



ELSEVIER

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

A set of SAS macros for calculating and displaying adjusted odds ratios (with confidence intervals) for continuous covariates in logistic B-spline regression models[☆]

Martin Gregory^a, Hanno Ulmer^b, Karl P. Pfeiffer^b,
Stefan Lang^c, Alexander M. Strasak^{b,*}

^a SAS Institute Inc., Heidelberg, Germany

^b Department of Medical Statistics, Informatics and Health Economics, Innsbruck Medical University, Austria

^c Institute of Statistics, Faculty of Economics and Statistics, University of Innsbruck, Austria

ARTICLE INFO

Article history:

Received 20 December 2007

Received in revised form

21 April 2008

Accepted 21 May 2008

Keywords:

B-splines

Confidence intervals

Logistic regression

Odds ratios

SAS macro

ABSTRACT

In clinical and epidemiologic research to investigate dose–response associations, 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. 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 effects of one or more continuous explanatory variable(s) on a binary disease outcome. In the present paper, we propose a set of SAS macros which perform non-parametric logistic regression analysis with B-spline expansions of an arbitrary number of continuous covariates, estimating adjusted odds ratios with respective confidence intervals for any given value with respect to a supplied reference value. Our SAS codes further allow to graphically visualize the shape of the association, retaining the exposure variable under consideration in its initial, continuous form while concurrently adjusting for multiple confounding factors. The macros are easily to use and can be implemented quickly by the clinical or epidemiological researcher to flexibly investigate any dose–response association of continuous exposures with the risk of binary disease outcomes. We illustrate the application of our SAS codes by investigating the effect of body-mass index on risk of cancer incidence in a large, population-based male cohort.

© 2008 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Conceivably due to a lack of easy-to-obtain alternative approaches available in standard statistical software packages, modern clinical and epidemiologic research, investigating risk ratios for the effects of explanatory variables on a

disease outcome, is widely based upon splitting the range of a continuous exposure variable into several categories, in order to handle non-linearity in dose–response [1–4]. Resulting step function analyses have, however, long been associated with several problems, most notably discontinuities in risks at category cut-points not being biologically plausible and violating

[☆] *Mode of availability:* The complete set of SAS macros, along with detailed documentation, can be obtained by contacting the corresponding author.

* *Corresponding author at:* Department of Medical Statistics, Informatics and Health Economics, Innsbruck Medical University, Schoepfstrasse 41, 6020 Innsbruck, Austria. Tel.: +43 512 9003 70921; fax: +43 512 9003 73922.

E-mail address: alexander.strasak@i-med.ac.at (A.M. Strasak).

0169-2607/\$ – see front matter © 2008 Elsevier Ireland Ltd. All rights reserved.

doi:10.1016/j.cmpb.2008.05.004

the assumption of actual risks varying smoothly with data [5–7]. Moreover, it has been shown that categorization may be associated with cut-point bias and considerable misclassification [1,3,8,9]. Step functions are known to provide a rather crude approximation of the true relationship by not using the full range of exposure data available to estimate associations; this, in turn, leads to a reduction of statistical power and less precise regression estimates [2,5,6].

Considering the above limitations, more efficient methods of statistical data analysis have been introduced in order to investigate joint covariate effects when linearity in dose–response cannot necessarily be assumed. Particularly, spline regression [10–12] has been proposed as a cogent alternative to conventional parametric regression approaches, since no underlying assumptions of linearity have to be fulfilled. In contrast to categorical models, spline models use the full range of exposure data to estimate the shape of the association and are known to be especially beneficial when covariate effects are more pronounced at extremes of the exposure distribution [13,14].

Despite obvious benefits of spline regression, to date, little standard statistical software is available to estimate any measure of risk, typically of interest when quantifying the effects of one or more continuous explanatory variable(s) on a binary disease outcome. In the present paper, we thus propose a set of SAS macros which perform non-parametric logistic regression analyses with B-spline expansions of an arbitrary number of continuous covariates, estimating adjusted odds ratios with their respective confidence intervals for any given value with respect to a supplied reference value. The SAS macros further allow to graphically visualize the shape of the association between a continuous exposure and the risk of a binary disease outcome, retaining the exposure variable under consideration in its initial, continuous form while concurrently adjusting for multiple confounding factors.

2. Calculation of odds ratios from logistic regression models using B-splines

Splines are lines or curves, which are usually required to be continuous and smooth; they are generally defined to be piecewise of polynomials of degree d whose function values and first $d-1$ derivatives agree at the points they join [15]. Using splines in simple or multiple regression analyses allows investigating non-linear effects of continuous covariates [16]. Let $(\xi_0 =) a < \xi_1 < \xi_2 < \dots < \xi_k < b(\xi_{k+1})$ be a subdivision by K distinct points on the interval $[a, b]$ on which the x variable is valued. These points are called “knots” of the spline function $s(x)$ used to transform the x variable. Knots correspond to break points in linearity which have to be optimized; in medical or epidemiological applications, knot points may represent threshold values of a risk factor for which the probability of a disease suddenly changes. The degree of polynomial pieces and the number as well as location of knots may vary, with knot points giving the curve freedom to bend and more closely follow the data. Because analyses using splines is often cumbersome and interpretation may be complex, it is necessary to compare the trade-off between model complexity (i.e., degrees of freedom of the spline and number of knots) and model fit in

order to assess whether a much more complex model provides a significantly better fit [13].

For a fixed sequence of knots $\xi = (\xi_1, \xi_2, \dots, \xi_k)$, the set of splines is a linear space of functions with $K + d + 1$ free parameters [16,17]. A useful basis $\{B_i(\cdot, \xi)\}_{i=1,2,\dots,K+d+1}$ for this linear space is given by Schoenberg’s B-splines, or basic-splines [18], which are numerically well-conditioned and achieve local sensitivity to the data. A more-detailed introduction to spline regression can be found elsewhere [10–12,15–17]; an extension of spline methods to logistic regression models was previously introduced by Devlin and Weeks [19].

The formal calculation of odds ratios from logistic regression models using a B-spline expansion of a continuous, independent variable was described in detail by Cao et al. [4] and implemented by them as an S-Plus function. The estimated odds ratio (or) for the predictor x with respect to a reference value x_{ref} is calculated as

$$\hat{\text{or}}(x, x_{\text{ref}}) = \exp \left(\sum_{i=1}^n \hat{\beta}_i [s_i(x) - s_i(x_{\text{ref}})] \right) \quad (1)$$

where n is the number of degrees of freedom of the spline expansion, $\hat{\beta}_i$ is the coefficient of the i th spline basis function estimated by the logistic regression and $s_i(x)$ is the value of the i th spline basis function at x . The limits of the respective 95% confidence interval, based on asymptotic normality of the estimates are calculated as

$$\exp(\log(\hat{\text{or}}(x, x_{\text{ref}})) \pm 1.96 \times \hat{\sigma}_{\log(\text{or})}) \quad (2)$$

where $\hat{\sigma}_{\log(\text{or})}^2$ is the variance of $\log(\hat{\text{or}}(x, x_{\text{ref}}))$ calculated as $s'Cs$ where s is the vector:

$$(S_1(x) - S_1(x_{\text{ref}}), \dots, S_n(x) - S_n(x_{\text{ref}})) \quad (3)$$

and C is the covariance matrix of the regression coefficients.

3. The SAS macros

Our SAS macros were developed in response to the problem of calculating and visualizing adjusted odds ratios for continuous covariates in a logistic regression when the covariates have been replaced by spline expansions because the relationship is non-linear. This task cannot be done using a single SAS procedure but must employ the TRANSREG [20] procedure to generate the B-spline expansions, the LOGISTIC [21] procedure to carry out the regression and a DATA Step in order to calculate odds ratios with their confidence intervals for the original covariate(s). The macros allow specification of a number of parameters influencing how the spline expansions are carried out, passing of model options to PROC LOGISTIC, producing listings and graphs in addition to the output datasets generated.

The set of macros is written in SAS Version 9.1 [22] and consists of three separate components. The first macro, %regspline, performs (1) the spline expansion(s), (2) the regression analyses and (3) estimates, with respect to a supplied reference value, adjusted odds ratios with their corresponding upper

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 odds ratios and 95% confidence intervals for the odds ratios. 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 “cf.” 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	Lists the independent variables which are to be replaced by a spline expansion. The list elements should be separated by a space
INDEPVARS	Names the remaining independent variables to be included in the model
KNOTS	Lists the knots to be used when generating the spline expansion. If multiple spline expansion variables have been specified, use a colon to separate the lists for each variable. The elements of the list for a variable are separated by spaces. The values should satisfy the requirements for the KNOTS model option of the TRANSREG procedure. You must specify a value for the knots parameter or for the nknots parameter but may not specify both
NKNOTS	Specifies the number of internal knots to be used when generating each spline expansion. Separate the values by a space. The value should satisfy the requirements for the NKNOTS model option of the TRANSREG procedure. You must specify a value for the NKNOTS parameter or for the knots parameter but may not specify both
DEGREE	Specifies the degree of the polynomial to use in generating each spline expansion. Separate the values by a space
XREF	Specifies a list of reference values for each spline variable to be used as denominator when calculating the odds ratios
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. You must specify a value for each variable specified in the SPLINEVAR parameter

Table 2 – Parameters required for the %regspline.plot SAS macro

Parameter	Description
DATA	Names the dataset containing the data to plot
SPLINEVAR	Names the numeric variable for which the spline expansion was calculated. This variable is used for the x axis
RISKVAR	Names a numeric variable containing the odds ratio values. This is the response variable
RISKLCL	Names a numeric variable containing the lower confidence limit for RISKVAR
RISKULC	Names a numeric variable containing the upper confidence limit for RISKVAR
XREF	Value for the vertical reference line corresponding to the reference value
XREF_LABEL	Label for the vertical reference line
LINEWIDTH	Specifies a relative thickness of the graphed lines. Default is 1.5
AXISHEIGHTV	Specifies a relative height for the font used for the values for the axis tickmarks for both axes. Default is 0.75
AXISHEIGHTL	Specifies a relative height for the font used for the axis labels for both axes. Default is 0.75
LEGENDHEIGHT	Specifies a relative height for the font used for the legend. Default is 0.5

and lower confidence limits for the exposure variable(s) under consideration. The second macro, `%regspline.plot`, produces a publication-quality graph of the estimated, adjusted odds ratios and associated confidence limits for a pre-specified range of values of the exposure variable(s) under consideration, referring to the supplied reference value with an odds ratio equaling 1.00. The third macro, `%regspline.subset`, writes estimated adjusted odds ratios and respective confidence intervals for given values of the exposure variable(s) under consideration to a SAS data set. Tables 1–3 provide a description of the parameters to be set by the user for each component of our set of SAS macros. The macros are easy to use and can be implemented quickly by the clinical or epidemiological researcher to flexibly investigate any dose–response association of one or more continuous exposure variable(s) with the risk of a binary disease outcome.

Table 3 – Parameters required for the %regspline.subset SAS macro

Parameter	Description
DATA	Names the input dataset
OUT	Names the output dataset to create
SUBSET	Specifies a list of the values of the expansion variable for which to extract records
OR	Names the variable containing the odds ratio
OR_LCM	Names the variable containing the lower 95% confidence limit for the odds ratio
OR_UCM	Names the variable containing the upper 95% confidence limit for the odds ratio

4. Application of the SAS macros

We demonstrate the use of our SAS codes in estimating the effect of body-mass index (BMI) on the risk of cancer incidence in a population-based male cohort. Our data is derived from the Vorarlberg Health Monitoring and Promotion Program (VHM&PP), one of the world's largest ongoing population-based risk factor surveillance programs, linked to disease specific morbidity and mortality outcomes [23–25]. The VHM&PP is routinely performed by the Agency for Social and Preventive Medicine and addresses all adults of the entire province. It routinely includes the recording of socio-demographic data and a physical examination with a fasting blood sample. A more-detailed description of the program methodology and study population has been reported previously [23–25]. Briefly, our study sample consisted of 73,624 men (mean age 41.6 years) and 5305 incident cancers during 19 years of follow-up. Baseline information on age and BMI (both continuous variables), smoking status (current, former, never) and occupational status (blue collar, white collar, self-employed) was collected for each study participant. In order to demonstrate the use of our SAS macros in simultaneously employing spline methods on more than one continuous covariate and to additionally gain sufficient flexibility, also in modelling of continuous confounders, in our example, we considered B-spline expansions of both, BMI and age. All models additionally were adjusted for smoking status and occupational status (both categorical variables). A BMI of 21.75 kg/m², as the mid-point of the reference range of normal BMI values [26], was used for the calculation of odds ratios in all models.

Considering the illustrative character of our analyses and the fact that issues of model selection in general remain a major challenge [13] far beyond the scope of the present paper, we confined our analysis on fitting linear (df = 1) and

we call the %regspline macro for each of the combinations of three and four knots with degrees 1 and 2. A representative call, for three knots and degree 2 of BMI and four knots and degree 1 of age, is

```
%regspline(data=sd.bmi_males,out=sd.k3d2_k4d1,
depvar=event,splinevar=bmi age,indepvars=smoker emptye,
nknots=3 4,degree=2 1,fit=fit_k3d2_k4d1,or=or_bmi or_age,
xref=21.75 39,xref_fuzz=0 0) ;
```

This call reads the input dataset, carries out a logistic regression of the binary-dependent variable on a second degree spline expansion with three knots of the BMI variable and a first degree spline expansion with four knots of age, with smoking status (three categories) and occupational status (three categories) as covariates. In addition to the printed output generated by PROC LOGISTIC, the macro call produces two SAS datasets as output. The first dataset contains the original variables along with the B-spline expansion variables, the predicted values, the estimated regression coefficients for each of the independent variables (original variables with a prefix "cf."), the estimated adjusted odds ratios compared to the specified BMI reference value of 21.75 (or.bmi) and the associated lower (or.bmi.lcm) and upper (or.bmi.ucm) confidence limits for each given value of BMI. The second dataset contains the AIC values for the model. Note that (1) one could also save other model fit statistics produced by PROC LOGISTICS, if required and (2) in case of simultaneously employing spline methods on more than one continuous covariate, it is necessary to specify hypothetic reference values for all spline variables in the model, even if the calculation of odds ratios is required for only one variable.

Using the second macro %regspline.plot, we can graph the estimated adjusted odds ratios for BMI along with the corresponding confidence limits, for the AIC-minimizing, best-fitting spline model (AIC = 31388.2; BMI: df = 2, knots = 3; age: df = 2, knots = 2), using the following call:

```
%regspline_plot(data=sd.k3d2_k2d2,color=,
xref=21.75,xref_label=BMI,
splinevar=bmi,riskvar=or_bmi,risklcl=or_bmi_lcm,riskuc1=or_bmi_ucm,
axisheightv=1.25,axisheightl=1.25,legendheight=0.5,
linewidth=1.75) ;
```

quadratic (df = 2) spline models with knots at consecutive quartiles (three knots) and quintiles (four knots) of the BMI and age distribution and abstained from fitting more complex spline models to our data. Note however, that it was reported that in practice, few well-located knots generally suffice in most cases, although deciding on their optimal location is difficult and no generally accepted criteria yet exist [16]. The best-fitting combination of spline expansions of BMI and age was defined as the one minimizing the Akaike's Information Criterion (AIC), a penalized likelihood that takes into account the number of parameters estimated in the model, compromising between a good fit and a parsimonious model [27]. It has been shown that AIC may perform poor for small samples or in presence of too many parameters in relation to the size of the sample [28]; however, none of these conditions affected our investigation.

In order to technically determine the AIC-minimizing, best-fitting combination of spline expansions of BMI and age,

The graph produced by this call is shown in Fig. 1. The input dataset was produced by the %regspline macro. The xref and xref.label parameters specify the position and label for the vertical reference line. The splinevar, riskvar, risklcl and riskuc1 parameters allow specification of the covariate, risk and confidence limit variables. The color parameter controls whether colored lines are used instead of dashed lines. The axisheightv, axisheightl and legendheight parameters control the text size of the axis values, labels and the legend text, respectively. These values specify a factor to be applied to the default text height. Finally, we have used the linewidth parameter to draw lines 75% thicker than the default used by SAS/GRAPH.

Table 4 containing selective numerically results on adjusted odds ratios for the association of BMI and cancer incidence from the AIC-minimizing, best-fitting spline model, was produced using the %regspline.subset macro. This macro creates a SAS data set containing a subset of an output dataset from the %regspline macro for pre-specified values of the spline

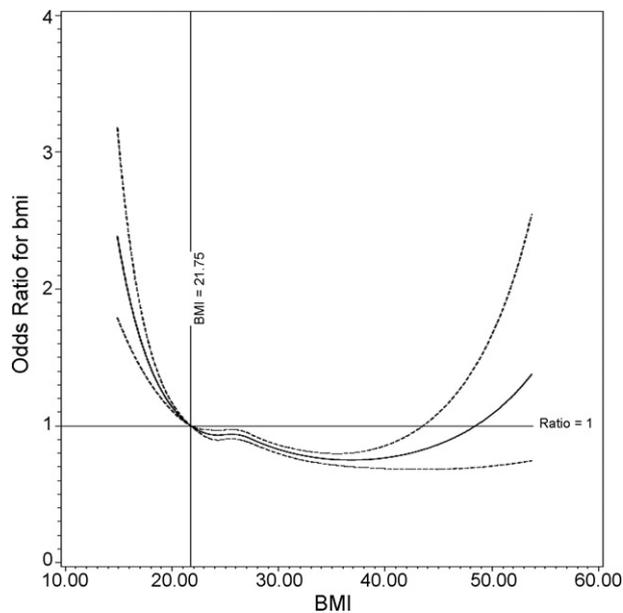


Fig. 1 – Estimated adjusted odds ratios (solid line) with 95% confidence intervals (dashed lines) for the association of BMI and overall cancer incidence from the AIC-minimizing, best-fitting regression spline model in 73,624 male VHM&PP participants. BMI was modeled by a B-spline expansion with $df=2$ and knots = 3. All estimates are adjusted for age (modeled as B-spline with $df=2$ and knots = 2), smoking status and occupational status. A BMI of 21.75 kg/m², as the mid-point of the reference range of normal BMI values, was used for the calculation of all odds ratios.

expansion variable. The call used to extract the values for BMI in Table 4 was

```
%regspline_subset(data=sd.k3d2_k2d2,out=rr_list,
subset=15 20 21.75 25 30 35 40 45 50,
splinevar=bmi,or=or_bmi,or_ucm=or_bmi_ucm,or_lcm=or_bmi_lcm);
```

Table 4 – Estimated adjusted odds ratios with 95% confidence intervals for the association of BMI and overall cancer incidence from AIC-minimizing, best-fitting regression spline model in 73,624 male VHM&PP participants

BMI	Odds ratio ^a	95% confidence Interval
15	2.35	1.77–3.11
20	1.14	1.11–1.16
21.75	1.00	Reference
25	0.94	0.90–0.97
30	0.83	0.80–0.86
35	0.75	0.71–0.80
40	0.77	0.69–0.85
45	0.86	0.68–1.09
50	1.07	0.70–1.62

^a BMI was modeled by a B-spline expansion with $df=2$ and knots = 3. All estimates are adjusted for age (modeled as B-spline with $df=2$ and knots = 2), smoking status and occupational status. A BMI of 21.75 kg/m², as the mid-point of the reference range of normal BMI values, was used for the calculation of all odds ratios.

The data step following the macro call produces a tab-separated file containing the values in a form which can be directly inserted into the table.

5. Requirements

Our macros require SAS release 8.2 or later; additionally, the modules SAS/STAT and SAS/GRAPH are required for the %regspline and %regspline.plot macros, respectively.

Conflicts of interest

None.

Acknowledgement

Funding sources: our work was supported by Austrian National Bank Grant OENB-12737 (to H.U.).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2008.05.004.

REFERENCES

- [1] D.G. Altman, Categorizing continuous variables, *Br. J. Cancer* 64 (1991) 975.
- [2] K.M. Boucher, M.L. Slattery, T.D. Berry, C. Quesenberry, K. Anderson, 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.
- [3] S. Greenland, Dose–response and trend analysis in epidemiology: alternatives to categorical analysis, *Epidemiology* 6 (1995) 356–365.
- [4] J. Cao, M.F. Valois, M.S. Goldberg, 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.
- [5] C. Weinberg, How bad is categorization? *Epidemiology* 6 (1995) 345–347.
- [6] S. Greenland, Avoiding power loss associated with categorization and ordinal scores in dose–response and trend analysis, *Epidemiology* 6 (1995) 450–454.
- [7] L.P. Zhao, L.N. Kolonel, Efficiency loss from categorizing quantitative exposures into qualitative exposures in case-control studies, *Am. J. Epidemiol.* 136 (1992) 464–474.
- [8] F.E. Harrell, K.L. Lee, B.G. Pollock, Regression models in clinical studies: determining relationships between predictors and response, *J. Natl. Cancer Inst.* 80 (1988) 1198–1202.
- [9] M. Maclure, S. Greenland, Tests for trend and dose response: misinterpretations and alternatives, *Am. J. Epidemiol.* 135 (1992) 96–104.
- [10] E.J. Wegman, I.W. Whright, Splines in statistics, *J. Am. Stat. Assoc.* 78 (1983) 351–365.
- [11] R.L. Eubank, *Spline Smoothing and Nonparametric Regression*, Marcel Dekker, Berlin, 1988.
- [12] K.J. Rothman, S. Greenland, *Modern epidemiology*, 2nd ed., Lippincott-Raven Publishers, Philadelphia, 1998.

- [13] P.S. Rosenberg, H. Katki, C.A. Swanson, L.M. Brown, S. Wacholder, R.N. Hoover, Quantifying epidemiologic risk factors using non-parametric regression: model selection remains the greatest challenge, *Stat. Med.* 22 (2003) 3369–3381.
- [14] J. Polesel, L. Dal Maso, V. Bagnardi, A. Zucchetto, A. Zambon, F. Levi, C. La Vecchia, S. Franceschi, Estimating dose–response relationship between ethanol and risk of cancer using regression spline models, *Int. J. Cancer* 114 (2005) 836–841.
- [15] F. Bessaoud, J.P. Daures, N. Molinari, Free knot splines for logistic models and threshold selection, *Comput. Methods Programs Biomed.* 77 (2005) 1–9.
- [16] N. Molinari, J.P. Daurés, J.F. Durand, Regression splines for threshold selection in survival data analysis, *Stat. Med.* 20 (2001) 237–247.
- [17] C. de Boor, *A Practical Guide to Splines*, Springer-Verlag, New York, 1978.
- [18] H.P. Curry, I.J. Schoenberg, On Polya frequency functions. IV. The fundamental splines and their limits, *Anal. Math.* 17 (1966) 71–107.
- [19] T.F. Devlin, B.J. Weeks, Spline functions for logistic regression modeling, in: *Proceedings of the 11th Annual SAS Users Group International Conference*, SAS Institute Inc., Cary, NC, 1986, pp. 646–649.
- [20] SAS Institute Inc., The TRANSREG Procedure, SAS Online Documentation, 9.1.3, 2003, <http://support.sas.com/onlinedoc/913/getDoc/en/statug.hlp/transregindex.htm>.
- [21] SAS Institute Inc., The LOGISTIC Procedure, SAS Online Documentation, 9.1.3, 2003, <http://support.sas.com/onlinedoc/913/getDoc/en/statug.hlp/logisticindex.htm>.
- [22] SAS Institute Inc., Cary, NC, 2007.
- [23] H. Ulmer, C. Kelleher, G. Diem, H. Concin, Long-term tracking of cardiovascular risk factors among men and women in a large population-based health system: the Vorarlberg Health Monitoring & Promotion Programme, *Eur. Heart J.* 24 (2003) 1004–1013.
- [24] A. Strasak, K. Rapp, W. Hilbe, W. Oberaigner, E. Ruttmann, H. Concin, G. Diem, K.P. Pfeiffer, H. Ulmer, The 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.
- [25] A. Strasak, E. Ruttmann, L. Brant, C. Kelleher, J. Klenk, H. Concin, G. Diem, K.P. Pfeiffer, H. Ulmer, The VHM&PP Study Group, Serum uric acid and risk of cardiovascular mortality: a prospective long-term study in 83683 Austrian men, *Clin. Chem.* 54 (2008) 273–284.
- [26] World Health Organization, http://www.who.int/bmi/index.jsp?introPage=intro_3.html.
- [27] H. Akaike, Information theory and an extension of the maximum likelihood principle, in: *Proceedings of the 2nd International Symposium on Information Theory*, Budapest, 1973, pp. 267–281.
- [28] N. Sugiura, Further analysis of the data by Akaike's information criterion and the finite corrections, *Commun. Stat. Theor. Methods* 7 (1978) 13–26.