Neuropsychiatric Disease and Treatment

Dovepress

open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

OPLS statistical model versus linear regression to assess sonographic predictors of stroke prognosis

This article was published in the following Dove Press journal:
Neuropsychiatric Disease and Treatment
29 August 2012
Number of times this article has been viewed

Kianoush Fathi Vajargah[1]
Homayoun Sadeghi-Bazargani[2,3]
Robab Mehdizadeh-Esfanjani[4]
Daryoush Savadi-Oskouei[4]
Mehdi Farhoudi[4]

[1]Department of Statistics, Islamic Azad University, Tehran, North Branch, [2]Neuroscience Research Center, Department of Statistics and Epidemiology, Tabriz University of Medical Sciences, Tabriz, Iran; [3]Public Health Department, Karolinska Institute, Stockholm, Sweden; [4]Neuroscience Research Center, Tabriz University of Medical Sciences, Tabriz, Iran

Abstract: The objective of the present study was to assess the comparable applicability of orthogonal projections to latent structures (OPLS) statistical model vs traditional linear regression in order to investigate the role of trans cranial doppler (TCD) sonography in predicting ischemic stroke prognosis. The study was conducted on 116 ischemic stroke patients admitted to a specialty neurology ward. The Unified Neurological Stroke Scale was used once for clinical evaluation on the first week of admission and again six months later. All data was primarily analyzed using simple linear regression and later considered for multivariate analysis using PLS/OPLS models through the SIMCA P+12 statistical software package. The linear regression analysis results used for the identification of TCD predictors of stroke prognosis were confirmed through the OPLS modeling technique. Moreover, in comparison to linear regression, the OPLS model appeared to have higher sensitivity in detecting the predictors of ischemic stroke prognosis and detected several more predictors. Applying the OPLS model made it possible to use both single TCD measures/indicators and arbitrarily dichotomized measures of TCD single vessel involvement as well as the overall TCD result. In conclusion, the authors recommend PLS/OPLS methods as complementary rather than alternative to the available classical regression models such as linear regression.

Keywords: Prognostic study, trans cranial doppler, partial least squares regression, orthogonal projections to latent structures, multicolinearity

Introduction

Multicolinearity, lower statistical power of study, and missing values are considered major threats to traditional regression models in classical statistics when there are a large number of variables and a small sample size. These various outcomes are common occurrences in brain imaging studies. Some methods have been developed to address these issues. For example, partial least squares (PLS) regression analysis is a known statistical method shown to attenuate the above mentioned problems. At the same time, PLS consists of some limitations such as interpretability problems, multicomponent results, and biased coefficients in some situations leading to a greater risk of overlooking real correlations.

Orthogonal projections to latent structures (OPLS) is a recent modification to the PLS regression analysis method. The main idea of the OPLS method is to separate the systematic variation in the X variable into two parts: (1) that which is linearly related to Y, and (2) that which is orthogonal to Y. This gives rise to a much better interpretability.[1] Compared to the traditional principal component analysis and factor analysis, both PLS and OPLS methods are known as supervised statistical methods.

Correspondence: Homayoun Sadeghi-Bazargani
Neuroscience Research Center, Department of Statistics and Epidemiology, Tabriz University of Medical Sciences, Department of Public Health Sciences, WHO Collaborating Center on Community Safety Promotion, Karolinska Institute, 2nd Floor, Norrbacka, SE-171 76, Stockholm, Sweden
Tel +46 7 3590 5877
Email homayoun.sadeghi@gmail.com

submit your manuscript | www.dovepress.com
Dovepress
http://dx.doi.org/10.2147/NDT.S33991

Supervised modeling techniques may be suitable alternatives or complementary options to the classical modeling techniques in that they can manage large numbers of variables for smaller sample sizes and simultaneously be less prone to threats from multicolinearity and missing values.[1–3] The implication of the OPLS statistical method is gaining interest in various research areas, ranging from its origin in chemometrics in 2006 to its most recent applications in areas such as diversified as public health research.[2,4–6] Imaging science (in particular, brain imaging) has become a recent field of application for OPLS modeling.[7–9] Trans cranial doppler (TCD) is a diagnostic imaging method used in stroke victims and has also been used in stroke risk prediction.[13]

Considering the limitations of classical regression models in managing large numbers of variables, the research oriented use of TCD produces a large number of variables that needs to be modeled through appropriate statistical methods. Limited knowledge is available regarding applicability of newer statistical methodology alternatives in modeling TCD findings as well as modeling the scale based assessment of stroke prognosis.

Thus, the aim of this study was to assess the applicability of the PLS/OPLS statistical models for investigating the role of TCD findings in order to predict ischemic stroke prognosis in comparison to traditional linear regression analysis.

Methods

The present study was conducted on 116 ischemic stroke patients admitted to a specialty neurology ward in Tabriz, Iran. TCD was performed on all patients in the first week of admission. In addition, vessel lesions as well as blood velocity in different vessels were determined. On the first week of admission, the Unified Neurological Stroke Scale (UNSS) function scale was used once for clinical evaluation and subsequently used again once more 6 months later. All data were primarily analyzed using simple linear regression analysis and later considered for multivariate analysis using new supervised models through SIMCA P+12 software, a known software package used for supervised modeling techniques (Umetrics AB, SE-90719, Umea, Sweden). OPLS is implemented in SIMCA P+12 such that the method is available under the standard PCA and PLS modeling framework. PLS and OPLS were considered to be the multivariate analysis methods. Categorical variables (after being converted into dummy variables) were entered into the model as other dichotomous and continuous measures. PLS was run prior to OPLS; however, only the OPLS results are presented in this study.

Four groups of variables were available to be analyzed. The first group included variables pertaining to demographic characteristics, past medical history, and medical examination. The second group of variables was based on laboratory measurements such as cholesterol. The third group of variables came from TCD of all patients and a binary result from computerized tomography scanning or magnetic resonance imaging for some patients. Lastly, the fourth group of variables consisted of scores from the UNSS function scale. UNSS measured after six months was considered to be the response variable that was predicted by other variables. A total of 147 variables were formatted for SIMCA P12 to be analyzed.

The methods of choice for multivariate analysis were considered to be PLS followed by OPLS. Before fitting the PLS/OPLS models, we transformed all variables with a low min/max ratio or high skewness. Log transformation for all respective variables were used and the R-square and Q-square measures were calculated for overall model assessment. A prediction set of 116 and validation set of 34 were both analyzed for validity assessment. Both coefficient variability analysis of variance and Hotelling's T2 range plot were used to assess whether fitted models were overall statistically significant. The models were also assessed using a Hotelling's T2 range plot. The normal probability plot of residuals was used to assess the normal distribution of standardized results and a lack of outliers. To make this report easy to read for clinicians, most of the mathematical and technical information have been omitted. However, details in this regard can be found elsewhere.[1]

This study was conducted in accordance with the ethical standards of the responsible committee of ethics in Tabriz University of Medical Sciences, Tabriz, Iran.

Results

Descriptive results

A total of 116 observations were analyzed as a prediction set. The mean age amongst the patients was approximately 62.1 years with a standard deviation of 11.9 years. Seventy-three percent of all participants were male. The mean body weight was 72.5 kg with a standard deviation of 10.4 amongst all patients. No pathologic findings were found in brain computerized tomography scans in 36% of the patients. Right hemiplegia was the most common complaint of patients followed by left hemiplegia and dizziness. The mean hemoglobin and hematocrit counts of all patients were 4.8 g/dl and 44.5%, respectively. The mean blood urea nitrogen and serum creatinine levels of the patients were 21.9 and 0.96 mg/dL, respectively.

Linear regression results

Due to linear regression analysis limitations, it was not possible to model all the measurements of blood velocity in TCD. The main limitation in this regard was the available high multicollinearity and large number of variables. Therefore, only the variables from TCD were modeled and were arbitrarily dichotomized to show if the blood velocity was abnormal in the TCD of single vessels. The results of the multivariate regression analysis showed that the baseline UNSS, right and left sided middle cerebral artery involvement, as well as right sided anterior cerebral artery involvement were statistically significant predictors of the UNSS score at six months.

OPLS regression analysis

Coefficient variability analysis of variance showed that the overall model was statistically significant. Hotelling's T2 range plot also showed the models to be acceptable. The normal probability plot of residuals, as shown in Figure 1, depicts the normal distribution of standardized results and lack of outliers. Distribution of residuals was shown to have an acceptable low skewness.

The regression model based on orthogonal projections to latent structures was able to identify 26 variables as statistically significant predictors of stroke prognosis. Figure 2 shows all of the 147 variables that were modeled being sorted based on their beta coefficients. The size of the coefficient represents the change in the response when the variable varies from 0 to 1, in coded units (one standard deviation when the data are scaled to unit variance), while the other variables are kept at their averages. These coefficients refer to the centered and scaled X-data or predictors, with Y (6th month UNSS score) just scaled, but not centered.[1] The coefficients basically express how strongly Y is correlated to the systematic part of each of the X-variables. The loadings plot of the model showing the relationship between the various variables is shown in Figure 3. As can be seen in the figure, having an overall normal TCD, versus an abnormal TCD, is quite correlated with the baseline UNSS score.

It was determined that the applied statistical models had good predictive power of approximately 80%. However, the baseline UNSS score had stronger predictability when compared to imaging predictors.

Discussion

The main outcome of interest for assessing the stroke functional prognosis in the present study was UNSS. It has been widely used to assess stroke prognosis in prognostic studies, or in assessing the efficacy of treatments in stroke clinical trials both in ischemic and hemorrhagic stroke.[10–19] TCD is both a diagnostic tool in stroke prognosis and has also been used in stroke risk prediction.[20–23]

In the present study, linear regression analysis results identified some TCD predictors of stroke prognosis, which were confirmed through the OPLS modeling method. Moreover, this methodology appeared to have a higher sensitivity in detecting the predictors of stroke prognosis and had detected several more predictors as well. Using this model, it was possible to use the single TCD measures or indicators, arbitrarily dichotomized measures of TCD single

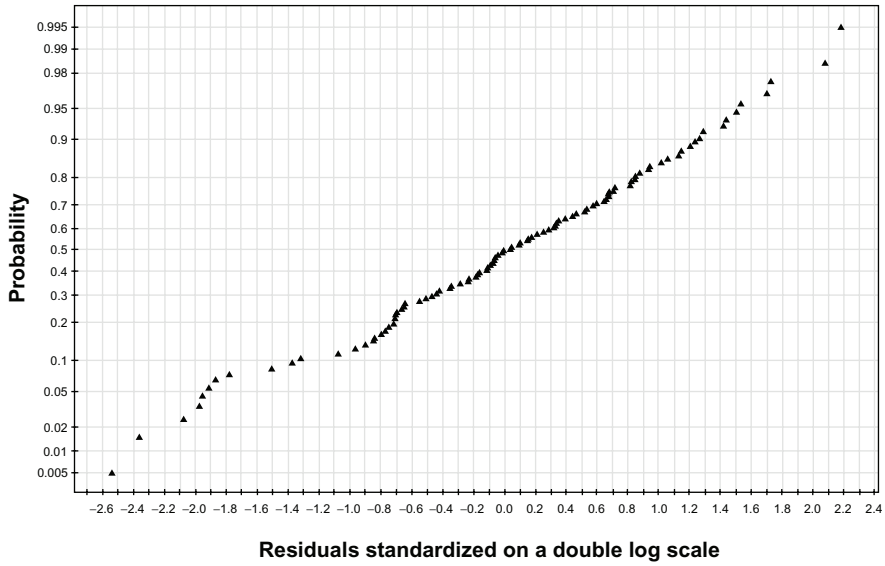

Figure 1 The normal probability plot of residuals.
Note: The normal probability plot of residuals of the regression model run using orthogonal projections to latent structure to investigate predictors of stroke prognosis.

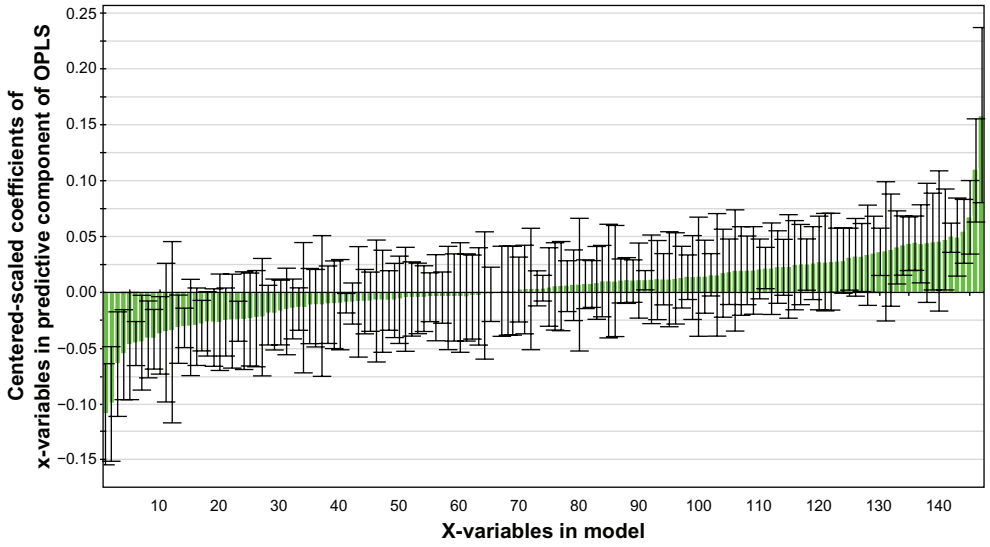

Figure 2 Plot of coefficients of the OPLS regression model.
Notes: Bars indicate the confidence intervals of the coefficients. The coefficient is significant when the confidence interval does not cross zero. The plot of coefficients of the OPLS regression model run to detect the predictors of stroke prognosis including TCD finding.
Abbreviations: OPLS, orthogonal projections to latent structures; TCD, transcranial doppler.

vessel involvement, as well as the overall TCD result. The inclusion of these variables in one single statistical model can be a major source of methodological concern, especially in the case of linear regression. Even when ignoring this problem, another source of concern associated with using linear regression analysis is the decreased statistical power of the study as encountered by increasing the number of parameters in a linear regression model. Another major concern in using baseline UNSS along with other possible predictors to predict the 6th month UNSS score in linear regression is that both scores are from the same type and possibly share common predictors. Thus, if not adequately cared for, this could be a strong source of multicollinearity.

The use of latent variable based models like PLS and OPLS could be a solution to the problem of multicolinearity and also provide higher statistical power in the study. One main finding in the present study was that the results of a modern complex model confirmed the predictors identified by a parsimonious linear regression model were based on a simple variable selection strategy. However, it was shown that other than the predictors

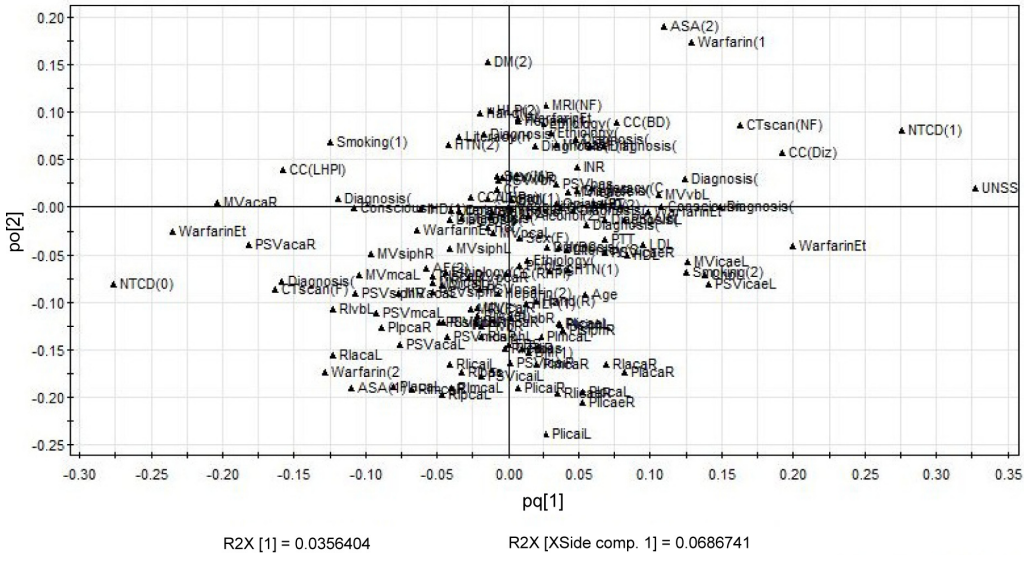

Figure 3 The loadings plot of the OPLS model.
Note: The loadings plot of the OPLS model to assess the predictors of 6th month UNSS score as a surrogate of stroke prognosis.
Abbreviations: OPLS, orthogonal projections to latent structures; UNSS, Unified Neurological Stroke Scale.

Table 1 Results of OPLS regression analysis in detecting predictors of stroke prognosis based on the UNSS score measured six months after a stroke attack

	MV		PSV		PI		RI	
	Right	Left	Right	Left	Right	Left	Right	Left
Part 1: crude velocity measures of blood velocity in TCD resulting in statistically significant predictors of the sixth month UNSS score								
ACA	*		*			*	*	*
MCA		*		*				
PCA						*		
Siphon	*							
VB		*						*
ICA (extra cranial)		*		*				
ICA (intra cranial)								
Part 2: other statistically significant predictors of a sixth month UNSS score								

Notes: Clinicians decision on involvement of LMCA, LMCAp and RMCA; left sided hemiplegia or dizziness as chief complaint; pathological findings in MRI, PTT, and serum LDL and HDL; having normal TCD; being a smoker, the reason for starting warfarin, and baseline UNSS score. *Statistically significant $P < 0.05$.

Abbreviations: OPLS, orthogonal projections to latent structures; UNSS, Unified Neurological Stroke Scale; TCD, transcranial doppler; ACA, anterior cerebral artery; MCA, middle cerebral artery; PCA, posterior cerebral artery; PI, Pulsatility index; RI, Resistance index; ICA, Internal carotid artery; LMCA, Left Middle Cerebral Artery; RMCA, Right Middle Cerebral Artery; MV, Mean velocity; PSV, Peak systolic velocity; LMCAp, Left Middle Cerebral Artery (proximal); MRI, magnetic resonance imaging; PTT, partial thromboplastin time; LDL, low-density lipoprotein; HDL, high-density lipoprotein.

determined by simple use of linear regression, several other predictors were identified using the OPLS model. This may be due to the higher power of the study in PLS/OPLS methods when compared to linear regression in classic statistics.

Nevertheless, two main disadvantages of such models should always be considered when using them. First is the higher complexity of such models, and secondly is the fact that these methods are rather new and need to be assessed in a substantial number of studies prior to recommending them as a single statistical methodology for modeling TCD findings or when being used in stroke prognostic studies. Therefore, the authors prefer to recommend them as complementary rather that alternative to the available classical regression models such as linear regression. Regardless of the main objective of this study, baseline UNSS score was a better predictor of stroke prognosis in comparison to the TCD predictors. This is suggestive of a lower prognostic predictive value of the TCD findings. Not specific to TCD, other imaging predictors have also been shown to have lower predictive values when compared to baseline clinical assessments;[24] however, this does not undermine the role of combined clinical imaging prediction sets.

Conclusion

In comparison to linear regression, the OPLS model appeared to have higher sensitivity in detecting the predictors of ischemic stroke prognosis and detected several more predictors. The study confirmed possible applicability of supervised models to predict some neurological outcomes.

Disclosure

The authors report no conflict of interest in this work.

References

1. Eriksson L, Johansson E, Wold N, Trygg J, Wikstrom C, Wold S. *Multi- and Megavariate Data Analysis: Advanced Applications and Method Extensions*. 1st ed. Umea: Umetrics AB; 2006.
2. Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS). *Journal of Chemom*. 2002;16(3):119–128.
3. Sadeghi-Bazargani H, Banani A, Mohammadi S. Using SIMCA statistical software package to apply orthogonal projections to latent structures modeling. World Automation Congress. Kobe, Japan; 2010:1–9.
4. Sadeghi-Bazargani H, Bangdiwala SI, Mohmmadi R. Applicability of new supervised statistical models to assess burn injury patterns, outcomes, and their interrelationship. *Ann Burns Fire Disasters*. 2011;24(4):191–198.
5. Sadeghi-Bazargani H, Bangdiwala SI, Mohammad K, Maghsoudi H, Mohammadi R. Compared application of the new OPLS-DA statistical model versus partial least-squares regression to manage large numbers of variables in a case-control study. *Scientific Research Essays*. 2011;6(20):4369–4377.
6. Molteni CG, Cazzaniga G, Condorelli DF, Fortuna CG, Biondi A, Musumarra G. Successful application of OPLS-DA for the discrimination of wild-type and mutated cells in acute lymphoblastic leukemia. *QSAR Comb Sci*. Aug 2009;28(8):822–828.
7. Sui J, Adali T, Yu Q, Chen J, Calhoun VD. A review of multivariate methods for multimodal fusion of brain imaging data. *J Neurosci Methods*. February 15, 2012;204(1):68–81.
8. Westman E, Muehlboeck JS, Simmons A. Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion. *Neuroimage*. August 1, 2012; 62(1):229–238.
9. Andersen AH, Rayens WS, Liu Y, Smith CD. Partial least squares for discrimination in fMRI data. *Magn Reson Imaging*. Apr 2012;30(3): 446–452.
10. Treves TA, Karepov VG, Aronovich BD, Gorbulev AY, Bornstein NM, Korczyn AD. Interrater agreement in evaluation of stroke patients with the unified neurological stroke scale. *Stroke*. Jun 1994;25(6): 1263–1264.

11. Edwards DF, Chen YW, Diringer MN. Unified Neurological Stroke Scale is valid in ischemic and hemorrhagic stroke. *Stroke*. Oct 1995;26(10):1852–1858.
12. Macciocchi SN, Diamond PT, Alves WM, Mertz T. Ischemic stroke: relation of age, lesion location, and initial neurologic deficit to functional outcome. *Arch Phys Med Rehabil*. Oct 1998;79(10):1255–1257.
13. Baracchini C, Manara R, Ermani M, Meneghetti G. The quest for early predictors of stroke evolution: can TCD be a guiding light? *Stroke*. Dec 2000;31(12):2942–2947.
14. Altieri M. SPASE-I: a multicenter observational study on pharmacological treatment of acute stroke in the elderly. The Italian Study of Pharmacological Treatment of Acute Stroke in the Elderly Group. *Neurol Sci*. Apr 2002;23(1):23–28.
15. Wohrle JC, Behrens S, Mielke O, Hennerici MG. Early motor evoked potentials in acute stroke: adjunctive measure to MRI for assessment of prognosis in acute stroke within 6 hours. *Cerebrovasc Dis*. 2004;18(2):130–134.
16. Dabrowski M, Bielecki D, Golebiewski P, Kwiecinski H. Percutaneous internal carotid artery angioplasty with stenting: early and long-term results. *Kardiol Pol*. Jun 2003;58(6):469–480.
17. D'Alisa S, Baudo S, Mauro A, Miscio G. How does stroke restrict participation in long-term post-stroke survivors? *Acta Neurol Scand*. Sep 2005;112(3):157–162.
18. Karepov VG, Gur AY, Bova I, Aronovich BD, Bornstein NM. Stroke-in-evolution: infarct-inherent mechanisms versus systemic causes. *Cerebrovasc Dis*. 2006;21(1–2):42–46.
19. Bajenaru O, Tiu C, Moessler H, et al. Efficacy and safety of Cerebrolysin in patients with hemorrhagic stroke. *J Med Life*. Apr–Jun 2010;3(2):137–143.
20. Jayasooriya G, Thapar A, Shalhoub J, Davies AH. Silent cerebral events in asymptomatic carotid stenosis. *J Vasc Surg*. Jul 2011;54(1):227–236.
21. Miller CM, Palestrant D, Schievink WI, Alexander MJ. Prolonged transcranial Doppler monitoring after aneurysmal subarachnoid hemorrhage fails to adequately predict ischemic risk. *Neurocrit Care*. Dec 2011;15(3):387–392.
22. King A, Serena J, Bornstein NM, Markus HS; for ACES Investigators. Does impaired cerebrovascular reactivity predict stroke risk in asymptomatic carotid stenosis? A prospective substudy of the asymptomatic carotid emboli study. *Stroke*. Jun 2011;42(6):1550–1555.
23. Persoon S, Luitse MJ, de Borst GJ, et al. Symptomatic internal carotid artery occlusion: a long-term follow-up study. *J Neurol Neurosurg Psychiatry*. May 2011;82(5):521–526.
24. Savadi-Oskouei D, Sadeghi-Bazargani H, Hashemilar M, DeAngelis T. Symptomatologic versus neuroimaging predictors of in-hospital survival after intracerebral haemorrhage. *Pak J Biol Sci*. May 1, 2010;13(9):443–447.