CLINICO-EPIDEMIOLOGICAL ALGORITHM FOR PREDICTING SYSTEMIC ARTERIAL HYPERTENSION AT HIGH ALTITUDE THROUGH MATHEMATICAL MODELLING

Lt Col RAJVIR BHALWAR*, Col HS SANDHU+,
Dr RC AHUJA#, Dr GK SINGH**, RP MISRA++

ABSTRACT

A population based hybrid design combining element of cohort and cross-sectional approach was used to develop a simple clinical algorithm to predict individual probability of developing hypertension (systolic BP > 140 mm Hg and /or diastolic BP > 90 mmHg). 3615 soldiers initially normotensive at the time of induction into high altitude, were studied by systematic random sampling. Multiple logistic regression analysis showed a high significant association between hypertension and age, body mass index (BMI), tobacco smoking and alcohol consumption. Using the constant/coefficient values obtained from the logistic model and the receiver operating characteristics (ROC) curve analysis, the following predictive rule was developed - "To the age in years, add (BMIx 3.86); also add 5.53 if he is a smoker; and add 19.81 if he consumes alcohol. If the total exceeds 142, the individual is at high risk of developing hypertension. This algorithm carries a sensitivity of 68.2% and specificity of 78.5%.

MJAFT 1994; 50 : 175-180

KEY WORDS : Hypertension; High altitude.

Introduction

During the previous three decades considerable interest has been generated in the epidemiological and clinical aspects of high altitude health problems. One of the major beneficiaries of this research have been our soldiers who guard the highest borders and the loftiest military posts in the world.

The troops, who move to high altitude are perfectly healthy, being in medical category SHAPE-1/AYE at the time of induction (i.e., considered fit for all duties after a thorough medical examination). However, scrutiny of results of periodic medical examinations gives an indication that raised systemic blood pressure is likely to be an important problem in these areas. At the same time, there seems to be a relative paucity of medical literature

in this field. Only a few studies are available in this area, and even these have been conducted on small samples in western countries, on certain selected groups like mountaineers or native highlanders.

It was against the above backdrop that the present study was undertaken to develop a simple clinico-epidemiological rule for predicting the probability of developing hypertension, at high altitude using multiple logistic regression modelling [1] and receiver operating characteristics (ROC) curve analysis [2].

Material and Methods

The present study was a population based hybrid design combining features of cohort and cross-sectional designs. The cohort was examined, at sea level, by a medical officer,

* Classified Specialist (PSM), Post-Doctoral Fellow, Clinical Epidemiology Unit, King George's Medical College, Lucknow; + Sr. Adv (PSM), ADH, HQ Central command, Lucknow. # Assoc Prof, Medicine and Clinical Epidemiology; ** Assoc Prof, Ortho and Clinical Epidemiology, ++ Biostatistician, Clinical Epidemiology Unit, King George's Medical College, Lucknow.

at the time of induction into high altitude, and only those found to be normotensive (i.e., free of the outcome of interest) were inducted. The second assessment was made, on a sample drawn from this cohort after they had stayed at high altitude for one year, for the presence or absence of the outcome, in a cross sectional manner. The setting was one of high altitude terrestrial environment, 3600 to 4200 metres above sea level, in Himalayan areas of North-West India. The reference population was soldiers of Indian Army, of low land origin, serving at an altitude of 3600 - 4200 metres, and normotensive at the time of induction.

Outcome variable of interest was systemic arterial hypertension, defined as systolic BP > 140 mm Hg and/or diastolic BP > 90 mm Hg [3]. The primary effect variables of interest were age (in years), body mass index (defined as weight in Kg/(height in mt)² [4], tobacco smoking (on more than three days a week and more than three cigarettes or their equivalent per week) and alcohol consumption (on more than 2 days a week and drinking more than 120 ml of rum or its equivalent in a week). Effect was measured in terms of Odds ratios, their 95% confidence limits and Chi Square test/ANOVA in univariate analysis; adjusted odds ratios in multiple logistic regression analysis; Beta-coefficients and constant (intercept) terms for developing a prediction rule; and diagnostic test characteristics, viz, sensitivity, specificity, predictive values and likelihood ratios for interpretation of the rule.

Alpha and Beta Errors were set at the conventional levels of $\alpha = 0.05$ (two tailed) (i.e., Type I error of wrongly rejecting a correct null hypothesis) and $\beta = 0.20$ (i.e., Type II error of wrongly accepting an incorrect null hypothesis). With an estimated prevalence of hypertension, as per criteria defined, of 5% (obtained from health examination data), Type I and Type II errors as above, Power=80%, and keeping a range of prevalence of 3.75% to 6.25% to be epidemiologically important for detection ($d=0.125$ on either side), the minimum sam-

ple size worked out to be 2384. It was decided to take an even larger sample size to ensure precision in estimating the parameter of interest. The sample size finally studied was 3615. The study sample was drawn after the soldiers had served at high altitude for one year, by systematic random sampling, after randomly selecting a starting point and a sampling interval. Such random selection also strengthened the variable as psychological and socioeconomic background, though these variables were not the variables of interest for this study.

Data collection was done as per guidelines given by WHO in their monograph "Cardio-vascular survey methods" [5]. A semistructured schedule was used to obtain data as regards age, tobacco smoking and alcohol consumption. Physical examination included measurement of height (without shoes, to the nearest cm) weight (in under garments, to the nearest kg) and blood pressure (right arm, subject reclining, first appearance of sound as systolic and phase V of sound as diastolic BP). Three consecutive readings of BP at 2-3 minute intervals were taken and the lowest was recorded. Subjects were interviewed using a personal interview technique after assuring them of confidentiality. Equipment used in the study was standardised against standard one. Pretesting of schedule was done on 100 subjects.

Initially, a dot plot of data in respect of continuously distributed variable, viz., age, BP and BMI was seen to ensure normality of distribution. Univariate analysis was done between outcome and study variable. Stratified analysis using Mantel-Haenszel procedure was done to evaluate any confounding or interaction. The correlation of the various predictor variables with each other was ruled out, to obviate any possibility of multi-collinearity in the mathematical model.

The variables found to be significantly associated in univariate analysis ($p < 0.05$) were taken for building up the multiple logistic regression model.

The value of constant and beta coefficients so obtained were used for calculating individual probabilities using the equation

$$\text{Prob. (D=1/X)} = \frac{\exp(a + \beta_1 x_1 + \dots + \beta_n x_n)}{1 + \exp(a + \beta_1 x_1 + \dots + \beta_n x_n)}$$

(where D assumes values of '1' if the individual develops the disease, otherwise zero. Vector $\beta' = (\beta_1 \dots \beta_n)$ measures the effect of exposure due to the associated variable vector $X = (X_1 \dots X_n)$).

The data was arrayed in ascending order of individual probabilities and subsequently divided into quartiles for working out the Hosmer Lemeshow test statistic for evaluating the 'goodness of fit' of the model. The individual probabilities were grouped and cross tabulated with the absence or presence of outcome variable. Using this table, the diagnostic test characteristics at various levels of probability cut off points were worked out. Using various levels of sensitivity and corresponding false positive fractions, a receiver operating characteristics (ROC) curve was plotted. The level of probability cut off point corresponding to the "optimum operating point" on the curve was used for developing the clinico-epidemiological predictive rule.

Results

Overall prevalence of hypertension: - Out of 3615 subjects, 179 had the outcome variable (4.95%).

Univariate Analysis : All the four study variable showed significant association with the outcome variable (Table -1)

Multiple logistic regression analysis : the following characteristics were observed (Table 2):-

- (i) *Age* :- Beta coefficient 0.958, SE = 0.0129, Z = 7.4133, $p < 0.001$, Highly significant. The odds ratio (OR) was 1.101 (95% confidence limits (CL) = 1.073 to 1.129). Thus, having adjusted for the linear relationship, for one year increase in age, the odds of developing hypertension increase by 10%.
- (ii) *BMI* :- Coefficient = 0.3701, $p < 0.001$, highly significant. OR = 1.448 (95% CL = 1.348 to 1.555). Thus, with one unit increase in BMI, the odds of having hypertension increase by almost 50%
- (iii) *Smoking* :- Coefficient = 0.5302, $p < 0.01$, highly significant. OR = 1.6999 (95% CL = 1.149 to 2.512). Thus the odds of smokers having hypertension are almost 1.7 times as compared to non smokers.

TABLE 1
Results of univariate analysis

Variable	Description	Test used	Test results	p value
Age (years)	Hypertension absent Mean = 29.30, SD = 6.44 Hypertension present Mean = 34.06, SD = 6.23	ANOVA	F = 92.57 DF (Num) = 1 DF (Denom) = 3613	< 0.001
BMI	Hypertension absent Mean = 21.572, SD = 1.88 Hypertension present Mean = 23.46, SD = 2.82	ANOVA	F = 162.07 DF (Num) = 1 DF (Denom) = 3613	< 0.001
Smoking	Hypertension absent Smoking +ve = 1906, -ve = 1530 Hypertension present Smoking +ve = 142, -ve = 37	Chi-Square	$X^2 = 39.44$ DF = 1	< 0.001
Alcohol use	Hypertension absent alcohol +ve = 2638, -ve = 798 Hypertension present Alcohol +ve = 172, -ve = 7	Chi-Square	$X^2 = 36.67$ DF = 1	< 0.001

TABLE 2
Details of multiple logistic regression analysis

Term	Beta coefficient	Standard deviation	Z-Score	Odds Ratio (95% conf limit)	p - Value
Age	0.0958	0.0129	7.4133	1.101 (1.073 - 1.129)	< 0.001
BMI	0.3701	0.0366	10.1138	1.448 (1.348 - 1.555)	< 0.001
Smoking	0.5302	0.1995	2.6578	1.699 (1.149 - 2.512)	< 0.01
Alcohol	1.8971	0.4007	4.7343	6.666 (3.039 - 14.621)	< 0.001
Constant	-16.3505	1.0043	-16.2799		< 0.001

-2 * Maximized Log Likelihood = 1165.2693; Likelihood Ratio = 259.67, df = 4, p < 0.001

TABLE 3
Hosmer-Lemeshow test for goodness of fit of model

	1st Quartile	2nd Quartile	3rd Quartile	4th Quartile	Total
Observed number of cases	16	13	28	122	179
Expected number from the model	11.29	16.50	31.37	118.64	177.80
Total number of subjects	904	904	904	903	3615

$\chi^2 = 3.24$; df = 2; p > 0.05 (not significant)

(there is no significant difference between observed and expected numbers. Thus the model fits the data well.)

Alcohol : Coefficient = 1.897, p < 0.001, highly significant. The odds of people who consume alcohol, vis a vis non-consumers, for developing hypertension are almost six and a half times (OR = 6.666, 95% CL = 3.039 - 14.621).

Test for 'Goodness of Fit' of the model: Details of Hosmer-Lemeshow test are presented in Table 3. No significant difference was found between the actual number of hypertensive and the expected number based on calculations from the model. Thus the 'goodness of fit' was proved; the model fitted well in the data.

Individual probabilities and presence/absence of outcome: Individual probabilities (calculated by the equation described earlier) were grouped using intervals of 0.02 and cross-tabulated with the presence/absence of hypertension (Table 4). This data was used to

develop the various test characteristics (Table 5).

ROC curve analysis: The ROC curve, plotting the sensitivity levels on Y axis and corresponding False Positive Fractions (FPF) on X axis is presented in Fig. 1. The optimum operating point on this curve corresponded to a sensitivity of 0.682 and FPF of 0.215; this, in turn, was at a probability cut off point of 0.06.

Development of the predictive rule:- The following steps were taken

$$\text{Log (odds of disease)} = a + \beta_1 X_1 + \dots + \beta_n X_n$$

Or

$$\log (p / (1-p)) = a + \beta_1 X_1 + \dots + \beta_n X_n$$

thus,

$$\log (0.06/(1-0.06)) = a + \beta_1 X_1 + \dots + \beta_n X_n$$

or

$$-2.75 = -16.3505 + 0.0958 \times \text{age} + 0.37 \times \text{BMI}$$

$$+ 0.5302 \times \text{smoking} + 1.8971 \times \text{alcohol}$$

Solving the above equation, we get

$$142 = \text{Age} + 3.86 \times \text{BMI} + 5.53 \times \text{smoking} + 19.81 \times \text{alcohol}$$

Fig. 1 Receiver operating characteristics (ROC) curve

TABLE 4

Frequency distribution at various levels of probability cut-off points and presence/absence of outcome variable (hypertension)

Probability cut-off point	Hypertension		Total
	Absent	Present	
0.02	1459	20	1479
0.04	847	23	870
0.06	390	14	404
0.08	261	23	284
0.10	166	12	178
0.12	57	3	60
0.14	88	6	94
0.16	15	3	18
0.18	17	0	17
0.20	17	2	19
0.22	28	14	42
0.24	15	11	26
0.26	7	3	10
0.28	41	13	54
0.30	11	0	11
0.32	3	0	3
0.34	14	32	46
Total	3436	179	3615

TABLE 5

Test characteristics at various levels of probability cut-off

Cut-off	SENS	SPEC	PPV	NPV	LR(+)	LR(-)
0.00	1.000	0.000	0.50	-	1.000	---
0.02	0.888	0.425	0.074	0.986	1.544	0.264
0.04	0.760	0.671	0.107	0.982	2.310	0.358
0.06	0.682	0.785	0.142	0.979	3.172	0.405**
0.08	0.553	0.881	0.171	0.974	3.978	0.519
0.10	0.486	0.909	0.217	0.971	5.341	0.565
0.12	0.469	0.925	0.247	0.971	6.253	0.574
0.14	0.436	0.951	0.317	0.970	8.898	0.593
0.16	0.419	0.955	0.329	0.969	9.311	0.608
0.18	0.419	0.960	0.355	0.969	10.47	0.605
0.20	0.408	0.965	0.380	0.969	11.66	0.613
0.22	0.330	0.974	0.393	0.965	12.69	0.688
0.24	0.268	0.978	0.387	0.962	12.18	0.748
0.26	0.251	0.980	0.395	0.962	12.55	0.764
0.28	0.179	0.992	0.533	0.959	22.37	0.828
0.30	0.179	0.995	0.653	0.959	35.80	0.825
0.32	0.179	0.996	0.696	0.959	44.75	0.824
0.34	0.000	1.000		0.950		1.000

Cut off = Probability cut-off point; Sens = Sensitivity; Spec = specificity; PPV = Positive predictive value; NPV = negative predictive value; LR(+) = Positive Likelihood ratio; LR(-) = Negative likelihood ratio; ** = probability cut-off point corresponding to optimum operating point on ROC curve.

Discussion

The above equation represents the clinico-epidemiological rule developed from this study. Stated otherwise, it means "to the age of an individual (in years), add (3.86 multiplied by BMI); add 5.33 if he is a smoker; add 19.81 if he consumes alcohol. If the total sum so obtained exceeds 142, the likelihood is that the individual will develop hypertension after one year stay at high altitude." This test carries a sensitivity of 68.2%, specificity of 78.5%, predictive values (+ve and -ve) as 14.2% and 97.7% respectively and likelihood ratios (+ve and -ve) as 3.172 and 0.405 respectively. Though the sensitivity (68.5%) and specificity (78.5%) are not very high, the present cut off point as indicated by ROC curve analysis can be gainfully utilised, for use in a community based field setting, especially because the treatment (i.e., surveillance and education) does not carry any extra cost or hazard and the disease does not carry any special "labelling" effect.

As an example, a young, 30 year old soldier, who is overweight (BMI = 25.5), consumes alcohol, though a non smoker, his score will be $30 + 3.86 \times 25.5 + 19.81 = 148.24$, i.e., > 142. Thus he is at high risk of developing hypertension. On the other hand, if we have a middle aged individual 50 years old, who smokes but does not drink and has normal body weight with a BMI of 22, his score will work out as 140.45. Thus he would be at low risk.

Extensive search on 'MEDLINE' using various relevant key words did not reveal similar studies in this field. Some studies have been done on systemic arterial blood pressure at high altitude but they have been limited to only a description of blood pressure, on rather

small samples, at much lower altitudes, and in the western countries (6-9).

It is possible that those subjects who are already susceptible to developing hypertension (due to risk factors like advancing age, overweight, smoking, drinking and various other factors like genetic predisposition, dietary habits etc.) may become more prone to express this outcome under effect of prolonged stay at high altitude, as suggested by the findings of this study. The algorithm developed above can be used as a simple method to identify those who are at "high risk" so that educational activities and surveillance can be targeted to this group.

REFERENCES

1. Klienbaum DG, Kupper LL, Morgenstern H. *Epidemiologic Research - Principles and quantitative methods*, 1st ed. New York: Van Nostrand Reinhold Company, 1982; 419-91.
2. Sackett DL, Haynes RB, Tugwell P. *Clinical Epidemiology - A basic science for clinical medicine*, 1st ed. Boston : Little, Brown and Company, 1985; 100-108.
3. World Health Organisation. Arterial Hypertension. *Tech Rep Ser* 1978; 628 : 8-9
4. World Health Organisation. Energy and Protein requirements. *Tech Rep Ser* 1985; 724 : 25-6.
5. Rose GA, Balckburn H, Gillum H, Prineas RJ. *Cardiovascular survey methods*, 2nd ed. Geneva: World Health Organisation 1982; 29-88.
6. Heath D, Williams DR. *Man at High Altitude - The pathophysiology of acclimatisation and adaptation*. Edinburgh : Churchill Livingstone, 1st Ed, 1977 : 161-3.
7. Appleton F. Possible influences of altitude on blood pressure (Abstr). *Circulation* 1967; suppl 11 : 55.
8. Ruiz L, Penaloza D. Altitude and hypertension (Abstr). *Mayo Clin Proc* 1957; 52 : 442.
9. Liu LS. Epidemiology of hypertension and cardiovascular disease - China experience. *Clin Exp Hypertens* 1990; 12(5) : 831-44.