



# A Set of SAS Macros for the Calculation of Relative Risks from Non-Parametric Logistic Regression Models using B-Splines<sup>1</sup>

Martin Gregory,<sup>1</sup> Karl-Peter Pfeiffer,<sup>2</sup> Hanno Ulmer,<sup>2</sup> and Alexander M. Strasak<sup>2</sup>

<sup>1</sup>SAS Institute Inc., Heidelberg, Germany

<sup>2</sup>Department of Medical Statistics, Informatics and Health Economics, Innsbruck Medical University, Austria

## Background

In clinical and epidemiologic research to investigate dose-response relationships, non-parametric spline regression has long been proposed as a powerful alternative to conventional parametric regression approaches, since no underlying assumptions of linearity have to be fulfilled.<sup>2,3</sup> For logistic spline models, however, to date, little standard statistical software is available to estimate any measure of risk, typically of interest when quantifying the effect of a continuous explanatory variable on a binary disease outcome.

## Formal calculation of relative risks from logistic regression models using b-splines

The formal calculation of relative risks from logistic regression models using a b-spline expansion of a continuous, independent variable was previously described by Cao et al.<sup>4</sup> and implemented by them as a S-Plus function. Specifically, the relative risk for the predictor  $x$  with respect to a reference value  $x_{ref}$  is calculated as

$$rr(x, x_{ref}) = \exp\left(\sum_{i=1}^n \beta_i [s_i(x) - s_i(x_{ref})]\right)$$

where  $n$  is the number of degrees of freedom of the spline expansion,  $\beta_i$  is the coefficient of the  $i^{\text{th}}$  spline basis function estimated by the logistic regression and  $s_i(x)$  is the value of the  $i^{\text{th}}$  spline basis function at  $x$ . 95% confidence intervals were computed based on asymptotic normality of the estimates.

## The SAS macros

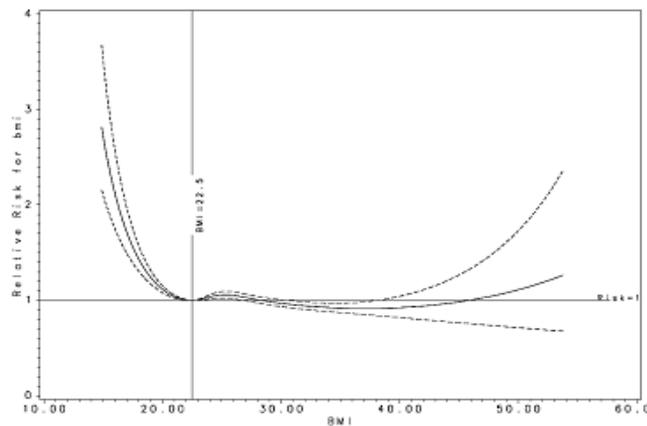
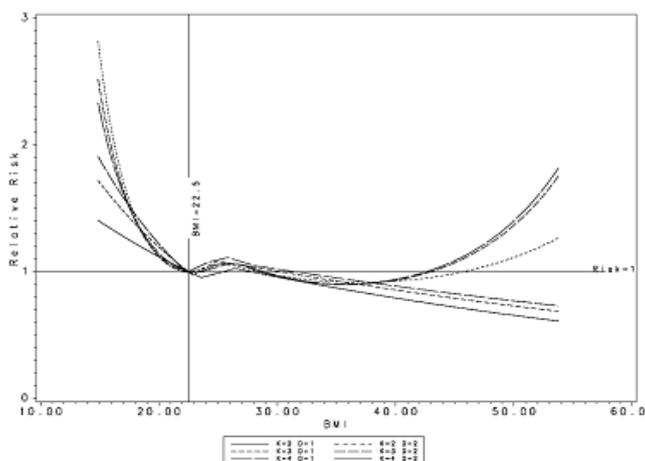
Our set of macros is written in SAS<sup>®</sup> Version 9.1 and consists of three separate components. The first macro, `%regspline`, performs (1) the spline expansion, (2) the regression analyses and (3) estimates, with respect to a supplied reference value, adjusted relative risks with their corresponding upper and lower confidence limits for the exposure variable under consideration. The second macro, `%regspline_plot`, produces a publication-quality graph of the estimated relative risks and associated confidence limits for a pre-specified range of values of the exposure variable under consideration, referring to the supplied reference value with a relative risk equaling 1.00. The third macro, `%regspline_subset`, writes estimated relative risks and respective confidence intervals for arbitrary values of the exposure variable under consideration to a SAS data set.

Table 1. Parameters required for the %regspline SAS macro

Parameter	Description
DATA	names the input dataset
OUT	names the output dataset and contains the variables of the spline expansion, the regression coefficients, the relative risk and 95% confidence intervals for the relative risk. The number of spline expansion variables is $k + d + 1$ where $k$ is the number of internal knots and $d$ the degree of the spline. The variables are named by adding a suffix of the form $n$ to the name of the spline variable, where $n \in \{0, 1, \dots, k + d\}$ . The coefficient variables are named by adding a prefix "coeff_" to the original variable names
OUTTEST	names an output dataset containing a single observation with the parameter coefficients. This dataset is produced by the OUTTEST option of the LOGISTIC procedure
DEPVAR	names the dependent variable on the input dataset
SPLINEVAR	names the independent variable which is to be replaced by a spline expansion
INDEPVAR	names the remaining independent variables to be included in the model
KNOTS	lists the knots to be used when generating the spline expansion. The value should satisfy the requirements for the KNOTS model option of the TRANSGEE procedure. You must specify a value for the knots parameter or for the knots parameter but may not specify both
KNKNOTS	specifies the number of internal knots to be used when generating the spline expansion. The value should satisfy the requirements for the KNKNOTS model option of the TRANSGEE procedure. You must specify a value for the knknots parameter or for the knots parameter but may not specify both
DEGREE	specifies the degree of the polynomial to use in generating the spline expansion
XREF	specifies a reference value of the spline variable to be used as denominator when calculating the relative risks
XREF_FUZZ	specifies how the macro should behave if the reference value specified by the xref parameter is not present on the input dataset. If xref fuzz = 0 and the specified value of xref is not present, the macro stops executing. Specifying a non-zero value for xref fuzz causes the macro to search for a value in the interval [xref - xref fuzz, xref + xref fuzz]. If any are found, the first on the input dataset is used. If none are found, the macro stops executing. The parameter is required and has a default value of 0.1

## Application of the SAS macros

We demonstrate the use of our macros in estimating the effect of body-mass index (BMI) on the risk of cancer incidence in a population-based cohort, consisting of 73624 Austrian men and 5305 incident cancers during 19 years of follow-up.<sup>5</sup> Baseline information on age, BMI, smoking status (current, former, never) and occupational status (blue collar, white collar, self-employed) was collected for each study participant.



## References

- Gregory M, Ulmer H, Pfeiffer KP, Lang S, Strasak AM. A set of SAS macros for calculating and displaying adjusted relative risks (with confidence intervals) for continuous covariates in logistic b-spline regression models. *Comput Methods Programs Biomed*, in Revision.
- Greenland S. Dose-response and trend analysis in epidemiology: Alternatives to categorical analysis. *Epidemiology* 6 (1995) 356-365.
- Boucher KM, Slattery ML, Berry TD, Quesenberry C, Anderson K. Statistical methods in epidemiology: a comparison of statistical methods to analyze dose-response and trend analysis in epidemiologic studies. *J Clin Epidemiol* 51 (1998) 1223-1233.
- Cao J, Valois MF, Goldberg MS, An S-Plus function to calculate relative risks and adjusted means for regression models using natural splines. *Comput Methods Programs Biomed*. 84 (2006) 58-62.
- Strasak AM, Rapp K, Hilde W, Oberaigner W, Ruttman E, Concin H, Diem G, Pfeiffer KP, Ulmer H; VHM&PP Study Group. Serum uric acid and risk of cancer mortality in a large prospective male cohort. *Cancer Causes Control* 18 (2007) 1021-1029.

## Contact Address:

Dr. Alexander M. Strasak  
Department of Medical Statistics, Informatics and Health Economics  
Innsbruck Medical University, Austria  
Schoepfstraße 41/1  
6020 Innsbruck - Austria  
Tel: +43/512-9003-70921  
Fax: +43/512-9003-73922  
e-mail: alexander.strasak@i-med.ac.at