

Meta analysis lab: Part I

In STATA, a comprehensive set of user-written commands is available for meta-analysis. Meta analysis of studies with binary (relative risk, odds ratio, risk difference) or continuous outcomes (mean differences) can be performed. We even can use meta-regression models to analyze association between treatment effect and study characteristics.

An overview of meta-analysis, with some discussion of meta-regression models (our advice right now is do not pay attention to the model formulation displayed on the site...we, the TAs, are not sure it is accurate. But the history and “file drawer” problem are interesting):

<http://en.wikipedia.org/wiki/Meta-analysis>

In this lab, two main commands, **meta** and **metareg**, will be covered. You need to download the .ado and .hlp files for the **meta** command from the course website to your personal .ado folder. (note, when you type “help meta” it will open a general help page. To get the specific help page for the **meta** command, click the blue link on that general help page, “help file for meta command”).

You need to run `ssc install metareg` to download the **metareg** command, or `ssc install metareg, replace` if need be. Same for **metan**: `ssc install metan`. (We will not discuss **metan** in this lab).

meta

- meta provides pooled estimates, confidence limits, and a test that the true pooled effect is zero, obtained from fixed and random effects meta-analysis. It also provides a test for heterogeneity between studies and an estimator of between studies variance, and, optionally, plots the individual and pooled estimates.
- The user provides the effect estimates as thetas (i.e., a log risk ratio, log odds ratio, or other measure of effect). Likewise, the user supplies a measure of each theta's variability (i.e., its standard error, se_theta, or its variance, var_theta).
- Alternatively, the user provides exp(theta) (i.e., a risk ratio or odds ratio) and its confidence interval, (ll, ul)).
- meta is most useful when data are extracted from published reports as effect measures and confidence intervals. If the raw data for the effect measure are available (i.e. all cell values in a 2x2 table), then metan is preferable to meta.
- Examples: meta logor selogor, eform gr(f) cline xline(1) id(trialnam) xlab
meta meandiff vardiff, var gr(r) eb print
meta rr ll ul, ci gr(e)

metareg

- metareg extends a random effects meta-analysis to estimate the extent to which one or more covariates, with values defined for each study in the analysis, explain heterogeneity in the treatment effects. Examples of such study-level covariates might be average duration of follow-up, some measure of study quality, or a measure of geographical location of each study.
- metareg fits models with two additive components of variance, one representing the variance within units, the other the variance between units, and therefore is applicable both to the meta-analysis situation, where each unit is one study, and to other situations such as multi-center trials, where each unit is one center.
- Examples: metareg logor covariate1 covariate2, wsse(selogor)
metareg logor dur, wsvar(vlor) bs(eb) noit
metareg meandiff qual avchol, wsse(sediff) bs(ml) tol(

Let us use the example in the lecture 2 to illustrate how **meta** command works. The data for six cities, Los Angeles, New York, Chicago, Dallas, Houston and San Diego (Listed as Table 1) is listed in the following table.

Table1: Log Relative Risks for six largest cities

City	Log RR estimate (% per 10 micrograms/ml)	Statistical standard error	Statistical variance
Los Angeles	0.25	0.13	0.0169
New York	1.40	0.25	0.0625
Chicago	0.60	0.13	0.0169
Dallas/Ft worth	0.25	0.55	0.3025
Houston	0.45	0.40	0.1600
San Diego	1.00	0.45	0.2025

Notice that the within-city variance is synonymous with “statistical variance” and is just the statistical standard error squared. (.13² = .0169, .25² = .0625, etc).

We can pretend that you got these estimates from a literature review. We want to get an overall estimate for the Log(RR). We could fit a **fixed** effects meta-regression which would treat city as a fixed effect. Essentially, you could do a regression/anova with a design (dummy) variable for city.

The fixed regression model would look like:

$$\text{Log}(\text{RR}_i) = \theta + b_i^* + \varepsilon_i$$

$$\varepsilon_i, \sim N(0, \sigma_i^2)$$

Where b_i^* is a fixed effect, as discussed in lecture 2, essentially the estimated difference between the city estimate and the “population” estimate. In the fixed effects analysis, the population mean will not take into account the variation of the b_i^* . The variable city in this analysis is not well situated to be a fixed effect because not all cities are represented; yet, we are looking for an overall population effect for all cities, using just a sampling of cities. Thus, cities would be better thought of as a random variable itself, leading us to fitting a random effects meta-regression.

Note the θ has no subscript. The above fixed effects meta-regression does not allow for θ to vary. If we suspected or wished to investigate if each city truly had its own effect accounting for the between city variation, we would need to conduct a **random** effects meta-regression that allows each city to have its true effect, θ_i , and we assume another normal distribution with variance component τ^2 to determine the variability of θ_i around θ

The random effects meta-regression model (two-stage normal-normal model) can be written as:

$$\begin{aligned} \text{Log}(\text{RR}_i) &= \theta + b_i + \varepsilon_i \\ \varepsilon_i &\sim N(0, \sigma_i^2) \\ b_i &\sim N(\theta, \tau^2) \end{aligned}$$

which if we allow $\theta_i = \theta + b_i$ (this is called “hierarchical centering”) we have:

$$\begin{aligned} \text{Log}(\text{RR}_i) &= \theta_i + \varepsilon_i \\ \varepsilon_i &\sim N(0, \sigma_i^2) \\ \theta_i &\sim N(\theta, \tau^2) \end{aligned}$$

Here, the σ_i^2 (within-city statistical variance) are known and listed on the above table. The parameters of interest are the overall estimated log RR, θ , and the between-city variance (a.k.a heterogeneity), τ^2 .

```
. meta logrr se, graph(e) cline xline(0) id(city) xlab
```

Meta-analysis

Method	Pooled Est	95% CI		Asymptotic		No. of studies
		Lower	Upper	z_value	p_value	
Fixed	0.546	0.385	0.707	6.662	0.000	6
Random	0.660	0.283	1.038	3.427	0.001	

Test for heterogeneity: Q= 18.391 on 5 degrees of freedom (p= 0.002)
Moment-based estimate of between studies variance = 0.134

Note: estimates and confidence limits are empirical Bayes

Study	Weights		Study Est	95% CI	
	Fixed	Random		Lower	Upper
Los Angeles	59.17	6.61	0.30	0.05	0.54
New York	16.00	5.08	1.17	0.74	1.59
Chicago	59.17	6.61	0.61	0.36	0.85
Dallas	3.31	2.29	0.53	-0.12	1.19
Houston	6.25	3.40	0.56	-0.00	1.13
San Diego	4.94	2.97	0.80	0.19	1.40

The estimate of τ^2 is 0.134 and the estimate of θ , the overall log RR is 0.66. The weights correspond to the weights used to get the overall pooled estimate.

```
display 59.17 + 16.00 + 59.17 + 3.31 + 6.25 + 4.94
148.84
```

```
display 59.17*.25 + 16.00*1.4 + 59.17*.6 + 3.31*.25 + 6.25*.45 + 4.94*1
81.2745
```

```
display 81.2745 / 148.84
.54605281
```

Note that **meta** performs both fixed and random effects analyses by default and the tabular output includes the weights from both analyses. Because the **meta** command requires only the estimated treatment effect and its standard error, it will be particularly useful in meta-analyses of studies in which the treatment effect is not derived from the standard 2 by 2 table.

Also notice that the fixed effect estimate, .546, is different from the random of .660. The fixed effect estimate will be biased if in actuality there is between study variation (which the fixed effect meta-regression ignores), and the difference suggests that fitting the random effect meta-regression analysis may be justified to account for the heterogeneity.

City	Log(RR) estimate	Statistical variance	Weight λ_i	λ_i	Empirical Bayes Estimate
Los Angeles	0.25	0.0169	6.61	0.89	0.30
New York	1.4	0.0625	5.09	0.68	1.16
Chicago	0.6	0.0169	6.61	0.89	0.61
Dallas/Ft worth	0.25	0.3025	2.29	0.31	0.53
Houston	0.45	0.1600	3.40	0.46	0.56
San Diego	1.0	0.2025	2.97	0.40	0.80
Over-all	0.66	0.037	27.0		

Note: The equations used to calculate the empirical Bayes estimates, weights, lambdas are listed in Lecture 2 notes.

Let's calculate by hand the Los Angeles (LA) and San Diego (SD) lambda's:

$$\lambda_1 = \tau^2 / (\tau^2 + \sigma_1^2) = .134 / (.134 + .0169) = .89$$

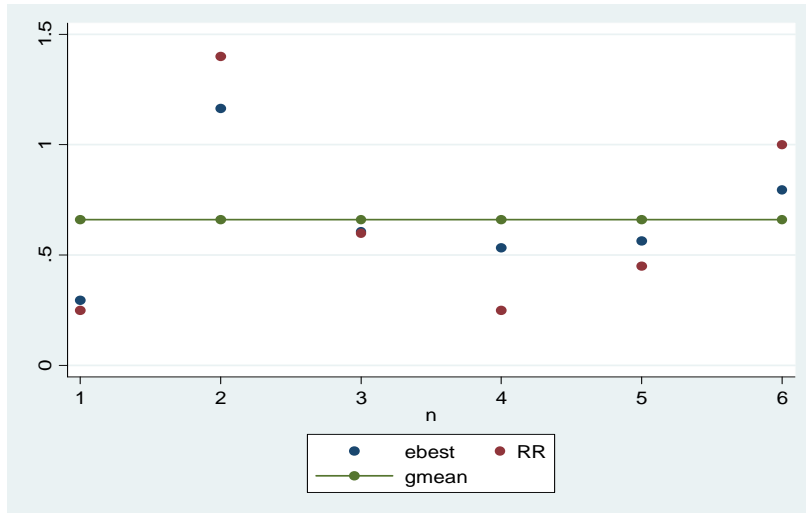
$$\lambda_6 = \tau^2 / (\tau^2 + \sigma_6^2) = .134 / (.134 + .2025) = .40$$

With the lambda's, we can now calculate the Empirical Bayes (EB) estimate:

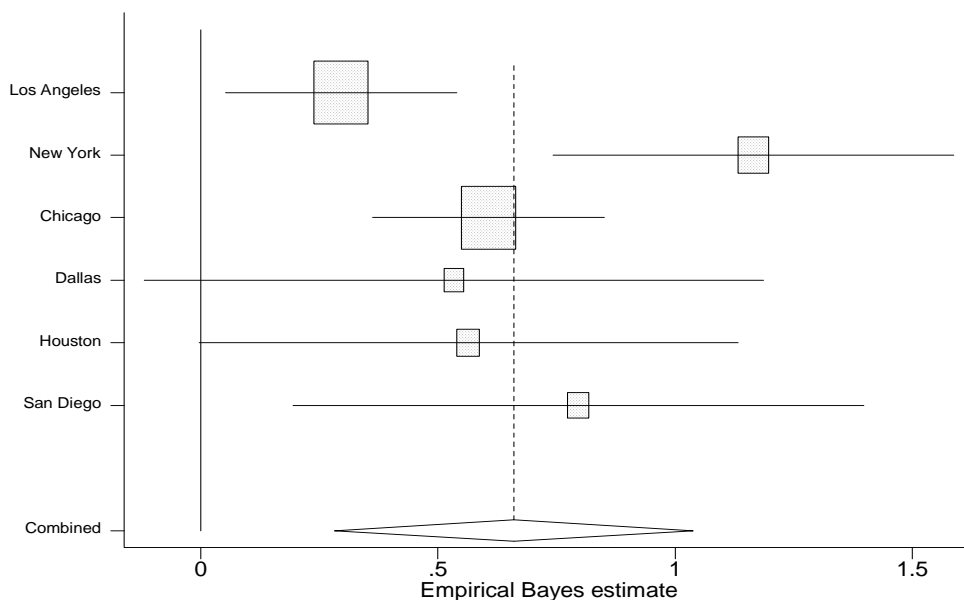
$$\theta_{EB1} = \lambda_1 * \theta_1 + (1 - \lambda_1) * \theta = 0.89 * 0.25 + 0.11 * 0.66 = .30$$

$$\theta_{EB6} = \lambda_6 * \theta_6 + (1 - \lambda_6) * \theta = 0.40 * 1.00 + 0.60 * 0.66 = .80$$

From the following graph, we can see that all empirical Bayes estimates have been shrunk towards the overall mean, 0.66. Here, red dots represent the original RR estimates and blue dots represent the empirical Bayes estimates. The green horizontal line stands for the estimated overall mean.



The **meta** command automatically produces a forest plot when the **graph** option is specified. In a forest plot the contribution of each study to the meta analysis (its weight) is represented by the area of a box whose centre represents the size of the point estimate of effect from that study. The confidence interval for the treatment effect from each study is also shown. The summary effect is shown by the middle of a diamond whose left and right extremes represent the corresponding confidence interval.



We may also use **metan** (its description is omitted in this lab, those who feel interested about it may check STATA online help documentations) and **metareg** on this dataset. The similar results can be obtained.

```
. metan logrr se,random
```

Study	ES	[95% Conf. Interval]		% Weight
1	0.250	-0.005	0.505	24.52
2	1.400	0.910	1.890	18.84
3	0.600	0.345	0.855	24.52
4	0.250	-0.828	1.328	8.49
5	0.450	-0.334	1.234	12.61
6	1.000	0.118	1.882	11.02
D+L pooled ES	0.660	0.283	1.038	100.00

```
Heterogeneity chi-squared = 18.39 (d.f. = 5) p = 0.002
I-squared (variation in ES attributable to heterogeneity) = 72.8%
Estimate of between-study variance Tau-squared = 0.1345
```

```
Test of ES=0 : z= 3.43 p = 0.001
```

```
. metareg logrr, wsvar(var)
```

```
Meta-regression                               Number of studies =
6
Fit of model without heterogeneity (tau2=0):   Q (5 df) =
18.3911                                         Prob > Q =
0.002
Proportion of variation due to heterogeneity   I-squared =
0.728
REML estimate of between-study variance:      tau2 =
0.1523
```

rr	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
_cons	.6620925	.2012365	3.29	0.022	.1447977

Note: the estimated tau-squared from **metareg** and **meta** are not very close to each other since one command used the method of moments method and the other used REML (Restricted Maximum Likelihood) method.

We will talk **metareg** command in more detail later.