, can use recycled predictions

Identity/Gauss Recycled Predictions

```
gen temp=treat
glm cost temp dissev blcost blqaly race,
  link(identity) family(gauss)
replace temp=0
predict polsc0
replace temp=1
predict polsc1

sum polsc*
```

Variable	Obs	Mean	Std. Dev.	Min	Max
polsc0	500	3016.503	703.866	1184.116	5527.065
polsc1	500	3038.497	703.866	1206.109	5549.059

DIFFERENCE: 22

Are We Using the Correct Family?

- The modified Parks test is a “constructive” test that recommends a family given a particular link function
- This test is included in the program we’ve titled glmdiag which is loaded by the following command:

```
do glmdiagnostic
```

- To perform the test, we first run the glm model and then run glmdiag:

```
replace temp=treat
 glm cost temp dissev blcost blqaly race,
link(identity) family(gauss)
 glmdiag
```

GLM Diagnostics, Identity/Gaussian

FITTED MODEL: Link = Identity ; Family = Gaussian

Results, Modified Park Test (for Family)

Coefficient: 1.391784

Family, Chi2, and p-value in descending order of likelihood

Family	Chi2	P-value
Poisson:	**1.4021**	**0.2364**
Gamma:	3.3790	0.0660
Gaussian NLLS:	17.6936	0.0000
Inverse Gaussian or Wald	23.6244	0.0000

Results of tests of GLM Identity link

Test		P-value
Pearson Correlation Test:		1.0000
Pregibon Link Test:		0.8913
Modified Hosmer and Lemeshow:		0.3487

eeict1.dta

GLMDIAG Saved Results

```
. return list
scalars:
           r(ln_coef) =  1.391784
          r(p_family) =  .2364
            r(p_gaus) =  .000026
            r(p_pois) =  .2363797
             r(p_gam) =  .0660326
           r(p_igaus) =  1.20000000000e-06
                r(N) =  500
         r(p_pearson) =  1
        r(p_pregibon) =  .8913000000000001
             r(p_h_m) =  .3487
                r(ll) =  -4253.36394877669
               r(aic) =  17.03745579510676
               r(bic) =  716710494.4875774
          r(deviance) =  716713564.503978
macros:
            r(family) : "poisson"
```

```
glm cost temp dissev blcost blqaly
race, link(identity) family(poisson)
```

Variance function: V(u) = u [Poisson]
Link function: g(u) = u [Identity]

Log likelihood = -113576
AIC 454.33
BIC 219210

cost	Coef	Std Err	z	P>|z|	95% CI	
temp	113.1149	4.798526	23.57	**0.000**	103.71	122.52
dissev	4008.434	22.67209	176.80	**0.000**	3964.00	4052.87
blcost	.3861272	.0036013	107.22	**0.000**	.3791	.3932
blqaly	-765.3726	16.58928	-46.14	**0.000**	-797.89	-732.86
race	-746.5739	5.324134	140.22	**0.000**	-757.01	-736.14
_cons	1925.985	16.49097	116.79	**0.000**	1893.664	1958.307

- PROBLEM WITH P-VALUES?

eeict1.dta

GLM Diagnostics, Identity/Poisson

FITTED MODEL: Link = Identity ; Family = Poisson

Results, Modified Park Test (for Family)

Coefficient: 1.436638

Family, Chi2, and p-value in descending order of likelihood

Family	Chi2	P-value
Poisson:	**1.7001**	**0.1923**
Gamma:	2.8301	0.0925
Gaussian NLLS:	18.4046	0.0000
Inverse Gaussian or Wald	21.7947	0.0000

Results of tests of GLM Identity link

Test		P-value
Pearson Correlation Test:		0.8818
Pregibon Link Test:		0.7021
Modified Hosmer and Lemeshow:		0.5134

rchapter5.dta

Predicted Cost

- As with OLS, coefficient from GLM, identify link, poisson family (113.11) equals predicted cost difference
- As with OLS, can use mean values for the other explanatory variables and calculate one difference in the predictions for treat = 0 and another for treat = 1
- As with OLS, can use recycled predictions
- Unlike OLS, standard errors for poisson family are wrong (we'll need to bootstrap the model if we want reasonable standard errors)

Identity/Poisson Recycled Predictions

```
glm cost temp dissev blcost blqaly race,link(identity)
  family(poisson)
replace temp=0
predict ppoisc0
replace temp=1
predict ppoisc1

sum ppoisc*
```

Variable	Obs	Mean	Std. Dev.	Min	Max
ppoisc0	500	2970.943	691.9996	1162.989	5450.039
ppoisc1	500	3084.057	691.9996	1276.104	5563.153

DIFFERENCE: 113

Change in Family Leads to Fairly Big Differences in Point Estimate (Not Sure About SE)

Cost	Coef.	Std Err	z	P>|z|	[95% Conf Interval]	
Gaussian / Identity						
temp	21.99	107.77	0.20	0.838	-189.2247	233.2112
Poisson / Identity						
temp	113.11	**4.80**	**23.57**	**0.000**	103.71	122.5198

- Change in family not "supposed" to affect coefficient dramatically (consistency)
- Change in coefficient may be due to:
  - Lack of significance of coefficients
  - Incorrect specification of link or covariates

Suppose We Started with GLM Log/Gamma

- Log link more commonly used in literature than identity link
- When we adopt the log link, we are assuming:

  $$\ln(E(y/x))=X\beta$$

- GLM with a log link differs from log OLS in part because in log OLS, we are assuming:

  $$E(\ln(y)/x)=X\beta$$

- $\ln(E(y/x) \neq E(\ln(y)/x)$

  i.e. log of the mean ≠ mean of the log costs

ln(E(y/x) ≠ E(ln(y)/x)

Variable	Group 2	Group 3
Observations		
1	15	35
2	45	45
3	75	55
Arithmetic mean	45	45
Log, arith mean cost	3.806662	3.806662 *
Natural log		
1	2.70805	3.555348
2	3.806662	3.806662
3	4.317488	4.007333
Arith mean, log cost	3.610734	3.789781 †

* Difference = 0; † Difference = 0.179047

Comparison of Results of GLM Gamma/Log and log OLS Regression

Variable	Coefficient	SE	z/T	*p* value
GLM, gamma family, log link				
Group 2	0.000000	0.405730	0.00	1.000
Constant	3.806662	0.286894	13.27	0.000
Log OLS				
Group 2	0.179048	0.492494	0.36	0.74
Constant	3.610734	0.348246	10.32	0.000

`glm cost temp dissev blcost blqaly race, link(log) family(gamma)`

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

Log likelihood = -4494.155729
AIC 18.00062
BIC -2988.518

cost	Coef	Std Err	z	P>|z|	95% CI	
temp	.0446782	.0356359	1.25	0.210	-.0251669	.1145232
dissev	1.409376	.1739606	8.10	0.000	1.06842	1.750333
blcost	.000122	.0000257	4.78	0.000	.0000724	.0017300
blqaly	-.2579657	.1223431	-2.11	0.035	-.4977537	-.0183796
race	-.2613111	.0395492	-6.61	0.000	-.3388262	-.1837961
_cons	7.610573	.1220851	62.34	0.000	7.371291	7.849856

eeict1.dta

Retransformation

- GLM avoids the problem that simple exponentiation of the results of log OLS yields biased estimates of predicted costs
- For the identity link, as for OLS, the coefficient represents the incremental cost
- For other (nonlinear) links such as the log, 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

Predicted Cost

- Coefficient from GLM, log link, gamma family (.0447) does not equal predicted cost difference
- Cannot use mean values for the other explanatory variables and calculate one difference in the predictions for treat = 0 and another for treat = 1
  - The mean of nonlinear retransformations does not equal the nonlinear retransformation of the mean
- Can use recycled predictions to create an identical covariate structure for the two groups

Log/Gamma Recycled Predictions

```
replace temp=0
predict pglmglc0
replace temp=1
predict pglmglc1

sum pglmglc*

 Variable| Obs     Mean Std. Dev.     Min        Max
---------+-----------------------------------------
pglmglc0 | 500  2964.034  733.7266 1542.916  6767.186
pglmglc1 | 500  3099.465  767.2515 1613.414  7076.388
```

DIFFERENCE: 135

Recycled vs Treatment-Specific Predictions

```
. replace temp=treat
. quietly glm cost temp dissev blcost blqaly race,
  link(log) family(gamma)
. predict pcost
(option mu assumed; predicted mean cost)
. tab treat,sum(pcost)
Treatment |     Summary of predicted mean cost
    group |       Mean   Std. Dev.       Freq.
----------+--------------------------------
        0 | 2973.8331   789.66446        250
        1 | 3089.2184   705.44167        250
----------+--------------------------------
    Total | 3031.5257   750.21371        500
```

DIFFERENCE: 115

Recycled vs Treatment-Specific Predictions (II)

- Difference between mean of the recycled predictions (135) and mean of treatment group-specific predictions (115) due to whether or not covariates are balanced
- Given the log link is a multiplicative model, If we want to hold all-else equal during both estimation AND prediction, must use method of recycled predictions

Is Gamma the Correct Family for Log Link?

FITTED MODEL: Link = Log ; Family = Gamma

Results, Modified Park Test (for Family)

Coefficient: 1.5912

Family, Chi2, and p-value in descending order of likelihood

Family	Chi2	P-value
Gamma:	**1.9560**	**0.1619**
Poisson:	4.0897	0.0431
Inverse Gaussian or Wald	23.2272	0.0000
Gaussian NLLS:	29.6281	0.0000

Results of tests of GLM Log link

Pearson Correlation Test:	.2460
Pregibon Link Test:	.1273
Modified Hosmer and Lemeshow:	.6199

eeict1.dta

What's the Appropriate Link?

- So far we have evaluated the identity link (with an "optimized" poisson family) and the log link (with an "optimized" gamma family)
- But what link should we use?

Selecting a Link Function

- There is no single test that identifies the appropriate link
- Instead can employ multiple tests of fit
  - Pregibon link test checks linearity of response on scale of estimation
  - Modified Hosmer and Lemeshow test checks for systematic bias in fit on raw scale
  - Pearson's correlation test checks for systematic bias in fit on raw scale
- Ideally, all 3 tests – which are also reported by glmdiag – will yield nonsignificant p-values

Rerun Identity/Poisson and Assess Fit Statistics

```
replace temp=treat

glm cost temp dissev blcost blqaly race,
link(identity) family(poisson)

glmdiag
```

rchapter5.dta

GLM Diagnostics, Identity/Poisson

FITTED MODEL: Link = Identity ; Family = Poisson

Results, Modified Park Test (for Family)

Coefficient: 1.436638

Family, Chi2, and p-value in descending order of likelihood

Family	Chi2	P-value
Poisson:	1.7001	0.1923
Gamma:	2.8301	0.0925
Gaussian NLLS:	18.4046	0.0000
Inverse Gaussian or Wald	21.7947	0.0000

Results of tests of GLM Identity link

Pearson Correlation Test:	**0.8818**
Pregibon Link Test:	**0.7021**
Modified Hosmer and Lemeshow:	**0.5134**

rchapter5.dta

Rerun Log/Gamma and Assess Fit Statistics

```
glm cost temp dissev bl* race, link(log)
 family(gamma)
  glmdiag
```

Run GLM DIAGNOSTICS, Gamma/Log

FITTED MODEL: Link = Log ; Family = Gamma

Results, Modified Park Test (for Family)

Coefficient: 1.5912

Family, Chi2, and p-value in descending order of likelihood

Family	Chi2	P-value
Gamma:	1.9560	0.1619
Poisson:	4.0897	0.0431
Inverse Gaussian or Wald	23.2272	0.0000
Gaussian NLLS:	29.6281	0.0000

Results of tests of GLM Log link

Pearson Correlation Test:	**.2460**
Pregibon Link Test:	**.1273**
Modified Hosmer and Lemeshow:	**.6199**

rchapter5.dta

Goodness of Fit Statistics

Test	Ident/Pois	Log/Gam
Pearson Correlation Test:	**0.8818**	.2460
Pregibon Link Test:	**0.7021**	.1273
Modified Hosmer and Lemeshow	0.5134	**.6199**

- Neither link dominates the other (less significant fit statistics for all 3 tests) and we haven't fully worked out how to trade-off among the tests, but identity/poisson model appears better than log/gamma model
- But can we improve the link?

rchapter5.dta

Can We Improve the Link?

- Stata's power link provides a flexible link function
- It allows generation of a wide variety of named and unnamed links, e.g.,
  - power 2: $\hat{u}_i = (B_iX_i)^{0.5}$
  - power 1 = Identity link; $\hat{u}_i = B_iX_i$
  - power .5 = Square root link; $\hat{u}_i = (B_iX_i)^2$
  - power .25: $\hat{u}_i = (B_iX_i)^4$
  - power 0 = log link; $\hat{u}_i = \exp(BiXi)$
  - power -1 = reciprocal link; $\hat{u}_i = (B_iX_i)^{-1}$
  - power -2 = inverse quadratic; $\hat{u}_i = (B_iX_i)^{-0.5}$

Can We Improve the Link? (2)

- Iteratively evaluate power links (in 0.1 intervals) between -2 and 2
  - Use the modified Park test to select a family
  - Rerun the GLM with the power and preferred link
  - Evaluate the fit statistics
  - Don't show you the results here, but we then fine tune the power link in 0.01 intervals within candidate regions of the power link

Power 0.65 Link / Poisson Family

Power 0.65 Link / Poisson Family

```
replace temp=treat

glm cost temp dissev blcost blqaly race,
link(power .65) family(poisson)
```

eeict1.dta

```
glm cost temp dissev blcost blqaly race,
    link(power .65) family(poisson)
```

Variance function: V(u) = u	[Poisson]
Link function: g(u) = u^(.65)	[Power]
Log likelihood = -113515.3	AIC 454.0853 BIC 219088.2

Cost	Coef	Std Err	z	P>|z|	95% CI	
temp	3.493932	.1927675	18.13	**0.000**	3.116115	3.87175
dissev	161.4855	.9285280	173.92	**0.000**	159.6656	163.3053
blcost	.0150344	.0001392	107.97	**0.000**	.0147615	.0153073
blqaly	-30.51369	.6645974	-45.91	**0.000**	-31.81628	-29.21111
race	-30.27	.2133011	-141.91	**0.000**	-30.68807	-29.85194
_cons	138.8326	.6584566	210.85	**0.000**	137.542	140.1231

eeict1.dta

Run GLM DIAGNOSTICS, .65/Poisson

FITTED MODEL: Link = Power .65 ; Family = Poisson

Results, Modified Park Test (for Family)

Coefficient: 1.495248

Family, Chi2, and p-value in descending order of likelihood

Family	Chi2	P-value
Poisson:	2.3212	0.1276
Gamma:	2.4111	0.1205
Gaussian NLLS:	21.1587	0.0000
Inverse Gaussian or Wald:	21.4285	0.0000

Results of tests of GLM Log link

Pearson Correlation Test:	**.9027**
Pregibon Link Test:	**.7469**
Modified Hosmer and Lemeshow:	**.5870**

rchapter5.dta

Power 0.65/Poisson Recycled Predictions

```
replace temp=0
predict pglmppc0
replace temp=1
predict pglmppc1
sum pglmppc*
 Variable |  Obs      Mean  Std. Dev.       Min       Max
----------+---------------------------------------------
pglmppc0  |  500  2983.316  704.3185  1338.796  5804.318
pglmppc1  |  500  3071.642  711.5133  1406.172  5916.306
```

DIFFERENCE: 88

Summary: GLM Analysis of Cost

	Id/Gau	Id/Pois	Log/Gam	0.65/Pois
Pearson	1.0000	0.8818	0.2460	0.9027
Pregibon	0.8913	0.7021	0.1273	0.7469
Mod H&L	0.3487	0.5134	0.6199	0.5870
Summary	0.4360	0.3394	1.4746	0.2441
Difference	22	113	135	88
P-value	0.84	0.26*	0.21	0.39*

* P-value derived from bootstrap (discussed next)

Bootstrapping the Multivariable Models

- Random draw with replacement from each treatment group, thus creating multiple bootstrap samples (also referred to as replicates)
- We bootstrap these models primarily to estimate nonparameteric p-values and CI on the cost (and QALY) scale AND to calculate standard errors for parametric tests
- In what follows, we use Stata's most basic bootstrap command, bsample

Structure of the Bootstrap

- Create a dataset to store estimates (bsmvpred.dta)
  - Each observation in the dataset represents the results from a separate bootstrap replicate
- Create a loop that will draw bootstrapped samples
  - Loop N times (we commonly use 2-3000 replicates, but in the current example we set N to 200
- Within each bootstrap sample:
  - Run the GLMs
  - Use method of recycled predictions to predict cost
  - Estimated the predicted means
  - Keep 1 observation; create variables that represent the predicted means; append the means to the dataset created to store the bootstrap results

bsmultiv.do

- We've provided the bootstrap program **bsmultiv.do** (listed in the appendix to these slides)
- **bsmultiv.do** is a purpose-built bootstrap program for the current dataset which estimates the 6 glm models we evaluated above in multiple bootstrapped datasets
- Current program set at 200 replicates (to save time in class), but 1000-3000 replicates recommended
- You can modify this program for your own dataset

Selected Boostrap Replicates, bsmvpred.dta

	pglmigc0	pglmigc1	pglmipc0	pglmipc1	pglmlgc0	pglmlgc1
1.	3108.104	3086.328	3061.173	3133.259	3055.328	3141.872
2.	2874.748	2848.656	2822.54	2900.865	2820.564	2906.826
3.	3046.532	3050.864	3002.845	3094.551	2998.789	3099.822
4.	2981.5	3046.561	2943.12	3084.94	2936.017	3115.354
5.	2947.865	3088.323	2897.962	3138.226	2887.306	3157.838
6.	2991.154	3111.779	2960.855	3142.076	2955.8	3164.625
7.	2922.351	2956.017	2868.459	3009.909	2869.413	3020.215
8.	3126.857	3076.667	3078.587	3124.937	3075.845	3140.477
9.	2978.72	2997.372	2957.52	3018.572	2963.5	3025.937
10.	3077.117	2985.335	3037.682	3024.77	3024.759	3046.886
11.	3103.544	3119.24	3066.059	3156.725	3049.399	3172.307
12.	2935.81	2977.378	2897.605	3015.583	2874.975	3044.03
13.	2919.418	2900.594	2874.812	2945.201	2868.466	2951.305

Summarize (3000 Draws), bsmvpred.dta

Variable	Obs	Mean	Std. Dev.	Min	Max
pglmigc0	3000	3017.178	90.61395	2719.115	3409.561
pglmigc1	3000	3038.906	71.1834	2789.372	3309.99
pglmipc0	3000	2972.174	87.65356	2662.726	3357.739
pglmipc1	3000	3083.909	70.30924	2820.354	3353.042
pglmlgc0	3000	2963.875	88.59197	2654.125	3350.047
pglmlgc1	3000	3099.931	73.44394	2834.418	3388.661
pglmppc0	3000	2984.217	88.75463	2677.809	3373.078
pglmppc1	3000	3071.829	70.75923	2811.923	3345.69
pglmigq0	3000	.5733505	.0135277	.5249925	.6199619
pglmigq1	3000	.6147949	.012695	.5737574	.6603948
pglmipq0	3000	.5733622	.0134999	.524086	.6191651
pglmipq1	3000	.6147833	.0126267	.5750274	.6578885
pglmppq0	3000	.5737368	.0134739	.5234635	.6183268
pglmppq1	3000	.6144159	.0125472	.5737772	.6558303

Summarize Differences

pglmigcd	3000	21.72763	106.684	-359.7065	359.5251
pglmipcd	3000	111.7347	100.3923	-256.0615	426.7947
pglmlgcd	3000	136.0555	106.5739	-237.0095	456.8499
pglmppcd	3000	87.6113	102.8321	-287.0056	409.1807
pglmigqd	3000	.0414444	.0179779	-.0197337	.098896
pglmipqd	3000	.0414211	.0178393	-.01814	.100709
pglmppqd	3000	.0406791	.0176908	-.0164816	.1013637

Bootstrap: Non-parametric Tests

- P-value: count the number of replicates for which the difference is above and below 0 (yielding a 1-tailed test of the hypothesis of a cost difference)
- CI: Order the differences from highest to lowest; identify the difference for the replicates that represent the 2.5^{th} and 97.5^{th} percentiles

Calculating Nonparametric P-Value

```
. use bsmvpred
. sum pglmigcd
Variable |  Obs       Mean  Std. Dev.      Min       Max
---------+---------------------------------------------
pglmigcd |  3000  21.72763   106.684 -359.7065  359.5251
. local den=r(N)
. sum pglmigcd if pglmigcd<0
Variable |  Obs       Mean  Std. Dev.      Min       Max
---------+---------------------------------------------
pglmigcd |  1273 -76.41349  61.71498 -359.7065 -.0378418
. local num=r(N)
. local p1=`num'/`den'
. if `p1'>0.5 {
. local p1=1-`p1'
. }
. local p2=2*`p1'
. display `p2'
.8486667
```

Calculating Nonparametric CI

```
. sort pglmigcd

. list pglmigcd if _n==round((.025*3000),1)+1
                 | _n==round((.975*3000),1)

      +-----------+
      |  pglmigcd |
      |-----------|
  76. | -190.7979 |
2925. |  230.9966 |
      +-----------+
```

Bootstrap: Parametric Tests

- Because each bootstrap replicate represents a mean difference, when one sums the replicates, the reported "standard deviation" is the standard error
  - P-value: Difference in means / SE = t statistic
  - CI: Difference in means $\pm$ 1.96 SE = 95% CI

Calculating Parametric P-Value

```
Required data:
  point estimate for difference (OLS):  22
  DOF:  498
  Bootstrapped SE:

. sum pglmigcd

Variable |  Obs     Mean Std. Dev       Min      Max
---------+------------------------------------------
pglmigcd | 3000 21.72763 106.684  -359.7065 359.5251

. local se=r(sd)

. display 2*ttail(494,(22/`se'))
.83670679
```

Calculating Parametric CI

```
Required data:
  point estimate for difference (OLS):  22
  T-statistic for DOF = 498
  Bootstrapped SE:  106.684

. local tstat=invttail(494,.025)

. display `tstat'
1.9647777

. display 22-(`tstat'*`se')
-187.61038

. display 22+(`tstat'*`se')
231.61038
```

Results (Principal Analysis and Bootstrapped SE)

GLM Link/fam	PE	P-val GLM	BS SE	P-val BS	Nonpar 95% CI	Par 95% CI
Cost						
id/gau	22	.838	107	.837	-191 to 231	-188 to 232
id/pois	113	.000	100	.264	-84 to 310	-86 to 312
log/gam	135	.210	107	.208	-74 to 344	-76 to 346
pow/pois	**88**	**.000**	**103**	**.393**	**-114 to 290**	**-114 to 291**
QALY						
id/gau	.0417	.024	.018	.021	.0065 to .0773	.0064 to .0770
id/pois	.0417	.465	.018	.021	.0071 to .0767	.0066 to .0768
pow/pois	**.0408**	**.4365**	**.018**	**.024**	**.0063 to .0754**	**.0060 to .0756**

Extended Estimating Equations

- Basu and Rathouz (2005) have proposed use of extended estimating equations (EEE) which estimate the link function and family along with the coefficients and standard errors
- Tends to need a large number of observations (thousands not hundreds) to converge
- Currently can't take the results and use them with a simple GLM command (makes bootstrapping resulting models cumbersome)

Special Cases (I)

- 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 : GLM model

Special Cases (II)

- 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

Multivariate Analysis: Summary/Conclusion

- Use mean difference in costs between treatment groups estimated from a multivariable model as the numerator for a cost-effectiveness ratio
- 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 our results to different specifications of the multivariable model

References

- Basu A. Rathouz PJ. Estimating Marginal and Incremental Effects on Health Outcomes Using Flexible Link and Variance Function Models. Biostatistics 2005;6:93-109.
- Duan N. Smearing estimate: a nonparametric retransformation method. J Am Stat Assoc. 1983;78:605-10.
- Manning WG. The logged dependent variable, heteroscedasticity, and the retransformation problem. Journal of Health Economics.1998;17:283- 295.
- Manning, WG. Mullahy J. Estimating log models: To transform or not to transform? Journal of Health Economics 2001; 20(4): 461-494.
- Manning WG, Basu A, Mullahy J. Generalized Modeling approaches to risk adjustment of skewed outcomes data. National Bureau of Economic Research, Technical Working Paper 293. http://www.nber.org/papers/T0293

Appendix 1: Bootstrap Program

Bootstrap Program: Creating bsmvpred

```
* drop _all is similar to clear, but maintains local variables, scalars, and matrices
drop _all

gen pglmigc0=.
gen pglmigc1=.
gen pglmipc0=.
gen pglmipc1=.
gen pglmlgc0=.
gen pglmlgc1=.
gen pglmppc0=.
gen pglmppc1=.
gen pglmigq0=.
gen pglmigq1=.
gen pglmipq0=.
gen pglmipq1=.
gen pglmppq0=.
gen pglmppq1=.
save bsmvpred,replace
```

Bootstrap Program: Starting the Bootstrap

```
set more off
* If you want to be able to replicate your results, set seed
*set seed 2345

* Major loop: runs N times: forvalues i=1/N {
forvalues i=1/200 {

* Displays a count every 50 iterations (to make sure something is happening)
if (`i'/50)==round((`i'/50),1) {
display `i'
}

drop _all
use rchapter5

* strata(treat): maintains sample size per group
* cluster(id): if participants had multiple observations, sample all of them
bsample,strata(treat)
```

Bootstrap Program: Cost Estimation (1)

```
gen temp=treat
quietly glm cost temp dissev bl* race,link(identity) family(gauss)
quietly replace temp=0
quietly predict pglmigc0
quietly replace temp=1
quietly predict pglmigc1

quietly replace temp=treat
quietly glm cost temp dissev bl* race,link(identity) family(poisson)
quietly replace temp=0
quietly predict pglmipc0
quietly replace temp=1
quietly predict pglmipc1
```

Bootstrap Program: Cost Estimation (2)

```
quietly replace temp=treat
quietly glm cost temp dissev bl* race male,link(log) family(gamma)
quietly replace temp=0
quietly predict pglmlgc0
quietly replace temp=1
quietly predict pglmlgc1

quietly replace temp=treat
quietly glm cost temp dis race blc blq,link(power .65) family(poisson)
quietly replace temp=0
quietly predict pglmppc0
quietly replace temp=1
quietly predict pglmppc1
```

Bootstrap Program: QALY Estimation

```
capture drop nqaly
sum qaly, meanonly
local rmax=r(max)
gen nqaly=r(max)-qaly
save temp,replace

quietly replace temp=treat
quietly regress nqaly temp dissev blc blq
quietly replace temp=0
quietly predict pglmigq0
quietly replace temp=1
quietly predict pglmigq1
quietly replace pglmigq0=`rmax'-pglmigq0
quietly replace pglmigq1=`rmax'-pglmigq1
```

Bootstrap Program: QALY Estimation

```
quietly replace temp=treat
quietly glm qaly temp dissev blcost blqaly,family(poisson) link(identity)
quietly replace temp=0
quietly predict pglmipq0
quietly replace temp=1
quietly predict pglmipq1
quietly replace pglmipq0=`rmax'-pglmipq0
quietly replace pglmipq1=`rmax'-pglmipq1

quietly replace temp=treat
quietly glm qaly temp dissev blc blq,family(poisson) link(power 1.56)
quietly replace temp=0
quietly predict pglmppq0
quietly replace temp=1
quietly predict pglmppq1
quietly replace pglmppq0=`rmax'-pglmppq0
quietly replace pglmppq1=`rmax'-pglmppq1
```

Bootstrap Program: Estimate Treatment Group Mean Costs

```
sum pglmigc0,meanonly
local pglmigc0=r(mean)
sum pglmigc1,meanonly
local pglmigc1=r(mean)
sum pglmipc0,meanonly
local pglmipc0=r(mean)
sum pglmipc1,meanonly
local pglmipc1=r(mean)
sum pglmlgc0,meanonly
local pglmlgc0=r(mean)
sum pglmlgc1,meanonly
local pglmlgc1=r(mean)
sum pglmppc0,meanonly
local pglmppc0=r(mean)
sum pglmppc1,meanonly
local pglmppc1=r(mean)
```

Bootstrap Program: Estimate Treatment Group Mean QALYS

```
sum pglmigq0,meanonly
local pglmigq0=r(mean)
sum pglmigq1,meanonly
local pglmigq1=r(mean)
sum pglmipq0,meanonly
local pglmipq0=r(mean)
sum pglmipq1,meanonly
local pglmipq1=r(mean)
sum pglmppq0,meanonly
local pglmppq0=r(mean)
sum pglmppq1,meanonly
local pglmppq1=r(mean)
```

Bootstrap Program: Keep 1 Row of Data, Substitute Group Means, Append and Save

```
quietly keep if _n==1

quietly replace pglmigc0=`pglmigc0'
quietly replace pglmigc1=`pglmigc1'
quietly replace pglmipc0=`pglmipc0'
quietly replace pglmipc1=`pglmipc1'
quietly replace pglmlgc0=`pglmlgc0'
quietly replace pglmlgc1=`pglmlgc1'
quietly replace pglmppc0=`pglmppc0'
quietly replace pglmppc1=`pglmppc1'
quietly replace pglmigq0=`pglmigq0'
quietly replace pglmigq1=`pglmigq1‘
quietly replace pglmipq0=`pglmipq0'
quietly replace pglmipq1=`pglmipq1'
quietly replace pglmppq0=`pglmppq0'
quietly replace pglmppq1=`pglmppq1'

keep pglmigc0-pglmppq1
quietly append using bsmvpred
quietly save bsmvpred,replace
}
```

Bootstrap Program:Use bsmvpred; Clean Up Empty Row; Calculate Mean Differences

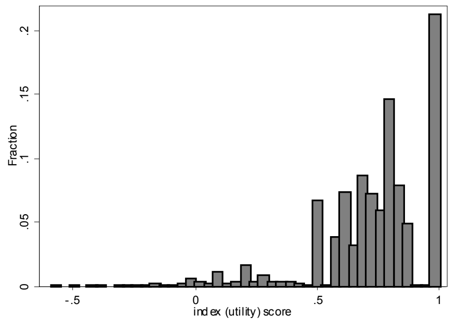
```
drop _all
use bsmvpred
drop if pglmigc0==.
capture drop pglmigcd
capture drop pglmipcd
capture drop pglmlgcd
capture drop pglmppcd
capture drop pglmigqd
capture drop pglmipqd
capture drop pglmppqd

gen pglmigcd=pglmigc1-pglmigc0
gen polsipcd=pglmipc1-pglmipc0
gen pglmlgcd=pglmlgc1-pglmlgc0
gen pglmppcd=pglmppc1-pglmppc0
gen pglmigqd=pglmigq1-pglmigq0
gen pglmipqd=pglmipq1-pglmipq0
gen pglmppqd=pglmppq1-pglmppq0

save,replace
```

Appendix 2: QALY Evaluation

QALY Evaluation

- While substantial attention has been paid to models for the evaluation of cost, substantially less has been paid to models for the evaluation of QALYs
- The QALY distribution shares certain complicating features with costs, but also has its own complicating features
  - Predictions should be confined to the theoretical range of the preference assessment instrument (e.g., –0.594 and 1.0 for the EQ-5D)
  - Long, heavy LEFT tails
  - (Particularly for pre-scored instruments) Often multi-modal (see Figure on next slide)
  - (Commonly) Large fraction of 1s

Sample EQ-5D Scores



Multivariable Approaches

- There are the beginnings of a literature on multivariable approaches
  - OLS (or GLM with identity link and gauss family) probably commonest
  - Alternatives
    - GLM with family (and link) diagnostics
    - GLM with a logit link and binomial 1 family or it's equivalent, beta regression (need specialized code for Stata)
    - (When there are a large fraction of 1s) 2-part models
- While we demonstrate some of these methods, more work is required before we will be able to identify best practice

Implemented Models

- Start with GLM gauss/identity
  - Evaluate GLM diagnostics
  - If necessary, reestimate GLM with better fitting family
- Also assess GLM gamma/log
  - Evaluate GLM diagnostics
  - If necessary, reestimate GLM with better fitting family

Common Starting Point: GLM with Gauss/Identity

```
glm qaly temp dissev blcost blqaly, link(identity) family(gauss)
```

Variance function: V(u) = 1 [Gaussian]
Link function: g(u) = u [Identity]

Log likelihood = 85.080395 AIC -.3203216
BIC -3055.401

qaly	Coef	Std Err	z	P>|z|	95% CI	
temp	.0627749	.0183515	3.42	0.001	.0268067	.0987432
dissev	-.1512017	.0831731	-1.82	0.069	-.314218	.0118147
blcost	-.0000359	.0000121	-2.96	0.003	-.000060	-.0000122
blqaly	.207374	.0633239	3.27	0.001	.0832614	.3314867
_cons	.511092	.0620345	8.24	0.000	.3895067	.6326773

eeict1.dta

GLM DIAGNOSTICS, Identity/Gauss

FITTED MODEL: Link = Identity ; Family = Gaussian
Results, Modified Park Test (for Family)
Coefficient: -.929485
Family, Chi2, and p-value in descending order of likelihood

Family	Chi2	P-value
Gaussian NLLS:	4.2582	0.0391
Poisson:	18.3496	0.0000
Gamma:	42.2987	0.0000
Inverse Gaussian or Wald	76.1054	0.0000

Results of tests of GLM Identity link

Pearson Correlation Test:	1
Pregibon Link Test:	.6741
Modified Hosmer and Lemeshow:	.8335

eeict1.dta

Troubling Findings

- Coefficient on the modified Park test is negative (we don't have any families that are negative) and p-value for the named families are all significantly rejected
- When confronted with coefficient < -0.5, consider subtracting all observations from maximum theoretically possible observation (e.g., 1.0 for most, if not all, instruments)

```
gen nqaly=1-qaly
sum qaly nqaly

Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+--------------------------------------------------------
    qaly |     500    .5941653    .2121148     .05679     .96882
   nqaly |     500    .4058347    .2121148     .03178     .94321
```

Estimate NQALY, GLM with Gauss/Identity

```
glm nqaly temp dissev blcost blqaly, link(identity) family(gauss)
```

Variance function: V(u) = 1 [Gaussian]
Link function: g(u) = u [Identity]

Log likelihood = 85.080395 AIC -.3203216
BIC -3055.401

nqaly	Coef	Std Err	Z	P>|z|	95% CI	
temp	**-.0627749**	**.0183515**	**-3.42**	**0.001**	-.0987432	-.0268067
dissev	**.1512017**	**.0831731**	**1.82**	**0.069**	-.0118147	.314218
blcost	**.0000359**	**.0000121**	**2.96**	**0.003**	.0000122	.000060
blqaly	**-.207374**	**.0633239**	**-3.27**	**0.001**	-.3314867	-.0832614
_cons	.488908	.0620345	7.88	0.000	.3673227	.6104933

eeict1.dta

```
* RECYCLED REDICTIONS

replace temp=0
predict pglmigq0
replace temp=1
predict pglmigq1
replace pglmigq0=1-pglmigq0
replace pglmigq1=1-pglmigq1

sum pglmipq*

Variable |     Obs        Mean    Std. Dev.       Min        Max
---------+-----------------------------------------------------
pglmigq0 |     500    .5627779    .0473131    .4202132   .6662163
pglmigq1 |     500    .6255528    .0473131    .4829882   .7289913
```

GLM DIAGNOSTICS, Identity/Gauss

FITTED MODEL: Link = Identity ; Family = Gaussian
Results, Modified Park Test (for Family)
Coefficient: .686724
Family, Chi2, and p-value in descending order of likelihood

Family	Chi2	P-value
Poisson	0.9443	0.3312
Gaussian NLLS:	4.5374	0.0332
Gamma:	16.5942	0.0000
Inverse Gaussian or Wald	51.4871	0.0000

Results of tests of GLM Identity link

Pearson Correlation Test:	1
Pregibon Link Test:	.6741
Modified Hosmer and Lemeshow:	.8335

eeict1.dta

Change Family to Poisson and Rerun Model

```
glm nqaly temp dissev blcost blqaly, link(identity) family(poisson)
```

Variance function: V(u) = u [poisson]
Link function: g(u) = u [Identity]

Log likelihood = -335.2046527
AIC 1.360819
BIC -3023.244

nqaly	Coef	Std Err	z	P>|t|	95% CI	
Temp	-.06313	.0566142	-1.12	0.265	-.1740918	.0478318
dissev	.16252	.2609842	0.62	0.533	-.3489997	.6740397
blcost	.0000373	.0000387	0.96	0.335	-.0000385	.0001132
blqaly	-.199954	.1926091	-1.04	0.299	-.5774608	.1775532
_cons	.477935	.190924	2.50	0.012	.1028309	.8512394

eeict1.dta

GLM DIAGNOSTICS, Identity/Poisson

FITTED MODEL: Link = Identity ; Family = Poisson

Results, Modified Park Test (for Family)

Coefficient: .703074

Family, Chi2, and p-value in descending order of likelihood

Family	Chi2	P-value
Poisson	0. 8796	0.3483
Gaussian NLLS:	4.9314	0.0264
Gamma:	16.7804	0.0000
Inverse Gaussian or Wald	52.6339	0.0000

Results of tests of GLM Identity link

Pearson Correlation Test:	.9396
Pregibon Link Test:	.6961
Modified Hosmer and Lemeshow:	.8949

eeict1.dta

```
* RECYCLED REDICTIONS

replace temp=0
predict pglmipq0
replace temp=1
predict pglmipq1
replace pglmipq0=1-pglmipq0
replace pglmipq1=1-pglmipq1

sum pglmipq*

Variable |    Obs       Mean  Std. Dev.        Min        Max
---------+--------------------------------------------------
pglmigq0 |    500   .5626003   .0479873   .4175745   .6685126
pglmigq1 |    500   .6258303   .0479873   .4807045   .7316426
```

Can We Improve the Link?

- Iteratively evaluate power links (in 0.1 intervals) between 1 and 2
  - Use the modified Park test to select a family
  - Rerun the GLM with the power and preferred link
  - Evaluate the fit statistics

Power 1.5 Link / Poisson Family

Power 1.5 Link / Poisson Family

```
glm nqaly temp dissev blcost blqaly, link(power 1.5) family(poisson)
```

Variance function: V(u) = u [Poisson]
Link function: g(u) = u^(1.5) [Power]

Log likelihood = -335.199289
AIC 1.360797
BIC -3023.255

nqaly	Coef	Std Err	z	P>|z|	95% CI	
Temp	-.059525	.053554	-1.11	0.266	-.164488	.045439
dissev	.156198	.244879	0.64	0.524	-.323756	.636152
blcost	.000036	.000037	0.97	0.331	-.000037	.000109
blqaly	-.185844	.180880	-1.03	0.304	-.540361	.168674
_cons	.322960	.180606	1.78	0.074	-.031021	.676941

eeict1.dta

```
* RECYCLED REDICTIONS

replace temp=0
predict pglm151pq0
replace temp=1
predict pglm151pq1
replace pglm151pq0=1-pglm151pq0
replace pglm151pq1=1-pglm151pq1

sum pglmpgq*

  Variable |  Obs        Mean  Std. Dev.         Min        Max
-----------+----------------------------------------------------
pglm151pq0 |  500   .5628606   .0458424   .4317995   .6701441
pglm151pq1 |  500   .6254807   .0496324   .485564    .7437032
```

Logit Link, Binomial 1 Family

- Alternatively, we can transform the QALY distribution so that it ranges between 0 and 1 and use a logit link and binomial 1 family (equivalent to beta regression)

```
local max=1
local min=0 (for EQ-5D, local min=-0.594)
local a=-`min'/(`max'-`min')
local b=1/(`max'-`min')
gen bqaly=`a'+(`b'*qaly)
sum qaly bqaly

Variable |     Obs        Mean    Std. Dev.       Min      Max
--------+----------------------------------------------------
    qaly |     500    .5941653    .2121148      .05679   .96822
   bqaly |     500    .5941653    .2121148      .05679   .96822
```

GLM with Binomial 1/Logit

```
glm bqaly temp dissev blcost blqaly, link(logit) family(binomial 1)
```

Variance function: V(u)=u*(1-u) [Bernoulli]
Link function: g(u)=ln(u/1-u) [Logit]

Log likelihood = -238.9699913 AIC .97588
BIC -2050.859

nqaly	Coef	Std Err	z	P>|z|	95% CI	
temp	.2626131	.1834617	1.43	0.152	-.0969653	.6221914
dissev	-.6328458	.832264	-0.76	0.447	-2.264053	.9983617
blcost	-.0001494	.0001208	-1.24	0.216	-.0003862	.0000875
blqaly	.8675488	.6338201	1.37	0.171	-.3747157	2.109813
_cons	.0373004	.6190775	0.06	0.952	-1.176069	1.25067

eeict1.dta

```
* RECYCLED REDICTIONS

replace temp=0
predict pglmlbq0
replace temp=1
predict pglmlbq1

sum pglmlbq*

Variable | Obs       Mean  Std. Dev.        Min        Max
---------+------------------------------------------------
pglmlbq0 | 500   .5628245   .048325   .4159106   .6653128
pglmlbq1 | 500   .6254634   .0461845   .4807669   .7210496
```

Run Link DIAGNOSTICS, Logit/Binomial 1

FITTED MODEL: Link = Logit ; Family = Binomial

Results of tests of GLM Identity link

Pearson Correlation Test:	.9914
Pregibon Link Test:	.5605
Modified Hosmer and Lemeshow:	.9242

eeict1.dta