

The SAS RELRISK9 Macro

Sally Skinner, Ruifeng Li, Ellen Hertzmark, and Donna Spiegelman

November 15, 2012

Abstract

The %RELRISK9 macro obtains relative risk estimates using PROC GENMOD with the binomial distribution and the log link. This is particularly useful when the odds ratio is not a good approximation to the rate ratio (e.g., because of high prevalence of the outcome or large relative risks).

Keywords: SAS, macro, proc genmod, repeated measures, relative risk, initial values
NOTE: do we want to have initial values as a keyword?

Contents

1	Description	2
2	Invocation and Details	2
3	Examples	5
3.1	Example 1. Basic macro call using default parameters: One observation per subject	5
3.2	Example 2. <i>INITMETH</i> =none, demonstrating <i>DICHEXP</i> , <i>ESTIMATE1</i>	6
3.3	Example 3. <i>INITMETH</i> =force	9
3.4	Example 4. Multiple observations per subject, also showing <i>DICHEXP</i> and printing the results of the Poisson model (<i>POISPRINT</i> =T)	10
3.5	Example 5. Errors in calls to %RELRISK9	14
4	Limitations	16
5	Known Bugs	16
6	Credits	16
7	Frequently Asked Questions	16
7.1	Q: My output does not say that I am running a GEE model, but I have repeated measures (multiple observations per subject).	16

7.2	Q: My .saslog says 'Out of Memory.'	16
7.3	Q: I have tried a lot of models and they do not converge properly. What should I do?	17
7.4	Q: How can I get .rtf output?	17

8 References **17**

1 Description

%RELRIK9 is a SAS macro that uses PROC GENMOD with the binomial distribution and the log link to compute relative risk estimates. Because the macro uses PROC GENMOD, it can handle repeated measures.

If the user specifies *EMPCAL=T*, the confidence intervals based on the empirical/robust estimates of the standard errors are given. Otherwise, model-based estimates are shown.

There are two main problems with this type of model: (1) it does not always converge; (2) the estimated probabilities derived from the model may not all be between 0 and 1. To help with convergence, the macro starts with initial values that are expected to be close to the "correct" values. This is done by running another procedure first, then using the results as initial values for the estimation procedure of proc genmod. Since we have found empirically that PROC GENMOD with the Poisson distribution and log link, using the empirical variance produces estimates fairly close to those that would be produced by PROC GENMOD with the binomial distribution and the log link, the default for the macro is to use this to obtain the initial values.

After the model is run, the macro checks whether all the predicted probabilities are below 1.

Error messages are printed if parameters or data do not conform to requirements. All error messages begin with "ERR'OR in macro". If any error occurs, the program terminates and no output is produced. **If the ERROR message does not contain the word 'macro,' it is a SAS error, not a macro error.**

2 Invocation and Details

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=132 replace formdlim='='
mautosource
sasautos=('/usr/local/channing/sasautos',
         <the location of the read macros for your data>);
```

This will allow you to use all the SAS read macros for the data sets as well as other public SAS macros, such as %LGTPHCURV9, %MIXCURV8, %PM, %INDIC3, %EXCLUDE, %MPHREG, %CALADJ, and %PCTL in /usr/local/channing/sasautos.

In the rest of this section, we will list all the input parameters, some of which are required and some of which are optional. If a parameter has a default value, it is given after the '='.

REQUIRED PARAMETERS

=====

DATA = the name of the input dataset
DEPEND = the name of the dependent variable, for the left side of
model equation in proc genmod
This should be a binary variable coded
0 for 'no' and 1 for 'yes.'
INDEPEND = the list of independent variable(s), for the right
side of the model equation in proc genmod
INITMETH = POISSON the method for finding the initial values
for each parameter

There are six choices:

1. POISSON: uses PROC GENMOD with the POISSON distribution.
2. NLIN: uses PROC NLIN to determine the initial values.
3. PREV: uses the previous estimates to initialize genmod. It can only be used if the previous run of RELRISK has converged and one or two variables have been replaced or added.
4. FORCE: allows the user to force either the intercept, the initial estimates, or both, using FINTERCEP and FINITVAL. Here is an example of the call:
INITMETH=force, FINTERCEPT=-5, FINITVAL=-1 -1 2
This sets the initial value of the intercept to -5 and the initial estimates for the first three variables in INDEPEND to -1, -1 and 2, respectively.
One might want to use this if a previous attempt did not converge, or if it converged, but resulted in predicted probabilities greater than 1.
5. LOGISTIC: uses PROC GENMOD with the logit link.
6. NONE: goes directly to the estimation of the log binomial model.
This is be useful (and a CPU-time-saver) if you know that a model will converge without problem to a set of legitimate values.
Default=Poisson

OPTIONAL PARAMETERS RELATING TO THE DATA

=====

WHERE = A subsetting clause to restrict the data used
EXTRAVAR = Additional variables to be kept in the working data set,
primarily for use in the WHERE parameter
NOTE: If you subset using a variable not in the model and do not list the variable in EXTRAVAR, the macro will not keep the variable in the working dataset and the working dataset is likely to have 0 observations.

OPTIONAL PARAMETERS RELATING TO THE MODEL

=====

CLASS = A list of class/categorical variables used in the model. If EMPCAL is T, the SUBJECT will automatically be included in CLASS.
NOTE: Use of CLASS variables can save typing time (and avoid errors), but, because of the possibly strange parameterization used for CLASS variables, it is advisable to use them only for covariates for which the parameter estimates are not of interest.
For covariates whose parameter estimates are of interest, make indicators.

EMPCAL = F Whether you want the empirical variance (GEE model)

SUBJECT = The subject identifier used for SUBJECT= in the REPEATED statement (REQUIRED if EMPCAL=T)

NOINT = F If T, no-intercept model is run. This cannot be run with INITMETH=NLIN

REPTYPE = IND The covariance structure of the data, such as IND, EXCH, UN, MDEP)
If you have repeated measures, you probably want a covariance structure other than IND.
Use of any covariance structure other than IND or EXCH requires the specification of a WITHINVAR (see next item) to fit observations properly into the data structure.

WITHINVAR = the index variable for within-group records
REQUIRED if EMPCAL=T and REPTYPE is not IND or EXCH.

WEIGHT = The name of the weight variable
OPTIONAL

OPTIONAL PARAMETERS RELATING TO THE OUTPUT

=====

DICHEXP = The name of the dichotomous exposure, if any of interest.
If you give this a value, the macro will produce a table giving the numbers of subjects, observations, and cases, as well as the percent of cases for each value of the dichotomous exposure, and for the study group overall.

COV = F If T, the covariance matrix is printed out and sent to ods with the name covb

CORR = F If T, the correlation matrix is printed out and sent to ods with the name corrb
Default=F

CONTRAST1 = First contrast statement
CONTRAST2 = Second contrast statement
CONTRAST3 = Third contrast statement
CONTRAST4 = Fourth contrast statement
ESTIMATE1 = First estimate statement
ESTIMATE2 = Second estimate statement
ESTIMATE3 = Third estimate statement
ESTIMATE4 = Fourth estimate statement

GMPRINT = F If T, the macro will print whatever results the PROC GENMOD gets. This is useful if you are having convergence problems, so you can have a reasonable place to start, using

```

INITMETH=FORCE.
OUTDAT = The name of a sas data set to keep the results, if desired
POISPRINT = F Whether to print the results of the Poisson model
              (if using INITMETH=POISSON)
TITLENUM = 4 The first "free" title number, after the titles
              you already have and wish to keep during this procedure ,
              More titles have been added to make it easier to navigate
              through the output. In order to avoid having these titles
              knock out your titles, use TITLENUM to state what line
              the MACRO titles should begin in. The default is 4.
              For example, if you are using title1 and title2, TITLENUM
              should be set to 3.

```

3 Examples

For the examples, we are using data from NHS II. The main dataset is N2DAT. In addition, we have dataset AGE18, which has one observation for each nurse and is primarily interested in the nurse's BMI at age 18. The outcomes of interest will be

```

OBESE      whether the nurse has bmi ge 30
OVERWT     whether the nurse has bmi ge 25 (includes obese)
OVWT18     whether the nurse reported being overweight at age 18

```

The predictors are

```

AGE group      as 5-year age group indicators; reference level
                is 25-29
BIRTH YEAR group as 5-year cohorts; reference level is 1945-1949
                (macro variable name is &byrgp_)
REGION         as indicators for MIDWEST SOUTH WEST; reference
                level is NORTHEAST

```

3.1 Example 1. Basic macro call using default parameters: One observation per subject

Here we model overweight at age 18 based on region and birth year group.

```

title2 'bare bones call (using default parameters) for one obs per subject';
%relrisk9(data=age18, depend=ovwt18, independ=midwest south west &byrgp_);

```

The macro output is

```

-----
/udd/stleh/doctn/relrisk  Program exrelrisk9  12JUN2012  12:57  stleh  1

```

bare bones call (using default parameters) for one obs per subject

NOTE: 5259 observations deleted because of missing values for dependent, independent, subject, or within variables.

/udd/stleh/doctn/relrisk Program exrelrisk9 12JUN2012 12:57 stleh 2
bare bones call (using default parameters) for one obs per subject

Results of running RELRISK9 on data age18
Outcome is ovwt18 . 12100 events in 111412 observations (10.9 %).

Parameter	ESTIMATE	Standard Error	P value	UNITS	Relative risk
Intercept	-2.1169	0.0231	<.0001	1	0.12040
MIDWEST	-0.0369	0.0203	0.0698	1	0.96381
SOUTH	-0.1413	0.0249	<.0001	1	0.86823
WEST	-0.3544	0.0293	<.0001	1	0.70158
BYRGP2	-0.0044	0.0252	0.8616	1	0.99562
BYRGP3	-0.0376	0.0254	0.1399	1	0.96314
BYRGP4	-0.0283	0.0290	0.3295	1	0.97213
Scale	1.0000	0.0000	-	1	2.71828

95% CI Low	95% CI high	RR (95% CI)
0.11507	0.12598	0.12 (0.12, 0.13)
0.92616	1.00299	0.96 (0.93, 1.00)
0.82680	0.91173	0.87 (0.83, 0.91)
0.66246	0.74301	0.70 (0.66, 0.74)
0.94771	1.04595	1.00 (0.95, 1.05)
0.91630	1.01238	0.96 (0.92, 1.01)
0.91844	1.02896	0.97 (0.92, 1.03)
2.71828	2.71828	2.72 (2.72, 2.72)

The macro first printed out a notice, much like those printed by SAS procedures, saying how many observations were deleted because of missing values. The regression output is preceded by information on the name of the dataset, the outcome variable, the number and percent of events, and the number of observations.

The space between the lines beginning 'bare bones' and 'Results' on page 2 of the output results from the fact that *TITLNUM*=4, so there is no *title3*.

3.2 Example 2. *INITMETH*=none, demonstrating *DICHEXP*, *ESTIMATE1*

The macro call is

```

title2 'overweight at age 18 by region and birth cohort';
title3 'initmeth=none';
%relrisk9(data=age18, depend=ovwt18, independ=midwest south west &byrgp_,
dichexp=west,
estimate1=%str('south/west' south 1 west -1 / exp),
initmeth=none);

```

Note that we did not give values for the parameters that have default values, except for the one we wanted to change (*INITMETH*).

Although *west* is just one of the set of region indicators, formally it is a binary variable, and we can use it to demonstrate *DICHEXP*.

In the *ESTIMATE1* parameter, we used *%str* to avoid problems that might arise because of the quotation marks or the slash. It is especially important to do this when there is a comma in the 'name' of the estimate. We used *exp* to get the relative risk of *south* compared to *west*.

```

-----
/udd/stleh/doctn/relrisk Program exrelrisk9 12JUN2012 12:52 stleh 1
overweight at age 18 by region and birth cohort
initmeth=none

```

NOTE: 5259 observations deleted because of missing values for dependent, independent, subject, or within variables.

```

-----
/udd/stleh/doctn/relrisk Program exrelrisk9 12JUN2012 12:52 stleh 2
overweight at age 18 by region and birth cohort
initmeth=none
Results of running RELRISK9 on data age18
Outcome is ovwt18 . 12100 events in 111412 observations (10.9 %).

```

Counting cases of ovwt18 by west

Obs	WEST	NS	NCASES	PCTCASES
1	0	94615	10704	11.3
2	1	16797	1396	8.3
3	.	111412	12100	10.9

```

-----
/udd/stleh/doctn/relrisk Program exrelrisk9 12JUN2012 12:52 stleh 3
overweight at age 18 by region and birth cohort
initmeth=none
Results of running RELRISK9 on data age18
Outcome is ovwt18 . 12100 events in 111412 observations (10.9 %).

```

Standard P Relative

Parameter	ESTIMATE	Error	value	UNITS	risk
Intercept	-2.1169	0.0231	<.0001	1	0.12040
MIDWEST	-0.0369	0.0203	0.0698	1	0.96381
SOUTH	-0.1413	0.0249	<.0001	1	0.86823
WEST	-0.3544	0.0293	<.0001	1	0.70158
BYRGP2	-0.0044	0.0252	0.8616	1	0.99562
BYRGP3	-0.0376	0.0254	0.1399	1	0.96314
BYRGP4	-0.0283	0.0290	0.3295	1	0.97213
Scale	1.0000	0.0000	-	1	2.71828

95% CI Low	95% CI high	RR (95% CI)
0.11507	0.12598	0.12 (0.12, 0.13)
0.92616	1.00299	0.96 (0.93, 1.00)
0.82680	0.91173	0.87 (0.83, 0.91)
0.66246	0.74301	0.70 (0.66, 0.74)
0.94771	1.04595	1.00 (0.95, 1.05)
0.91630	1.01238	0.96 (0.92, 1.01)
0.91844	1.02896	0.97 (0.92, 1.03)
2.71828	2.71828	2.72 (2.72, 2.72)

```

/udd/stleh/doctn/relrisk Program exrelrisk9 12JUN2012 12:52 stleh 4
overweight at age 18 by region and birth cohort
initmeth=none
Results of running RELRISK9 on data age18
Outcome is ovwt18 . 12100 events in 111412 observations (10.9 %).

```

Obs	LABEL	MEANESTIMATE	MEANLOWERCL	MEANUPPERCL	LBETAESTIMATE
1	south/west	1.2375	1.1602	1.3200	0.2131
2	Exp(south/west)	-	-	-	1.2375

Obs	STDERR	ALPHA	LBETALOWERCL	LBETAUPPERCL	CHISQ	PROBCHISQ
1	0.0329	0.05	0.1486	0.2776	41.93	<.0001
2	0.0407	0.05	1.1602	1.3200	-	-

The results of *DICHEXP* came before those of the model.

The results for *ESTIMATE1* print last. The first line gives all the results you need. the 'MEAN' variables are the relative risk and its 95% confidence limits. the 'LBETA' variables are the estimate of the log(RR) and its 95% confidence bounds. PROBCHISQ is the p-value.

3.3 Example 3. *INITMETH=force*

Here are three (3) examples using *INITMETH=force*. The first gives initial values for all the effects. The second gives initial values for the intercept and the first effect. The third gives an initial value for the intercept only (and note that it is much lower than those in the first two macro calls). Note also that, since *INITMETH=force* just gives proc genmod a place to start, it is not necessary to be precise with the values used.

```
title3 'initmeth=force, all parameters initialized';
%relrisk9(data=age18, depend=ovwt18, independ=midwest south west &byrgp_,
initmeth=force, fintercept=-5, finitval=.1 0 -.2 0 -.1 0);
```

```
title3 'initmeth=force, intercept and first parameter initialized';
%relrisk9(data=age18, depend=ovwt18, independ=midwest south west &byrgp_,
initmeth=force, fintercept=-5, finitval=.1 );
```

```
title3 'initmeth=force, only intercept initialized ';
%relrisk9(data=age18, depend=ovwt18, independ=midwest south west &byrgp_,
initmeth=force, fintercept=-10);
```

Since all three macro calls gave the same results, we show only one.

```
-----
/udd/stleh/doctn/relrisk  Program exrelrisk9   110CT2011   11:14   stleh   1
```

```
initmeth=force, all parameters initialized
```

```
NOTE: 5259 observations deleted because of missing values for
       dependent, independent, subject, or within variables.
```

```
-----
/udd/stleh/doctn/relrisk  Program exrelrisk9   110CT2011   11:14   stleh   2
```

```
initmeth=force, all parameters initialized
```

```
Results of running RELRISK9 on data age18
```

```
Outcome is ovwt18 . 12100 events in 111412 observations (10.9 %).
```

Parameter	ESTIMATE	Standard Error	P value	UNITS	Relative risk
Intercept	-2.1169	0.0231	<.0001	1	0.12040
MIDWEST	-0.0369	0.0203	0.0698	1	0.96381
SOUTH	-0.1413	0.0249	<.0001	1	0.86823
WEST	-0.3544	0.0293	<.0001	1	0.70158
BYRGP2	-0.0044	0.0252	0.8616	1	0.99562
BYRGP3	-0.0376	0.0254	0.1399	1	0.96314
BYRGP4	-0.0283	0.0290	0.3295	1	0.97213
Scale	1.0000	0.0000	-	1	2.71828

95% CI Low	95% CI high	RR (95% CI)
0.11507	0.12598	0.12 (0.12, 0.13)
0.92616	1.00299	0.96 (0.93, 1.00)
0.82680	0.91173	0.87 (0.83, 0.91)
0.66246	0.74301	0.70 (0.66, 0.74)
0.94771	1.04595	1.00 (0.95, 1.05)
0.91630	1.01238	0.96 (0.92, 1.01)
0.91844	1.02896	0.97 (0.92, 1.03)
2.71828	2.71828	2.72 (2.72, 2.72)

The results of *DICHEXP* came before those of the model.

3.4 Example 4. Multiple observations per subject, also showing *DICHEXP* and printing the results of the Poisson model (*POISPRINT=T*)

In this example, we examine obesity over time in our longitudinal dataset, N2DAT. Since we have repeated measures, we need to use the *EMPCAL*, *SUBJECT*, and *REPTYPE* parameters. In addition, we get two contrasts for the global effects of the region and birth cohort variables, use a *WHERE* parameter, get the fit statistics, and show the effect of *TITLENUM*. Furthermore, we show the similarity/difference between the Poisson and the log-binomial estimates.

```

title2 'longitudinal obesity by region and birth cohort';
title3 'among those not overweight at age 18';
title4 'using where parameter, empcal=t';
title5 'showing use of contrast, fitprint, and titlenum parameters';
%relrisk9(data=n2dat, depend=obese, independ=midwest south west &byrgp_,
empcal=t, reptype=exch, subject=id,
fitprint=t, titlenum=7,
poisprint=t,
dichexp=west,
contrast1=%str('effect of region' midwest 1, south 1, west 1),
contrast3=%str('effect of birth cohort' byrgp2 1, byrgp2 1, byrgp4 1),
where=ovwt18 eq 0);

```

Here are the results.

```

/udd/stleh/doctn/relrisk Program exrelrisk9 12JUN2012 13:07 stleh 1
longitudinal obesity by region and birth cohort
among those not overweight at age 18
using where parameter, empcal=t
showing use of contrast, fitprint, and titlenum parameters

```

NOTE: 1615 observations deleted because of missing values for dependent, independent, subject, or within variables.

 /udd/stleh/doctn/relrisk Program exrelrisk9 12JUN2012 13:07 stleh 2
 longitudinal obesity by region and birth cohort
 among those not overweight at age 18
 using where parameter, empcal=t
 showing use of contrast, fitprint, and titlenum parameters

Results of running RELRISK9 on data n2dat
 Results of Poisson regression with empirical variance
 Outcome is obese . 53419 events in 451954 observations (11.8 %).

Parameter	ESTIMATE	Empirical Standard Error Estimates	P value	UNITS	relative risk	95% CI Low	95% CI high
Intercept	-1.8768	0.0186	<.0001	1	0.15308	0.14759	0.15878
MIDWEST	0.1942	0.0181	<.0001	1	1.21438	1.17212	1.25816
SOUTH	0.1586	0.0211	<.0001	1	1.17192	1.12440	1.22144
WEST	-0.1846	0.0250	<.0001	1	0.83142	0.79173	0.87309
BYRGP2	-0.1949	0.0193	<.0001	1	0.82292	0.79234	0.85469
BYRGP3	-0.5019	0.0207	<.0001	1	0.60537	0.58131	0.63042
BYRGP4	-0.8100	0.0264	<.0001	1	0.44484	0.42239	0.46849

 /udd/stleh/doctn/relrisk Program exrelrisk9 12JUN2012 13:07 stleh 3
 longitudinal obesity by region and birth cohort
 among those not overweight at age 18
 using where parameter, empcal=t
 showing use of contrast, fitprint, and titlenum parameters

Results of running RELRISK9 on data n2dat
 Outcome is obese . 53419 events in 451954 observations (11.8 %).
 where ovwt18 eq 0
 Counting cases of obese by west

Obs	WEST	NS	NOBS	NCASES	Cases as a percent of obs
1	0	83893	382745	46841	12.2
2	1	15395	69209	6578	9.5
3	.	99288	451954	53419	11.8

/udd/stleh/doctn/relrisk Program exrelrisk9 12JUN2012 13:07 stleh 4
 longitudinal obesity by region and birth cohort
 among those not overweight at age 18
 using where parameter, empcal=t
 showing use of contrast, fitprint, and titlenum parameters

Results of running RELRISK9 on data n2dat
 Outcome is obese . 53419 events in 451954 observations (11.8 %).
 where ovwt18 eq 0

Obs	CRITERION	VALUE
1	QIC	323667.0955
2	QICu	323634.8028

/udd/stleh/doctn/relrisk Program exrelrisk9 12JUN2012 13:07 stleh 5
 longitudinal obesity by region and birth cohort
 among those not overweight at age 18
 using where parameter, empcal=t
 showing use of contrast, fitprint, and titlenum parameters

Results of running RELRISK9 on data n2dat
 Outcome is obese . 53419 events in 451954 observations (11.8 %).
 where ovwt18 eq 0

Parameter	ESTIMATE	Empirical Standard Error Estimates	P value	UNITS
Intercept	-1.8765	0.0187	<.0001	1
MIDWEST	0.1940	0.0181	<.0001	1
SOUTH	0.1575	0.0211	<.0001	1
WEST	-0.1847	0.0250	<.0001	1
BYRGP2	-0.1951	0.0193	<.0001	1
BYRGP3	-0.5018	0.0207	<.0001	1
BYRGP4	-0.8098	0.0264	<.0001	1

Relative risk	95% CI Low	95% CI high	RR (95% CI)
0.15313	0.14763	0.15883	0.15 (0.15, 0.16)
1.21404	1.17175	1.25786	1.21 (1.17, 1.26)

1.17055	1.12303	1.22008	1.17 (1.12, 1.22)
0.83137	0.79166	0.87308	0.83 (0.79, 0.87)
0.82277	0.79220	0.85453	0.82 (0.79, 0.85)
0.60544	0.58139	0.63049	0.61 (0.58, 0.63)
0.44494	0.42248	0.46860	0.44 (0.42, 0.47)

```

/udd/stleh/doctn/relrisk Program exrelrisk9 12JUN2012 13:07 stleh 6
longitudinal obesity by region and birth cohort
among those not overweight at age 18
using where parameter, empcal=t
showing use of contrast, fitprint, and titlenum parameters

```

```

Results of running RELRISK9 on data n2dat
Outcome is obese . 53419 events in 451954 observations (11.8 %).
where ovwt18 eq 0

```

Obs	CONTRAST	DF	CHISQ	PROBCHISQ	TYPE
1	effect of region	3	326.46	<.0001	Score
2	effect of birth cohort	1	95.80	<.0001	Score

The space between the lines beginning 'showing' and 'Results' comes from the fact that *TITLENUM=7*.

The first part of the output (page 2) tells you that it is based on the Poisson estimates (because we set *POISPRINT=T*). The second part (page 4) is based on the log-binomial estimates. Note that when there are multiple observations per subject, *DICHEXP* produces a table based on observations, not subjects. The results of *DICHEXP* printed before the log-binomial results.

Below we do a trend test on birth cohort using the ordered categorical variable. We set *TITLENUM=6* to avoid writing over the titles we gave before the macro call.

```

title2 'longitudinal obesity by region and birth cohort';
title3 'among those not overweight at age 18';
title4 'using where parameter, empcal=t';
title4 'trend test for birth cohort';
%relrisk9(data=n2dat, depend=obese, independ=midwest south west bygp,
empcal=t, reptype=exch, subject=id,
titlenum=6,
where=ovwt18 eq 0);

```

Here are the results.

```

/udd/stleh/doctn/relrisk Program exrelrisk9 12JUN2012 13:10 stleh 1
longitudinal obesity by region and birth cohort
among those not overweight at age 18

```

trend test for birth cohort

NOTE: 1615 observations deleted because of missing values for
dependent, independent, subject, or within variables.

/udd/stleh/doctn/relrisk Program exrelrisk9 12JUN2012 13:10 stleh 2
longitudinal obesity by region and birth cohort
among those not overweight at age 18
trend test for birth cohort

Results of running RELRISK9 on data n2dat
Outcome is obese . 53419 events in 451954 observations (11.8 %).
where ovwt18 eq 0

Parameter	ESTIMATE	Empirical Standard Error Estimates	P value	UNITS
Intercept	-1.5734	0.0216	<.0001	1
MIDWEST	0.1938	0.0181	<.0001	1
SOUTH	0.1575	0.0211	<.0001	1
WEST	-0.1838	0.0250	<.0001	1
BYGP	-0.2678	0.0075	<.0001	1

Relative risk	95% CI Low	95% CI high	RR (95% CI)
0.20733	0.19873	0.21630	0.21 (0.20, 0.22)
1.21382	1.17156	1.25761	1.21 (1.17, 1.26)
1.17055	1.12303	1.22009	1.17 (1.12, 1.22)
0.83211	0.79236	0.87386	0.83 (0.79, 0.87)
0.76503	0.75380	0.77644	0.77 (0.75, 0.78)

3.5 Example 5. Errors in calls to %RELRIK9

In this macro call, a required parameter has been omitted;

```
%relrisk9(data=n2dat, independ=midwest south west);
```

The diagnostics in the .log and the .lst are

/udd/stleh/doctn/relrisk Program exrelrisk9 12JUN2012 14:13 stleh 1

ERR''OR in macro call: The name of the dependent variable was omitted.
The RELRISK9 macro stopped.

In the macro call below, a variable is misspelled.

```
title2 'a variable name misspelled';  
%relrisk9(data=age18, depend=ovwt18, independ=midwest south wes);
```

The diagnostics in the .log and the .lst are

/udd/stleh/doctn/relrisk Program exrelrisk9 12JUN2012 14:16 stleh 1
a variable name misspelled

ERR''OR in macro run:

```
    An error occurred during the creation of the working dataset,  
    while keeping DEPEND and INDEPEND from age18  
    The names of your variables may be incorrect.  
        dependent vbl is ovwt18 .  
        predictors are  midwest south wes .
```

The RELRISK9 macro stopped.

Finally, we show an example of setting *EMPCAL=T* without giving a value for *SUBJECT*.

```
title2 'empcal=t without specifying a subject variable';  
%relrisk9(data=age18, depend=ovwt18, independ=midwest south west,  
empcal=t);
```

The diagnostics in the .log and the .lst are

/udd/stleh/doctn/relrisk Program exrelrisk9 12JUN2012 14:17 stleh 1
empcal=t without specifying a subject variable

```
ERR''OR in macro call: You asked for a GEE model (empcal=T),  
    but did not specify a variable to serve as the subject (SUBJECT).  
    The macro will run a non-GEE model (with REPTYPE=IND)
```

```
/udd/stleh/doctn/relrisk Program exrelrisk9 12JUN2012 14:17 stleh 2
empcal=t without specifying a subject variable
```

NOTE: 5259 observations deleted because of missing values for dependent, independent, subject, or within variables.

The diagnostics in the .log and the .lst are

4 Limitations

These models do not always converge. They are especially likely to have trouble for very large data sets, very high overall prevalences of the outcome, large numbers of covariates, or continuous covariates.

5 Known Bugs

At the moment, *INITMETH=prev* and *INITMETH=nlin* do not work. *INITMETH=poisson* and *INITMETH=force* should be adequate substitutes.

6 Credits

This program was written by Sally Skinner from Tufts University and modified by Ruifeng Li and Ellen Hertzmark for the Channing Laboratory. The documentation was written by Ruifeng Li and Ellen Hertzmark. Questions can be directed to Ruifeng Li, strui@channing.harvard.edu, 617-432-6321.

7 Frequently Asked Questions

7.1 Q: My output does not say that I am running a GEE model, but I have repeated measures (multiple observations per subject).

A: You probably did not specify *EMPCAL=T*.

7.2 Q: My .saslog says 'Out of Memory.'

A: Run your job with extra memory by submitting with the following shell command:

```
qbs -q 1 -o memsize=2048M program name
```

7.3 Q: I have tried a lot of models and they do not converge properly. What should I do?

A: Indicators that represent small numbers of observations can be a problem. If you have these, either combine categories (if that makes sense) or delete the observations in the small categories (this may make sense if these are 'missing' categories). You may also need to pare down your models to the essential variables. If you can get those running, you can try adding one of the other variables at a time, and using the *INITMETH=FORCE* option.

If you have made a serious effort and you still don't have acceptable results, you may use the Poisson model with the empirical variance. *INITMETH=POISSON*, *POISPRINT=T*, *EMPCAL=T*. The macro will still try to fit a log-binomial model, but before it does, it will print the results of the Poisson model.

7.4 Q: How can I get .rtf output?

A: Make an output dataset *OUTDAT=name of output dataset you want*. Then
ods rtf file='name of .rtf to make';
proc print data=*name of output dataset*;
ods rtf close;

Note: you might prefer to make a permanent sas dataset in *outdat*, so you can determine which variables you want to print and in which order.

8 References

References for using the log-binomial model are

```
AUTHOR   = Wacholder, Sholom  
TITLE    = Binomial regression in GLIM: Estimating risk ratios and risk  
          differences  
JOURNAL  = American Journal of Epidemiology  
VOLUME  = 123  
PAGES   = 174-184  
YEAR    = 1986
```

Skov T, Deddens J, Petersen MR, Endahl L. Prevalence proportion ratios: estimation and hypothesis testing. *International Journal of Epidemiology* 1998;27;91-95.

In the methods section for a cross-sectional study, you write

Prevalence ratios and 95% confidence intervals were estimated by binomial regression with the log link function.

If you are using log-Poisson models with the empirical covariance structure because some of the log-binomial models did not converge, then use the following reference, in addition to the other two.

Zou G. A modified Poisson regression approach to prospective studies with binary data. Am J Epidemiol 2004; 159: 702-706.

The methods section should read

Log-binomials models were used to estimate relative risks and confidence intervals (ref). In a few instances, the models didn't converge and log-Poisson models, which provide consistent but not fully efficient estimates of the relative risk and its confidence intervals, were used (ref Zou paper).