

## A SAS Macro for the Calculation of Relative Risks from Non-Parametric Logistic Regression Models using B-Splines

Martin Gregory<sup>1</sup>, Karl P. 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

Due to a lack of alternative procedures, available in statistical standard software packages, current epidemiologic research, estimating risk ratios for the effects of independent exposure variables in studies of disease aetiology, is largely based upon the practice of categorizing continuous risk factors, in order to handle possible non-linearity in dose-response. However, it was long reported that this approach is associated with several problems, most notably, jumps in risks at category cut points not being biologically plausible and violating the assumption of actual risks varying smoothly with data. Moreover, it has been shown that these approaches are typically associated with a considerable loss of power and possible cut-point bias, introduced by categorization. We provide a SAS-based macro, performing non-parametric logistic regression analyses using a b-spline expansion of an independent variable under consideration and calculating relative risks with respective confidence intervals for each predicted value with respect to a supplied reference value. The macro allows specification of a number of parameters influencing how the spline expansion is carried out, producing listings and graphs in addition to the output datasets generated. The calculations are based on an algorithm previously described and implemented as an S-Plus function. 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  $x$ . Beside estimating relative risks with respective confidence intervals, our macro allows to graphically determine the shape of the association between a given, continuous exposure and the risk of a binary outcome, retaining the independent variable under consideration in its initial, continuous form, concurrently adjusting for multiple confounding factors. As the SAS macro is based upon non-parametric spline regression, it avoids common problems associated with non-linearity, typically present in conventional parametric regression approaches.