The SAS MEDIATE Macro

Ellen Hertzmark, Mathew Pazaris, and Donna Spiegelman

June 6, 2012

Abstract

The %MEDIATE macro calculates the point and interval estimates of the percent of treatment (exposure) effect (PTE) explained by one or more intermediate variables. The macro is designed for treatment effects estimated as relative risks in Cox regression survival analysis using PROC PHREG and for treatment effects from generalized linear models using PROC GENMOD. When fitting log-binomial models with PROC GENMOD, an option is available to improve model convergence. **Keywords: SAS, macro, Cox models, proportional hazards regression, intermediate variable, percent of treatment effect explained, mediation proportion, generalized linear models, log-binomial models**

Contents

1 Description

The %MEDIATE macro computes the mediation proportion of one or more treatment or exposure effect(s) that are explained by the effect of the treatment or exposure on one or more intermediate variables. For example, the mediation proportion quantifies the extent to which the preventive effect of Lipitor on MI risk is mediated through Lipitor's effect on LDL cholesterol.
The macro is designed for exposure effects estimated as relative risks in survival analysis using PROC PHREG in SAS or with effects estimated as regression slopes, or function thereof in generalized linear models using PROC GENMOD.

2 Invocation and Details

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

```
mautosource sasautos= <directories where macros are located>.
```

For example, at the Channing Lab, an options statement might be

```
        options nocenter ps=78 ls=125 replace formdlim='['
        mautosource
        sasautos=('/usr/local/channing/sasautos',
           '/proj/nhsass/nhsas00/nhstools/sasautos);
```

This will allow you to use all the SAS read macros for the data sets (`/proj/nhsass/nhsas00/nhstools/sasautos`), as well as other public SAS macros, such as %PM, %INDIC3, %EXCLUDE, %LOGITR, and %MPHREG9.

The macro call is

```
%mediate(

General Options

  DATA     = The name of the dataset
             REQUIRED
  ID       = The name(s) of >= 1 variable(s) that uniquely
             identifies each record (e.g
             id or id period)
             REQUIRED
  EXPOSURE = The main exposure or treatment variable of interest,
             expressed as ONE VARIABLE, or alternately you may use a set of
             indicators for an EXPOSURE.
             REQUIRED
  INTERMED = The intermediate variable(s).
             This can be a set of indicators, or any other
             representation of the intermediate variable, such
             as a set of spline indicators.
             REQUIRED
  COVARS   = List of covariates in the model, if any
             OPTIONAL
  INTMISS  = Whether you want to use missing indicators for
             unknown values of the INTERMED variable vs the
             model-specific complete case analysis.
             Default=F
             OPTIONAL
  WHERE    = A sub setting clause, if desired
             NOTE:  if any of the variables named in WHERE is not
             among TIME, EVENT, EXPOSURE, INTERMED, COVARS, then
             they should be listed in EXTRAV (see next).
             OPTIONAL
  EXTRAV   = A list of variables used in the WHERE clause that
             are not part of the model or strata (see above)
             OPTIONAL
  SURV     = If this is a survival analysis set to T, if a
             generalized linear model, set to F (default=T)
             OPTIONAL

SURV=T Options

  TIME     = The survival time variable (time to outcome or censoring)
             REQUIRED if surv=T
  EVENT    = The event variable (1=yes, 0=no)
             REQUIRED if surv = T
  STRATA   = Strata for the PROC PHREG, if desired.
             These would usually be the same as the strata used
             in the original PROC PHREG or MPHREG9 analysis,
             typically AGEMO and year of questionnaire return.
             OPTIONAL
  MODPRINT = Whether you want to print the results of the
             PROC PHREG used in the macro
             Default=F
             OPTIONAL
  TIES     = Ties option for phreg (default=breslow)
             OPTIONAL
  PROCOPT  = Procedure options for phreg
             OPTIONAL
  MODOPT   = Model options for phreg
             OPTIONAL
```

```
SURV=F Options

  OUTCOME  = The name of the dependent variable when surv (see above) = F
             REQUIRED if surv = F
  TYPE     = If using a log-binomial(relative risk) regression model,
             indicates if relrisk9 macro should be used to help with
             convergence.  type=1 indicates that relrisk9 should be used.
             Type = 0 indicates relrisk9 should not be used. (default=1).
             OPTIONAL
  DIST     = proc genmod distribution option for use with type=0 (default=nor)
             OPTIONAL
  LINK     = proc genmod distribution option for use with type=0 (default=identity)
             OPTIONAL
  RR2      = If using a log-binomial(relative risk) regression model, PTE is
             normally calculated from the beta coefficients and is 1-(b/a).  Setting
             this option to 1 tells the macro to calculate the PTE from the relative
             risks   using a method described in the literature(RRa-RRb)/(RRa-1),
             however this is not symmetric. This is displayed in addition to the PTE
             and is labeled PTE_RR. Also, a third alternative calculation method is
             used, 1-RRb/RRa, where the PTE is calculated using The relative risks.
             This method is symmetric, but is a new idea, not described in the
             literature. This is labeled PTE_RR2_alt.
  debugdv  = Option is used for debuging.  Tells macro to print the genmod dist and
             type options used in the final model.  This is useful for determining
             if the log-binomial model converged or if the log-poisson model was used
             instead. (optional).
);
```

The macro checks for two inconsistencies in the results. First, it makes sure that the model has converged. (If it does not, a ERROR message is printed instead of the usual output.) Second, it checks whether the calculated percent of treatment effect explained is between 0% and 100%. If it is not, then the model estimates are printed without the calculated effect, and the macro gives an ”ERROR in macro run” message, saying ”INTERMED is not intermediate to EXPOSURE.”

3 PHREG examples

In the examples, we are interested in the effect of whole grain consumption on the incidence of type II diabetes in NHS, and how much of that effect is mediated by the effect of whole grain consumption on lowering BMI.
The original analysis was done in %MPHREG9 (stratified by age in years and time period). The variable for the number of grams per day of whole grain (divided by 40) is `WGRAIN40`. The variable for BMI is the set of indicators `&BMIC_`. The covariates were sets of indicators for coffee consumption (`COF`), alcohol consumption (`ALCC`), activity (`PACUMQ`), family history of diabetes (`FAMDB`), smoking (`SMKC`), hormone replacement therapy (`PMHC`), polyunsaturated to saturated fat ratio (`PSRAT`), OC use (`OC_EVER`), energy (`CAL`), processed meat (`PRMEAT`), carbonated beverage consumption (`SOD`), non-carbonated soft drink consumption (`PUN`). Continuous variables were modeled as quintile indicators.

Before using the whole grain variable as a continuous variable, we checked that there were no outliers by running %DESCR8, which shows the outlier bounds, as well as the minimum and maximum values of the variable(s).

Just to see what would happen, we ran the macro with the whole grain variable in its original (continuous) form, as well as using the trend variable from the quintile analysis.To show the effect of INTMISS, we ran the macro with INTMISS=T and INTMISS=F.We also used the whole data set and the dataset restricted to the 92% of the observations with non-missing BMI (`BMI_UP`).

3.1 Example 1. Original continuous variable, INTMISS=T, all data

The first macro call was

```
title2 'all data, wgrain40, missing indicator';
%mediate(data=all1, exposure=wgrain40, intermed=bmi_up, id=id interval,
time=tdb, event=dbcase,
intmiss=t,
covars= &calr_ famdb &pmhc_ &smkc_ &pacumq_ &alcc_ oc_ever
&cof_ &prmeat_ &sod_ &pun_ &psrat_   t86 t88 t90 t92 t94 t96 t98 t00
agemo agemo1);
```

The result was

```
[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[

The SAS System                          16:09 Wednesday, December 13, 2006   7
wgrain40, missing indicator


Relative risk for outcome dbcase with exposure wgrain40
Calculating the proportion of treatment effect explained by bmi_up
Adjusted for calr1 calr2 calr3 calr4 famdb pmhc2 pmhc3 pmhc4 pmhcm smkc2 smkc3
 smkc4 smkc5 smkcm pacumq2 pacumq3 pacumq4 pacumq5 pacumqm alcc2 alcc3 alcc4
oc_ever cof1 cof2 cof3 cof4 prmeat1 prmeat2 prmeat3 prmeat and other variables
Missing indicator used for missing values of bmi_up
      628234 observations read.
      628233 observations used without bmi_up in the model.
      628233 observations used with    bmi_up in the model.

Exposure coefficient (RR) unadjusted for bmi_up: 0.54 ( 0.47 --  0.61)
Exposure coefficient (RR)   adjusted for bmi_up: 0.68 ( 0.60 --  0.77)
Proportion of wgrain40 effect explained by bmi_up:
                             39.2% ( 30.7% --  47.1%)   p = <.0001
[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[
```

So the effect of whole grain consumption on BMI is responsible for 39% of the protective effect of whole grain consumption on diabetes incidence.

3.2 Example 2. Original continuous variable, INTMISS=F, all data

The macro call is

```
title2 'all data, wgrain40, no missing indicator';
%mediate(data=all1, exposure=wgrain40, intermed=bmi_up, id=id interval,
time=tdb, event=dbcase,
intmiss=f,
covars= &calr_ famdb &pmhc_ &smkc_ &pacumq_ &alcc_ oc_ever
&cof_ &prmeat_ &sod_ &pun_ &psrat_   t86 t88 t90 t92 t94 t96 t98 t00
agemo agemo1);
```

and the PTE results are

```
[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[

The SAS System                          11:27 Thursday, December 14, 2006   7
wgrain40, no missing indicator


Relative risk for outcome dbcase with exposure wgrain40
Calculating the proportion of treatment effect explained by bmi_up
Adjusted for calr1 calr2 calr3 calr4 famdb pmhc2 pmhc3 pmhc4 pmhcm smkc2 smkc3
 smkc4 smkc5 smkcm pacumq2 pacumq3 pacumq4 pacumq5 pacumqm alcc2 alcc3 alcc4
oc_ever cof1 cof2 cof3 cof4 prmeat1 prmeat2 prmeat3 prmeat and other variables
      628234 observations read.
      628233 observations used without bmi_up in the model.
      577813 observations used with    bmi_up in the model.

Exposure coefficient (RR) unadjusted for bmi_up: 0.54 ( 0.47 --  0.61)
Exposure coefficient (RR)   adjusted for bmi_up: 0.68 ( 0.59 --  0.77)
Proportion of wgrain40 effect explained by bmi_up:
                             37.8% ( 27.1% --  47.7%)   p = <.0001
```

Because INTMISS=F, different numbers of observations were used for the two analyses (models with and without bmi_up). The reported RRs are nevertheless the same, to two decimal places, compared to Example 1, where the missing indicator method was used for BMI_UP and the same data was used to estimate the RRs in both models.

3.3 Example 3. Excluding all observations with missing BMI from the data

We next made a new dataset, ALL2, from which all observations with missing BMI were excluded. The results of running %MEDIATE (with INTMISS=F, of course) on this datasets were

```
[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[

The SAS System                          11:27 Thursday, December 14, 2006   9
wgrain40, no missing bmi_up


Relative risk for outcome dbcase with exposure wgrain40
Calculating the proportion of treatment effect explained by bmi_up
Adjusted for calr1 calr2 calr3 calr4 famdb pmhc2 pmhc3 pmhc4 pmhcm smkc2 smkc3
 smkc4 smkc5 smkcm pacumq2 pacumq3 pacumq4 pacumq5 pacumqm alcc2 alcc3 alcc4 o
c_ever cof1 cof2 cof3 cof4 prmeat1 prmeat2 prmeat3 prmeat and other variables
      577813 observations read.
      577813 observations used without bmi_up in the model.
      577813 observations used with    bmi_up in the model.

Exposure coefficient (RR) unadjusted for bmi_up: 0.52 ( 0.45 --  0.59)
Exposure coefficient (RR)   adjusted for bmi_up: 0.68 ( 0.59 --  0.77)
Proportion of wgrain40 effect explained by bmi_up:
                             41.3% ( 32.3% --  49.6%)   p = <.0001
```

3.4 Example 4. Using trend variable instead of original continuous variable

This is an alternative approach when there are missing data in the intermediate. There is no general right answer as to which method is better - each makes somewhat different empirically unverifiable assumptions about the nature of confounding and of the missing data mechanisms. In this example, we again use data ALL2, but the exposure variable is WGRAINT40 (the median score of quintile–divided by 40). The result is

```
[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[

The SAS System                           11:27 Thursday, December 14, 2006   8
wgraint40, no missing bmi_up


Relative risk for outcome dbcase with exposure wgraint40
Calculating the proportion of treatment effect explained by bmi_up
Adjusted for calr1 calr2 calr3 calr4 famdb pmhc2 pmhc3 pmhc4 pmhcm smkc2 smkc3
 smkc4 smkc5 smkcm pacumq2 pacumq3 pacumq4 pacumq5 pacumqm alcc2 alcc3 alcc4 o
c_ever cof1 cof2 cof3 cof4 prmeat1 prmeat2 prmeat3 prmeat and other variables
      577813 observations read.
      577813 observations used without bmi_up in the model.
      577813 observations used with    bmi_up in the model.

Exposure coefficient (RR) unadjusted for bmi_up: 0.47 ( 0.41 --  0.54)
Exposure coefficient (RR)   adjusted for bmi_up: 0.64 ( 0.55 --  0.74)
Proportion of wgraint40 effect explained by bmi_up:
                             41.4% ( 32.2% --  49.8%)   p = <.0001
```

In this case, there were only minor differences between the results with the original continuous variable and those with the trend variable.

4 GENMOD Examples

In the examples, we are interested in the effect of sexual orientation on the incidence of early onset of tobacco use by age 14 in NHS II, and how much of that effect is mediated through the effect of childhood physical and sexual abuse (Jun et al. 2010). We will examine both a binary outcome for smoking before age 15, as well as, a continuous outcome for average number of cigarettes per day before age 15.

The binary variable for smokers who start smoking before age 15 is agesmk13. The continuous variable for cigarettes per day before age 15 is cigcon1589. The variable for sexual orientation is the set of indicators lesbia and bisexu. The variable for abuse is the set of indicators vioc01v2, vioc01v3, vioc01v4,sexc01v2 and sexc01v3. Other included covariates are age89,
black5, hispa5, asian5, orace5, racem, st15midw2,

`st15south2, st15west2, st15mnus, dadsm99, momsm99, botsm99, missm99, msomcol, mhigrad, mlthigr and mdkmiss`.

The dataset was restricted to observations with non-missing abuse data, so INTMISS remained at its default value(F).

4.1 Example 1. Binary outcome

Here we calculate what proportion of the relative risk of early smoking initiation in relation to sexual orientation (lesbian, bisexual) is mediated by experiences of childhood physical and sexual abuse, after adjusting for confounders by other risk factors for early smoking. We do so using the binary smoking variable `agesmk13`.

The macro call was

```
%include '/udd/stmjp/SAS_Projects/Macros/mediate/docs/examples/master.so.smkini.sas';

%mediate(data=smkini_13, id=id, exposure= lesbia bisexu,
intermed= vioc01v2 vioc01v3 vioc01v4 sexc01v2 sexc01v3,
covars=age89 black5 hispa5 asian5 orace5 racem st15midw2 st15south2
st15west2 st15mnus dadsm99 momsm99 botsm99 missm99 msomcol mhigrad
mlthigr mdkmiss,  intmiss=F, outcome=agesmk13,  modprint=t,
notes=nonotes, where=, extrav=, procopt=, modopt=, type = 1, surv=0);
```

The result was

```
[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[

Relative risk for outcome: agesmk13, exposure:  lesbia
Calculating the proportion of treatment effect mediated by vioc01v2 vioc01v3 vio
c01v4 sexc01v2 sexc01v3
Adjusted for:  age89 black5 hispa5 asian5 orace5 racem st15midw2 st15south2 st15
west2 st15mnus dadsm99 momsm99 botsm99 missm99 msomcol mhigrad mlthigr mdkmiss
Exposure effect unadjusted for the hypothesized intermediates
     vioc01v2 vioc01v3 vioc01v4 sexc01v2 sexc01v3:
            2.053 (     1.576 --      2.674)
Exposure effect adjusted for the hypothesized intermediates
     vioc01v2 vioc01v3 vioc01v4 sexc01v2 sexc01v3:
            1.824 (     1.398 --      2.381)
Proportion of lesbia effect mediated by
      vioc01v2 vioc01v3 vioc01v4 sexc01v2 sexc01v3:
                 16.402% (     8.117% --     24.687%)   p = 0.0001

Relative risk for outcome: agesmk13, exposure:  bisexu
Calculating the proportion of treatment effect mediated by vioc01v2 vioc01v3 vio
c01v4 sexc01v2 sexc01v3
Adjusted for:  age89 black5 hispa5 asian5 orace5 racem st15midw2 st15south2 st15
west2 st15mnus dadsm99 momsm99 botsm99 missm99 msomcol mhigrad mlthigr mdkmiss
Exposure effect unadjusted for the hypothesized intermediates
     vioc01v2 vioc01v3 vioc01v4 sexc01v2 sexc01v3:
            2.852 (     2.088 --      3.895)
```

```
Exposure effect adjusted for the hypothesized intermediates
     vioc01v2 vioc01v3 vioc01v4 sexc01v2 sexc01v3:
           2.617 (     1.925 --      3.559)
Proportion of bisexu effect mediated by
      vioc01v2 vioc01v3 vioc01v4 sexc01v2 sexc01v3:
                  8.185% (     2.617% --     13.753%)   p = 0.0040

[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[
```

So the effect of childhood physical activity and sexual abuse could explain 16% of the effect of sexual orientation on early onset smoking for lesbian women and is highly significant with p = 0.0001. Additionally, the effect of childhood physical and sexual abuse could explain 8% of the effect in bisexual women and again is highly significant with p = 0.004.

4.2 Example 2. Continuous outcome

As in example 1, here we again calculate what proportion of the relative risk of early smoking initiation in relation to sexual orientation (lesbian, bisexual)is mediated by experiences of childhood physical and sexual abuse, after adjusting for confounders by other risk factors for early smoking. This time however, we use the continuous variable `cigcon1589` (average number of cigarettes smoked per day before age 15).

The macro call was

```
%mediate(data=smkrs_13, id=id, exposure= lesbia bisexu,
intermed= vioc01v2 vioc01v3 vioc01v4 sexc01v2 sexc01v3,
covars=age89 black5 hispa5 asian5 orace5 racem st15midw2 st15south2 st15west2 st15mnus
dadsm99 momsm99 botsm99 missm99 msomcol mhigrad mlthigr mdkmiss,  intmiss=F, outcome
=cigcon1589, notes=nonotes, type = 0, surv=0);
```

the result was

```
[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[

vioc01v2 vioc01v3 vioc01v4 sexc01v2 sexc01v3 is not intermediate to lesbia

note: proc genmod was run using dist = nor and type = IND

Relative risk for outcome: cigcon1589, exposure:  bisexu
Calculating the proportion of treatment effect mediated by vioc01v2 vioc01v3 vioc01v4
sexc01v2 sexc01v3 Adjusted for:  age89 black5 hispa5 asian5 orace5 racem st15midw2
st15south2 st15west2 st15mnus dadsm99 momsm99 botsm99 missm99 msomcol mhigrad mlthigr
mdkmiss
Exposure coefficient unadjusted for the hypothesized intermediates
     vioc01v2 vioc01v3 vioc01v4 sexc01v2 sexc01v3:
           1.126 (    -1.541 --      3.793)
Exposure coefficient adjusted for the hypothesized intermediates
     vioc01v2 vioc01v3 vioc01v4 sexc01v2 sexc01v3:
            0.844 (    -1.820 --      3.507)
Proportion of bisexu effect mediated by
      vioc01v2 vioc01v3 vioc01v4 sexc01v2 sexc01v3:
                 25.057% (   -38.422% --     88.536%)   p = 0.4391

[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[
```

In this example, among lesbian women, (exposure1 / exposure2) was not between 0 and 1. That is, the effect of being a lesbian on early initiation of smoking was larger when adjusting for the hypothesized intermediates than not. Since intermediate variables must always attenuate the effect of an exposure, this analysis suggests that childhood violence is not intermediate in this case. The effect of sexual orientation on childhood physical and sexual abuse was responsible for 25% of the effect of sexual orientation on cigarettes per day before age 15 for bisexual women, however, but this result was not significant (p = 0.44).

4.3 Example 3. Use of Type=1 option to improve convergence in relative risk regression

Here, we will examine what happens when the log-binomial model fails to converge when using the type=1 option to use the the log-poisson model is used instead, with robust variance to guarantee valid asymptotic inference. For this example, we will examine the effect of sexual orientation on the incidence of use of Alcohol at age 15-17 (variable alco1589) in NHS II, and again, will look at how much of that effect is mediated by the effect of childhood physical and sexual abuse.

Note: For this example, we have set `debugdv=1`. This option tells the macro to print the genmod `dist` and `type` options used in the final model. We will use this option here in order to better understand what is going on.

The macro call was

```
[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[


%Mediate(Data=Alcoh_15, Id=Id, Exposure= Lesbia Bisexu,
Intermed= Vioc01v2 Vioc01v3 Vioc01v4 Sexc01v2 Sexc01v3,
Covars=Age89 Black5 Hispa5 Asian5 Orace5 Racem St15midw2 St15south2
St15west2 St15mnus Dadsm99 Momsm99 Botsm99 Missm99 Msomcol Mhigrad Mlthigr Mdkmiss,
Intmiss=F, Outcome=Alco1589,  Modprint=T, Notes=Nonotes, Where=, Extrav=, Procopt=,
Modopt=, Type = 1, Surv=0, debugdv=1);

[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[

The result is

note: proc genmod was run using dist = poisson and type = IND

Relative risk for outcome: Alco1589, exposure:  Lesbia
Calculating the proportion of treatment effect mediated by Vioc01v2 Vioc01v3 Vio
c01v4 Sexc01v2 Sexc01v3
Adjusted for:  Age89 Black5 Hispa5 Asian5 Orace5 Racem St15midw2 St15south2 St15
west2 St15mnus Dadsm99 Momsm99 Botsm99 Missm99 Msomcol Mhigrad Mlthigr Mdkmiss
```

```
Exposure effect unadjusted for the hypothosized intermediates
     Vioc01v2 Vioc01v3 Vioc01v4 Sexc01v2 Sexc01v3:
            1.505 (     1.324 --      1.711)
Exposure effect adjusted for the hypothesized intermediates
     Vioc01v2 Vioc01v3 Vioc01v4 Sexc01v2 Sexc01v3:
            1.462 (     1.287 --      1.660)
Proportion of Lesbia effect mediated by
      Vioc01v2 Vioc01v3 Vioc01v4 Sexc01v2 Sexc01v3:
                  7.169% (     3.517% --     10.821%)   p = 0.0001

note: proc genmod was run using dist = poisson and type = IND


Relative risk for outcome: Alco1589, exposure:  Bisexu
Calculating the proportion of treatment effect mediated by Vioc01v2 Vioc01v3 Vio
c01v4 Sexc01v2 Sexc01v3
Adjusted for:  Age89 Black5 Hispa5 Asian5 Orace5 Racem St15midw2 St15south2 St15
west2 St15mnus Dadsm99 Momsm99 Botsm99 Missm99 Msomcol Mhigrad Mlthigr Mdkmiss
Exposure effect unadjusted for the hypothosized intermediates
     Vioc01v2 Vioc01v3 Vioc01v4 Sexc01v2 Sexc01v3:
            1.542 (     1.296 --      1.834)
Exposure effect adjusted for the hypothesized intermediates
     Vioc01v2 Vioc01v3 Vioc01v4 Sexc01v2 Sexc01v3:
            1.505 (     1.267 --      1.789)
Proportion of Bisexu effect mediated by
      Vioc01v2 Vioc01v3 Vioc01v4 Sexc01v2 Sexc01v3:
                  5.507% (     1.754% --      9.261%)   p = 0.0040
```

By choosing the option `debugdv=1`, we see a note in the log (note: proc genmod was run using `dist = poisson` and `type = IND`), indicates that because the log-binomial model did not converge, the log-poisson model was used instead. These errors may be ignored and the final results provided by the macro are valid.

Further examination of the log reveals a set of warnings and errors resulting from the log-binomial run

```
WARNING: The relative Hessian convergence criterion of 0.0134081216 is greater
         than the limit of 0.0001. The convergence is questionable.
WARNING: The procedure is continuing but the validity of the model fit is
         questionable.
ERROR: The mean parameter is either invalid or at a limit of its range for some
       observations.
ERROR: Error in parameter estimate covariance computation.
ERROR: Error in estimation routine.
WARNING: Output 'geencov' was not created.  Make sure that the output object
         name, label, or path is spelled correctly.  Also, verify that the
         appropriate procedure options are used to produce the requested output
         object.  For example, verify that the NOPRINT option is not used.
WARNING: Output 'GEENCorr' was not created.  Make sure that the output object
         name, label, or path is spelled correctly.  Also, verify that the
         appropriate procedure options are used to produce the requested output
         object.  For example, verify that the NOPRINT option is not used.
```

```
WARNING: Output 'GEERCorr' was not created.  Make sure that the output object
         name, label, or path is spelled correctly.  Also, verify that the
         appropriate procedure options are used to produce the requested output
         object.  For example, verify that the NOPRINT option is not used.
WARNING: Output 'geercov' was not created.  Make sure that the output object
         name, label, or path is spelled correctly.  Also, verify that the
         appropriate procedure options are used to produce the requested output
         object.  For example, verify that the NOPRINT option is not used.
WARNING: Output 'geencov' was not created.  Make sure that the output object
         name, label, or path is spelled correctly.  Also, verify that the
         appropriate procedure options are used to produce the requested output
         object.  For example, verify that the NOPRINT option is not used.
WARNING: The quoted string currently being processed has become more than 262
         characters long.  You may have unbalanced quotation marks.
WARNING: The quoted string currently being processed has become more than 262
         characters long.  You may have unbalanced quotation marks.
WARNING: Data set WORK.PPAR was not replaced because this step was stopped.
```

4.4 Example 4. Using the Relative Risk PTE option

Here, we will we will use the same model used in example 3, where we looked at the effect of sexual orientation on the incidence of Alcohol at age 15-17 (variable alco1589) in NHS II, to determine how much of that effect is mediated through the effect of childhood physical and sexual abuse. Here however, we will use the RR PTE option.

```
[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[

%mediate(data=alcoh_15, id=id, exposure= lesbia bisexu,
intermed= vioc01v2 vioc01v3 vioc01v4 sexc01v2 sexc01v3,
covars=age89 black5 hispa5 asian5 orace5 racem st15midw2 st15south2
st15west2 st15mnus dadsm99 momsm99 botsm99 missm99 msomcol mhigrad mlthigr mdkmiss,
intmiss=F, outcome =alco1589,  modprint=t, notes=nonotes, type = 1, surv=0,  RR2=1);

[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[
```

```
Relative risk for outcome: alco1589, exposure:  lesbia
Calculating the proportion of treatment effect mediated by vioc01v2 vioc01v3 vio
c01v4 sexc01v2 sexc01v3
Adjusted for:  age89 black5 hispa5 asian5 orace5 racem st15midw2 st15south2 st15
west2 st15mnus dadsm99 momsm99 botsm99 missm99 msomcol mhigrad mlthigr mdkmiss
Exposure effect unadjusted for the hypothesized intermediates
     vioc01v2 vioc01v3 vioc01v4 sexc01v2 sexc01v3:
            1.505 (     1.324 --      1.711)
Exposure effect adjusted for the hypothesized intermediates
     vioc01v2 vioc01v3 vioc01v4 sexc01v2 sexc01v3:
            1.462 (     1.287 --      1.660)
Proportion of lesbia effect mediated by
      vioc01v2 vioc01v3 vioc01v4 sexc01v2 sexc01v3:
         PTE =      7.169% (     3.517% --     10.821%)   p = 0.0001
         PTE_RR =      8.608% (5.3021492802 % -- 13.677246848 %)   p = <.0001
         PTE_RR_alt =     2.889% (1.8759899612 % -- 4.425303042 %)   p = <.0001


Relative risk for outcome: alco1589, exposure:  bisexu
Calculating the proportion of treatment effect mediated by vioc01v2 vioc01v3 vio
c01v4 sexc01v2 sexc01v3
Adjusted for:  age89 black5 hispa5 asian5 orace5 racem st15midw2 st15south2 st15
west2 st15mnus dadsm99 momsm99 botsm99 missm99 msomcol mhigrad mlthigr mdkmiss
Exposure effect unadjusted for the hypothesized intermediates
     vioc01v2 vioc01v3 vioc01v4 sexc01v2 sexc01v3:
            1.542 (     1.296 --      1.834)
Exposure effect adjusted for the hypothesized intermediates
     vioc01v2 vioc01v3 vioc01v4 sexc01v2 sexc01v3:
            1.505 (     1.267 --      1.789)
Proportion of bisexu effect mediated by
      vioc01v2 vioc01v3 vioc01v4 sexc01v2 sexc01v3:
         PTE =      5.507% (     1.754% --      9.261%)   p = 0.0040
         PTE_RR =      6.705% (3.490225665 % -- 12.496651597 %)   p = 0.0022
         PTE_RR_alt =     2.356% (1.2933547852 % -- 4.2532503905 %)   p = 0.0010
```

5 Frequently Asked Questions

5.1 Q: In the variable have RR=1.00, but the PTE is not 0%

A: This is a rounding issue. The macro output rounds the RRs to 2 decimal places. If both RRs are at least .995, they round to 1.00. But one might be .996 and the other might be .999. This is why we used *wgrain40* in our models rather than the original variable *wgrain* (i.e. to give reasonable values away from 1 for the point estimates and to show they were different from each other).

6 Computational Methods

The macro compares a full model (Model 1) that includes the exposure, an intermediate variable and any covariates with a partial model (Model 2) that leaves out the intermediate variable. Following Lin et al.(1997) the macro first makes a duplicate of each record, which results in one record for model 1 and one record for model 2. It also makes a duplicate of the exposure, intermediate and covariate variables in each of the above records. This creates, for example, the variables `exposure1` and `exposure2`, `intermed1` and `intermed2` , etc. In the records for model 1, the first set of variables have the original values. The second set of variables are all set to zero. Conversely, in the records for model 2, the first set of variables are all set to zero. All of the second set of variables are set to the original values with one exception: the intermediate variable (*intermed2*) is always set to zero. If INTMISS=T, a missing indicator is added to the full model when the intermediate variable is missing in the full model (model 1). The intermediate variable is set to zero, rather than missing. The missing indicator is set to one if the intermediate variable was missing and zero otherwise. Both variables are zero in the partial model (model 2).

The basic PHREG call is written as follows:

```
    proc PHREG data = duplicated_dataset outest = estims;
    strata modelnum;
    model time * event(0) =
       exposure1 exposure2 intermed1 intermed2 covar1 covar2;
    output out = dfbs dfbeta = dt1 - dt2;
    id ids;
```

In this code, the variable `modelnum` is the model number and has a value of 1 or 2. By this method, both the models (with and without the intermediate variable) are run simultaneously. The PHREG call produces two output datasets. The first, `ESTIMS`, contains the estimates of the regression coefficients. The second output dataset, `DFBS`, contains the dfbetas for `exposure1` and `exposure2`. The dfbetas are used to calculate the covariance matrix between the two models. Briefly, this is achieved by converting the data into a matrix and then, using PROC IML, the transposition of this matrix is multiplied by the original matrix to produce the 2 X 2 covariance matrix.
This above processing is adapted from Example 54.8 (Analysis of Recurrent Events Data) in the SAS manual chapter for PROC PHREG [SAS Manual, SAS V9 Stat Users Guide Volume 5, pp 3304-3314]. From the coefficient estimates in `ESTIMS` the mediated proportion by the intermediate variable can be calculated by the following formula.

```
PTE = (1 - (estimate for exposure2 / estimate for exposure1)) * 100
```

To calculate the standard error of the above estimate for PTE we use equation 5 on page 1519 of Lin et al. [Statistics in Medicine, Vol. 16, 1515-1527 (1997)]. This calculation uses the values from `ESTIMS` as well as the covariance matrix described above. To improve asymptotic behavior, we transform using Fisher's z transformation and the delta method to get the 95% confidence limits of the transformed variable, then back-transform to report the 95% confidence interval on the original scale.

6.1 Additional calculations

The macro can create a missing indicator for the intermediate variable. In the model 1 record, if the intermediate variable is missing, it will instead be set to zero and the missing indicator variable will have a value of one. If the intermediate variable is not missing, it will keep its value and the missing indicator will have a value of zero. In the model 2 record, both the intermediate variable and its missing indicator are always set to zero.

The macro checks for two inconsistencies in the results. First, it makes sure that the model has converged. (If it does not, a warning message is printed instead of the usual output.) Second, it checks whether the calculated percent of treatment effect explained is between 0% and 100%. If it is not, then the model estimates are printed without the calculated effect.

6.2 Computer time

Each of the example runs took about 20 minutes of cpu time on machines with processing speeds of .75 to 1.2 GHz. More than 80% of this time was used in the PROC PHREG step in which the dfbetas were made.

7 Credits

Written by Mathew Pazaris, Ellen Hertzmark, Jim Fauntleroy, Sally Skinner, Denise Jacobson
(currently or formerly of Tufts University), and Donna Spiegelman for the Channing Laboratory. Questions

can be directed to Mathew Pazaris stmjp@channing.harvard.edu
or Ellen Hertzmark stleh@channing.harvard.edu,(617) 432-4597.

8 References

Lin, DY, Fleming, TR, DeGruttola, V: Estimating the proportion of treatment effect explained by a surrogate marker.
Statistics in Medicine 16(13): 1515-1527, 1997.
(1997).

Rothman, KJ, Greenland S. Measures of effect and association.
Modern Epidemiology. Philadelphia, pa: Lippincott Williams & Wilkins;
1998:47-64.

Jun HJ, Austin SB, Wylie SA, Corliss HL, Jackson B, Spiegelman D, Pazaris MJ, Wright RJ. The mediating effect of childhood abuse in sexual orientationdisparities in tobacco and alcohol use during adolescence: results from theNurses' Health Study II. Cancer Causes Control. 2010:[Epub ahead ofprint]