American Journal of Epidemiology
Vol. 168, No. 11
DOI: 10.1093/aje/kwn278
Advance Access publication October 25, 2008

Practice of Epidemiology

A Simple Approach for Fitting Linear Relative Rate Models in SAS

David B. Richardson

Initially submitted March 27, 2008; accepted for publication August 4, 2008.

The linear relative rate model has been employed in epidemiologic analyses of a variety of environmental and occupational exposures. In contrast to an exponential rate model, the linear relative rate model implies that the excess relative rate of disease changes in an additive fashion with exposure. The linear relative rate model may be fitted using EPICURE (HiroSoft International Corporation, Seattle, Washington), a specialized statistical software package widely used for such analyses. In this paper, the author presents a simple approach to fitting the linear relative rate model to epidemiologic data using PROC NLMIXED in the SAS statistical software package (SAS Institute Inc., Cary, North Carolina). This approach is illustrated via analyses of data from a study of mortality in a cohort of South Carolina asbestos textile workers (1940–2001).

cohort analysis; dose-response function; epidemiologic methods; linear trend; models, statistical; Poisson regression; software

Abbreviation: ICD, *International Classification of Diseases*.

Epidemiologic data derived from occupational and environmental studies are often analyzed using exponential rate models of the form Rate $= e^{(\beta_0+\beta_1 d)}$, where d is an exposure variable of interest (1). The antilogarithms of β_0 and $(\beta_0 + \beta_1)$ provide estimates of the rate of disease in the absence and presence of exposure, respectively.

Estimation of a dose-response trend under an exponential rate model implies that for every 1-unit increase in the exposure metric, the rate of disease increases (or decreases) in a multiplicative fashion by a factor of $e^{(\beta_1)}$. An exponential change in disease rates with exposure to an occupational or environmental agent may not conform to observed data, in which case model misspecification may lead to standard errors for coefficient estimates that are biased, loss of power for a model-based test of a dose-response association, and the possibility that estimates of effect for extreme or uncommon exposure levels are substantially distorted (2).

A linear relative rate model offers an important alternative to the exponential rate model (3). The linear relative rate model has the form Rate $= e^{(\beta_0)}(1 + \beta_1 d)$, where β_1 represents the excess relative rate per unit of exposure. Estimation of a dose-response trend under a linear relative rate model implies that for every 1-unit increase in the exposure metric, the rate of disease increases (or decreases) in an additive fashion. The linear relative rate model has been used in analyses of many different factors: exposure to radon in the home (4) and in underground uranium mines (5, 6), external exposure to ionizing radiation among atomic bomb survivors (7, 8), patients treated by radiotherapy (9), and workers at nuclear weapons facilities and nuclear power plants (10, 11), and in relation to nonradiologic carcinogens, including chrysotile asbestos (12) and benzene (13).

One obstacle to fitting the linear relative rate model has been implementation using standard statistical packages. Data analysts have tended to use specialized software that was written specifically for fitting models of this form to epidemiologic data (14). In this paper, I illustrate how the linear relative rate model may be readily fitted using the SAS statistical software package (SAS Institute Inc., Cary, North Carolina).

MATERIALS AND METHODS

PROC NLMIXED in SAS produces likelihood estimates that are based on adaptive Gaussian quadrature. A variety of

Correspondence to Dr. David Richardson, Department of Epidemiology, CB 7435, School of Public Health, University of North Carolina, Chapel Hill, NC 27599 (e-mail: david.richardson@unc.edu).

optimization techniques are available for carrying out the maximization, the default being a dual quasi-Newton algorithm (15). PROC NLMIXED can handle a wide variety of dependent variables. In this paper, I focus on application of PROC NLMIXED to the fitting of linear relative rate Poisson regression models.

Linear relative rate model

Consider a cohort study in which incident cases of disease have been ascertained over a period of follow-up. An analytical data structure can be generated for the purposes of Poisson regression analyses consisting of counts of person-time and events, represented by the variables *pyr* and *events*, cross-classified by levels of explanatory variables. A linear relative rate model of the form

$$\text{Rate} = e^{(\beta_0+\beta_1 z_1+\beta_2 z_2)}(1+\beta_3 d),$$

where z_1 and z_2 are covariates and d is the exposure variable of interest, may be fitted via SAS as follows:

```
proc nlmixed data= ;
eta = Beta0 + Beta1*z1 + Beta2*z2 ;
lambda = exp(eta)*(1 + Beta3*d);
model events ~ poisson (pyr*lambda); run;
```

The term ''eta'' defines the log-linear term of the model, while ''lambda'' specifies that the rate of disease is the product of exp(eta) and the linear term (1 + Beta3*d). Lastly, the model statement specifies that the number of events follows a Poisson distribution with the expected number of events equal to the product of the observed person-time and the modeled rate of disease, lambda. The parameters Beta0, Beta1, Beta2, and Beta3 are estimated from the data.

Likelihood-based confidence intervals for parameters in the linear relative rate model are generally preferred over asymptotic confidence intervals, since they give better coverage (16). The numerical integration maximum likelihood method employed by PROC NLMIXED generates a true log-likelihood fit statistic that can be used to compare nested models or to derive likelihood-based confidence intervals. A likelihood-based confidence interval can be derived through a series of likelihood ratio tests; a data analyst compares the residual deviance of a model in which all parameters are allowed to vary with the residual deviance of a model in which a parameter of interest is fixed at a specified level while allowing the other model parameters to vary. The residual deviance of this model will differ from that of the original or ''null'' model. The 2 values that fix the parameter of interest and result in a change in the residual deviance by 3.84 represent the upper and lower bounds of the 95% confidence interval for this parameter. An iterative search provides these confidence bounds; a simple SAS macro can be used to automate this search and efficiently obtain likelihood-based confidence bounds (see Appendix).

Linearity in the dose-response function can be assessed, for example, by comparing a model with a linear-quadratic dose-response function to a model with a purely linear dose-response function. Fitting of a piecewise constant function offers another method for assessing the shape of the dose-response function. One approach to assessing modification of the effect of exposure in linear relative rate models by a study covariate, m, is inclusion of a log-linear subterm for the linear exposure effect (7, 17), implying a model of the form

$$\text{Rate} = e^{(\beta_0+\beta_1 z_1+\beta_2 z_2)}\left(1+\beta_3 d \times e^{\beta_4 m}\right).$$

This type of model is readily estimated in SAS as follows:

```
proc nlmixed data= ;
eta = Beta0 + Beta1*z1 + Beta2*z2 ;
lambda = exp(eta)*(1 + Beta3*d*exp(Beta4*m) );
model events ~ poisson (pyr*lambda); run;
```

Empirical example

To illustrate the fitting of linear relative rate models via SAS, I use data from a recent analysis of mortality among asbestos textile workers discussed by Hein et al. (12). The cohort included all workers employed in asbestos textile production at a plant in South Carolina between January 1, 1940, and December 31, 1965. Vital status was ascertained through December 31, 2001. The outcome of interest, lung cancer mortality, was defined on the basis of underlying cause of death, coded according to the revision (Fifth through Tenth) of the *International Classification of Diseases* (ICD) in effect at the time of death (ICD-5 codes 047B–047F; ICD-6 codes 162 and 163; ICD-7 codes 162.0, 162.1, 162.8, and 163; ICD-8/-9 code 162; and ICD-10 codes C33 and C34). The primary exposure of interest was defined as cumulative asbestos exposure, expressed in fiber-years per mL, and was computed for each worker as the product of the length of employment in each job in a year and the estimated asbestos exposure rate for that job. Person-time and events were tabulated with cross-classification by attained age (<50, 50–<55, 55–<60, 60–<65, 65–<70, 70–<75, 75–<80, or ≥80 years), sex, race (nonwhite vs. white), and cumulative exposure (<0.73, 0.73–<2.96, 2.96–<14.05, 14.05–<95, 95–<146.85, 146.85–<195.51, or ≥195.51 fiber-years/mL) (18). Cumulative exposure categories were defined by the 10th, 25th, 50th, 75th, 90th, and 95th percentiles of the exposure distribution among cases; scores were assigned to exposure categories on the basis of the person-time weighted mean value of exposures accrued in each cell of the person-time table defined by the cross-classification of categories of covariates and cumulative exposure.

Following the notation of Hein et al. (12), I fitted a linear relative rate model of the form

$$\text{Rate} = e^{(\beta_0+\beta_1 z_1+\beta_2 z_2+\beta_3 z_3+\beta_4 z_4+\beta_5 z_5+\beta_6 z_6+\beta_7 z_7+\beta_8 z_8+\beta_9 z_9)} \times(1+\beta_{10} d),$$

where z_1–z_9 represent female sex, nonwhite race, and categories of attained age (50–<55, 55–<60, 60–<65, 65–<70, 70–<75, 75–<80, or ≥80 years) and d represents estimated

Table 1. Estimated Poisson Regression Coefficients Obtained via SAS[a] PROC NLMIXED and the AMFIT Program of the EPICURE[b] Statistical Software Package[c]

Term	Parameter	SAS		EPICURE	
		Estimate	SE	Estimate	SE
Intercept	β_0	2.793	0.3583	2.792	0.3582
Sex					
Male	Referent				
Female	β_1	−1.006	0.1607	−1.006	0.1607
Race					
White	Referent				
Nonwhite	β_2	−1.022	0.2378	−1.022	0.2377
Age, years					
<50	Referent				
50–<55	β_3	2.051	0.4449	2.051	0.4448
55–<60	β_4	2.657	0.4107	2.657	0.4106
60–<65	β_5	3.266	0.3922	3.266	0.3920
65–<70	β_6	3.603	0.3886	3.603	0.3884
70–<75	β_7	3.588	0.4021	3.588	0.4019
75–<80	β_8	3.856	0.4155	3.856	0.4154
≥80	β_9	4.178	0.4325	4.178	0.4323
Asbestos exposure, fiber-years/mL	β_{10}	0.01974	0.0052	0.01974	0.0051

Abbreviation: SE, standard error.

[a] SAS Institute Inc., Cary, North Carolina.

[b] HiroSoft International Corporation, Seattle, Washington.

[c] Fittings of a linear relative rate model for lung cancer mortality (1940–2001) among workers in a South Carolina asbestos textile plant. The linear relative rate model takes the form

$$\text{Rate} = e^{(\beta_0+\beta_1 z_1+\beta_2 z_2+\beta_3 z_3+\beta_4 z_4+\beta_5 z_5+\beta_6 z_6+\beta_7 z_7+\beta_8 z_8+\beta_9 z_9)}(1+\beta_{10} d),$$

where $z1$–$z9$ represent female sex, nonwhite race, and categories of attained age (50–<55, 55–<60, 60–<65, 65–<70, 70–<75, 75–<80, or ≥80 years) and d represents estimated cumulative asbestos exposure (in fiber-years/mL) lagged by 10 years.

cumulative asbestos exposure lagged by 10 years. Tabulated person-years were divided by 100,000 so that the model intercept represented the estimated lung cancer mortality rate (per 100,000 person-years) at the referent level for all model covariates. A likelihood-based 95% confidence interval was determined for parameter β_{10}. Results obtained via SAS PROC NLMIXED were compared with those obtained using the AMFIT module of the computer software package EPICURE (HiroSoft International Corporation, Seattle, Washington), a specialized statistical software package widely used for linear relative rate analyses (14).

RESULTS

Analyses of the relation between lung cancer mortality and cumulative asbestos exposure under a 10-year lag were conducted via fitting of a linear relative rate model. Table 1 reports the coefficients obtained from fitting the model. Parameter estimates and associated standard errors obtained via SAS PROC NLMIXED are nearly identical to those obtained via the AMFIT module of EPICURE.

The 95% confidence interval for the parameter β_{10} was derived via the SAS macro presented in the Appendix (95% confidence interval: 0.01119, 0.03216). The 95% confidence interval is very similar to the confidence interval derived via EPICURE (95% confidence interval: 0.01119, 0.03216). Figure 1 shows the predicted mortality rate as a function of cumulative exposure for white males aged 60–64 years; the fitted linear relative rate model is shown along with estimates obtained via a model that included 6 indicator variables for the 7 cumulative exposure categories.

In order to evaluate effect modification by sex, I fitted a model that included a log-linear term for female sex, of the form $\left(1+\beta_{10} d \times e^{\beta_{11} m}\right)$. The estimated coefficient for males ($\beta_{10} = 0.01692$) was identical to the value obtained via the EPICURE package. The estimated value for β_{11} and its associated standard error were very similar when derived via EPICURE ($\beta_{11} = 0.5171$; standard error, 0.5431) and SAS PROC NLMIXED ($\beta_{11} = 0.5176$; standard error, 0.5455). There was no evidence that inclusion of a log-linear term to allow for effect modification by sex improved model fit (likelihood ratio test: $\chi^2 = 0.857$ (1 df), with identical test

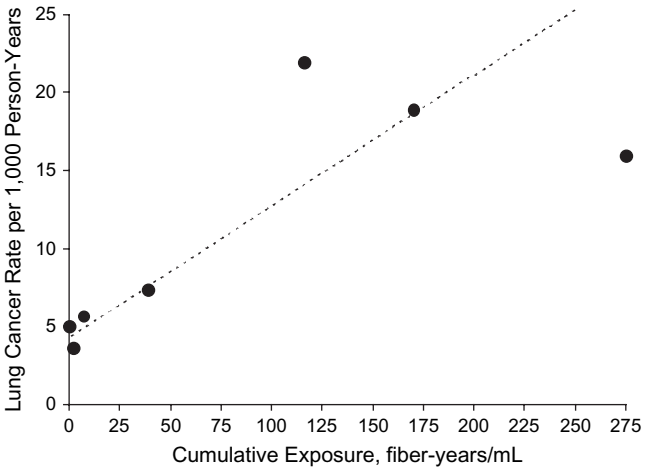

Figure 1. Estimated lung cancer mortality rate for white males aged 60–64 years as a function of cumulative exposure to chrysotile asbestos (based on the model in Table 1). The points show estimated rates for categories of cumulative exposure, plotted at the category-specific mean exposure level.

statistic values being obtained via EPICURE and SAS PROC NLMIXED).

DISCUSSION

This paper illustrates how exposure-disease analyses conducted using a linear relative rate model can be easily implemented via SAS PROC NLMIXED. I have focused on the fitting of linear relative rate models via Poisson regression methods. This approach is widely used in the occupational and environmental literature, and the linear relative rate model is one of the important models considered in the Poisson regression context. Linear odds ratio models, of the form Odds $= e^{(\beta_0)}(1 + \beta_1 d)$, may be of interest in some settings (19). SAS PROC NLMIXED is highly flexible, and it is also possible to implement an approach analogous to the fitting of a linear odds ratio model.

The results shown in Table 1 are very similar to those reported previously by Hein et al. (12), despite some differences in approach (such as categorization of variables). Importantly, however, the results obtained via fitting of the linear relative rate model via SAS PROC NLMIXED are nearly identical to those obtained via the AMFIT module of the EPICURE statistical package, demonstrating how linear relative rate Poisson regression models can be fitted via a few lines of SAS code.

The value for the model deviance obtained via the AMFIT module of the EPICURE package was not identical to the -2 log likelihood obtained via SAS PROC NLMIXED. The deviance obtained via the AMFIT module is calculated as minus twice the difference between the log likelihood for the current fitted model and the log likelihood obtained in a saturated model (14). The deviance obtained via the AMFIT module therefore differs from the -2 log likelihood obtained via SAS PROC NLMIXED by a constant amount (i.e., by an amount that is constant across fittings of nested models to the same data structure). Consequently, the reported deviance from either statistical package can be used to conduct valid likelihood ratio tests.

The linear relative rate model is commonly used in radiation epidemiology (20), and it has a number of appealing attributes. An estimate of a radiation dose-response trend obtained via a linear relative rate model is easily communicated: The parameter for the radiation dose effect describes the excess relative rate of disease per unit dose. Furthermore, for studies of populations that include people who received high doses of radiation, such as the cohort of Japanese atomic bomb survivors, the linear relative rate model for solid cancers provides a good fit to the observed data. Nonetheless, there are also important limitations to the linear relative rate model. The linear term of the model $(1 + \beta\text{dose})$ cannot be negative; if it is, this would imply a negative rate of disease. Exponential models, in contrast, have the desirable property that these estimated rates are necessarily positive quantities, regardless of the values of the linear predictor in the regression model.

By default, PROC NLMIXED assigns all model parameters an initial value of 1. While not required, model convergence may be facilitated by providing NLMIXED with more plausible starting values for some model parameters, including the intercept. This is done via the PARMS statement. Alternatively, convergence of regression models in PROC NLMIXED can be facilitated by scaling the model intercept (e.g., dividing person-years by 100,000) so that the estimated parameter takes a value near the starting value.

Linear relative rate models may be of interest in evaluations of dose-response relations with a single continuous exposure variable and in studies of interactions between multiple exposure variables (21). This paper should facilitate evaluation of linear relative rate models by demonstrating how such models are easily fitted using the SAS statistical package.

ACKNOWLEDGMENTS

Author affiliation: Department of Epidemiology, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.

This project was supported by grant K01-OH008635 from the National Institute for Occupational Safety and Health.

The author thanks Dr. Misty Hein of the National Institute for Occupational Safety and Health for her comments on the manuscript and for her support of these analyses, which made use of data derived from her research.

Conflict of interest: none declared.

REFERENCES

1. Pearce N, Checkoway H, Dement J. Exponential models for analyses of time-related factors, illustrated with asbestos textile worker mortality data. *J Occup Med*. 1988;30(6):517–522.
2. Greenland S. Modeling and variable selection in epidemiologic analysis. *Am J Public Health*. 1989;79(3):340–349.

3. Greenland S. Introduction to regression models. In: Rothman K, Greenland S, eds. *Modern Epidemiology.* 2nd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 1998:359–399.
4. Darby S, Hill D, Auvinen A, et al. Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. *BMJ*. 2005;330(7485):223.
5. Lubin JH, Boice JD Jr, Edling C, et al. Radon-exposed underground miners and inverse dose-rate (protraction enhancement) effects. *Health Phys.* 1995;69(4):494–500.
6. Committee on the Biological Effects of Ionizing Radiation, National Research Council. *Health Risks of Radon and Other Internally Deposited Alpha-Emitters: BEIR IV.* Washington, DC: National Academy Press; 1988.
7. Preston DL, Shimizu Y, Pierce DA, et al. Studies of mortality of atomic bomb survivors. Report 13: solid cancer and noncancer disease mortality: 1950–1997. *Radiat Res.* 2003;160(4):381–407.
8. Pierce DA, Shimizu Y, Preston DL, et al. Studies of the mortality of atomic bomb survivors. Report 12, part I. Cancer: 1950–1990. *Radiat Res.* 1996;146(1):1–27.
9. Boice JD Jr, Blettner M, Kleinerman RA, et al. Radiation dose and leukemia risk in patients treated for cancer of the cervix. *J Natl Cancer Inst*. 1987;79(6):1295–1311.
10. Cardis E, Gilbert ES, Carpenter L, et al. Effects of low doses and low dose rates of external ionizing radiation: cancer mortality among nuclear industry workers in three countries. *Radiat Res.* 1995;142(2):117–132.
11. Howe GR, Zablotska LB, Fix JJ, et al. Analysis of the mortality experience amongst U.S. nuclear power industry workers after chronic low-dose exposure to ionizing radiation. *Radiat Res.* 2004;162(5):517–526.
12. Hein MJ, Stayner LT, Lehman E, et al. Follow-up study of chrysotile textile workers: cohort mortality and exposure-response. *Occup Environ Med*. 2007;64(9):616–625.
13. Rinsky RA, Hornung RW, Silver SR, et al. Benzene exposure and hematopoietic mortality: a long-term epidemiologic risk assessment. *Am J Ind Med*. 2002;42(6):474–480.
14. Preston DL, Lubin JH, Pierce DA, et al. *Epicure: User's Guide*. Seattle, WA: HiroSoft International Corporation; 1993.
15. SAS Institute Inc. SAS OnlineDoc 9.1.2. Cary, NC: SAS Institute Inc; 2003.
16. Moolgavkar SH, Venzon DJ. Confidence regions in curved exponential families: application to matched case-control and survival studies with general relative risk function. *Ann Stat*. 1987;15:346–359.
17. Ron E, Lubin JH, Shore RE, et al. Thyroid cancer after exposure to external radiation: a pooled analysis of seven studies. *Radiat Res.* 1995;141(3):259–277.
18. Wood JW, Richardson DB, Wing S. A simple program to create exact person-time data in cohort analyses. *Int J Epidemiol*. 1997;26(2):395–399.
19. Greenland S. Introduction to regression modeling. In: Rothman K, Greenland S, eds. *Modern Epidemiology.* 2nd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 1998:401–432.
20. Committee to Assess Health Risks from Exposure to Low Levels of Ionizing Radiation, National Research Council. *Health Risks from Exposure to Low Levels of Ionizing Radiation: BEIR VII Phase 2*. Washington, DC: National Academies Press; 2006.
21. Thomas DC. General relative-risk models for survival time and matched case-control analysis. *Biometrics*. 1981;37(4):673–686.

APPENDIX

Assume that we have data set *DS*, which is an analytical data structure consisting of counts of person-time and events, represented by the variables *pyr* and *events*, cross-classified by categories of explanatory variables. The SAS macro below can be used to derive likelihood-based 95% confidence intervals for parameters in a linear relative rate model fitted to these data. The model is invoked by the command *%bounds* (data=*DS*, case=*events*, pyr=*pyr*, eta= , lambda= , param=). The term "eta" of the command invoking the macro allows the user to specify the log-linear term of the regression model. The term "lambda" of the command invoking the macro allows the user to specify that the rate of disease is the product of exp(eta) and a linear term of the model. Lastly, the term "param" of the command invoking the macro allows the user to specify the regression model parameter for which he or she wishes to estimate the 95% confidence interval.

```
%macro bounds (data= , case= , pyr= , eta= , lambda= , param= );
%* initialize values;
ods output fitstatistics = fitstatistics ParameterEstimates=ParameterEstimates;
proc nlmixed data=&data ;
eta = &eta ;
lambda = &lambda ;
model &case ~ poisson(&pyr* lambda);

data fitstatistics;
set fitstatistics;
if (Descr = "-2 Log Likelihood") then do;
call symput ('LL', put(value,best16.));
call symput("refLL",put(value,best16.)); end;
run;

data ParameterEstimates;
set ParameterEstimates;
if (PARAMETER = "&param") then do;
call symput("istep",put(STANDARDERROR,best16.));
call symput("ibeta",put(ESTIMATE,best16.)); end;
```

```
run;

data lci uci;
set _null_;

data lci uci;
null= 0; neglogl=0; difference=0; param=0; step=0;

%DO I = 1 %TO 2; %* I=1 is Lower Bound;
%let conv=0;
%let step=&istep;
%let beta=&ibeta;

%DO %WHILE (&CONV=0);
ods output fitstatistics = fitstatistics ParameterEstimates=ParameterEstimates;
proc nlmixed data=&data ;
&param=&beta;
eta = &eta ;
lambda = &lambda;
model &case ~ poisson(&pyr* lambda);

data fitstatistics;

set fitstatistics;
format value best16.;
if (Descr = "-2 Log Likelihood") then call symput ('LL', put(value,best16.));
run;

%let diff=%sysevalf(&LL-&refLL);
%if %sysevalf(3.8413 <= &diff) %then %do;
%if %sysevalf(&diff <= 3.8415) %then %do;
%let CONV=1; %end; %end;
%if %sysevalf(&diff > 3.8415) %then %do;
            %IF &I=1 %then %let beta=%sysevalf(&beta+&step);
            %IF &I=2 %then %let beta=%sysevalf(&beta-&step);
              %let step=%sysevalf(&step*0.5); %end;

data tmp;
null= &refLL; neglogl=&LL; difference=&diff; param=&beta; step=&step; run;

%IF &I=1 %then %do; proc append base =lci data = tmp;run; %end;
%IF &I=2 %then %do; proc append base =uci data = tmp;run; %end;

%IF &I=1 %then %let beta=%sysevalf(&beta-&step);
%IF &I=2 %then %let beta=%sysevalf(&beta+&step);
%end; %*<=Do while loop;

%end; %*<= end Do loop;

dm 'out; clear; pgm';
proc print data=lci; title1 "95% Lower Confidence Bound - Iterations";
proc print data=uci; title1 "95% Upper Confidence Bound - Iterations";
data lb (keep=PARAM); set lci end=eof; if eof then output lb;
data ub (keep=PARAM); set uci end=eof; if eof then output ub;
data bd (rename=(param=Bound)); set lb ub;
proc print data=bd NOOBS; title1 "Likelihood-Based 95% Lower and Upper
Confidence Bounds for Parameter &param";     title2 "Point Estimate is &ibeta";
run;

%mend bounds;
```

For the model shown in Table 1, the 95% confidence interval for the parameter β_{10} would be obtained by invoking the macro with the following syntax.

```
%bounds (data=DS, case=events, pyr=pyr, eta=b0 + b1*female + b2*black + b3*ag2 +
b4*ag3 + b5*ag4 + b6*ag5 + b7*ag6 + b8*ag7 + b9*ag8,
lambda=exp(eta)*(1 + b10*cds), param=b10);
```