# The SAS %BLINPLUS Macro

Roger Logan and Donna Spiegelman

April 10, 2012

**Abstract**

The macro %blinplus corrects for measurement error in one or more model covariates logistic regression coefficients, their standard errors, and odds ratios and 95% confidence intervals for a biologically meaningful difference specified by the user (the "weights"). Regression model parameters from Cox models (PROC PHREG) and linear regression models (PROC REG) can also be corrected. A validation study is required to empirically characterize the measurement error model. Options are given for main study/external validation study designs, and main study/internal validation study designs (Spiegelman, Carrol, Kipnis; 2001). Technical details are given in Rosner et al. (1989), Rosner et al. (1990), and Spiegelman et all (1997).

**Keywords: SAS, macro, measurement error**

## Contents

## 1 Description

This method of correction for logistic regression coefficients and related quantities is strictly valid only under the following conditions:

1. Outcome is rare (overall observed disease probability $< 0.05$)

2. All model covariates measured with error are continuous (perfectly measured covariates may be either categorical or continuous)

3. Measurement error is not severe (e.g., reliability coefficient $> 0.5$)

The sensitivity of the method to departures from assumptions 1 and 3 were carefully studied in Rosner et al. (1989). The sensitivity of the method to departures from the normality assumption of the measurement error model were studied by Carroll et al. (JRSS-B, 1990), where it was found that the method is quite robust to departures from this assumption.

**No datasets can contain any missing data.**

# 2    Invocation

To call %blinplus, your program must know where to look for it. The most efficient way is to include the statement in your SAS program.

```
options mautosource sasautos='/usr/local/channing/sasautos';
```

You must load in a main study and validation study data set to be used in the regression and for determining the error model. A data set of weights, for use in calculating the odds ratios, must be defined. The weights for the covariates should be defined in the same order as those listed in the model statement used the the main study regression.

To obtain the uncorrected estimates of the model parameters you must run your regression using the main study, making an output data set using the covout option. This can be done using either PROC REG, PROC LOGISTIC, or PROC PHREG. For example using a logistic regression ,

```
proc logistic data=<main study data set name> covout outest=bvarbm;
  model <dependent variable> = <covariates> ;
```

This makes a (non-permanent) SAS data set (**bvarbm**) with the estimates and the variance-covariance matrix for the parameters.

In the case where there is an internal validation study, the user will also need to perform the same regression on the validation study data. In this case, the user would include the command

```
proc logistic data=<validation study data set name> covout outest=bvarbv;
  model <dependent variable> = <covariates> ;
```

This makes a (non-permanent) SAS data set (**bvarbv**) with the estimates and the variance-covariance matrix for the parameters.

You then type

```
%blinplus(
    type     = < type of regression, can be LOGISTIC, REG, or PHREG   >,
    valid    = < name of dataset with Validation Study data  >,
    main_est = < dataset containing estimates from main study regression >,
    weights  = < Dataset with weights  used in odds ratio calculations >,
    err_var  = < List of the variables measured with error. These ae the
                   variables used in right hand side of the model statement
                   in the regression. >,
    true_var = < List of the variables measured without error. These are listed in
                    the same order as the variables listed in err_var >,
    change   = < Percent Change Scale: T/F (need type=REG for T) >,
    internal = <  If validation study is Internal, do you wish to
                    combine estimates? T/F  (if T, need to provide
                    specify val_est argument. Default value is FALSE. >,
    val_est  = < Dataset containing Internal validation estimates. Same format
                    as in main_est option.>
        );
```

# 3   Examples

Examples for running the macro blinplus for the various types of regressions permitted. In the following examples we will use proc logistic. The variable x is measured with error and y is the true value. S is a perfectly measured covariate, while case is the outcome indicator.

The data sets are

- main = main study data set.

- ext_val = external validation data set. In a main study/external validation study design, the external validation study does not have data on the outcome variable for the primary regression models or there are so few cases, the regression parameters of interest cannot be estimated in the valdation study.

- int_val = internal validation data set. In a main study/internal validation study design, the internal validation study has data on the outcome, and the parameters of interest can be estimated in the validation study.

```
data main;
    input y x s case age;
    cards;
    1 2 2  0  10
    1 2.1 5 1 10
    2 3  4 0  11
    3 4 3  1  12
    2 2.9 3 0 13
    3 4.1 5 0 13
;
```

```
data  ext_val;
 input y x s;
cards;
  1  1.5 0
  1  0.8 1
  2  1.4 1
  2  2.3 0
;

data int_val;
 input case y x s;
cards;
1  0    1.25 0
0  1    0.5  2
0  1.5 1.4  1
0  2    2.3  0
0  3    2.5  1
1  2    1.6  1
;

data wts;
input name $ weight;
cards;
x 1.0
s 1.0
;

proc logistic descending covout outest=bvarb data=main noprint;
model case = x s;

%blinplus(
 type      = logistic,
 valid     = ext_val,
 main_est  = bvarb,
 weights   = wts,
 err_var   = x s,
 true_var  = y s,
 internal  = f,
 val_est   =
 );
```

The output for the previous SAS program for the main study/external validation study design in included on the next page.

MEASUREMENT ERROR CORRECTION OF REGRESSION ESTIMATES VIA THE METHOD OF REGRESSION CALIBRATION
 References: Rosner BA, Spiegelman D, Willett WC. Amer J Epi 1990. 132:734-745.
             Spiegelman D, Carroll R, Kipnis V.  Stat in Med 2001. 20:139-160.
 Programmers: Donna Spiegelman, Sc.D., Harvard School of Public Health, Boston MA,
              Roger Logan Ph.D., Harvard School of Public Health, Boston MA,
              Doug Grove, M.S.,  Aidan McDermott, M.A.,  Carrie Wager, B.S.,
                  of The Channing Laboratory, Boston MA.


                        TYPE OF REGRESSION: LOGISTIC

                              OPTIONS USED:

Validation Study Regression Coeffiecients:

         INTERCPT        x        s

x           -1.16      1.4     1.12

Variance matrix of error-variables:
              x

x          0.02

Main study regression coeffiecients: Uncorrected


   --------   ----------------------------------------------------------------

     WT        B          SE         OR          95% CI            p
   --------   ----------------------------------------------------------------

x   1.00  -0.04791    1.10300    0.95322    0.10972  -  8.28112   0.96535
s   1.00   0.42994    0.83453    1.53717    0.29947  -  7.89029   0.60642
   --------   ----------------------------------------------------------------

Main study regression coefficients: Corrected


   --------   ----------------------------------------------------------------

     WT        B          SE         OR          95% CI            p
   --------   ----------------------------------------------------------------

x   1.00  -0.03422    0.78788    0.96636    0.20629  -   4.52680  0.96535
s   1.00   0.46827    1.32017    1.59723    0.12012  -  21.23840  0.72281
   --------   ----------------------------------------------------------------

Example using logistic regression with an internal validation study.

```
proc logistic descending covout outest=bvarbm data=main noprint;
model case = x s;
proc print data=bvarb;

proc logistic descending covout outest=bvarbv data=int_val noprint;
model case = y s;

%blinplus(
 type      = logistic,
 valid     = int_val,
 main_est  = bvarbm,
 weights   = wts,
 err_var   = x s,
 true_var  = y s,
 internal  = t,
 val_est   = bvarbv
 );
```

The output is on the following page.

MEASUREMENT ERROR CORRECTION OF REGRESSION ESTIMATES VIA THE METHOD OF
REGRESSION CALIBRATION
 References: Rosner BA, Spiegelman D, Willett WC. Amer J Epi 1990. 132:734-745.
            Spiegelman D, Carroll R, Kipnis V.   Stat in Med 2001. 20:139-160.
Programmers: Donna Spiegelman, Sc.D., Harvard School of Public Health, Boston MA,
             Roger Logan Ph.D., Harvard School of Public Health, Boston MA,
             Doug Grove, M.S.,  Aidan McDermott, M.A.,  Carrie Wager, B.S.,
                  of The Channing Laboratory, Boston MA.

                      TYPE OF REGRESSION: LOGISTIC
                            OPTIONS USED:

Validation Study Regression Coeffiecients:

          INTERCPT        x        s

x         -1.8265 1.602633 1.030767


Variance matrix of error-variables:
              x

x         0.051023


Main study regression coeffiecients: Uncorrected


  --------   ----------------------------------------------------------------

             WT        B        SE        OR          95% CI         p
  --------   ----------------------------------------------------------------

x            1.00   -0.04791   1.10300   0.95322   0.10972  -  8.28112   0.96535
s            1.00    0.42994   0.83453   1.53717   0.29947  -  7.89029   0.60642
--------   ----------------------------------------------------------------

Main study regression coefficients: Corrected


--------   ----------------------------------------------------------------

             WT        B        SE        OR          95% CI         p
--------   ----------------------------------------------------------------

x            1.00   -0.02989   0.68825   0.97055   0.25186  -  3.73998   0.96535
s            1.00    0.46076   1.18954   1.58528   0.15401  - 16.31776   0.69850
--------   ----------------------------------------------------------------

Internal validation regression
--------   ----------------------------------------------------------------

             WT        B        SE        OR          95% CI         p
--------   ----------------------------------------------------------------

x            1.00   -1.12947   1.32039   0.32320   0.02430  -  4.29944   0.39232
s            1.00   -0.92784   1.41757   0.39541   0.02457  -  6.36357   0.51277
--------   ----------------------------------------------------------------

Pooled estimate: Internal


--------   ----------------------------------------------------------------

             WT        B        SE        OR          95% CI         p
--------   ----------------------------------------------------------------

x            1.00   -0.02473   0.55155   0.97557   0.33095  -  2.87573   0.96423
s            1.00    0.11285   0.87670   1.11947   0.20079  -  6.24126   0.89758
--------   ----------------------------------------------------------------

A real example

An example of using %blinplus on data from the Nurses Health Study using a logistic regression with an external validation study. In this example alcohol intake is measured with error (alc) and the true measurement is called alc_dr.

```
options ps=72 ls=100 nodate nonumber center;

libname NHSD "/Data";
libname tmp  "/tmp";

data main;
   set NHSD.bcanc97(keep= id brc80_88 case alc age80 bmeno2 amenar1
                         amenar3 par2afb2 par2afb3 par2afb4
                         par3afb2 par3afb3 par3afb4 BBD FHX
                         BMI2 BMI3 BMI4 BMI5 );
   rename age80=age;
run;

data valid;
  set NHSD.valid(keep=id alc_dr alc_ffq);
  rename alc_ffq=alc;
run;

data valid;
  merge valid(in=inv) tmp.main(drop=alc in=inm);
  by id;
  if inv and inm;
run;

data wts;
  length name $ 8;
  name = 'alcohol'; weight = 12; output;
  name = 'age'; weight = 1;    output;
  name = 'Postdbm'; weight = 1;  output;
  name = 'Amen<=12'; weight = 1;  output;
  name = 'Amen>=14'; weight = 1;  output;
  name = 'p2b<=23'; weight = 1;  output;
  name = 'p2b23-25'; weight = 1;  output;
  name = 'p2b>=26'; weight = 1;   output;
  name = 'p3b<=23'; weight = 1;   output;
  name = 'p3b23-25'; weight = 1;  output;
  name = 'p3b>=26'; weight = 1;   output;
  name = 'BBD'; weight = 1;    output;
  name = 'FHist'; weight = 1;   output;
  name = 'BMI21-23'; weight = 1;   output;
  name = 'BMI23-25'; weight = 1;   output;
  name = 'BMI25-29'; weight = 1;   output;
  name = 'BMI>29'; weight = 1;  output;
run;
```

```
proc logistic data= main noprint descending covout outest=bvarbm;
    model case =  alc age bmeno2 amenar1 amenar3 par2afb2
                  par2afb3 par2afb4 par3afb2 par3afb3 par3afb4
                  BBD FHX BMI2 BMI3 BMI4 BMI5;

  run;

%include 'blinplus.sas';

%blinplus(
      type     = logistic,
      valid    = valid,
      main_est = bvarbm,
      weights  = wts,
      err_var  = alc age bmeno2 amenar1 amenar3 par2afb2
                  par2afb3 par2afb4 par3afb2 par3afb3
                  par3afb4 BBD FHX BMI2 BMI3 BMI4 BMI5,
      true_var = alc_dr age bmeno2 amenar1 amenar3 par2afb2
                  par2afb3 par2afb4 par3afb2 par3afb3
                  par3afb4 BBD FHX BMI2 BMI3 BMI4 BMI5,
      internal = f,
      val_est  =
      );
```

MEASUREMENT ERROR CORRECTION OF REGRESSION ESTIMATES VIA THE METHOD OF REGRESSION
CALIBRATION
    References: Rosner BA, Spiegelman D, Willett WC. Amer J Epi 1990. 132:734-745.
               Spiegelman D, Carroll R, Kipnis V.  Stat in Med 2001. 20:139-160.
    Programmers: Donna Spiegelman, Sc.D., Harvard School of Public Health, Boston MA,
                 Roger Logan Ph.D., Harvard School of Public Health, Boston MA,
                 Doug Grove, M.S.,  Aidan McDermott, M.A.,  Carrie Wager, B.S.,
                        of The Channing Laboratory, Boston MA.


                        TYPE OF REGRESSION: LOGISTIC

                             OPTIONS USED:


Validation Study Regression Coeffiecients:


         INTERCPT  alcohol      age  Postdbm Amen<=12 Amen>=14  p2b<=23 p2b23-25
:         p2b>=26 p3b<=23 p3b23-25  p3b>=26      BBD    FHist BMI21-23 BMI23-25
:         BMI25-29    BMI>29


alcohol  0.318576 0.653848 0.087162 -0.37537 -0.18334 1.068832 0.554026 2.183137
:         1.257233 1.199313 1.892999  1.54797  0.19046 -1.31281 -2.79668 -1.91837
:         -4.95709  -4.5956


Variance matrix of error-variables:
          alcohol


alcohol  25.47038


Main study regression coeffiecients: Uncorrected


             --------   ----------------------------------------------------------------

                 WT        B        SE        OR          95% CI            p

             --------   ----------------------------------------------------------------


alcohol      12.00   0.00696   0.00206   1.08712   1.03565  -  1.14115   0.00074
age           1.00   0.07023   0.00588   1.07275   1.06047  -  1.08518   0.00000
Postdbm       1.00  -0.40050   0.07864   0.66999   0.57428  -  0.78164   0.00000
Amen<=12      1.00  -0.03995   0.05929   0.96084   0.85542  -  1.07925   0.50046
Amen>=14      1.00  -0.34039   0.07965   0.71149   0.60865  -  0.83170   0.00002
p2b<=23       1.00  -0.12824   0.12448   0.87964   0.68919  -  1.12271   0.30293
p2b23-25      1.00  -0.12738   0.10797   0.88040   0.71248  -  1.08788   0.23807
p2b>=26       1.00   0.31307   0.12130   1.36762   1.07823  -  1.73469   0.00985
p3b<=23       1.00  -0.18724   0.10281   0.82924   0.67790  -  1.01436   0.06856
p3b23-25      1.00  -0.25602   0.10209   0.77412   0.63374  -  0.94560   0.01215
p3b>=26       1.00   0.03885   0.13511   1.03961   0.79774  -  1.35482   0.77373
BBD           1.00   0.48656   0.08430   1.62671   1.37897  -  1.91896   0.00000
FHist         1.00   0.59813   0.08382   1.81872   1.54319  -  2.14344   0.00000
BMI21-23      1.00   0.19391   0.07834   1.21398   1.04118  -  1.41546   0.01332
BMI23-25      1.00  -0.05286   0.08564   0.94851   0.80195  -  1.12185   0.53704
BMI25-29      1.00   0.07857   0.08343   1.08174   0.91856  -  1.27392   0.34629
BMI>29        1.00  -0.17933   0.09936   0.83583   0.68792  -  1.01553   0.07110


                                      10

```
                --------   ----------------------------------------------------------------

Main study regression coefficients: Corrected


                --------   ----------------------------------------------------------------

                  WT        B         SE        OR            95% CI              p

                --------   ----------------------------------------------------------------


alcohol         12.00    0.01065   0.00320   1.13628    1.05388  -  1.22511    0.00088
age              1.00    0.06930   0.00597   1.07176    1.05930  -  1.08437    0.00000
Postdbm          1.00   -0.39650   0.07990   0.67267    0.57516  -  0.78672    0.00000
Amen<=12         1.00   -0.03800   0.06019   0.96271    0.85558  -  1.08327    0.52787
Amen>=14         1.00   -0.35177   0.08076   0.70344    0.60046  -  0.82408    0.00001
p2b<=23          1.00   -0.13414   0.12639   0.87447    0.68259  -  1.12029    0.28855
p2b23-25         1.00   -0.15063   0.10947   0.86017    0.69407  -  1.06602    0.16882
p2b>=26          1.00    0.29969   0.12322   1.34944    1.05990  -  1.71807    0.01501
p3b<=23          1.00   -0.20001   0.10430   0.81872    0.66736  -  1.00442    0.05514
p3b23-25         1.00   -0.27618   0.10357   0.75868    0.61930  -  0.92943    0.00766
p3b>=26          1.00    0.02237   0.13725   1.02262    0.78142  -  1.33826    0.87055
BBD              1.00    0.48453   0.08557   1.62342    1.37274  -  1.91987    0.00000
FHist            1.00    0.61211   0.08667   1.84432    1.55619  -  2.18579    0.00000
BMI21-23         1.00    0.22368   0.07995   1.25067    1.06927  -  1.46285    0.00515
BMI23-25         1.00   -0.03244   0.08706   0.96808    0.81621  -  1.14821    0.70944
BMI25-29         1.00    0.13135   0.08721   1.14037    0.96120  -  1.35293    0.13202
BMI>29           1.00   -0.13040   0.10344   0.87774    0.71667  -  1.07501    0.20741

                --------   ----------------------------------------------------------------
```

# 4   Warnings

- The variables listed in err_var and true_var need to be in the same order as those listed in the model statement used in the main study or validation study regression. The names for the variables do not need to be the same, just their order in the list. For example, in the above examples the main study used the variables x and s in the model statement, where x is measured with error and s is a perfectly measured covariate. The order in the err_var was listed as err_var = x s (same is in the model statement and in the definition of the weights data set). The corresponding statement for true_var was true_var = y s , where y is the gold standard measurement for x and s has the same meaning as in the err_var statement. We could have used true_var = y sv, where sv is the validation study variable corresponding to the main study variable s.

- In the case of an internal validation study the order of the variables in the model satement should be the same as in the model statement for the main study. These should also agree with the order in which the weights are defined. (See the second example presented above.)

- These methods are not appropriate for Nurses' Health Study and Health Professionals' Followup Study data using cumulatively updated averages.

- For simple updated variables, Cox models (PROC PHREG) can be run and it reasonable assumed that after adjusting for age, the measurement error model is constant through successive follow-up cycles.

# 5 References

Rosner B, Willett WC, Spiegelman D "Correction of logistic relative risk estimates and confidence intervals for systematic withinperson measurement error". *Statistics in Medicine* 8: 1051-1069, 1989.

Rosner B, Spiegelman D, Willett WC "Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error". *American Journal of Epidemiology* 1990;132: 734-735.

Spiegelman D, McDermott A, Rosner B "The many uses of the 'regression calibration' method for measurement error bias correction in nutritional epidemiology". *American Journal of Clinical Nutrition*, 1997; 65:1179S-1186S.

Spiegelman D, Carroll RJ, Kipnis V "Efficient regression calibration for logistic regression in main study/internal validation study designs with an imperfect regerence instrument". *Statistics in Medicine* 2001;20:139-160.

# 6 Credits

Written by Donna Spiegelman, Sc.D., Harvard School of Public Health, Boston MA, Roger Logan Ph.D., Harvard School of Public Health, Boston MA, Doug Grove, M.S., Aidan McDermott, M.A., Carrie Wager, B.S., Ruifeng Li M.S., for the Channing Laboratory.

Questions can be directed to Xiaomei Liao (stxia@channing.harvard.edu), Ruifeng Li (strui@channing.harvard.edu) and Donna Spiegelman(stdls@hsph.harvard.edu).