Eur J Epidemiol (2009) 24:297–306
DOI 10.1007/s10654-009-9336-9

CARDIOVASCULAR DISEASE

Nonlinear association between serum testosterone levels and coronary artery disease in Iranian men

Nader Fallah · Kazem Mohammad · Keramat Nourijelyani · Mohammad Reza Eshraghian · Seyyed Ali Seyyedsalehi · Maria Raiessi · Maziar Rahmani · Hamid Reza Goodarzi · Soodabeh Darvish · Hojjat Zeraati · Gholamreza Davoodi · Saeed Sadeghian

Received: 29 June 2008 / Accepted: 23 March 2009 / Published online: 9 April 2009

Abstract Previous studies have shown controversial results about the role of androgens in coronary artery disease (CAD). We performed this study to examine and compare the relationship between androgenic hormones and CAD using conventional linear statistical techniques as well as novel non-linear approaches. The study was conducted on 502 consecutive men who were referred for selective coronary angiography at Tehran Heart Center due to different indications. We studied the relationship between androgenic hormones and CAD by using the generalized linear models, generalized additive models, and neural networks. Free testosterone (fT), total testosterone (tT) and dehydroepiandrosterone sulfate levels in patients with significant CAD versus normal individuals were 6.69 ± 3.20 pg/ml, 16.60 ± 6.66 nm/l, and 113.38 ± 72.9 μg/dl versus 7.12 ± 3.58 pg/ml, 15.82 ± 7.26 nm/l, and 109.03 ± 68.19 μg/dl, respectively ($P > 0.05$). The Generalized linear models was unable to show any significant relationship between androgenic hormones and CAD, while generalized additive model and neural networks supported the significant effect of androgenic hormones on CAD. This finding suggests a nonlinear association of tT levels with CAD: lower levels have a preventive effect on CAD, whereas higher values increase the risk of CAD. Emphasizing the non-linearity of the variables may provide new insight into the possible explanation of the effect of androgenic hormones on CAD.

N. Fallah · K. Mohammad (✉) · K. Nourijelyani · M. R. Eshraghian · H. Zeraati
Department of Epidemiology and Biostatistics, University of Tehran/Medical Sciences, Tehran, Iran
e-mail: mohamadk@tums.ac.ir

N. Fallah
e-mail: nfallah@mathstat.dal.ca

K. Nourijelyani
e-mail: nouri4@yahoo.com

M. R. Eshraghian
e-mail: eshraghian@yahoo.com

H. Zeraati
e-mail: zeraatih@tums.ac.ir

S. A. Seyyedsalehi
Department of Biomedical Engineering, Amirkabir University of Technology, Tehran, Iran
e-mail: ssalehi@cic.aut.ac.ir

M. Raiessi · H. R. Goodarzi · G. Davoodi · S. Sadeghian
Tehran Heart Center, University of Tehran/Medical Sciences, Tehran, Iran

M. Raiessi
e-mail: mari_d81@yahoo.com

H. R. Goodarzi
e-mail: hami_nag@yahoo.com

G. Davoodi
e-mail: ghdavoodi@hotmail.com

S. Sadeghian
e-mail: sadeghian15@yahoo.com

M. Rahmani
Canada's Michael Smith Genome Sciences Center, British Columbia Cancer Research Center, Vancouver, BC, Canada
e-mail: mrahmani@bcgsc.ca

S. Darvish
Department of Genecology and Obstetrics, University of Tehran/Medical Sciences, Tehran, Iran
e-mail: s_darvish_n@yahoo.com

Keywords Androgens · Risk factors · Coronary artery disease · Angiography · Nonlinear models · Neural network

Abbreviations

AIC	Akaike information criteria
ANOVA	Analysis of variance
BMI	Body mass index
BIC	Bayesian information criteria
CAD	Coronary artery disease
CHOL	Cholesterol
CRP	C-reactive protein
DHEAS	Dehydroepiandrosterone sulfate
DM	Diabetes mellitus
EF	Ejection fraction
ELISA	Enzyme-linked immunosorbent assay
fT	Free testosterone
GLM	Generalized linear models
HDL	High density lipoprotein
HTN	Hypertension
LDL	Low density lipoprotein
Lp(a)	Lipoprotein(a)
MLP	Multi-layer perceptron
MSE	Mean square error
ROC	Receiver operating characteristic
SCA	Selective coronary angiography
SD	Standard deviation
SLP	Single layer perceptron
TC	Total cholesterol
TGs	Triglycerides
tT	Total testosterone
edf	Equally degree of freedom

Introduction

Despite wide variations in absolute rates of cardiovascular diseases, global statistics reveal that in different populations, men are 2.2 times (range, 1.2–4.5) as likely as women to develop cardiovascular disease [1]. This has been largely attributed to the established cardioprotective role of endogenous estrogens [2–6]. Likewise, testosterone has been proposed to have a protective role for the cardiovascular system in men. In vitro studies have demonstrated that testosterone may exert anti-atherogenic effects through different mechanisms [7–12]. However, despite the established role of testosterone in vascular reactivity and coronary artery disease (CAD) risk profiles, its effect on CAD has not been fully identified and present reports are contradictory. Some reports have suggested testosterone to have a neutral or even protective effect on cardiovascular system [13]. While other reports have shown that use of anabolic steroids may be responsible for sudden cardiac death, myocardial infarction, and hypertension [14].

In our previous study, we found no association between androgen levels and CAD, using the linear models [15]. In the present study, we used generalized linear model, generalized additive model, and neural networks as some alternative techniques to investigate association between serum androgen levels with presence and severity of CAD. Furthermore, we indicated how not using such models may result a misleading conclusions. The central idea of the additive model is to replace the usual linear function of a covariate with a smoother function. The additive model includes the sum of several appropriately chosen functions. Neural networks, on the other hand, represent a broad class of nonlinear models capable of learning from data. These are classes of nonparametric regression models which in fact relax the traditional regression assumptions of linearity for examining nonlinear relationships between variables of interest.

In this study, after using some ordinary parametric and nonparametric methods with careful controlling for the effect of other confounder, we examined generalized additive models and neural networks to assess the association between serum levels of testosterone and CAD in a group of Iranian male patients. The association between androgens and severity of CAD were also studied by means of generalized additive logistic regression, generalized additive Poisson regression and ordinary additive models. Additionally, we used nonlinear ordinary regression and nonlinear logistic regression based on neural networks (multilayer perceptron). Finally, we defined and applied a new model for nonlinear Poisson regression based on multilayer perceptron [16–20].

Patients and methods

Design and study subjects

Between May and August 2005, out of 719 consecutive male patients undergoing selective coronary angiography (SCA) at Tehran Heart Center, a total of 502 were selected. Patients were referred for coronary angiography for typical chest pain, positive exertion stress test, positive myocardial perfusion scan, and history of myocardial infarction. Patients with chronic renal failure requiring dialysis or those on any medications known to affect sex hormone levels, including antiandrogen treatment for prostatic carcinoma or androgens preparations, were excluded. Since C-reactive protein (CRP) is measured simultaneously, patients with fever, major trauma, or infections, and myocardial infarction within the preceding 14 days were also excluded from the study because of the possible influence of these conditions on CRP levels. Informed, written consent was obtained from all participants. The

study was approved by the Tehran Heart Center Ethics Committee.

Clinical and lifestyle characteristics

Clinical data was collected using a common protocol on a preceded standardized questionnaire administered one day prior to catheterization. Baseline clinical data collected at this time included age, sex, smoking history, history of previous CAD, stroke, hypertension and diabetes mellitus (DM), taking medications including beta blockers and diuretics or alcohol. Hypertension was defined as having a history of high blood pressure, as confirmed by a physician, or as being on antihypertensive drugs. A current smoker was defined as a patient who was smoking ($\geq$1 cigarette/day) during the preceding month of the study. Weight and height were measured using a standard scale. Body mass index (BMI) was calculated by the following formula: BMI = weight (kg)/height^2 (m). DM was defined as a fasting blood glucose $\geq$140 mg/dl or a diagnosis of diabetes needing dietary intervention or drug therapy.

Measurement of androgens and biochemical variables

Blood specimen was collected after overnight fasting for at least 12 h on the day of catheterization, prior to coronary angiography between 8:30 and 9:30 a.m., and before injection of any contrast material or heparin. Blood samples were sent to the research laboratory of Tehran Heart Center where all laboratory measurements were performed. Serum was isolated by centrifugation (2,500 rpm, 30 min, 4°C) and multiple aliquots were frozen at −80°C for measurement of biochemical variables. Patients were assessed for triglycerides (TGs), total cholesterol (TC), low density lipoprotein (LDL-C) and high density lipoprotein (HDL-C), serum free testosterone (fT), total testosterone (tT), dehydroepiandrosterone sulfate (DHEAS), Lipoprotein(a) [Lp(a)], and CRP. Sample analyses for lipids and lipoproteins were performed using Selectra 2 auto-analyzer (Vital Scientific, Spankeren, Netherlands); TC and TGs kits (Pars Azmon Inc., Tehran, Iran) were used. TC and TGs were assayed using enzymatic colorimetric tests with cholesterol esterase and cholesterol oxidase and glycerol phosphate oxidase, respectively. HDL-C was measured after precipitation of the apo B-containing lipoproteins with phosphotungistic acid. LDL-C was calculated using Friedewald formula [21]. LDL-C was not calculated when TGs concentration was $\geq$400 mg/dl. Lp(a) was measured with enzyme-linked immunosorbent assay (ELISA; Biopool, Sweden). C-reactive protein (CRP) was measured using highly sensitive immunoenzymometric assays (Diagnostic Biochem, Toronto, Canada). The analysis of the samples was performed by Sunrise ELISA reader (TECAN Co, Salzburg, Austria). Free testosterone was measured by ELISA, and tT and DHEAS were measured by radioimmunoassay (Immunotech, Marseille, France).

Angiographic assessment

Coronary angiograms were obtained with the Judkins percutaneous retrograde femoral artery technique [22]. Two cardiologists without prior knowledge of patients' conditions and laboratory results reviewed all films. Coronary artery involvement was defined as $\geq$50% diameter obstruction of a major coronary vessel. Further, the lesions were classified according to Gensini Score [23]. In addition, number of involved vessels with lesion more than 50% considered as count data: one-vessel, two-vessel, and three-vessel disease [24]. Also here as a hospital based sampling we have not zero inflated problem. The angiography films were re-reported by the cardiologists, which yielded an intra- and inter-observer reliability of more than 95%.

Statistical analysis

Data were analyzed using SPSS software version 15, R software version 2.5, and MTLAB version 7.2. To analyze the data, we used Chi-square (or Fisher's exact tests, if required) for categorical variables, and Student's *t* test for continuous variables. One way ANOVA was also used for comparison of hormone levels between patients with single-, two-, and three-vessel disease. In addition, Pearson's correlation tests were used to find any correlation between continuous variables. Later, multiple linear regression, logistic regression, and Poisson regression were used to eliminate the effect of confounders. Generalized additive regression, generalized additive logistic regression, generalized additive Poisson regression models were used to investigate the nonlinearity relation between CAD and androgenic hormones concentration, with control for probable confounders. Finally, for more investigation of nonlinearity we used neural networks in the above three regression [16–20, 25–30]. More technical details in this regard can be found in Appendices 1 and 2. The *P* value $<$0.05 was considered to be statistically significant and the mean values of continuous variables are reported along with their $\pm$ standard deviations (SD).

Results

From 502 patients, 83 (16.5%) had single-, 108 (21.5%) two-, and 197 (39.3%) three-vessel disease. The remaining 114 patients (22.7%) had normal angiogram or minimal lesions. In addition, 27 (5.4%) of all patients with CAD had a left main stem lesion. Table 1 represents the baseline

characteristics and Table 2 represents the concentration of androgenic hormones, CRP, lipids, and lipoproteins in two groups (with and without significant CAD). As shown in Table 2, in men with CAD, the concentration of androgens is similar to those without significant CAD. In correlation studies, fT showed a weakly negative correlation with CRP ($P = 0.03$, $r = -0.096$) and Lp(a) ($P = 0.01$, $r = -0.12$). Moreover, age had a strong negative correlation with DHEAS ($P < 0.001$, $r = -0.33$). However, TGs, TC, LDL-C, and HDL-C did not correlate with tT, fT, and DHEAS. Furthermore, we did not find significant correlation between serum androgen levels with Gensini score as determinant of severity of CAD ($P > 0.05$). Serum concentrations of DHEAS, fT, and tT did not differ between patients with normal coronary, single-, two-, and three-vessel disease.

Based on the measure of our dependent variables, we considered a nominal scale for CAD, a count scale for number of vessel, and an interval scale for Gensini score. Then, we used logistic, Poisson, and multiple linear regression models, respectively. In the multivariable regression model, subjects were adjusted for these variables: age, hyperlipidemia, DM, Lp(a), smoking, TC, and LDL-C. The influence of androgen levels on response variables were assessed after adjustment for these variables. However, there were no associations between androgenic hormones and the presence and severity of CAD (Table 3).

In the next step, we used generalized additive regression with Spline smooth function to investigate the correlation of response variables and androgen levels after controlling for confounding risk factors. Table 4 shows the results of the above fitted models, indicating that CAD and vessel scores are not significantly related with serum androgenic hormone levels (P value > 0.05); however, based on generalized additive regression, Gensini score and testosterone

Table 1 Baseline characteristics in patients with coronary artery disease and normal coronary

	CAD[a]	NCAD[b]	P value
Age (years)	57.0 ± 10.5	52.9 ± 11.4	<0.001
Diabetes mellitus	88 (23.1%)	12 (11%)	0.006
Hypertension	121 (31.8%)	20 (18.2%)	0.006
Hyperlipidemia	147 (38.6%)	20 (18.3%)	<0.001
Current smoking	125 (32.8%)	24 (21.8%)	0.027
Ejection fraction (%)	50 ± 11%	56 ± 10%	<0.001
Body mass index (Kg/m^2)	28.05 ± 4.37	28.75 ± 4.75	0.14

Plus minus values are mean ± SD

[a] Patients with coronary artery disease

[b] Patients with normal coronary arteries

Table 3 Results of using logistic, Poisson and multiple regression between hormone level and CAD, VESSEL and GENSINI score

	Logistic regression		Poisson regression		Multiple regression	
	Estimate	P value	Estimate	P value	Estimate	P value
Intercept	−0.4382	0.80	−0.215	0.63	41.49	0.11
LDL	0.0132	0.07	0.002	0.26	0.03	0.728
CHOL	−0.0004	0.94	0.001	0.59	0.18	0.051
CRP	0.0086	0.58	0.001	0.83	0.36	0.05
LPA	0.0084	0.46	0.002	0.46	0.01	0.96
AGE	0.0451	0.001	0.013	0.001	0.81	0.001
FREET	−0.0338	0.51	−0.001	0.93	−0.29	0.72
TEST	0.0497	0.06	0.006	0.31	0.43	0.27
BMI	−0.0431	0.16	−0.007	0.38	−0.87	0.07
EF	−0.0495	0.001	−0.001	0.001	−1.33	0.001
HLP	1.3102	0.001	0.23	0.005	9.00	0.07
DM	0.4148	0.34	0.09	0.30	2.44	0.67
SMOKE	1.0930	0.001	0.24	0.005	4.08	0.42
HTN	0.3888	0.24	−0.001	0.99	0.11	0.98

Table 2 Hormone profile and other laboratory tests in study populations

	CAD[a]	NCAD[b]	P value
Free testosterone (pg/ml)	6.69 ± 3.20	7.12 ± 3.58	0.22
Total testosterone (nm/l)	16.60 ± 6.66	15.82 ± 7.26	0.28
Dehydroepiandrosterone sulfate (μg/dl)	113.38 ± 72.9	109.03 ± 68.19	0.57
Lipoprotein a (mg/dl)	19.95 ± 14.14	16.34 ± 11.43	0.013
C-Reactive protein (mg/dl)	11.52 ± 12.90	9.57 ± 6.65	0.12
Total cholesterol (mg/dl)	203.00 ± 46.00	190.35 ± 49.87	0.012
Low-density lipoprotein cholesterol (mg/dl)	129.43 ± 38.75	116.42 ± 39.12	0.002
High- density lipoprotein cholesterol (mg/dl)	37.84 ± 9.47	37.54 ± 9.33	0.77
Triglyceride (mg/dl)	186.33 ± 101.38	177.19 ± 144	0.45

Plus minus values are mean ± SD

[a] Patients with coronary artery disease

[b] Patients with normal coronary arteries

levels show a significant relationship (P value < 0.05). In three models, all risk factors were controlled as confounders. Figure 1 suggests a nonlinear effect of some independent variables on dependent variables, especially hormone levels on CAD.

Table 4 Results of fitting separate generalized additive regression between hormone level and CAD, VESSEL and GENSINI score as dependent variables

	Generalised additive					
	Logistic regression		Poisson regression		Multiple regression	
	Estimate	*P* value	Estimate	*P* value	Estimate	*P* value
(Intercept)	1.084	<0.001	0.375	<0.001	44.95	<0.001
HLP	1.321	0.001	0.247	0.003	8.373	0.088
DM	0.472	0.32	0.092	0.33	3.409	0.550
SMOKE	1.229	0.001	0.232	0.007	3.443	0.493
HTN	0.264	0.47	−0.035	0.68	−2.186	0.663
	edf	*P* value	edf	*P* value	edf	*P* value
s(LDL)	4.380	0.03	2.09	0.14	1.00	0.37
s(CHOL)	1.001	0.72	1.00	0.73	1.59	0.22
s(CRP)	1.000	0.61	1.00	0.90	1.60	0.09
s(LPA)	1.341	0.66	4.39	0.23	1.00	0.83
s(AGE)	3.513	0.001	1.78	0.002	1.80	0.002
s(FREET)	1.000	0.38	1.00	0.79	1.00	0.71
s(TEST)	2.877	0.05	3.08	0.19	3.97	0.005
s(BMI)	6.781	0.06	1.00	0.28	1.00	0.05
s(EF)	6.201	0.002	2.46	0.005	4.04	0.001

Testosterone hormone showed a preventive effect in low level values and a risk factor effect in high levels. Based on the testosterone curve, we divided testosterone level into 3 groups; low (0 to 7), medium (7.1 to 22), and high (22.1 to 43.3). The chi-square test showed a significance statistical difference ($P < 0.05$). The odds ratio for low and high level of testosterone is 0.48 and 1.32, respectively for medium group as a baseline; this proves the hypothesis about preventive and risk factors role of testosterone. Interestingly fT did not show significant correlation with the presence of CAD in this study.

To further investigate the nonlinearity nature of variables, we used three models of neural networks. We randomly divided half of our data for training, and the rest for testing. In each model, we used Akaike (AIC) and Bayesian (BIC) information criteria as primary measures of fit. We used six schemes on neural networks topology with 2, 3, 4, 5, 10, and 20 nodes in hidden layer and the momentum rate was 0.012. Based on the least AIC and BIC, we chose a multilayer perceptron with two nodes in hidden layer as the best model (Table 5). For an accuracy rate of forecasting in logistic regression, we used the misclassification index and receiver operating characteristic (ROC). Furthermore, we used the mean square error (MSE) for the multiple and Poisson regression methods. Table 5 shows that in each model we have nonlinearity in data, and the neural networks prediction showed different

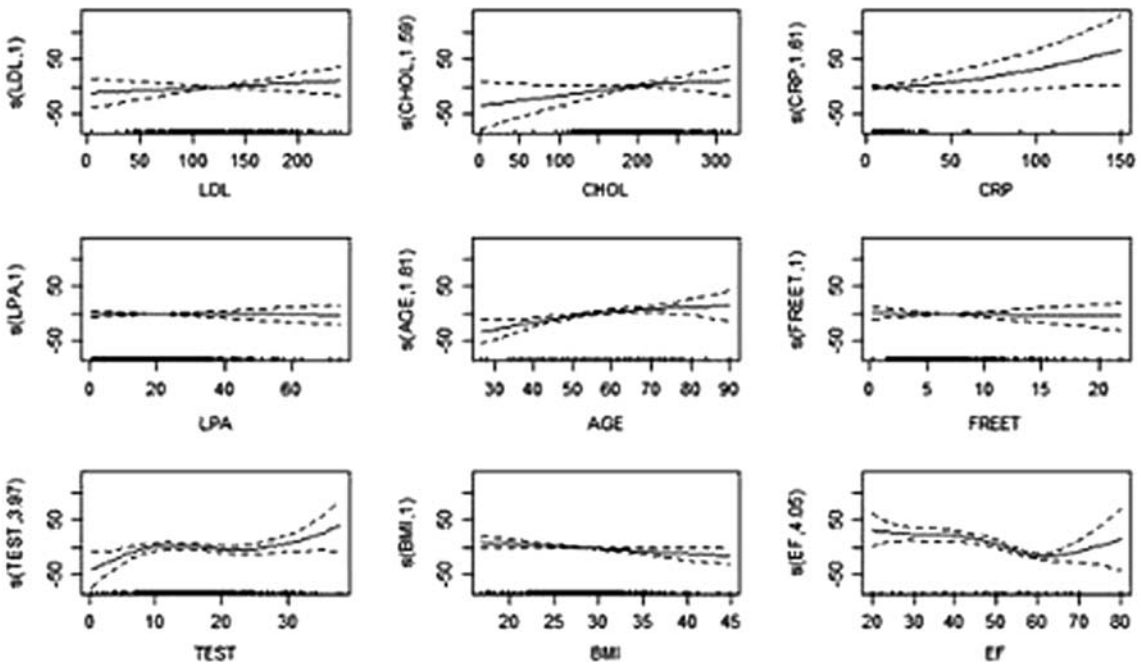

Fig. 1 Generalized additive multiple regression graphs for LDL (low density lipoprotein), CHOL (Cholesterol), CRP (C-reactive protein), LPA (lipoprotein A), AGE, FREET (free testosterone), TEST (testosterone), BMI (body mass index), and EF (ejection fraction). *Solid line* represents the point estimation; *dotted lines* represent upper and lower 95% confidence bands

Table 5 Results of fitting separate regression models between testosterone level and CAD VESSEL and GENSINI score and neural networks (NN)

	# parameter	AIC	BIC	Misclassification	ROC	MSE
Logistic regression	2	314	300	0.15	0.50	–
Poisson regression	2	805.9	790	–	–	1.39
Multiple regression	2	711.5	694	–	–	1.06
Logistic regression by NN	7	318.3	343.2	0.10	0.66	–
Poisson regression by NN	7	436.4	461.5	–	–	1.32
Multiple regression by NN	7	252.4	277.1	–	–	0.96

improvement of prediction level in each of the three models.

Discussion and conclusion

Previous studies have shown controversial results about the role of androgens in CAD. In our previous study, we showed no relationship between androgenic hormone levels and the presence of CAD [15]. We performed this study to investigate the possibility of relationship between androgen levels and presence and severity of CAD, focusing on nonlinear relationship between variables. Using ordinary linear regression models, we did not find any significant association between androgen levels and CAD, even after controlling for confounders. Table 3 indicates the results of widely used models, namely the logistic, Poisson and multiple linear regression models, to assess the effect of androgen levels on CAD. We controlled for confounding risk factors, these models did not reveal any significant association between androgen levels and CAD. However, using the more complicated model of generalized additive regression with Spline smooth function enabled us to detect a nonlinear effect of the covariates. We showed that, when other risk factors are controlled in a generalized additive regression models (Table 4), testosterone level is significantly related to CAD ($P < 0.05$). Although no significant association between the Poisson and logistic model were detected. Figure 1 suggested a nonlinear effect of testosterone on CAD, when lower levels indicate preventive roles, it appears that for greater values, the risk of CAD continuously increase as testosterone levels rise. We considered the cut point in testosterone levels intuitively, but some other techniques such as CART [18] or two segmented logistic regression [25] was also applied to determine the exact cut point. These techniques could be considered as motivation for further work. In addition, for a more thorough investigation, we used the neural networks technique in each of the three classical models, confirming the nonlinear nature of our data.

The nonlinear modeling procedures described here are useful for two reasons; first, they help to prevent model misspecification, which can lead to incorrect conclusions regarding treatment efficacy; second, they provide information about the relationship between prognostic factors and disease risk that are not revealed by the use of standard modeling techniques. Linearity is always a special case, thus, simple linear relationships can be easily confirmed by flexible modeling of covariate effects, similar to some previous research. These findings may provide new evidence of an association between androgen levels and CAD. We believe that our study contributes to a better understanding of the relationship of the severity of CAD and androgen levels by using previously undiscussed applications of generalized additive models and neural networks. Furthermore, we introduce a new nonlinear Poisson regression by neural networks. We suggest that reanalysis of previous data in this field using these methods, may achieve different results and we recommend that future researchers investigate the nonlinearity nature of variables using a generalized additive model, neural networks or other nonlinear models.

Acknowledgments This study was sponsored by Tehran University/Medical Sciences. Part of this work was carried out during a sabbatical year (student visitor period) of Nader Fallah at the Mathematics and Statistics Department of Dalhousie University and University of British Columbia. Authors wish to thank R. Ripley for her help and I. Nabney for his Matlab Code. The authors thank Janet Brush, Parveer Pannu, and Catherine Pretty for editing this manuscript.

Appendix 1

Generalized additive models

Generalized additive models and generalized linear models can be applied in similar situations, but they serve different analytic purposes. Generalized linear models emphasize estimation and inference for the parameters of the model, while generalized additive models focus on exploring data non-parametrically. Therefore, this model could find nonlinearity relation between predictor and response variables. Generalized additive models permit the response probability distribution to be any exponential family of distributions. Many widely used statistical models belong to this general class, including additive models for Gaussian data, binary data, and nonparametric log-linear models for Poisson data.

Suppose that Y is a response random variable and X_1, …, X_p is a set of predictor variables. A regression procedure can be viewed as a method for estimating how the value of Y depends on the values of X_1, …, X_p. Given a sample of values for Y and X, estimates of β_0, β_1, …, β_n are often obtained by the least squares method. In regression models the effects of prognostic factors x_i in terms of a linear predictor of the form $\sum x_j\beta_j$, where the β_j are parameters. The generalized additive model replaces $\sum x_j\beta_j$ with $\sum f_j(x_j)$ where f_j is a unspecified (non-parametric) function. This function is estimated in a flexible manner using a scatter plot smoother. The estimated function $\hat{f}_j(x_j)$ can reveal possible non linearity in the effect of the x_j. Suppose y is a response or outcome variable, and x is a prognostic factor. We wish to fit a smooth curve $f(x)$ that summarizes the dependence of y on x. If we were to find the curve that simply minimizes $\sum (y_i - f(x_i))$, the result would be an interpolating curve that would not be smooth at all. The cubic spline smoother imposes smoothness on $f(x)$. We seek the function $f(x)$ that minimizes

$$\sum (y_i - f(x_i))^2 + \lambda \int f''(x)^2. \tag{1}$$

Notice that $\int f''(x)^2$ measures the "Curvature'' of the function f: λ is a non-negative smoothing parameter that must be chosen by the data analyst. The λ has a direct relation with degree of freedom. Larger values of λ force f to be smoother. More detailed have been presented elsewhere [18].

Appendix 2

Artificial neural networks

Neural network

A neural network is a set of interconnected simple processing elements called neurons, where each connection has an associated weight. The *neuron* or *unit* processes its inputs to create an output. The network consists of a number of input units representing the predictors, one or more output units corresponding to the predicted variables and possibly some internal units to increase the model complexity or flexibility. The weights associated with the interconnections between the units will be optimized in fitting the model to the data. The most commonly used form of neural network is the multi-layer perceptron (MLP). A MLP consists of one input layer of units, one output layer of units and possibly one or more layers of 'hidden' units. The input units pass their inputs to the units in the first hidden layer or directly to the output units. Each of the hidden layer units adds a constant (termed as 'bias') to a weighted sum of its inputs and calculates an activation function ϕ_h of the result. This is then passed to hidden units in the next layer or to the output unit(s). The activation function is usually chosen in advance. Common choices of the activation function include the logistic function, tangent hyperbolic function or other monotonic functions. In this paper we fix the activation function as tangent hyperbolic function. The output units apply a linear, logistic, thresholds or other function ϕ_0 to the weighted sum of their inputs plus its 'bias'. In this paper we use exponential function in output layer.

Denote the inputs as x_i's and the outputs t_k's, for MLP with one hidden layer

$$t_k = \phi_0\left(\alpha_k + \sum_{j\to k}\omega_{jk}\phi_h\left(\alpha_j + \sum_{i\to j}\omega_{ij}x_i\right)\right). \tag{2}$$

If we have only one output node, k will be equal to one. The weights can be determined by optimizing some proper criterion function such as minimizing the sum of squared errors of the predicted variable or maximizing the log-likelihood of the data in cases where a distribution of the response variable can be assumed. The structure of MLP made it possible to fit very general non-linear functional relationships between inputs and outputs. Research results have shown that neural networks with enough hidden units can approximate any arbitrary functional relationships [26, 27]. However, over-fit can be a serious problem in such a framework. This problem usually is overcome either by stopping the optimization early or more often by using *regularization* techniques to penalize the optimization criterion. By adding a penalty term to the optimization criterion, the estimates of the weights will be shrunk which is also termed as shrinkage method. The following smoothness penalty is often used in shrinkage method:

$$L = -\log \text{likelihood} + \lambda \sum_{\text{weights}} \omega_{ij}^2. \tag{3}$$

This process is also known as *weight decay* in neural network literatures. The tuning parameter λ can be chosen by cross-validation. For fixed number of hidden units, we minimize this penalized log-likelihood in (3) to get the weights estimated. To control the complexity of the model due to the number of the hidden units, criteria such as AIC and BIC are used.

Optimization criteria

Given a training set comprising a set of input vectors$\{x_n\}$, where $n = 1, \ldots, N$, together with the corresponding target vector $\{y_n\}$, if we assume that data points y_n ($n = 1, \ldots, N$) are independent conditional on x_n, the likelihood function can be written as:

$$P(y|x) = \prod_{n=1}^{N} p(y_n|x_n) \tag{4}$$

or

$$P(y_1, \ldots, y_N|x_1, \ldots, x_N) = \prod_{n=1}^{N} p(y_n|x_n).$$

The error function can be defined as the negative log-likelihood:

$$E = -\log P(y_1, \ldots, y_N|x_1, \ldots, x_N) = -\sum_{n=1}^{N} \log p(y_n|x_n). \tag{5}$$

Linear and logistic regression

For regression problems with normality assumption, this can be reduced to the most commonly used squared error criterion:

$$E(w) = \frac{1}{2}\sum_{n=1}^{N} \{y_n - t_n(x_n; w)\}^2. \tag{6}$$

For classification problems, it is often advantageous to associate the network outputs to the posterior probabilities of each class. For a problem with two classes (such as normal and CAD), the target variable $\{y_n\}$ is binary and can be assumed to follow binomial distribution with its probability as $t_n(x_n; w)$. The error function in (5) then yields the cross-entropy error function:

$$E = -\sum \{y_n \ln t_n + (1 - y_n) \ln(1 - t_n)\}. \tag{7}$$

This definition extended to other generalized linear models (GLM) by other researcher such as in multinomial logistic regression and ordinal logistic regression or Cox regression for survival models [16–20, 26–30]. We will consider the Poisson regression in the following.

Poisson regression

Suppose we have a single target variable with count response, we consider the non-linear Poisson regression for neural networks as an extension of generalized linear models. It seems this model has not been introduced in literatures before.

The Poisson probability distribution for count data is given by:

$$P[Y_n = y_n] = \frac{e^{-\lambda_n}\lambda_n^{y_n}}{y_n!}, y_n = 0, 1, 2, \ldots. \tag{8}$$

In linear Poisson regression, the most commonly used formulation is the log-linear link function: $\ln \lambda_n = x_n'\beta$. Thus the expected value for y_n is given by $E[y_n|x_n] = \lambda_n = e^{x_n'\beta}$.

Here we model λ_n as a function of x_n by an MLP neural network:

$$t_n = \hat{\lambda}_n = \phi_0\left(\alpha + \sum_j \omega_j \phi_h\left(\alpha_j + \sum_{i\to j} \omega_{ij} x_n\right)\right) \tag{9}$$

where ϕ_0 is fixed as an exponential function.

Substituting Poisson probability function in (5) and using (9) as Poisson means, the negative log-likelihood criterion can be obtained as:

$$E = -\sum_{n=1}^{N} [-t_n + y_n \log t_n - \ln y_n!]. \tag{10}$$

Eliminating the last term which is not related to the model fitting, we have:

$$E = -\sum_{n=1}^{N} [-t_n + y_n \log t_n]. \tag{11}$$

Model fitting

We compare the performances of different models using simulations. Likelihood error criterion functions such as that in (11) are used to fit models with fixed number of units in hidden layer. To guard against over-fitting, a penalized version of (11) given as below is used in the non-linear model fitting.

$$E_r = E + \lambda \sum_{\text{weights}} \omega_{ij}^2. \tag{12}$$

For each neural networks model, to identify the number of units in hidden layer, both criteria Akaike Information Criterion (AIC) and Schwarz Bayesian Information Criterion (BIC) are calculated:

$$\text{AIC} = -2\,\text{Log likelihood} + 2m \tag{13}$$

$$\text{BIC} = -2\,\text{Log likelihood} + m\log(N) \tag{14}$$

where m is the number of the estimated parameters and N is the number of the observations. The model with the smallest value of the information criterion is considered to be the best. However, it should be noticed that in our neural network model fittings, for each setting of fixed number of hidden units, the negative log-likelihood score we get is suboptimal since the weights are optimized on a penalized version of (11). We thus can only get approximations of the AIC and BIC values. We also calculated MSE for testing set as a reference measure for accuracy, where MSE is defined as

$$\frac{1}{N}\sum_{n=1}^{N} (\lambda_n - t_n)^2. \tag{15}$$

The predictions by different models are ranked by MSE.

The models considered include 2, 3, 4, 5, 10, 20 hidden units. To save the computation time, the weight decay parameter is pre-fixed at 0.012 in our computation. This value is chosen based on some empirical study for different choices of weight decay parameter.

Error gradient calculation

Back propagation is a general computing technique to fit parameters in MLP. The computation involves the numerical evaluation of derivatives of the error function with respect to the weights and biases. The general form of back propagation is described elsewhere [18]. Here we use a special algorithm based on the article by Pearlmutter [30] for computation of Hessian Matrix, similar to Nabney [29] approach. The scaled conjugate gradient algorithm is used for optimization. The code is written in R 2.5 and Matlab 7.2.

Clinical application

Reports in medical literature suggest that neural network models are applicable in diagnosing such as ricket disease [31] myocardial infarction [32] pulmonary emboli [33] and gastrointestinal hemorrhage [34], using waveform analysis of EKGs [35], prediction of health outcome [36, 37], and radiographic images [38]. Neural networks have also been successfully applied in clinical outcome prediction of trauma mortality [39], surgical decision making on traumatic brain injury patients [40], recovery from surgery [41, 42], pediatric meningococcal disease [43], transplantation outcome [44] Alzheimer's [45] and Dementia [46]. In addition some more technical comparison between statistical methods and artificial intelligence techniques for medical data exist [45–55].

References

1. Liu PY, Death AK, Handelsman DJ. Androgens and cardiovascular disease. Endocr Rev. 2003;24:313–40. doi:10.1210/er.2003-0005.
2. Callies F, Stromer H, Schwinger RH, et al. Administration of testosterone is associated with a reduced susceptibility to myocardial ischemia. Endocrinology. 2003;144:4478–83. doi:10.1210/en.2003-0058.
3. Channer KS, Jones TH. Cardiovascular effects of testosterone: implications of the "male menopause"? Heart. 2003;89:121–2. doi:10.1136/heart.89.2.121.
4. Dobs AS, Bachorik PS, Arver S, et al. Interrelationships among lipoprotein levels, sex hormones, anthropometric parameters, and age in hypogonadal men treated for 1 year with a permeation-enhanced testosterone transdermal system. J Clin Endocrinol Metab. 2001;86:1026–33. doi:10.1210/jc.86.3.1026.
5. Malkin CJ, Pugh PJ, Jones TH, Channer KS. Testosterone for secondary prevention in men with ischaemic heart disease? QJM. 2003;96:521–9. doi:10.1093/qjmed/hcg086.
6. Manson JE, Bassuk SS, Harman SM, et al. Postmenopausal hormone therapy: new questions and the case for new clinical trials. Menopause. 2006;13:139–47. doi:10.1097/01.gme.0000177906.94515.ff.
7. Costarella CE, Stallone JN, Rutecki GW, Whittier FC. Testosterone causes direct relaxation of rat thoracic aorta. J Pharmacol Exp Ther. 1996;277:34–9.
8. Deenadayalu VP, White RE, Stallone JN, Gao X, Garcia AJ. Testosterone relaxes coronary arteries by opening the large-conductance, calcium-activated potassium channel. Am J Physiol Heart Circ Physiol. 2001;281:H1720–7.
9. English KM, Jones RD, Jones TH, Morice AH, Channer KS. Testosterone acts as a coronary vasodilator by a calcium antagonistic action. J Endocrinol Invest. 2002;25:455–8.
10. Malkin CJ, Pugh PJ, Jones RD, Jones TH, Channer KS. Testosterone as a protective factor against atherosclerosis-immunomodulation and influence upon plaque development and stability. J Endocrinol. 2003;178:373–80. doi:10.1677/joe.0.1780373.
11. Wu FC, von Eckardstein A. Androgens and coronary artery disease. Endocr Rev. 2003;24:183–217. doi:10.1210/er.2001-0025.
12. Yue P, Chatterjee K, Beale C, Poole-Wilson PA, Collins P. Testosterone relaxes rabbit coronary arteries and aorta. Circulation. 1995;91:1154–60.
13. Kamischke A, Heuermann T, Kruger K, et al. An effective hormonal male contraceptive using testosterone undecanoate with oral or injectable norethisterone preparations. J Clin Endocrinol Metab. 2002;87:530–9. doi:10.1210/jc.87.2.530.
14. Zitzmann M, Nieschlag E. Testosterone levels in healthy men and the relation to behavioural and physical characteristics: facts and constructs. Eur J Endocrinol. 2001;144:183–97. doi:10.1530/eje.0.1440183.
15. Davoodi G, Amirezadegan A, Borumand MA, Dehkori MR, Kazemisaeid A, Yaminisharif A. The relationship between level of androgenic hormones and coronary artery disease in men. Cardiovasc J Afr. 2007;18:362–6.
16. Bishop CM. Pattern recognition and machine learning: Springer, 2006.
17. Faraggi D, Simon R. The maximum likelihood neural network as a statistical classification model. J Stat Plan Inference. 1995;46:93–104. doi:10.1016/0378-3758(95)99068-2.
18. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. New York: Springer; 2001.
19. Ripley BD. Pattern recognition and neural networks. Cambridge: Cambridge University Press; 1996.
20. Ripley RM, Harris AL, Tarassenko L. Non-linear survival analysis using neural networks. Stat Med. 2004;23:825–42. doi:10.1002/sim.1655.
21. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem. 1972;18:499–502.
22. Judkins MP. Selective coronary arteriography. I. A percutaneous transfemoral technic. Radiology. 1967;89:815–24.
23. Gensini GG. A more meaningful scoring system for determining the severity of coronary heart disease. Am J Cardiol. 1983;51:606. doi:10.1016/S0002-9149(83)80105-2.
24. Pollak A, Rokach A, Blumenfeld A, Rosen LJ, Resnik L, Dresner Pollak R. Association of oestrogen receptor alpha gene polymorphism with the angiographic extent of coronary artery disease. Eur Heart J. 2004;25:240–5. doi:10.1016/j.ehj.2003.10.028.
25. Pastor R, Guallar E. Use of two-segmented logistic regression to estimate change-points in epidemiologic studies. Am J Epidemiol. 1998;148:631–42.

26. Funahashi K. On the approximate realization of continuous mapping by neural networks. Neural Netw. 1989;2:183–92. doi:10.1016/0893-6080(89)90003-8.
27. Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Netw. 1989;2:359–66. doi:10.1016/0893-6080(89)90020-8.
28. Mathieson MJ. Ordinal models for neural networks. In neural networks in financial engineering. In: Refences A-P, Abu-Mostafa Y, Moody J, Weigend A, editors. Proceedings of the Third International Conference on Neural Networks in the Capital Markets. Singapore: World Scientific; 1996. p. 523–36.
29. Nabney IT. Netlab: algorithms for pattern recognition. London: Springer; 2001.
30. Pearlmutter BA. Fast exact multiplication by the Hessian. Neural Comput. 1994;6:147–60. doi:10.1162/neco.1994.6.1.147.
31. Fallah N, Faghihzadeh S, Mahmoudi M. Comparing and Contrasting Fuzzy Min-Max Neural Network with the Classical Statistical Clustering Methods in classification of Rickets Disease. Bulletin of the 53rd session of the International Statistical Institute. 2001;2:445–6.
32. Baxt WG. Use of an artificial neural network for the diagnosis of myocardial infarction. Ann Intern Med. 1991;115:843–8.
33. Kemeny V, Droste DW, Hermes S, et al. Automatic embolus detection by a neural network. Stroke. 1999;30:807–10.
34. Das A, Ben-Menachem T, Cooper GS, et al. Prediction of outcome in acute lower-gastrointestinal haemorrhage based on an artificial neural network: internal and external validation of a predictive model. Lancet. 2003;362:1261–6. doi:10.1016/S0140-6736(03)14568-0.
35. Vijaya G, Kumar V, Verma HK. ANN-based QRS-complex analysis of ECG. J Med Eng Technol. 1998;22:160–7.
36. Song X, Mitnitski A, MacKnight C, Rockwood K. Assessment of individual risk of death using self-report data: an artificial neural network compared with a frailty index. J Am Geriatr Soc. 2004;52(7):1180–4. doi:10.1111/j.1532-5415.2004.52319.x.
37. Song X, Mitnitski A, Cox J, Rockwood K. Comparison of machine learning techniques with classical statistical models in predicting health outcomes. Medinfo. 2004;11:736–9.
38. Penedo MG, Carreira MJ, Mosquera A, Cabello D. Computer-aided diagnosis: a neural-network-based approach to lung nodule detection. IEEE Trans Med Imaging.. 1998;17:872–80. doi:10.1109/42.746620.
39. Izenberg SD, Williams MD, Luterman A. Prediction of trauma mortality using a neural network. Am Surg. 1997;63:275–81.
40. Li YC, Liu L, Chiu WT, Jian WS. Neural network modeling for surgical decisions on traumatic brain injury patients. Int J Med Inform. 2000;57:1–9. doi:10.1016/S1386-5056(99)00054-4.
41. Grigsby J, Kooken R, Hershberger J. Simulated neural networks to predict outcomes, costs, and length of stay among orthopedic rehabilitation patients. Arch Phys Med Rehabil. 1994;75:1077–81. doi:10.1016/0003-9993(94)90081-7.
42. Tu JV, Guerriere MR. Use of a neural network as a predictive instrument for length of stay in the intensive care unit following cardiac surgery. Comput Biomed Res. 1993;26:220–9. doi:10.1006/cbmr.1993.1015.
43. Nguyen T, Malley R, Inkelis S, Kuppermann N. Comparison of prediction models for adverse outcome in pediatric meningococcal disease using artificial neural network and logistic regression analyses. J Clin Epidemiol. 2002;55:687–95. doi:10.1016/S0895-4356(02)00394-3.
44. Dorsey SG, Waltz CF, Brosch L, Connerney I, Schweitzer EJ, Bartlett ST. A neural network model for predicting pancreas transplant graft outcome. Diabetes Care. 1997;20:1128–33. doi:10.2337/diacare.20.7.1128.
45. Buscema M, Grossi E, Snowdon D, Antuono P. Auto-contractive maps: an artificial adaptive system for data mining, an application to Alzheimer disease. Curr Alzheimer Res. 2008;5:481–98. doi:10.2174/156720508785908928.
46. Rossini PM, Buscema M, Capriotti M, Grossi E, Rodriguez G, Del Percio C, et al. Is it possible to automatically distinguish resting EEG data of normal elderly vs. mild cognitive impairment subjects with high degree of accuracy? Clin Neurophysiol. 2008;119:1534–45. doi:10.1016/j.clinph.2008.03.026.
47. Allore H, Tinetti ME, Araujo KL, Hardy S, Peduzzi P. A case study found that a regression tree outperformed multiple linear regression in predicting the relationship between impairments and social and productive activities scores. J Clin Epidemiol. 2005;58:154–61. doi:10.1016/j.jclinepi.2004.09.001.
48. DiRusso SM, Chahine AA, Sullivan T, et al. Development of a model for prediction of survival in pediatric trauma patients: comparison of artificial neural networks and logistic regression. J Pediatr Surg. 2002;37:1098–104. discussion 1098–104. doi:10.1053/jpsu.2002.33885.
49. Eftekhar B, Mohammad K, Ardebili HE, Ghodsi M, Ketabchi E. Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data. BMC Med Inform Decis Mak. 2005;5:3. doi:10.1186/1472-6947-5-3.
50. Kattan MW, Hess KR, Beck JR. Experiments to determine whether recursive partitioning (CART) or an artificial neural network overcomes theoretical limitations of Cox proportional hazards regression. Comput Biomed Res. 1998;31:363–73. doi:10.1006/cbmr.1998.1488.
51. Costanza MC, Paccaud F. Binary classification of dyslipidemia from the waist-to-hip ratio and body mass index: a comparison of linear, logistic, and CART models. BMC Med Res Methodol. 2004;4:7. doi:10.1186/1471-2288-4-7.
52. Marble RP, Healy JC. A neural network approach to the diagnosis of morbidity outcomes in trauma care. Artif Intell Med. 1999;15:299–307. doi:10.1016/S0933-3657(98)00059-1.
53. Flouris AD, Duffy J. Application of artificial intelligence systems in the analysis of epidemiological data. Eur J Epidemiol. 2006;21:167–70. doi:10.1007/s10654-006-0005-y.
54. Grassi M, Villani S, Marinoni A. Classification methods for the identification of 'case' in epidemiological diagnosis of asthma. Eur J Epidemiol. 2001;17(1):19–29. doi:10.1023/A:1010987521885.
55. Wolfe R, McKenzie DP, Black J, Simpson P, Gabbe BJ, Cameron PA. Models developed by three techniques did not achieve acceptable prediction of binary trauma outcomes. J Clin Epidemiol. 2006;59:26–35. doi:10.1016/j.jclinepi.2005.05.007.