Clin Chem Lab Med 2009;47(4):464–470 © 2009 by Walter de Gruyter • Berlin • New York. DOI 10.1515/CCLM.2009.104

Minimizing the effects of multicollinearity in the polynomial regression of age relationships and sex differences in serum levels of pregnenolone sulfate in healthy subjects

Milan Meloun[1,*], Martin Hill[2] and Helena Včeláková-Havlíková[2]

[1] Department of Analytical Chemistry, Faculty of Chemical Technology, Pardubice University, Pardubice, Czech Republic

[2] Institute of Endocrinology, Prague, Czech Republic

Abstract

Background: Pregnenolone sulfate (PregS) is known as a steroid conjugate positively modulating N-methyl-D-aspartate receptors on neuronal membranes. These receptors are responsible for permeability of calcium channels and activation of neuronal function. Neuroactivating effect of PregS is also exerted via non-competitive negative modulation of $GABA_A$ receptors regulating the chloride influx. Recently, a penetrability of blood-brain barrier for PregS was found in rat, but some experiments in agreement with this finding were reported even earlier. It is known that circulating levels of PregS in human are relatively high depending primarily on age and adrenal activity.

Methods: Concerning the neuromodulating effect of PregS, we recently evaluated age relationships of PregS in both sexes using polynomial regression models known to bring about the problems of multicollinearity, i.e., strong correlations among independent variables. Several criteria for the selection of suitable bias are demonstrated. Biased estimators based on the generalized principal component regression (GPCR) method avoiding multicollinearity problems are described.

Results: Significant differences were found between men and women in the course of the age dependence of PregS. In women, a significant maximum was found around the 30th year followed by a rapid decline, while the maximum in men was achieved almost 10 years earlier and changes were minor up to the 60th year. The investigation of gender differences and age dependencies in PregS could be of interest given its well-known neurostimulating effect, relatively high serum concentration, and the probable partial permeability of the blood-brain barrier for the steroid conjugate.

Conclusions: GPCR in combination with the *MEP* (mean quadric error of prediction) criterion is extremely useful and appealing for constructing biased models. It can also be used for achieving such estimates with regard to keeping the model course corresponding to the data trend, especially in polynomial type regression models.

Clin Chem Lab Med 2009;47:464–70.

*Corresponding author: Prof. RNDr. Milan Meloun, DrSc, Department of Analytical Chemistry, University Pardubice, 532 10 Pardubice, Czech Republic
Phone: +42-46-6037026, Fax: +42-46-6037068,
E-mail: milan.meloun@upce.cz

Received November 28, 2008; accepted January 9, 2009; previously published online March 12, 2009

Keywords: age relationship of pregnenolone sulfate; multicollinearity; principal component regression; steroids.

Introduction

Pregnenolone sulfate (PregS) is known as a steroid conjugate positively modulating N-methyl-D-aspartate receptors on neuronal membranes. These receptors are responsible for permeability of calcium channels and activation of neuronal function (1–6). Neuroactivating effect of PregS is also exerted via non-competitive negative modulation of $GABA_A$ receptors regulating the chloride influx (1, 7). Recently, a penetrability of blood-brain barrier for PregS was found in rat (8), but the experiments in agreement with this finding were reported even earlier (9–11). It is known that circulating levels of PregS in human are relatively high depending primarily on age and adrenal activity (12, 13). The use of PregS (with significant responsiveness to ACTH stimulation) was also suggested as a marker of adrenal dysfunction in children (13). Taken together, complete information about the age dependence of PregS, including the determination of reference limits in the age groups, should be of interest not only in the diagnostics of adrenal disorders but also in the diagnostics of some disturbances of the central nervous system as anxiety-depressive syndrome (14), or the diseases connected with aging, such as Alzheimer's disease. de Peretti and Mappus described in detail the age dependence of PregS during childhood and adolescence, but the changes during adulthood were not investigated (12). We recently concentrated on this problem evaluating the age relationships of PregS in both sexes in the age range from 4 to 70 years using a polynomial regression model (15).

In this paper, we attempt to demonstrate a correct approach for the analysis of age dependencies of the hormones, which are commonly represented by the data with a non-Gaussian distribution and non-constant variance. In addition, multicollinearity (strong correlations between independent variables) is characteristic in the commonly used ordinary polynomial regression models. The difficulties as mentioned above could result in a complete misinterpretation of the data. Accordingly, we suggest an approach based on one of the biased regression models. A generalized principal component polynomial regression (GPCR) of the data transformed by power transformation was chosen as a suitable method capable of coping with the aforementioned problems. The aims of this paper are to emphasize the importance of understanding the nature of any near-singularities in the data that might cause problems with ordinary least squares regression, and to describe the algorithm, the principal component regression. Several criteria for the selection of suitable bias are demonstrated and biased estimators based on the generalized principal component regression (GPCR) method are described.

Methodology

The polynomial linear regression model with n observations of the m-th order polynomial variable and for an additive model of measurements errors is assumed, $\boldsymbol{y}=\boldsymbol{X}\,\boldsymbol{\beta}+\boldsymbol{\varepsilon}$. Vector $\boldsymbol{y}$ has dimensions $(n\times 1)$ and matrix $\boldsymbol{X}$ has dimensions $(n\times m)$. Random errors ε_i in dependent variable y should have a normal distribution $N(0,\ \sigma^2)$. When the least squares assumptions are valid, the parameter estimates $\boldsymbol{b}$ found by minimization of the sum of squared residuals RSS

$$RSS=\sum_{i=1}^{n}\left[y_i-\sum_{j=0}^{m}x_{ij}b_j\right]^2\approx\text{minimum} \qquad [1]$$

are the best linear unbiased estimators (16, 17). The conventional least-squares estimator $\boldsymbol{b}$ has the form $\mathbf{b}=(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$ with corresponding variance $D(\mathbf{b})=\sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$. However, some difficulties arise when the matrix $\mathbf{X}^{\mathrm{T}}\,\mathbf{X}$ appears to be singular. In some cases, especially with polynomial models, the parameter estimates may be without physical meaning. The multicollinearity problem in regression refers to the set of problems created when there are near-singularities among the columns of the $\mathbf{X}$ matrix and certain linear combinations of the columns of $\mathbf{X}$ are nearly zero (18).

a) The condition number $K=\lambda_{max}/\lambda_{min}$ contains λ_{max} and λ_{min}, the largest and the smallest eigenvalues of a matrix $\mathbf{R}$ (19). The condition number provides a measure of the sensitivity of the solution to the normal equations to small changes in $\mathbf{X}$ or $\mathbf{y}$. A large condition number indicates that a near-singularity is causing the matrix to be poorly conditioned. If $K>1000$, very strong multicollinearity is detected.
b) The variance inflation factor for the j-th regression parameter VIF_j is defined as the ratio of the variance of the jth regression coefficient to the same variance for orthogonal variables when $\mathbf{R}$ is the unit matrix. If $VIF_j>10$, strong multicollinearity is detected. If there is a near-singularity involving X_j and the other independent variables, R_j^2 will be near 1.0 and VIF_j will be large. If X_j is orthogonal to the other independent variables, R_j^2 will be 0 and VIF_j will be 1.0.

Biased regression refers to that class of regression methods in which unbiasedness is no longer required. GPCR solves the collinearity problem by elimination of those dimensions of the X-space that induce the problem and has been described previously (18).

One of the main properties of regression models is a good predictive ability. This predictive ability can also be adopted for the selection of an, in some sense, optimum criterion parameter P. Various criteria for testing prediction ability may be used; one of the most efficient seems to be the mean quadratic error of prediction MEP (in the literature it is also referred to as the mean squared error of prediction $MSEP$) defined by the relationship

$$MEP=\frac{\sum_{i=1}^{n}(y_i-\mathbf{x}_i^{\mathrm{T}}\mathbf{b}_{(i)})^2}{n} \qquad [2]$$

where $\mathbf{b}_{(i)}$ is the estimate of regression parameters when all points except the ith one were used and $\boldsymbol{x}_i$ is the ith row of matrix $\mathbf{X}$. The statistic MEP uses a prediction $\hat{y}_i$ from an estimate constructed without including the ith point. The most suitable model is that which gives the lowest value (minimum) of the mean quadratic error of prediction MEP. Beyond the MEP, the predicted coefficient of determination R^2_p (maximum) and the Akaike information criterion AIC (minimum) can also be used. The MEP can be used to express the predicted determination coefficient:

$$R_P^2\ (\times 100\%)=1-\frac{n\times MEP}{\sum_{i=1}^{n}y_i^2-n\times\bar{y}^2} \qquad [3]$$

The Akaike information criterion AIC is thus defined:

$$AIC=n\ln\left(\frac{U(\mathbf{b})}{n}\right)+2m \qquad [4]$$

The most suitable model gives the lowest value of the mean quadratic error of prediction MEP and the Akaike information criterion AIC, and the highest value of the predicted determination coefficient, R_p^2. The calculated P does not generally correspond to a global minimum but parameter estimates and statistical characteristics are greatly improved.

Transformation in cases of non-normality of variable distributions

There are two basic reasons for transforming variables in regression: transformation of the dependent variable is a possible remedy for non-normality and for heterogeneous variances of the random errors $\boldsymbol{\varepsilon}$. Transformations to improve normality generally have lower priority than transformation to simplify relationship or stabilize variance. Fortunately, transformations to stabilize variance often have the effect of improving normality as well.

Transformation for symmetry of error distribution is carried out by a simple power transformation:

$$y_i^{(\lambda)} = \begin{cases} y^{\lambda} & \text{for parameter } \lambda > 0 \\ \ln y & \text{for parameter } \lambda = 0 \\ -y^{-\lambda} & \text{for parameter } \lambda < 0 \end{cases} \qquad [5]$$

which does not retain the scale, is not always continuous, and is suitable only for positive y. Optimal estimates of parameter $\hat{\lambda}$ are sought by minimizing the absolute values of particular characteristics of an asymmetry of residuals $\hat{e}$.

Transformation leading to approximate normality may be carried out by the use of the family of Box-Cox transformations (20), defined as:

$$y_i^{(\lambda)} = \begin{cases} (y^{\lambda} - 1)/\lambda & \text{for parameter } \lambda \neq 0 \\ \ln y & \text{for parameter } \lambda = 0 \end{cases} \qquad [6]$$

where y is a positive variable and λ is a real number or in a form with variable standardization. The maximum likelihood solution is obtained by the least squares analysis on the transformed data for several choices of λ from, say $\lambda = -3$ to $+3$. Let $RSS(\lambda)$ be the residual sum of squares from fitting the model to transformed dependent variable $y^{(\lambda)}$ for the given choice of λ and let $s^2(y) = RSS(y)/n$.

Box-Cox transformation leads to y when λ is equal to 1 and log y being the limiting form of the function as λ tends to 0. There is no reason to suppose that either of these values of λ is optimal, and it thus makes sense to try a range of values and see which yields the minimum of RSS. If it is to try ten values of λ, ten new dependent variables are generated within the regression application using the functional form and the different values of λ. The resulting y is regressed separately on the explanatory variables.

Materials and methods

Subjects and plasma samples

Blood was taken from 230 healthy females aged between 10 and 70 years, and from 179 healthy males aged between 4 and 69 years, who had been invited by random selection for a survey of iodine deficiency in the district of Cheb in West Bohemia, within the framework of a study on iodine deficiency in the Czech Republic. Blood was withdrawn from the cubital vein between 08:00 h and 10:00 h into heparin-coated vials. After no more than 2 h, serum was separated and stored in a freezer at –20°C until processed.

Determination of pregnenolone sulfate

PregS was determined using the specific radioimmunoassay as described previously (15). The sensitivity of analysis was 32 pg/tube, and the inter- and intra-assay coefficients of variation were 10.9% and 4.3%, respectively.

Procedure of statistical data treatment

The procedure for the construction of a best polynomial regression subset $\{1, x, x^2, x^3, \ldots, x^m\}$ comprises the following steps:

Step 1 Proposal of a model.

Step 2 Examination of the multicollinearity and statistical significance of the parameter estimates.

Step 3 Construction of a more accurate model using GPCR with an optimum value of a criterion P minimizing characteristics *MEP*, *AIC* and maximizing R_P^2 can be used: on the base of *MEP* or *AIC* the most convenient regression model of the transformed data is determined. If some parameters are statistically non-significant, the most suitable parameter P is sought with the use of *MEP* and *AIC*.

Step 4 Examination of variable normality and transformation, recalculation of results.

Software used

The algorithm in S-Plus was written for the creation of the computation of the GPCR and the linear regression module of the ADSTAT package was also used, cf. (21).

Results and discussion

Statistical evaluation of data

We attempt to describe age relationships and sex differences in serum levels of PregS using principal component regression on a polynomial model. For the solution of these types of problems, GPCR with an optimum value of a criterion P can be used. The main aim is to find a degree of polynomial regression model m which describes the content of PregS for male and female patients, respectively, in dependence on age, and also to estimate all the polynomial parameters, $E(\varepsilon/x) = \beta_0 + \beta_1 x + \ldots + \beta_m x^m$. The purpose of the least squares analysis will influence the manner in which the model is constructed. There are potential uses of the regression equations given in providing a good description of the behavior of the response variable, the prediction of future responses and the estimation of mean responses, the estimation of parameters β and development of realistic models of the process. Each objective has different implications on how much emphasis is placed on eliminating variables from the model, as to how important it is that the retained variables be causally related to the response variable, and on the amount of effort required to make the model realistic.

In **step 1** the proposal of a regression model for male patients' data is used: Figure 1 presents the *MEP* statistics for an increasing degree of polynomial m and the ordinary least squares (OLS) used. The lowest *MEP* value was achieved for a polynomial of the 7th degree. Although the polynomial of the 6th degree differs only slightly, the 6th degree polynomial was preferred here. The parameter estimates from β_0 through β_6 are not significantly different from zero, which here results from strong multicollinearity.

In **step 2** the exploratory data analysis in regression is applied, and the scatter plot of the regression curve

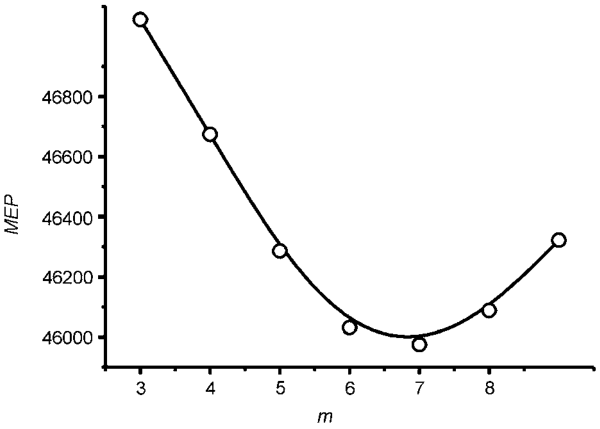

Figure 1 A search for the optimum polynomial degree *m* obtained for the lowest value of the mean error of prediction *MEP* and for the highest value of the predicted determination coefficient R_P^2 regarding the ordinary least squares (OLS) polynomial regression of the age dependencies of pregnenolone sulfate in the serum of 179 males aged 4–69 years.

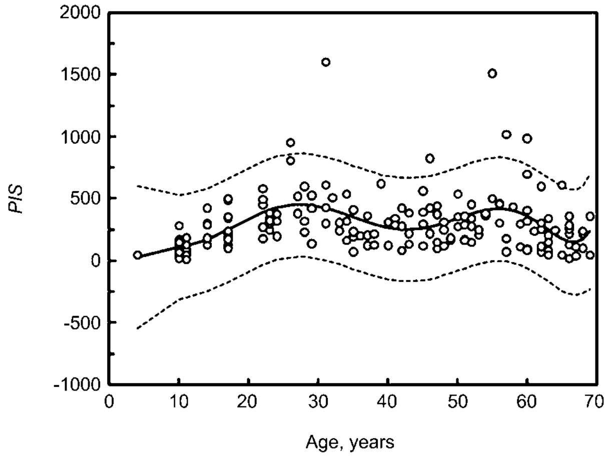

Figure 2 The scatter plot of polynomial regression of the age dependencies of pregnenolone sulfate (nmol/L) in the serum of 179 males aged 4–69 years from Figure 1 calculated with the use of the ordinary least squares (OLS) method and original data.
The curves of the mean prediction (full line) and the 95% Working-Hotelling interval bands of prediction (dashed lines) are symmetrical and rather broad at the ends of the data interval and therefore do not permit prediction of *y* outside the data interval. All of the parameters of the polynomial were statistically non-significant.

of PregS data in dependence on age of male patients (Figure 2) shows a skewed asymmetric distribution of random errors in variable *y* as shown in Table 1. The rankit *Q-Q* plot of jackknife residuals proves the non-normal distribution. The ordinary residuals exhibit strong heteroscedasticity (Figure 3) and some influential points.

In **step 3** the examination of multicollinearity comprises an estimation of the maximum condition number $K=3.63\times10^8$ which is higher than 1000 and the largest value of the variance inflation factor $VIF=3.05\times10^7$ which is higher than 10; and both criteria indicate strong multicollinearity. Since the test criterion $F_R=6.75$ is greater than the corresponding quantile of the Fisher-Snedecor *F*-distribution $F_{0.95}(5, 179)=2.15$, the proposed regression model is statistically significant. In contrast, the quantile of the Student *t*-distribution, $t_{0.975}(179\text{-}6)=1.974$ is greater than all $t_0=1.435$, $t_1=-1.790$, $t_5=-1.953$, $t_6=1.788$ but not than $t_2=2.141$, $t_3=-2.186$, and $t_4=2.097$; therefore, parameters β_0, β_1, β_5, and β_6 are statistically non-significant, while β_2, β_3, and β_4 are significant. It may be concluded that the OLS method is not convenient for parameter estimation in cases of strong multicollinearity in the data and that GPCR should be used instead.

In **step 4** a trial-and-error search of the most suitable value of the criterion parameter *P* using the mean quadratic error of prediction *MEP*, the Akaike information criterion *AIC* and the predicted determination coefficient R_P^2 was applied (Figure 4). The lowest *MEP* and *AIC* value and the highest value for R_P^2 is for $P=2.0\times10^{-4}$, which means that for $P>2.0\times10^{-4}$ all parameter estimates are statistically significant and therefore different to zero. While the OLS method, with $P=10^{-34}$, found the polynomial $y=810.2(564.8, \text{N})-247.9(138.5, \text{N})\ x+27.54(12.86, \text{S})\ x^2-1.26(0.58, \text{S})\ x^3+2.81(1.34, \text{S})\times10^{-2}\ x^4-2.99(1.53, \text{N})\times10^{-4}\ x^5+1.23(0.68, \text{N})\times10^{-6}\ x^6$ (where the parameter standard

Table 1 Search for the optimum power λ in the power transformation of the dependent variable y^{λ}.

	Search for the optimum power λ in the power transformation				
	−0.25	−0.20	−0.15	−0.10	−0.01
R	0.5544	0.5562	0.5569	0.5567	0.5542
$100R^2$ (%)	30.74	30.93	31.02	31.00	30.31
$100R^2_P$ (%)	50.79	51.04	51.18	51.22	51.03
MEP	7285.7	7220.2	7219.0	7280.4	7548.1
AIC	1593.6	1592.1	1592.2	1593.9	1600.6
$RSS(\lambda)\times10^{-6}$	1.2170	1.2070	1.2080	1.2190	1.2660
Skewness $g_1(e)$	−1.43	−1.21	−1.00	−0.82	−0.50
Kurtosis $g_2(e)$	7.42	6.32	5.42	4.73	3.89
Found λ can be	Not used	Not used	Used	Not used	Not used

The search for the optimum power λ in the power transformation of the dependent variable y^{λ} is based on the maximum of the curve $R_P^2=f(\lambda)$ for the age dependencies of pregnenolone sulfate in the serum of 179 males aged 4–69 years from Figure 2, calculated with the use of generalized principal component regression (GPCR) based on an examination of the following variables and test results: $n=179$, $m=6$, $P=0.0002$ (GPCR), $R=f(\lambda)$, $100R^2=f(\lambda)$, $100R_P^2=f(\lambda)$; for normal distribution, the skewness g_1 should be equal to zero and the kurtosis g_2 equal to 3.

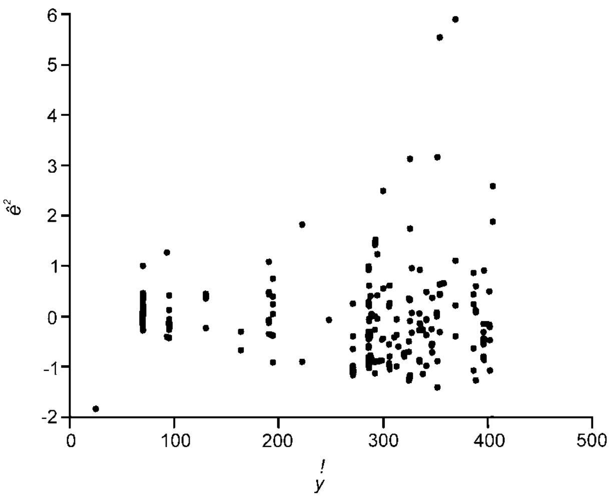

Figure 3 The sector pattern shape for the polynomial dependence of pregnenolone sulfate in serum of 179 men calculated with the ordinary least squares (OLS) analysis from Figure 1 proves heteroscedasticity and therefore indicates a need for data transformation.

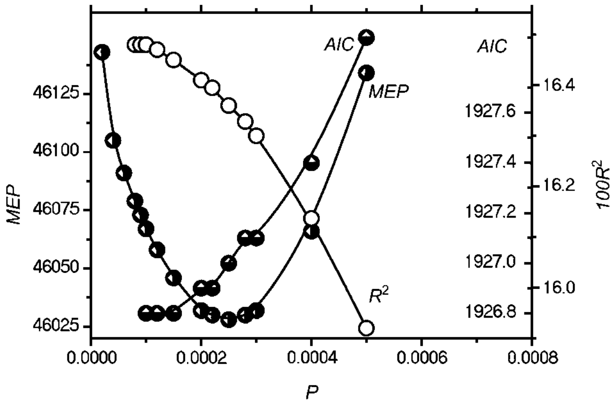

Figure 4 A search for the optimum value of the criterion *P* for the age dependencies of pregnenolone sulfate in the serum of 179 men calculated with generalized principal component regression (GPCR) from Figure 1 according to which the terms corresponding to small eigenvalues are omitted. The optimum value concerns the minimum on curve of the mean error of prediction *MEP* in dependence on the GPCR criterion *P*.

deviation is given in parentheses and N shows that the parameter estimate is statistically non-significant, while S shows that it is significant) with MEP=51192, AIC=1921.1 and $R_P^2\times100\%=14.91\%$, GPCR, with $P=2.0\times10^{-4}$ found $y=-395.8(155.1,\ S)+64.93(17.65,\ S)\ x-1.50(0.50,\ S)\ x^2-1.39(0.91,\ N)\times10^{-3}\ x^3+2.27(0.85,\ S)\times10^{-4}\ x^4+2.41(0.85,\ S)\times10^{-6}\ x^5-4.92(0.19,\ S)\times10^{-8}\ x^6$ with a lower value of the criterion MEP=46032, AIC=1926.9 and a higher value of $R_P^2\times100\%=34.75\%$. All parameters estimated by GPCR are statistically significant and are therefore acceptable, even though an excellent curve fitting was achieved in both cases.

To examine the normality of random error distribution in the dependent variable *y* and to find the most convenient variable transformation, the *RSS*(λ) for different values of power λ were sought out and the best power estimated (Table 1, Figure 5). Several

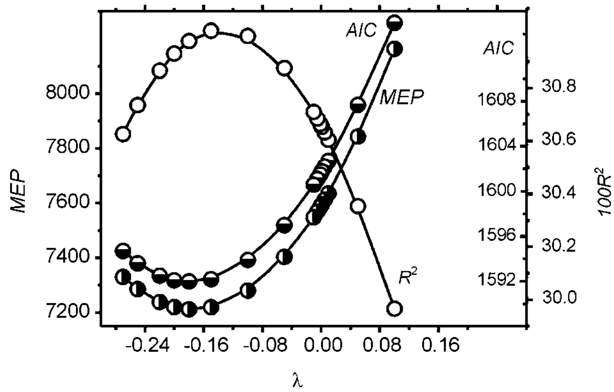

Figure 5 A search for the optimum power λ in the power transformation of the dependent variable y^λ based on the maximum of the curve $R_P^2=f(\lambda)$ for the age dependencies of pregnenolone sulfate in the serum of 179 men calculated with generalized principal component regression (GPCR) from Figure 2.

resolution criteria were applied to find the optimal power λ, but the most important was such λ for which the normality of residual distribution was achieved, i.e., for which the skewness g_1 is nearly zero and the kurtosis g_2 is nearly equal to 3. The resulting power was λ=–0.15. Figure 6 shows the scatter plot of the polynomial found (m=6) through transformed data and the criterion parameter P=0.0002, and a rankit *Q-Q* plot of jackknife residuals then proves a normal distribution and a homoscedasticity of residuals. The GPCR method with $P=2.0\times10^{-4}$ and using transformed data found $y^{(\lambda=-0.15)}=539.5(60.9,\ S)+36.58(6.93,\ S)\ x-0.81(0.20,\ S)\ x^2-1.03(0.36,\ S)\times10^{-3}\ x^3+1.19(0.34,\ S)\times10^{-4}\ x^4+1.29(0.33,\ S)\times10^{-6}\ x^5-2.54(0.75,\ S)\times10^{-8}\ x^6$ with the values of MEP=7219.0, AIC=1592.2 and $R_P^2\times100\%=51.12\%$. All parameters estimated by GPCR are statistically significant and are

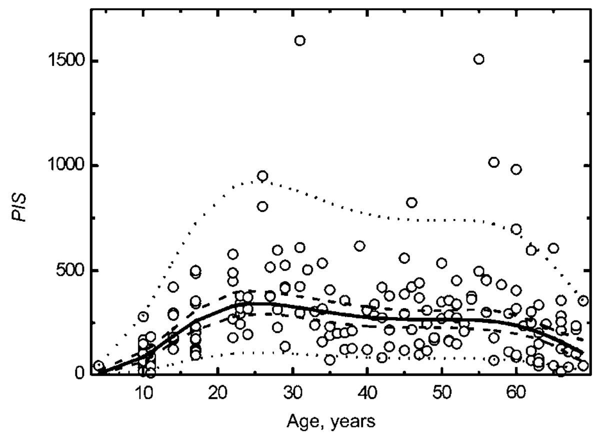

Figure 6 The scatter plot of the 6th polynomial regression of the age dependencies of pregnenolone sulfate (nmol/L) in the serum of 179 men from Figure 2 calculated with the use of the transformed dependent variable y^λ and generalized principal component regression (GPCR) analysis.
The curves of the mean prediction (full line) and the 95% Working-Hotelling interval bands of prediction (dashed lines) permit prediction of *y* outside the data interval. All of the parameters of the polynomial were statistically significant (*t*-tests).

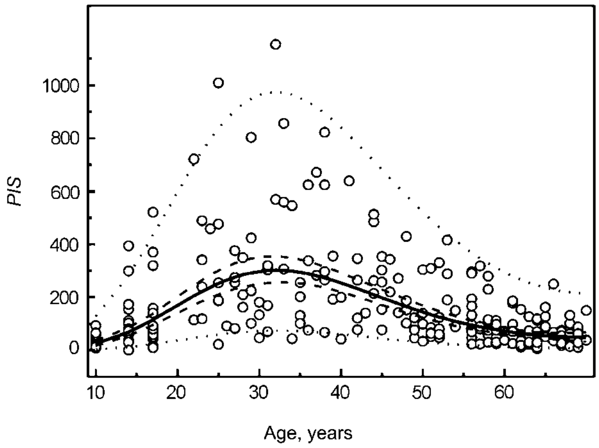

Figure 7 The scatter plot of the 4th polynomial regression of the age dependencies of pregnenolone sulfate (nmol/L) in the serum of 230 females aged 10–70 years calculated with the use of the transformed dependent variable y^{λ} and ordinary least squares (OLS) analysis.
The curves of the mean prediction (full line) and the 95% Working-Hotelling interval bands of prediction (dashed lines) permit prediction of y outside the data interval.

therefore acceptable, even though an excellent curve fitting was achieved in both cases.

Similarly, an analysis of female patients' data was conducted and the optimum degree of polynomial was found $m=4$ (Figure 7). As for the male data, a non-normality of random errors in dependent variable y and the heteroscedasticity of ordinary residuals may be proven. The best power found for power transformation was $\lambda=0.2$. When the GPCR method on transformed data was applied, the optimum polynomial ($m=4$) with parameter $P=0.0004$ results. While the OLS method (OLS with $P=10^{-35}$) found the polynomial $y=-254.3(211.7, \text{N})+27.0(29.5, \text{N})\ x+0.225(1.333, \text{N})\ x^2-2.17(2.40, \text{N})\times 10^{-2}\ x^3+2.00(1.50, \text{N})\times 10^{-4}\ x^4$ (where the parameter standard deviation is given in parentheses and N indicates that the parameter estimate is statistically non-significant, while S indicates that it is significant) with $MEP=24335$, $AIC=2324.3$ and $R_P^2\times 100\%=54.71\%$, GPCR, with $P=4.0\times 10^{-4}$ and $\lambda=+0.20$ found $y^{0.2}=0.661(0.217, \text{S})+0.148(0.017, \text{S})\ x-1.940(0.287, \text{S})\times 10^{-3}\ x^2-1.903(0.239, \text{S})\times 10^{-5}\ x^3+3.007(0.538, \text{S})\times 10^{-7}\ x^4$ with a lower value of $MEP=0.189$, $AIC=-380.0$ and a higher value of $R_P^2=62.54$. All parameters estimated by GPCR are statistically significant and therefore are acceptable, even though an excellent curve fitting was achieved in both cases. The ordinary residuals exhibit a normal distribution and an obvious homoscedasticity.

Comparison of age relationships in men and in women

The course of the age dependence in males differs from that in females (Figures 6 and 7). In women, a pronounced maximum after the 30th year was followed by a relatively rapid decline up to senescence, while in men the maximum after the 20th year was succeeded by a minor decline up to the 40th year and a plateau up to the 60th year followed by more a rapid decrease. As confirmed using two-way analysis of variance (ANOVA) [the details of statistical analysis with ANOVA were published elsewhere (15)] with sex and age group as the first and second factors, respectively, both the sex and the age differences were highly significant ($p<0.0001$), as were the differences in the shapes of the age dependencies (age/sex interactions).

Conclusions

GPCR in combination with the *MEP* criterion is extremely useful and appealing for constructing biased models. It can also be used for achieving such estimates with regard to keeping the model course corresponding to the data trend, especially in polynomial type regression models. In the search for the best degree of polynomial, several statistical characteristics of regression quality should be considered together. Significant differences were found between men and women in the course of the age dependence of PregS. In women, a significant maximum was found around the 30th year followed by a rapid decline, while the maximum in men was achieved almost 10 years earlier and changes were minor up to the 60th year. The investigation of gender differences and age dependencies in PregS could be of interest given its well-known neurostimulating effect, relatively high serum concentration, and the probable partial permeability of the blood-brain barrier for the steroid conjugate, which is reflected, e.g., in the correlation of complaints by patients suffering from premenstrual syndrome with serum levels of the conjugate. With regard to the method of data analysis, principal component regression is an extremely useful tool for the investigation of curvilinear dependencies, especially in polynomial regression models.

Acknowledgements

Financial support from the Grant Agency of the Czech Republic (grant no. GAČR 303/06/1817) is gratefully acknowledged.

References

1. Murray HE, Gillies GE. Differential effects of neuroactive steroids on somatostatin and dopamine secretion from primary hypothalamic cell cultures. J Neuroendocrinol 1997;9:287–95.
2. Wakerley JB, Richardson CM. The neurosteroid pregnenolone sulphate enhances NMDA-induced phasic firing of vasopressin neurons in the rat supraoptic nucleus. Neurosci Lett 1997;226:123–6.
3. Mathis C, Vogel E, Cagniard B, Criscuolo F, Ungerer A. The neurosteroid pregnenolone sulfate blocks deficits induced by a competitive NMDA antagonist in active avoidance and lever-press learning tasks in mice. Neuropharmacology 1996;35:1057–64.

4. Monnet FP, Mahe V, Robel P, Baulieu EE. Neurosteroids, via Sigma receptors, modulate the [3H]norepinephrine release evoked by N-methyl-D-aspartate in the rat hippocampus. Proc Natl Acad Sci USA 1995;92:3774–8.
5. Flood JF, Morley JE, Roberts E. Memory-enhancing effects in male mice of pregnenolone and steroids metabolically derived from it. Proc Natl Acad Sci USA 1992;89:1567–71.
6. Flood JF, Morley JE, Roberts E. Pregnenolone sulfate enhances post-training memory processes when injected in very low doses into limbic system structures: the amygdala is by far the most sensitive. Proc Natl Acad Sci USA 1995;92:10806–10.
7. Majewska MD, Demirgoren S, London ED. Binding of pregnenolone sulfate to rat brain membranes suggests multiple sites of steroid action at the GABAA receptor. Eur J Pharmacol 1990;189:307–15.
8. Wang MD, Wahlstrom G, Backstrom T. The regional brain distribution of the neurosteroids pregnenolone and pregnenolone sulfate following intravenous infusion. J Steroid Biochem Mol Biol 1997;62:299–306.
9. Wang M, Seippel L, Purdy RH, Backstrom T. Relationship between symptom severity and steroid variation in women with premenstrual syndrome: study on serum pregnenolone, pregnenolone sulfate, 5 alpha-pregnane-3,20-dione and 3 alpha-hydroxy-5 alpha-pregnan-20-one. J Clin Endocrinol Metab 1996;81:1076–82.
10. Corpechot C, Collins BE, Carey MP, Tsouros A, Robel P, Fry JP. Brain neurosteroids during the mouse oestrous cycle. Brain Res 1997;766:276–80.
11. Rajkowski KM, Robel P, Baulieu EE. Hydroxysteroid sulfotransferase activity in the rat brain and liver as a function of age and sex. Steroids 1997;62:427–36.
12. de Peretti E, Mappus E. Pattern of plasma pregnenolone sulfate levels in humans from birth to adulthood. J Clin Endocrinol Metab 1983;57:550–6.
13. de Peretti E, Forest MG, Loras B, Morel Y, David M, Francois R, et al. Usefulness of plasma pregnenolone sulfate in testing pituitary-adrenal function in children. Acta Endocrinol Suppl 1986;279:259–63.
14. Bičíková M, Tallová J, Hill M, Krausová Z, Hampl R. Serum concentrations of some neuroactive steroids in women suffering from mixed anxiety-depressive disorder. Neurochem Res 2000;25:1623–7.
15. Havliková H, Hill M, Hampl R, Stárka L. Sex- and age-related changes in epitestosterone in relation to pregnenolone sulfate and testosterone in normal subjects. J Clin Endocrinol Metab 2002;87:2225–31.
16. Meloun M, Militký J, Forina M. Chemometrics for analytical chemistry, vol. 2. PC-aided regression and related methods. Chichester: Ellis Horwood, 1994.
17. Meloun M, Militký J, Hill M, Brereton RG. Crucial problems in regression modelling and their solutions. Analyst 2002;127:433–50.
18. Meloun M, Hill M, Militký J, Vrbíková J, Stanická S, Škrha J. New methodology of influential point detection in regression model building for the prediction of metabolic clearance rate of glucose. Clin Chem Lab Med 2004;42:311–22.
19. Belsey DA, Kuh E, Welsch RE, editors. Regression diagnostics: identifying influential data and sources of collinearity. New York: Wiley, 1980.
20. Rawlings JO, Pantula SG, Dickey DA. Applied regression analysis. New York: Springer Verlag, 1998.
21. Kupka K. ADSTAT. Pardubice: Trilobyte Statistical Software, 1999.