The SAS %METADOSE Macro

Ruifeng Li and Donna Spiegelman

August 26, 2010

Abstract

The %metadose macro is a SAS macro for meta-analysis of linear and nonlinear dose-response relationships. It is used when research reports studying the same dose-response relationship have different exposure or treatment levels. It is a two step macro: First, for each study, it uses either the Greenland method (AJE, 1992) or Hamling method (SIM, 2008) to get estimated cell counts of the 2X2 table adjusted for counfounding, then it estimates the asymptotic correlation between the adjusted log odds ratio estimates for each exposure level relative to the referent level, from which we can get the estimated covariance matrix for these study-specific estimates. After this step, we get a single pooled estimate and its variance estimate across different exposure or treatment levels. Then, meta-analysis is performed analysis for all the studies using the single study-specific trend estimate, in common units across studies. An option also exists to explore and graph non-linearity in the poooled results.

Keywords: meta analysis, pooled analysis, dose-response, exposure-response, Greenland, Hamling, pooling, non-linearity, splines

Contents

1 Description

The %METADOSE macro is a SAS macro to do meta-analysis for a dose-response relationship. It is a two-step procedure:

First, obtain a single estimate for trend in the common units across studies and its variance estimate for each study. To study dose-response relationships, different published studies often use different exposure or treatment levels (doses) to report the effects of the same exposure, so we need to pool these estimates across levels to get one estimate in common units for each study. Since these estimates for separate exposure levels depend on the same reference group, they are not independent, but are correlated in some degree. The meta-analysis must take this correlation into account.

In Greenland et al.'s paper(1992), the authors developed an approach that yields an efficient point estimator and a consistent variance estimator under assumptions likely to be approximated in practice. Their approach is based on constructing an approximate covariance estimate for the adjusted log odds/rate/risk ratios from a fitted table that conforms to the adjusted log odds/rate/risk ratios, and matches the crude 2X2 table margins.

In Hamling et al.'s paper(2008), the authors developed another approach for reconstructing the cell counts of the original 2x2 tables for each study. Their fitted 2X2 table takes confounding into account more directly, and the crude 2X2 table margins don't match these reported, but the variances of the estimated log odds/rate/risk ratios do match the reported ones. This macro uses Hamling;s method to get estimated cell counts, then use the same approach as Greenland's to obtain covariance estimates and the weighted least squares estimates.

Then, we use another in-house Channing SAS macro (see Ref #5 for detail), %meta to do meta-analysis across studies based on these single pooled estimates and the corresponding variance-covariance matrix.

This macro also has an option to assess if the log-linear exposure-response relationships are non-linear graphically and through a formal statistical hypothesis test (see Ref #6 for detail)

2 Invocation and Detail

In order to run this macro, your program must know where to look for it. You can tell SAS where to look for macros by using the options

```
options nocenter ps=78 ls=80 replace formdlim='=';
mautosource
sasautos=('/usr/local/channing/sasautos', OTHER DITECTORIES);
```

This will allow you to use %lgtphcurv8 as well as other public SAS macros, such as %PM, %INDIC3, %EXCLUDE, %MPHREG, %CALADJ, and %PCTL in **/usr/local/channing/sasautos**.

Outside the Channing system, you can download metadose.sas via Prof. Spiegelman's website: `http://www.hsph.harvard.edu/faculty/spiegelman/metadose.html` (both the SAS program and the documentation in pdf format)

In the rest of this section, we will list all the input parameters, some of which are required and some of which are optional, but strongly suggested, and some of which are truly optional.

Note, if a value is given to the right of the "=", that is the default.

```
/** The following are related with how the dataset should be.
    They are all REQUIRED except you only need to provide names for one of the two: UB or LB**/
    dat       =,  <the name of the input dataset >
                  DAT should contain one line for each exposure level
                  given for each study, including the reference level.
                  Each observation in DAT should include
                  variables for study name, study type (see
                  below), the (median) dose level, the number of
                  observations and the number of cases
                  for the dose level, and for the RR or OR
                  and its 95% lower and/or upper confidence
                  bounds.
                  For each study, the reference level should
                  be the first observation for the study with ratio (RR or OR) 1.00,
                  even it is not the smallest or largest value of the exposure.>

    ratio     =, <the variable name for the odds ratio for case-control,
                  risk ratio for cohort, or rate ratio for person-time study>
                  NOTE:  for the reference level, this variable
                  should have value 1.00.
                  If there is another level with OR or RR=1,
                  write it as 1.0000001.
    UB        =,  <variable name for 95% confidence upper bound
                    Note, this can be blank if LB is not>
    LB        =,  <variable name for 95% confidence lower bound.
                    Note, this can be blank if UB is not>
    dose      =,  <median dose value for each category,including the reference group>
    Ncase     =,  <variable name for number of cases for each category,
                    including the reference group>
    Ntotal    =,  <variable name for total number of
                  person-time (cohort) or subjects (case-control) for each category,
                  including the reference group>
    studyname =,  <variable name for study name>
    studytype =,  <1: incidence rate ratio, for cohort study with person-time, default;
                   2: risk ratio for cohort study;
                   3: odds ratio for case-control study >
                  Note that you can combine different types of studies in one analysis.
                  >

/** The following are related with meta analysis **/
    meta      =T, <do meta analysis? can do when there is more than 1 study>
    wt        =1, <the increment for which you want to report the final meta-analysis estimate>
    unit_wt   =,  <units of the increment, e.g. gram/day>

/** The following is the method options to choose from with default values listed here,
    so they are optional **/
    var_covar=GH,  <4 options with default with both G and H.
                     G: Greenland method ; K: known var_covar;
```

```
                    H: Hamling method;    GH: both Greenland & Hamling>

    realCov=   <if var_covar is K,  you need to provide the var-covar matrix as a dataset where
                the variable list is: STUDY, COV1, COV2, etc. For example, if you have a 4
                exposure levels besides reference group, you need to have 5 variables:
                study cov1-cov4, with observations 1-4 have the same study name, and the
                rest just like a 4X4 var-covar matrix output from proc phreg, with the
                variances on diagonal, and covariances off diagonal, then the next 4
                observations will be for the next study>


/** The following are related to the  linearity assessment, optional with default values
    listed here **/

    linearCheck=0, <check non-linearity, 1: yes, 0: no>
    nk=4,

/** The following are related to the graph to examine non-linearity between dose and response,
if linearCheck=1. All are optional with default values listed here **/
    graphtitle=,<graph title>
    graphname_g=greenland.&dat..ps, <for greenland method to get var-covar plot, default name
                would be three parts: first is greenland, the second is the data set name
                provided by user for the first parameter dat=, and the third is ps>
    graphname_h=hamling.&dat..ps,<for Hamling method to get var-covar plot, otherwise,see above>
    graphname_k=realcov.&dat..ps,<for real var-covar plot, otherwise, see above>

    cutoff=F, <if not F, it should have two values with space:  the second is the value
               at which to truncate the vertical axis, the first number is either 1 or 2 with
                1: just truncate  95%CI upper limit
                2: truncate 95%CI upper limit and spline curve
               >
    vlabel=, <optional, vertical label>
    hlabel=, <optional, horizontal label>
    ci=2,  <1: clouds for CI, 2: dotted line for CI, 0: no CI>
    axordv   = < range of the vertical axis and
                 tick-mark spacing for odds ratio or rate ratio
                 plots, <low> to <high> by <increment>) >,
    axordh   = < range of the vertical axis and tick mark spacing
                 for incidence rate plots,
                 (<low> to <high> by <increment>) >,
    displayx=T,          <3 values. T: smooth histgram, RUGPLOT, and F>
    densnum=F,     <whether to show numbers on vertical axis of smoothed hist.>
    outplot=PS,   <default format: ps file>
```

3 Example with output

Here is an example to illustrate what is required and how to use this macro. This is from Heiki A. Bischoff-Ferrari's calcium intake and bone density meta-analysis for men. Five studies were used in this meta-regression (see ref #7 for detail) and we would like to do a dose response meta analysis. They are all cohort studies with total subjects provided.

First, you need a dataset with the required variables: study name, median of each category in common units across studies (dose), risk ratio, upper and/or lower bound, study type, total number of subjects, and total number of cases for each level. Here is the dataset:

```
data cal_men;
input
citation $     Cit_No    calcium    effect    lower    upper    type    N    A;
datalines;
   Owusu          1       359      1.00       .        .        2     8613    8
   Owusu          1      1334      1.19      0.42     3.35      2     8613   12
   Owusu          1      1040      0.75      0.25     2.27      2     8613    7
   Owusu          1       708      1.54      0.60     4.00      2     8613   14
   Owusu          1       596      1.78      0.73     4.33      2     8613   15
   Holbrook       2       396      1.00       .        .        2      142    6
   Holbrook       2       853      0.30      0.08     2.13      2      284    9
   Meyer          4       436      1.00       .        .        2     4907   15
   Meyer          4      1092      0.64      0.28     1.45      2     4907   10
   Meyer          4       925      1.08      0.53     2.21      2     4906   16
   Meyer          4       723      0.96      0.46     2.00      2     4909   14
   Paganini-Hill  5       283      1.00       .        .        2     1202   21
   Paganini-Hill  5      1017      1.11      0.76     6.46      2      705   13
   Paganini-Hill  5       641      0.79      0.14     2.46      2     1059   16
   Looker         8       283      1.00       .        .        2      529   16
   Looker         8      1107      0.53      0.20     1.20      2      529    9
   Looker         8       829      0.67      0.30     1.50      2      529   10
   Looker         8       530      0.52      0.20     1.20      2      529    9
;

run;
```

Note that the reference level is the first line for each study with the value of effect 1.00 and 95% CIs set to missing.

Second, call the macro:

```
%include "/udd/strui/metadose/Macros/metadose.sas";
title "Dose response meta analysis among men for calcium and bone density studies";
%metadose(  dat=cal_men,
            ratio=effect,
            UB=upper,
            Ncase=A, Ntotal=N,
            dose=calcium,
            studyname=citation,
            studytype=type,
            meta=T, wt=300, unit_wt=mg/day,
            var_covar=GH, linearCheck=1, ci=2, graphtitle=calcium effect on men
);
```

Here is the output:

These are the input data plus the estimated cell counts using both methods

Obs	citation	Median dose	RR/OR	UL	Cases	Greenland Est. Cases	Hamling Est. Cases	Actual Person-time /# of subjects	Hamling Est. Person -time/Subjects
1	Holbrook	396	1.00	.	5.1425	9.3750	5.1425	142	28.2117
2	Holbrook	853	0.30	1.16	3.0855	5.6250	3.0855	284	56.4234
3	Looker	283	1.00	.	11.2211	16.1765	11.2211	529	38.8918
4	Looker	1107	0.53	1.20	5.8209	8.5735	5.8209	529	38.0658
5	Looker	829	0.67	1.50	7.6704	10.8382	7.6704	529	39.6796
6	Looker	530	0.52	1.20	5.8407	8.4118	5.8407	529	38.9299
7	Meyer	436	1.00	.	12.4590	14.9450	12.4590	4907	82.4330
8	Meyer	1092	0.64	1.45	8.3734	9.5648	8.3734	4907	86.5641
9	Meyer	925	1.08	2.21	12.9625	16.1373	12.9625	4906	79.4111
10	Meyer	723	0.96	2.00	11.8022	14.3530	11.8022	4909	81.3406
11	Owusu	359	1.00	.	6.2165	8.9457	6.2165	8613	53.0282
12	Owusu	1334	1.19	3.35	6.2086	10.6454	6.2086	8613	44.5046
13	Owusu	1040	0.75	2.27	5.2202	6.7093	5.2202	8613	59.3722
14	Owusu	708	1.54	4.00	8.8859	13.7764	8.8859	8613	49.2199
15	Owusu	596	1.78	4.33	12.3149	15.9233	12.3149	8613	59.0161
16	Paganini	283	1.00	.	3.3134	21.3033	3.3134	1202	6.7412
17	Paganini	1017	1.11	6.46	3.1428	13.8693	3.1428	705	5.7604
18	Paganini	641	0.79	2.46	1.6047	14.8274	1.6047	1059	4.1326

These are the point estimates and input data using both methods of meta-analysis for dose-response
RR/OR is given in an increment of 300 mg/day

Obs	studyname	G_BSTAR	G_SESTAR	G_RRSTAR	H_BSTAR	H_SESTAR	H_RRSTAR
1	Holbrook	-.001411457	.000799743	0.65479	-.001411457	.000799743	0.65479
2	Looker	-.000543779	.000369408	0.84948	-.000554078	.000364986	0.84686
3	Meyer	-.000168184	.000329286	0.95080	-.000167815	.000328672	0.95090
4	Owusu	-.000081962	.000375165	0.97571	-.000094242	.000373156	0.97212
5	Paganini	0.000084623	.000535312	1.02571	0.000080429	.000534486	1.02442

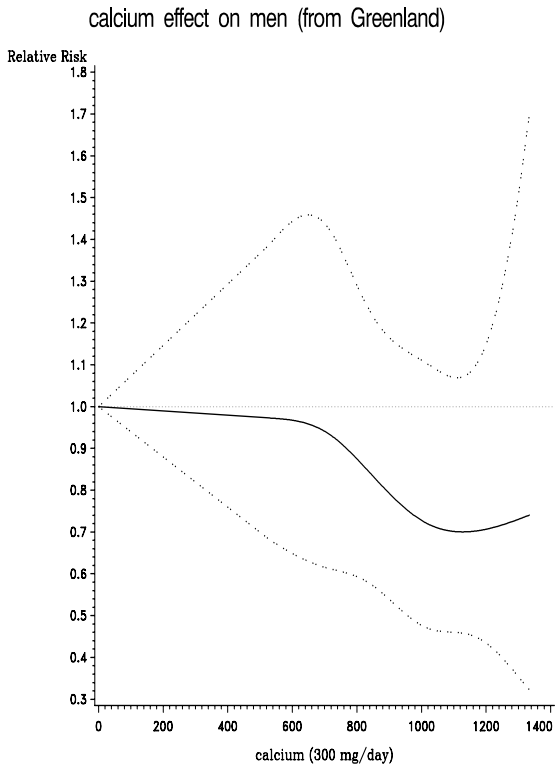
The following are the results from the meta regression

Model	Method	Pooled Est (SE)	RR/OR(CI)	Z-score	linear trend p-value	test for heterogeneity tau**2	p-value	Q	df
fixed	Greenland	-0.0003(0.0002)	0.9195(0.8239, 1.0262)	-1.4986	0.1340		0.4980	3.3694	4
fixed	Hamling	-0.0003(0.0002)	0.9176(0.8227, 1.0235)	-1.5438	0.1226		0.4958	3.3833	4
Random	Greenland	-0.0003(0.0002)	0.9195(0.8239, 1.0262)	-1.4986	0.1340	0.0000			
Random	Hamling	-0.0003(0.0002)	0.9176(0.8227, 1.0235)	-1.5438	0.1226	0.0000			

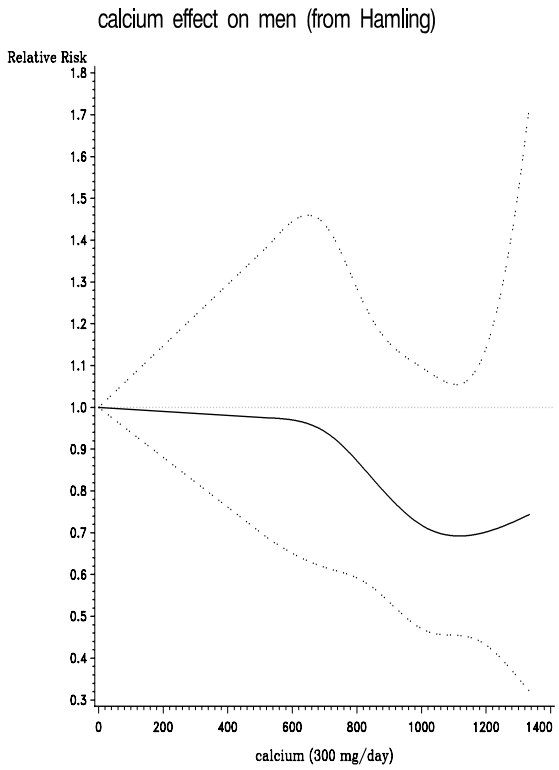
```
====================================
Assessment of non-linearity - pvalues
====================================
                                        Greenland          Hamling
Test for non-linearity                  0.7304586          0.7037282
Test for overall significance of curve  0.2376305          0.2137401
```

Since we want to do nonlinearity check, the SAS macro also produced publication quality graphs. Note however there is no evidence for significant non-linearity regardless of the method used.

The following is the graph using Greenland method:



The following is the graph using Hamling method:



See the reference #3 for more examples.

4 Credits

This macro is written by Ruifeng Li with the help of Prof. Donna Spiegelman, and Ellen Hertzmark written the submacro %meta.

Any questions should be addressed to Ruifeng Li via email strui@channing.harvard.edu or via phone 617-432-6321.

5 References

1. Greenland S, Longnecker MP. Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. Am J Epidemiol 1992; 135:1301-1309.

2. Hamling J, Lee P. etc. Facilitating meta-analyses by deriving relative effect and precision estimates for alternative comparisons from a set of estimates presented by exposure level or disease category. Statistics in Medicine 2008, Volume 27, Issue 7, Date: 30 March 2008, Pages: 954-970

3. Orisni N., Li R., Wolk A., Spiegelman D.: Meta-analysis for linear and non-linear dose-response relationships: examples, an evaluation of approximations, and software. Submit to AJE for publication

4. More references indirectly related with this macro can be seen in the references section of #2

5. Reference for meta macro:
   DerSimonian R, Laird NM (1986).
   Meta-analysis in clinical trials. Controlled Clinical Trials 7: 177-188.

6. Durrleman, S. and R. Simon, Flexible regression models with cubic splines. Statistics in Medicine, 1989. 8(5): P. 551-561.

7. Govindarajulu US, Spiegelman D, Thurston SW, Ganguli B, Eisen EA. Comparing smoothing techniques in Cox models for exposure-response relationships. Statistics in Medicine, 2007; 26: 3735-3752.

7. Bischoff-Ferrari HA et al. Calcium Intake and Risk of Hip Fracture in Men and Women: A Meta-Analysis of Prospective Cohort Studies and Randomized Controlled Trials. American Journal of Clinical Nutrition 2007.