Multivariate dose-response meta-analysis: an update on `glst`

Nicola Orsini

Unit of Biostatistics
Unit of Nutritional Epidemiology
Institute of Environmental Medicine
Karolinska Institutet
http://www.imm.ki.se/biostatistics/

2013 Nordic and Baltic Stata Users Group meeting

September 27, 2013

Outline

- Background of the method
- Present new possibilities when specifying the covariance structure of exposure-disease relative risks
- Illustrate the steps for a flexible multivariate dose-response meta-analysis

Background

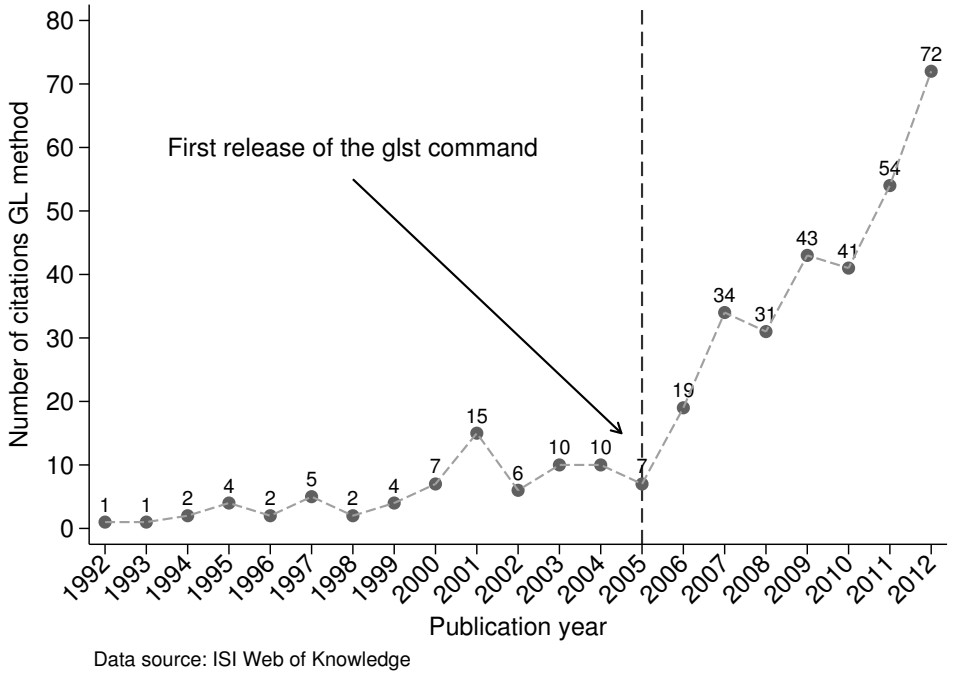
- Quantitative review of summarized dose-response data
- Method first formalized by Greenland and Longnecker (*American J Epi*, 1992)
- `glst` available on SSC Archive since May 2005
- 1st Nordic and Baltic Stata User Meeting in September 2005

Background

- Orsini N, Bellocco R, Greenland S (2006) Generalized least squares for trend estimation of summarized dose-response data, *Stata Journal*, 6(1), pp.40-57

- Advanced method (Part 4) in Jonathan A. C. Sterne (Editor) *Meta-Analysis in Stata*: An Updated Collection from the Stata Journal, 2009, StataPress

Increasing number of dose-response meta-analyses



Data source: ISI Web of Knowledge

- 80% published after 2005
- n=42 published during the first 4 months of 2013 ($\sim$ 2 every week)
- Majority is using Stata

Who is using glst?

- The `glst` method was used and/or cited in leading international journals (i.e., *JAMA*, *Lancet*, *Stroke*, *Gastroenterology*, *Annals of Oncology*, *American J of Medicine*, *American J of Clinical Nutrition*, *American J Epidemiology*, *International J Epidemiology*, *Journal National Cancer Institute*, *International J of Cancer*, *Cancer Causes and Control*, *Computational Statistics and Data Analysis*, *Statistics in Medicine*).

- World Cancer Research Fund/American Institute for Cancer Research (WCRF/AICR) for the ongoing review of cancer prevention research.

Current limitations of `glst`

- Only one way to estimate the covariance of log relative risks requiring a table of unadjusted data

- Only univariate random-effect dose-response meta-analysis

Example 1: Alcohol intake and breast cancer risk

Case-control data. *Int J Cancer*. 1988 15;41(3):390-393.

dose	rr	lb	ub	n	case
0	1.00	1.00	1.00	337	165
2	0.80	0.51	1.27	167	74
6	1.16	0.73	1.85	186	90
11	1.57	0.99	2.51	212	122

Model

We specify a log linear model

$$y = \mathbf{X}\beta + \epsilon$$

y is the vector of non referent log relative risks

$\mathbf{X}$ the corresponding design matrix (dose or some transform of it)

The model has no intercept. The trend is forced to go through the origin.

Generalized Least Squares Estimation

Relative risks are estimated using a common reference exposure level

ϵ are not independent, $Cov(\epsilon) = \Sigma$

Σ can be obtained from published or primary data

We minimize $(\mathbf{y} - \mathbf{X}\beta)'\Sigma^{-1}(\mathbf{y} - \mathbf{X}\beta)$ with respect to β.

A GLS estimator $\mathbf{b}$ of β is

$$\mathbf{b} = (\mathbf{X}'\Sigma\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{y}$$

$$\mathbf{V} = Cov(\mathbf{b}) = (\mathbf{X}'\Sigma\mathbf{X})^{-1}$$

Approximating covariance matrix

Diagonal elements Σ are the variances of y

Non-diagonal elements Σ, covariances, can be approximated in different ways

- Table of pseudo counts corresponding to y (*American J Epidemiology* 2012, GL method)
- Table of pseudo counts corresponding to y and diag(Σ) (*Statistics in Medicine* 2007, Hamling method)
- Directly from published floated confidence intervals (*Statistics in Medicine* 1991, Easton method)
- Directly from regression models on primary data

Example 1: Alcohol intake and breast cancer risk

Covariances obtained using the **GL** method

```
.   gen double logrr = log(rr)
.   gen double se =  ((log(ub) - log(lb))/(2*invnorm(.975)))
.   replace dose = dose/5
.   glst logrr dose, se(se) data(n case) cc eform

Generalized least-squares regression                 Number of obs    =         3

------------------------------------------------------------------------------
       logrr |     exb(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        dose |   1.255011   .1296668     2.20   0.028     1.024948    1.536714
------------------------------------------------------------------------------

.   mat list e(Sigma)

symmetric e(Sigma)[3,3]
           c1          c2          c3
r1   .05417235
r2   .01881768   .05627467
r3   .01943145   .02068682   .05632754
```

The risk of colorectal cancer increased by 26% for every 5-g/day increase in alcohol intake.

Example 1: Alcohol intake and breast cancer risk

Covariances obtained using the **Hamling** method

```
.   glst logrr dose, se(se) data(n case) cc eform h

Generalized least-squares regression                 Number of obs    =         3

------------------------------------------------------------------------------
       logrr |      exb(b)   Std. Err.       z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        dose |   1.257904   .1290452     2.24   0.025     1.028786    1.538049
------------------------------------------------------------------------------

.   mat list e(Sigma)

symmetric e(Sigma)[3,3]
            c1          c2          c3
r1   .05417235
r2   .02002802   .05627467
r3   .02002802   .02002802   .05632754
```

Details and evaluations of the GL and Hamling method can be found in Orsini *et al American J Epidemiology* 2012 1;175(1):66-73.

Floating absolute risk method

- Easton DF, Peto J, Babiker AG. (*Stat Med* 1991;10:1025-1035) proposed to publish directly a transformation of the standard variance/covariance matrix estimated on primary data.

- The method introduced the concept of floated variances (or standard errors) so that one can compare any two exposure levels using only the floated variances because the covariances are approximately zero.

- The average covariance of the published relative risks is approximately the floated variance of the chosen reference exposure level.

Example 2: Alcohol intake and endometrial cancer risk

- A total of 1,280,296 middle-aged women in the United Kingdom enrolled in the Million Women Study were routinely followed for incident cancer (*J Natl Cancer Inst* 2009;101:296-305).

- Relative risks were estimated using the proportional hazard regression model and 95% confidence intervals were derived using the floating absolute risk method.

- One solution is to go from floated to conventional confidence intervals (Orsini *Computer Methods and Programs in Biomedicine*, 2010, 98, 90-93).

```
ssc install ftocci
```

Example 2: Alcohol intake and endometrial cancer risk

Table 1 – Multivariable adjusted relative risks about alcohol intake and endometrial cancer risk from the Million Women Study [6] presented according to different methods (floating absolute risk and conventional) based on original and published data. Used with permission of Oxford University Press.

Exposure levels	Alcohol intake	Based on original data		Based on published data
		Floating absolute risk RR (FCI)[a]	Conventional RR (CI)[a]	Conventional RR (CI)[a]
1	Non and past drinkers	1.06 (1.00–1.12)	1.00 (1.00–1.00)[b]	1.00 (1.00–1.00)[b]
2	≤2 drinks/week	1.00 (0.95–1.06)[b]	0.95 (0.87–1.02)	0.94 (0.87–1.02)
3	3–6 drinks/week	0.99 (0.92–1.05)	0.93 (0.85–1.02)	0.93 (0.86–1.02)
4	7–14 drinks/week	0.90 (0.83–0.97)	0.85 (0.77–0.94)	0.85 (0.77–0.93)
5	≥15 drinks/week	1.05 (0.91–1.22)	1.00 (0.85–1.16)	0.99 (0.85–1.16)

[a] RR, relative risk; FCI, floated confidence interval; CI, confidence interval.
[b] Reference exposure level.

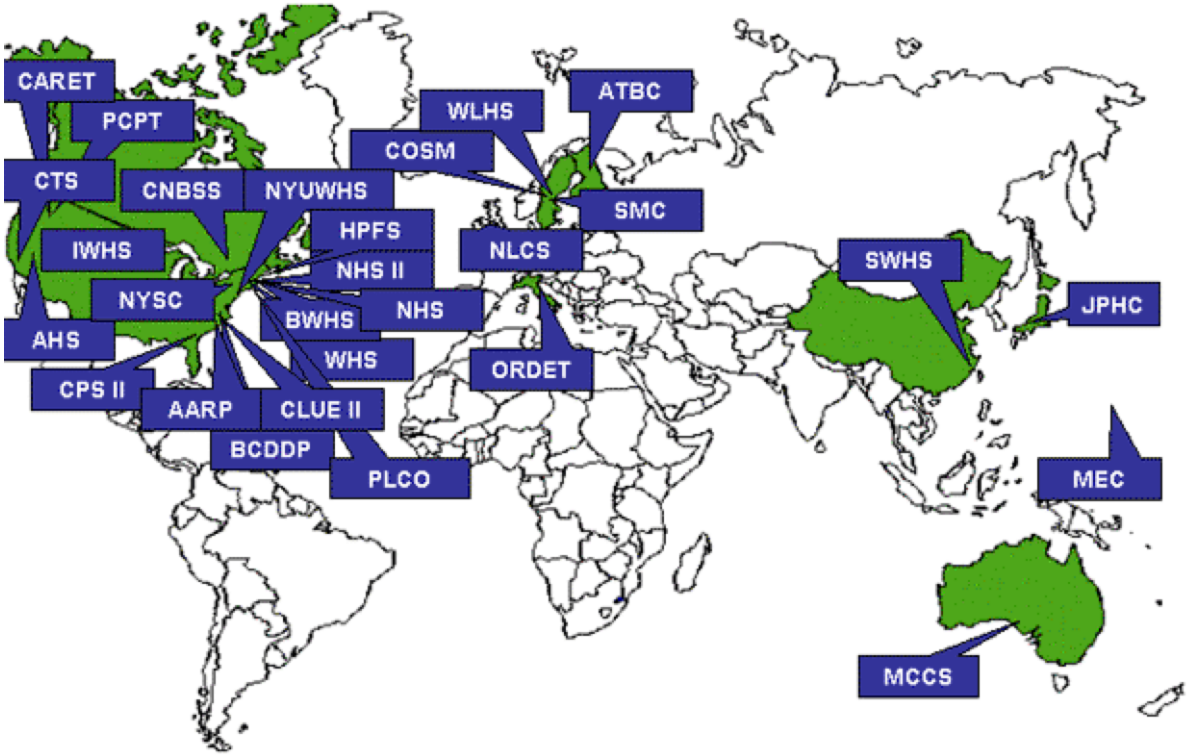
```
gen covariance = ((log(1.06)-log(.95))/(2*invnormal(.975)))^2

glst logrr dose, se(se) ci eform acov(covariance)
```

Covariance matrix available from primary data

- Pooling projects and Pooling results of standardised analysis are increasingly popular in medical research.
- There is no need to approximate Σ based on available published information
- Use Σ obtained directly from the principal investigator

Pooling Project of Prospective Studies of Diet and Cancer



http://www.hsph.harvard.edu/poolingproject

Example 4: Alcohol intake and colorectal cancer risk - Single study

Below are the findings obtained from a standard regression analysis on individual data.

```
       rr          lb          ub        dose       cases    peryears
        1           1           1           0          28    22185.73
    .6592   .4029785    1.078329      1.8286          38    43030.54
   .67331   .4160625     1.08961      9.1992          43    53088.96
    .6136   .3661768    1.028189     22.8571          32    45348.09
   .75652   .4034072    1.418711     35.6667          16    19790.79
  1.22424   .6986875    2.145112     58.4257          27    19919.85

. mat list VCM

symmetric VCM[5,5]
          c1        c2        c3        c4        c5
r1   .063049
r2   .036555   .060316
r3   .036471   .037203    .06937
r4   .036379   .037422    .03863    .10292
r5   .036048    .03748     .0393    .04146    .08189

.    display ((log(2.145112)-log(.6986875))/(2*invnormal(.975)))^2
.08189
```

Example 4: Alcohol intake and colorectal cancer risk - Single study

```
. glst logrr dose, se(se) eform mcov(VCM)

Generalized least-squares regression                  Number of obs    =         5

------------------------------------------------------------------------------
       logrr |     exb(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        dose |   1.082282   .0528902     1.62   0.106     .9834286    1.191071
------------------------------------------------------------------------------

. glst logrr dose, se(se) data(peryears cases) ir eform h

Generalized least-squares regression                  Number of obs    =         5

------------------------------------------------------------------------------
       logrr |     exb(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        dose |   1.078982   .0508713     1.61   0.107     .9837442     1.18344
------------------------------------------------------------------------------
```

In this specific cohort study the risk of colorectal cancer increased by 8% for every 12-g/day increase in total alcohol intake.

Multivariate dose-response meta-analysis

- Define transformations of the quantitative exposure.
- First stage. Fit the dose-response model within each study.
- Second stage. Pool study-specific trends using fixed or random effect models.
- Test of hypothesis.
- Graphical presentation of the pool dose-response relation.

Example 4: Alcohol intake and colorectal cancer risk - Multiple studies

Ann Intern Med 2004;140(8):603-613.

8 eligible prospective cohort studies participating in the Pooling Project of Prospective Studies of Diet and Cancer.

6 exposure intervals (from 0 grams/day to $\geq$ 45 grams/day) for each study

$6 \times 8 = 48$ data points

3,646 cases and 2,511,424 person-years were included in the dose-response analysis.

Relative risks were adjusted for smoking status, smoking duration for past and current smokers (years), number of cigarettes smoked daily for current smokers, educational level, body mass index, and energy intake (kcal/day).

Transformations of the quantitative exposure

We model total alcohol intake using restricted cubic splines with 3 knots at fixed percentiles (10th, 50th, 90th) of the aggregated exposure.

```
. use http://www.imm.ki.se/biostatistics/data/ex_alcohol_crc, clear

. mkspline doses = dose, nk(3) cubic

. clist study dose doses1 doses2 in 1/6

     study       dose     doses1     doses2
1.     atm          0          0          0
2.     atm     1.8286     1.8286   .0018429
3.     atm     9.1992     9.1992   .2346417
4.     atm    22.8571    22.8571    3.34394
5.     atm    35.6667    35.6667   9.741412
6.     atm    58.4257    58.4257   25.58743
```

Fixed-effects multivariate dose-response meta-analysis

Assuming a common underlying dose-response relation applies to all studies.

```
. glst logrr doses1 doses2, se(se) data(peryears cases) id(study type)

Two-stage fixed-effects dose-response model      Number of studies   =       8

Generalized least-squares regression                 Number of obs   =      48
------------------------------------------------------------------------------
       logrr |      Coef.   Std. Err.       z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      doses1 |  -.0004926   .0034609    -0.14   0.887    -.0072759    .0062906
      doses2 |   .0191275   .0087931     2.18   0.030     .0018933    .0363617
------------------------------------------------------------------------------
```

Random-effects multivariate dose-response meta-analysis

Taking into account (statistical) heterogeneity of the dose-response relation across studies.

```
. glst logrr doses1 doses2 , se(se) data(peryears cases) id(study type) random

Two-stage random-effects dose-response model     Number of studies   =        8
Method of Moments
Generalized least-squares regression                 Number of obs   =       48

------------------------------------------------------------------------------
       logrr |      Coef.   Std. Err.       z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      doses1 |   -.001442   .0042411    -0.34   0.734    -.0097544    .0068705
      doses2 |   .0215935   .0109739     1.97   0.049     .0000851    .0431019
------------------------------------------------------------------------------
```

Several approaches to estimation of multivariate meta-analysis (`mvmeta`) have been developed (maximum likelihood, restricted maximum likelihood, methods of moments). See paper (and references therein) by Ian White on Multivariate random-effects meta-regression. *Stata Journal*. 2011, 11(2): 255-270.

Testing hypothesis

A Wald-type test for the hypothesis of no exposure-disease association can be obtained by testing simultaneously both regression coefficients equal to zero.

```
. testparm doses1 doses2
           chi2(  2) =   26.22
         Prob > chi2 =    0.0000
```

A Wald-type test for non-linearity can be obtained by testing the regression coefficient of the second spline equal to zero (shown in the output). The small p-value (0.049) indicates departure from linearity.

Point estimates for the predicted dose-response trend

Define $\mathbf{Z}$ the matrix containing transformations (`doses1`, `doses2`) of $\mathbf{X}$ (`dose`).

Define $\mathbf{Z}_{ref}$ the row vector containing the values of $\mathbf{Z}$ at the reference level.

Define $\mathbf{b}$ the vector of estimated regression coefficients with a dose-response meta-analysis.

The exposure-disease relative risks as function of the quantitative exposure is given by

$$exp((\mathbf{Z} - \mathbf{Z}_{ref})\mathbf{b})$$

Interval estimates for the predicted dose-response trend

The approximate pointwise 95% confidence interval of the predicted relative risks is then calculated as follows

$$exp\left((\mathbf{Z}-\mathbf{Z}_{ref})\mathbf{b} \pm 1.96\sqrt{diag((\mathbf{Z}-\mathbf{Z}_{ref})Cov(\mathbf{b})(\mathbf{Z}-\mathbf{Z}_{ref})^{T})}\right)$$

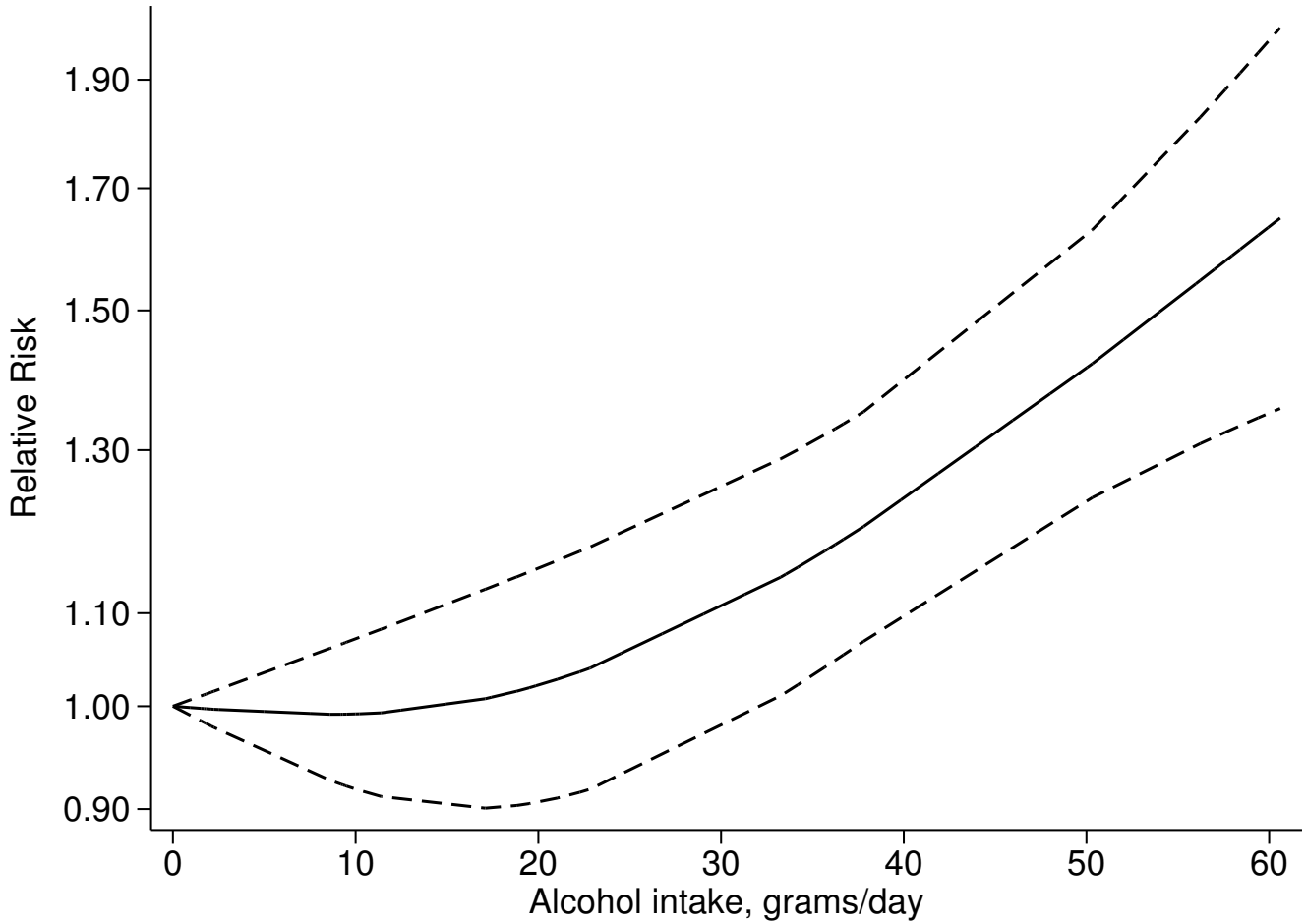
One can use the post-estimation command `predictnl`.

The `xblc` command greatly facilitate the tabular and graphical presentation of predicted outcomes (Orsini & Greenland, *Stata Journal*. 2011. 11(1): 1-29).

Graphical presentation of the dose-response relation

```
levelsof dose

xblc doses1 doses2 , covname(dose) at(‘r(levels)’) ref(0) eform line
```



Updated information, examples, and references

`http://www.imm.ki.se/biostatistics/glst`

The revision of the `glst` command is on-going.