TRAKYA ÜNİVERSİTESİ
TIP FAKÜLTESİ DERGİSİ
THE JOURNAL OF SCHOOL OF MEDICINE OF TRAKYA UNIVERSITY

Mevlüt Türe, Dr.
Trakya Üniversitesi Tıp Fakültesi
Biyoistatistik Anabilim Dalı
22030 Edirne

ref: 2006/1

Sayın Mevlüt Türe,

Trakya Üniversitesi Tıp Fakültesi Dergisi'nin 2006/1. sayısında yayınlanması planlanan yazınızı dergide basılacak şekilde hazırlayarak kontrol için size gönderiyoruz.

Yazınızla ilgili olarak:

- **Yazı içinde yapılan düzeltmeleri kontrol ediniz. istemeden bir yanlışlığa yol açmış olmayalım.**
- **İngilizce özeti içerik ve kelime sayısı olarak Türkçe özete uygun şekilde lütfen yeniden düzenleyiniz.**
- **Yazınıza Türkçe başlık ekleyiniz.**
- **Şekilleri ve Tabloları lütfen inceleyiniz.**
- Kaynaklar içinde Index Medicus'ta yer alan dergilerden gösterdikleriniz kontrol edildi, gerekli düzeltme ve ekler yapıldı.
- Yazarlarla iletişimde ana araç olarak Interneti kullanıyoruz. Yazınız üzerinde yapılan duzeltmeleri inceleyerek ve dizilmiş halini kontrol ederek bize e-posta yoluyla ulaşabilir ve gerekli gördüğünüz düzeltmeleri bildirebilirsiniz. Türkçe font ve yazı karakteri ile ilgili problemlerden dolayı, kapsamli düzeltmeler ayrı ayrı belirtilerek bize Word dosyasına yazılmış şekilde e-postaya iliştirilerek gonderilmelidir.

  Bu uygulamayla yazarlardan istediğimiz birşey daha var. Kontrol edildikten sonra yanıtlarını, ek ve düzeltmeleri bize iletirlerken, gönderecekleri e-posta mesajında yazının düzeltilmiş halini ve dizilmiş halini ayrı ayrı incelediklerini ve bu düzeltmeleri ve dizilmiş şeklini onayladıklarını bize aynı e-posta mesajı içinde mutlaka bildirmelidirler.

Kontrol ve düzeltmelerinizi yaptıktan sonra yanıtınızı Ekin'e iletmenizi rica ederiz. Derginin en kısa zamanda yayına hazırlanması gerektiğinden yanıtınızı göndermede lütfen zaman kaybetmeyiniz.

Çalışmalarınızda başarılar dileriz.
02.05.2006
Adnan Arif Baycan

İletişim adresi:
EKİN Tıbbi Yayıncılık,
Osmanağa Mah. Nüzhet Efendi Sok. No: 49/3
34714 Kadıköy, İstanbul.
Tel: 0216 - 337 70 73 - 345 77 61 Faks: 0216 - 345 71 30
e-posta: ekinegitim@superonline.com

EĞİTİM ARAŞTIRMASI

The Effect of Health Status, Nutrition, and Some Other Factors on Low Performance at School Using Induction Technique

Türkçe başlık

Mevlüt TÜRE, Zekeriya AKTÜRK, İmran KURT, Nezih DAĞDEVİREN

***Başvuru tarihi** / **Submitted:** 16.05.2005 **Kabul tarihi** / **Accepted:** 28.07.2005*

Objectives: In this study it was aimed to evaluate the importance of some hypothetical factors (nutrition, health indicators, and risky behaviors, and other factors such as personal characteristics and family indicators) on academic achievement using Logistic Regression (LR) and Chi-squared Automatic Interaction Detection (CHAID) method.

Study Design: Participants were 873 secondary school and high school students selected randomly from 12150 students after stratification by school populations in Edirne, in 2003.

Results: The Chi-squared Automatic Interaction Detection had a better sensitivity and predictive rate (61.19% and 67.70% respectively) than LR (50.00% and 64.29% respectively). However, CHAID had slightly lower specificity (74.25%) compared with LR (75.69%). Father's educational level was the most important determining factor in CHAID method. Smoking status, time reserved for homework and nutrition were the next important factors predicting low school performance according to CHAID method.

Conclusion: The classification tree algorithm could be used in risk analysis and target segmentation for academic achievement management. The results of this study will contribute to developguidelines for persons involved in the education of secondary school and high school students.

Key Words: CHAID; logistic regression; academic achievement; health indicators; nutrition, smoking.

Amaç: Bu çalışmada Lojistik Regresyon (LR) ve Chi-squared Automatic Interaction Detection (CHAID) yöntemleri kullanılarak bazı faktörlerin (beslenme, sağlık göstergeleri, riskli davranışlar, kişilik özellikleri, aile göstergeleri vb.) okul başarısı üzerindeki etkileri araştırıldı.

Çalışma Planı: Çalışma örneklemi, 2003 yılında Edirne'de okuyan 12150 öğrenciden oluşan çalışma evreninden, tabakalı örneklemeyle rasgele seçilen 873 ortaokul ve lise öğrencisinden oluşturuldu.

Bulgular: Chi-squared Automatic Interaction Detection'nın, duyarlılık ve doğruluk oranları (sırasıyla %61.19 ve %67.70), LR'nin duyarlılık ve doğruluk oranlarından (sırasıyla %50.00 ve %64.29) daha yüksek bulundu. Buna karşın, CHAID'nin özgüllük oranı (%74.25) LR'nin özgüllük oranından (%75.69) biraz düşüktü. Babanın eğitim düzeyi CHAID yönteminde en önemli faktör olarak bulundu. Yine CHAID yöntemine göre sigara kullanımı, ev ödevi için ayrılan süre ve beslenme faktörleri, başarısızlığı tahmin eden diğer önemli faktörler olarak saptandı.

Sonuç: Sınıflandırma ağacı algoritması, okul başarısının kontrolü için risk analizi ve hedef belirlemede kullanılabilir bir yöntemdir. Bu çalışmanın sonuçlarının, ortaokul ve lise öğrencilerinin eğitimiyle ilgili kişilere bir kılavuz olarak katkıda bulunacağını umuyoruz.

Anahtar Sözcükler: CHAID; lojistik regresyon; okul başarısı; sağlık göstergeleri; beslenme; sigara kullanımı.

Trakya Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı (Türe, Yrd. Doç. Dr.; Kurt, Araş. Gör.), Aile Hekimliği Anabilim Dalı (Aktürk, Doç. Dr.; Dağdeviren, Yrd. Doç. Dr.).

İletişim adresi: Dr. Mevlüt Türe. Trakya Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, 22030 Edirne.
Tel: 0284 - 235 76 41 / 1631 Faks: 0284 - 235 76 52 e-posta: ture@trakya.edu.tr

School performance, a general assessor of intelligence and cognition, is influenced by factors such as biological, social and environmental ones. Several determinants have been shown to be important in the academic performance of children. Intelligence, self-esteem, sleep, parental education, overcrowded housing, personality, nutrition, alcohol consumption, smoking status and family environment are some of the possible variables.[1-12] The importance of smoking in school success with special emphasis on the transition from secondary to high school (age 15) has been underlined for Turkish students.[13] Although many of these and other factors have been extensively investigated, there is still no perfect model to predict school performance. On the other hand, literature search shows a negligence of some parameters such as general health indicators and substance abuse.

Since many of the variables thought to be important in academic achievement are also interrelated, more sophisticated statistical analyses have to be implemented in the analysis of this problem. The methods used for analysis so far are mainly bivariate correlations and regression analyses.

The advantage of decision tree as a statistical analysis is that it selects the independent variable having the strongest association with the dependent variable according to a specific criterion.[14] We expect that the results of a decision tree method will be more concrete and easy to understand in supporting parents and educators for the detection and prevention of poor achievers at school.

İn this study we aimed to evaluate the importance of some hypothetical factors on academic achievement using logistic regression (LR) and Chi-squared Automatic Interaction Detection (CHAID) method.

MATERIALS AND METHODS

Participants

Participants were 873 students selected randomly from secondary school and high school after stratification by school populations from 12150 students in Edirne, in 2003, a Turkish city with 140000 inhabitants.

Data collection

Determining the procedure, the potential participants of each school were collected in one classroom and the researchers made a brief presentation explaining the aim of the study. Those not wanting to participate were allowed to leave after the presentation. The following instruments were used in data collection.

A self administered questionnaire, which contained well structured questions on the child's nutritional history, time reserved for homework and habits. It also collected demographic data such as sex, school type, family type, and family's education. The questionnaire was applied within the classrooms in an anonymous atmosphere.

A standard standing scale for measurement of weight (in kilogram, kg) and height (in centimeter, cm) SECA model.

A peak flow meter, peak flow was defined as the peak air flow from the mouth by forced expiration. Peak flow meter recordings were performed with Personal Best (Healthscan Products Inc., USA) according to the instructions of the manufacturer.

A sphygmomanometer was used to measure blood pressure. Systolic and diastolic blood pressures were recorded by a licensed nurse using an aneroid sphygmomanometer (ERKA, Kallmeyer Medizintechnik GmbH & Co. KG, Germany) by auscultation method from the right arms in sitting position.

School report cards of the last year were retrieved in order to determine academic achievement. Academic achievement was divided into two categories: high performance (no poor grades in the last record card) and low performance (one or more poor grades in the last report card).

Variables used in the study

One dependent and 18 independent variables were used in the study. The study framework is shown in Fig. 1.

Dependent variable

School academic performance as to the last report card, which served as the outcome variable.

Independent variables

Sex: (girl or boy)

School type: School type was reported assecondary school and high school. According to the educational context in Turkey, 8 years of obligatory schooling was divided into primary school (5 years) and secondary school (3 years) followed by voluntary high school education (3 years).

Time reserved for homework: The daily average time reserved by the student for doing homework was reported by the student.

Alcohol consumption: Self-reported alcohol consumption status of the student defined as "yes" or "no".

Smoking status: Self-reported smoking status of the student defined as "smoker" and "non-smoker".

Time reserved for watching TV: The average daily TV watching time as reported by the student.

Father's educational level: Father's education was categorized as "illiterate plus just literate", "primary school graduate", "secondary school graduate", "high school graduate", and "university degree".

Mother's educational level: Mother's education was categorized as "illiterate plus just literate", "primary school graduate", "secondary school graduate", "high school graduate", and "university degree".

Number of siblings: The number of living siblings of the student.

Core family: Core family was defined as "mother, father, and children living together" and was categorized as "yes" and "no".

Body mass index: Body mass index (BMI) was calculated by the formula BMI= weight (kg)/ height^2 (m). BMI was categorized as thin (≤18.5), normal (18.5-29.9), and obese (≥30).

Systolic blood pressure: The measured value of systolic blood pressure (mmHg).

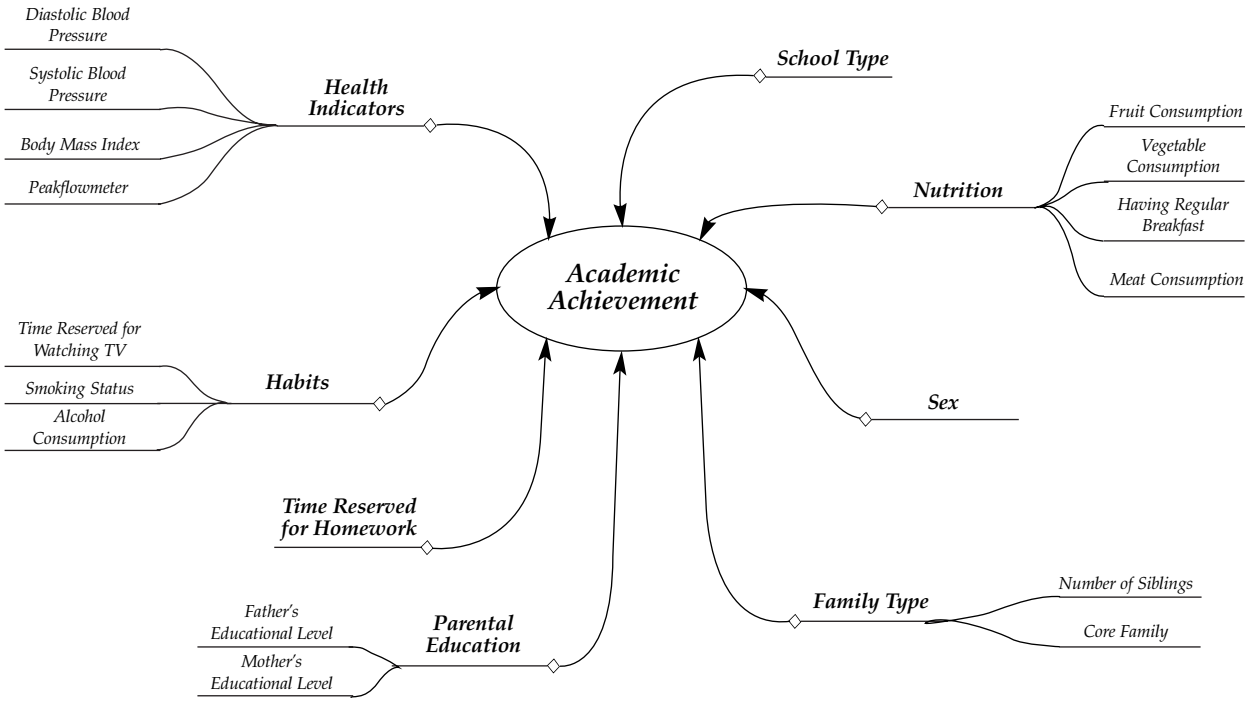

Fig. 1. *Framework of the analysis.*

Diastolic blood pressure: The measured value of diastolic blood pressure (mmHg).

Peak flow meter: The measured peak flow meter value of the student.

Fruit consumption: Fruit consumption status of the student categorized as "every day", "frequently", or "rarely".

Vegetable consumption: Average consumption of vegetable meals per week as reported by the student.

Having regular breakfast: Self-reported breakfasting behavior defined as "having regular breakfast" and "not having regular breakfast".

Meat consumption: Average consumption of meat meals per week as reported by the student.

Logistic regression

Logistic regression is a regression method for predicting a dichotomous dependent variable.[15-17] Logistic regression was performed to identify risk factors for academic achievement using sex, family type, parental education, time reserved for homework, habits, health indicators, school type, and nutrition as independent variables and the academic achievement as the dependent variable. In producing the LR equation, the maximum-likelihood ratio was used to determine the statistical significance of the variables. Logistic regression is an effective way of estimating probabilities from dichotomous variables.

In this study, forward conditional LR was performed with risk factors as covariates, to assess the independent effect of each factor. Odds ratio (OR) with 95% confidence intervals (CI) were calculated to examine the strength and precision of the statistical associations with risk factors for academic achievement.

Decision tree

A decision tree is a non-linear discrimination method, which uses a set of independent variables to split a sample into progressively smaller subgroups. The procedure is iterative at each branch in the tree, it selects the independent variable that has the strongest association with the dependent variable according to a specific criterion.[14,16,17]

Logistic regression and decision tree induction have different underlying assumptions. For LR, it is assumed that the influence of a variable on the outcome is uniform across all subjects unless specific interactions with other variables are included. However, the decision tree assumes that the effect of a variable in the subset is unrelated to the effect of the variable in other subsets of subjects. In this study, the decision tree categorized all subjects according to whether or not they were likely to have low academic performance.

Chi-squared Automatic Interaction Detection method, based on the chi-square test of association, was used in this study. A CHAID tree is a decision tree that is constructed by repeatedly splitting subsets of the space into two or more child nodes, beginning with the entire data set.[14,16,17] To determine the best split at any node, any allowable pair of categories of the predictor variables is merged until there is no statistically significant difference within the pair with respect to the target variable. This CHAID method naturally deals with interactions between the independent variables that are directly available from an examination of the tree. The final nodes identify subgroups defined by different sets of independent variables.

Cross-validation involves splitting the sample into a number of smaller samples. Trees are than generated, excluding the data from each subsample in turn. For each tree, misclassification risk is estimated by applying the tree to the subsample excluded in generating it. The cross-validated risk estimate for the overall tree is calculated as the average of the risks for all of these trees.[18]

In this paper, the CHIAD algorithm with growing criteria of the likelihood ratio chi-square statistic was used for building the tree and evaluating the splits. To identify the nodes of interest, i.e. the nodes with a relatively high probability, a gains chart was constructed showing the nodes sorted by the number of cases in the target category for each node.

Research model

We analyzed the simultaneous relationship among the independent variables for academic

Table 1. Characteristics of study subjects

Characteristics		Academic Achievement				
		High Performance (n=435)		Low Performance (n=438)		
		Median	Inter-quartile	Median	Inter-quartile	*p* value
Time reserved for homework (hour/day)		3	2-4	2	2-3	0.001
Time reserved for watching TV (hour/day)		2	1-3	2	1.45-3	0.649
Vegetable consumption (portion/week)		4	2-5	3	2-5	<0.001
Meat consumption (portion/week)		3	2-4	3	2-4	0.062
Systolic Blood Pressure (mmHg)		110	100-120	110	100-120	0.466
Diastolic Blood Pressure (mmHg)		70	60-80	70	60-80	0.620
Peak flow meter		300	250-380	310	250-406	0.227
		n	%	n	%	*p* value
School type	(1) Primary	233	76.9	279	81.8	0.002
	(2) Secondary	202	23.1	159	18.2	
Sex	(1) Girl	256	29.3	227	26.0	0.037
	(2) Boy	179	70.7	211	74.0	
BMI	(1) Thin	172	19.7	151	17.3	0.092
	(2) Normal	223	25.5	256	29.3	
	(3) Obese	40	4.6	31	3.6	
Core family	(1) Yes	65	7.7	73	8.7	0.456
	(2) No	370	92.3	365	91.3	
Alcohol consumption	(1) Yes	40	5.2	62	8.0	0.016
	(2) No	395	94.8	376	92.0	
Having regular breakfast	(1) Yes	331	45.5	308	42.4	0.751
	(2) No	104	54.5	130	57.6	
Fruit consumption	(1) Every day	160	18.3	151	17.3	0.414
	(2) Frequently	196	22.5	192	22.0	
	(3) Rarely+Never	79	9.0	95	10.9	
Smoking status	(1) Smoker	49	6.2	118	14.9	<0.001
	(2) Non-smoker	386	93.8	320	85.1	
Number of sibblings	≤1	42	4.8	47	5.4	0.448
	2	261	29.9	237	27.1	
	3	80	9.2	96	11.0	
	4	33	3.8	33	3.8	
	5≤	19	2.2	25	2.9	
Mother's education	(1) Illit.+Lit.	22	2.5	35	4.0	<0.001
	(2) Primary	197	22.6	260	29.8	
	(3) Secondary	61	7.0	56	6.4	
	(4) High	108	12.4	68	7.8	
	(5) University	47	5.4	19	2.2	
Father's education	(1) Illit+Lit	10	1.1	14	1.6	<0.001
	(2) Primary	124	14.2	190	21.8	
	(3) Secondary	63	7.2	88	10.1	
	(4) High	146	16.7	102	11.7	
	(5) University	92	10.5	44	5.0	

Illit.: Illiterate, Lit.: Literate.

achievement (Fig. 1). This study compared the relative effects of each risk factor for academic achievement in the multivariate analysis model. We tried to discover the significant patterns and

relationship among the risk factors and make decision rules for the management of academic achievement.

RESULTS

Comparison of characteristics between high and low performance.

The characteristics of the study population are shown in Table 1. We performed the classical statistical analysis to examine the difference in the distribution of variables between the high and low performance. Numeric variables were tested for normal distribution by the Kolmogorov-Smirnov test. Table 1 shows the variables that were significantly different between the two groups based on the Mann-Whitney U-test because the distribution of each continuous variable was non-normal and chi-square test at 5% level for high and low performance.

Eight variables (school type, sex, time reserved for homework, vegetable consumption, alcohol consumption, smoking status, father's educational level and mother's educational level) were independently significant ($p<0.05$).

Analysis of the effect of risk factors on low performance.

The results of LR show that school type, mother's educational level, alcohol consumption, time reserved for homework and vegetable consumption were excellent predicting variables of academic achievement (Table 2).

Secondary school students had a better performance compared with high school students (OR=1.58; 95% CI: 1.03, 2.44). Decreasing maternal education results in poor performance. Alcohol consumers are more prone to poor performance compared with those not consuming alcohol (OR=0.42; 95% CI: 0.22, 0.81). As time reserved for homework increases, the chance for low performance at school decreases (OR=0.84; 95% CI: 0.73, 0.95). Higher vegetable consumption is associated with better school performance (OR=0.91; 95% CI: 0.83, 0.97).

Decision tree and rules for the prediction of low performance.

Of the eighteen variables that were entered in the CHAID method, ten were selected by the program for the decision tree, and a total of twelve subgroups were created. The ten variables were: father's educational level, time reserved for homework, smoking status, vegetable consumption, school type, meat consumption, time reserved for watching TV, fruit consumption, BMI, and sex.

In the decision tree method, we identified the variables that play important roles in explaining low performance (Fig. 2). This indicated that the father's educational level was the

Table 2. Odds Ratio of significant risk factors on low performance

	95% C.I.for OR			
Characteristics	OR	Lower	Upper	*p*
School type (1)	1.58	1.03	2.44	0.038
Mother's educational level	–	–	–	0.003
Mother's educational level (1)	0.55	0.21	1.44	0.225
Mother's educational level (2)	0.28	0.10	0.83	0.021
Mother's educational level (3)	0.32	0.11	0.90	0.031
Mother's educational level (4)	0.15	0.04	0.53	0.003
Alcohol consumption	0.42	0.22	0.81	0.010
Time reserved for homework	0.84	0.73	0.95	0.008
Vegetable consumption	0.91	0.83	0.97	0.022
Constant	18.12	–	–	0.001

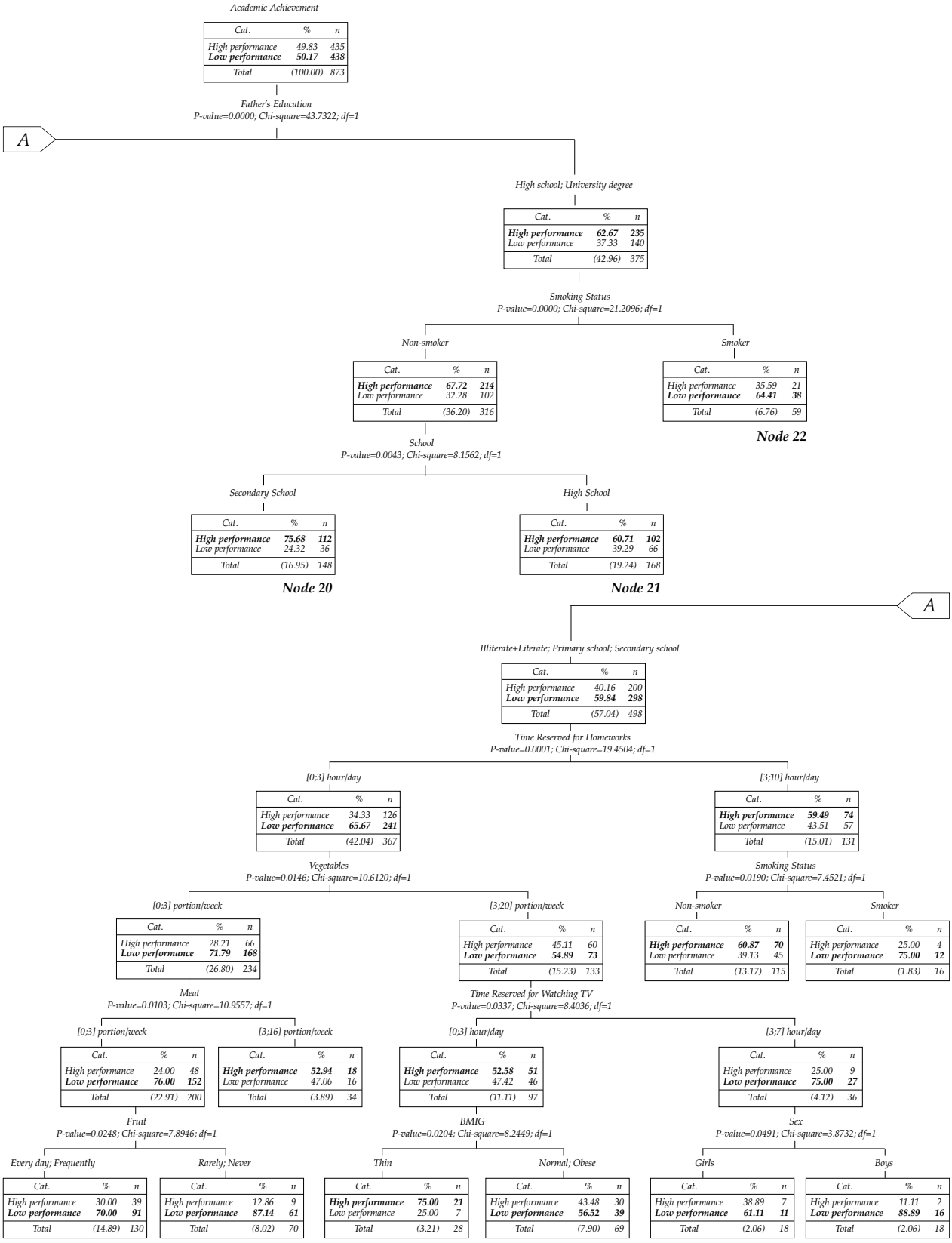

Fig. 2. *Decision tree by chi-squared automatic interaction detection algorithm.*

most important determining factor. This first-level split produced the two initial branches of the decision tree: illiterate + literate, primary or secondary school (unadjusted low performance percentage = 59.84%) versus high school or university (unadjusted low performance percentage = 37.33%).

The best predictor for fathers educational level "illiterate + literate, primary or secondary school" was time reserved for homework while the best predictor for father's educational level "high school or university" was smoking status. While vegetable consumption was the best predictor for those making 3 hours or less homework, smoking status was the best predictor for those making more than 3 hours homework. The best predictor for those consuming 3 or less portions of vegetable per week was meat consumption whereas the best predictor for those more than 3 portions of vegetables per week is time reserved for watching TV. Fruit consumption was found as the best predicting variable for those consuming 3 or less portions of meat per week. While BMI was the best predicting variable for students 3 or less hours of TV per day, sex was the best predictor for those watching more than 3 hours TV per day. The best predictor of those with fathers educational level "high school or university" and non-smoker was school type.

Decision trees are charts that illustrate decision rules (Table 3). The decision rules provide specific information about risk factors based on the rule induction. They begin with one root node that contains all of the observations in the sample. As shown in Figure 2, the decision tree has 22 leaf nodes, of which 12 are terminal nodes. Each node depicted in the decision tree can be expressed in terms of an 'if-then' rule.

Target segmentation for the management of low performance.

The gains chart produced by the decision tree can be used for a risk analysis for low performance management. The gain summary shows which nodes have the highest and lowest proportions of a target category within the node. There are two parts to the gains chart: node-by-node statistics and cumulative statistics (Table 4). In the gains chart, nodes were sorted by the number of cases in the target category for each node. The first node in the table, node 14, contains 16 low performance cases out of 18 subjects, i.e. a low performance rate of 88.89%. For this type of gains chart, with a categorical target variable, the gain score equals the percentage of cases with the target category-in this case, low performance-for the node. The index score shows how the proportion of low performance for this particular node compares to the overall proportion of low performance. For node 6, the index score is about 173.69%, indicating that the proportion of respondents for this node is about 1.7 times the low performance rate for the overall sample. The cumulative statistics demonstrate how well we do at finding low performance cases by taking the best segments of the sample. If we only take the best node (node 14), we reach 3.65% of low performance cases by targeting only 2.06% of the sample. If we include the next best node as well (node 6), then we get 17.58% of the low performance from only 10.08% of the sample. Including node 17 increases these values to 20.32% of low performance cases from 11.91% of the sample. If we include until node 11, we get 61.19% of low performance cases and we must contact 43.53% of the sample to get them. At this stage, we are at the crossover point, where we start to see diminishing returns.

The gains chart also provides valuable information about which segments to target and which to avoid. We might base the decision on the number of prospects we want, the desired low performance rate for the target sample, or the desired proportion of all potential low performance cases we want to contact.

The overall risk estimate in classification tree was 0.3230 (standard error of risk estimate 0.0158), indicating that 67.7% of the cases will be classified correctly by using the decision rule based on the current tree. However, the cross-validated risk estimate was 0.3357 (standard error of risk estimate 0.0164).

Table 3. Decision rules for the prediction of low performance

Node	Father's education	Time reserved for homework (h/day)	Smoking status	Vegetable consumption	Meat	Time reserved for watching TV	Fruit consumption	BMI	Sex	School type	Probability of low performance (%)
14	Illit. + Lit., primary or secondary	0≤ HW ≤3	*	3<Veg.≤20	*	3<TV≤7	*	*	B	*	88.89
6	Illit. + Lit., primary or secondary	0≤ HW ≤3	*	0≤Veg.≤3	0≤Meat≤3	*	Rarely + Never	*	*	*	87.15
17	Illit. + Lit., primary or secondary	3< HW ≤10	Smoker	*	*	*	*	*	*	*	75.00
5	Illit. + Lit., primary or secondary	0≤ HW ≤3	*	0≤Veg.≤3	0≤Meat≤3	*	Everyday or frequently	*	*	*	70.00
22	High or university	*	Smoker	*	*	*	*	*	*	*	64.41
13	Illit. + Lit., primary or secondary	0≤ HW ≤3	*	3<Veg.≤20	*	3<TV≤7	*	*	G	*	61.11
11	Illit. + Lit., primary or secondary	0≤ HW ≤3	*	3<Veg.≤20	*	0≤TV≤3	*	Normal or obese	*	*	56.52
7	Illit. + Lit., primary or secondary	0≤ HW ≤3	*	0≤Veg.≤3	3<Meat≤16	*	*	*	*	*	47.06
21	High or university	*	Non-smoker	*	*	*	*	*	*	High school	39.29
16	Illit. + Lit., primary or secondary	3< HW ≤10	Non-smoker	*	*	*	*	*	*	*	39.13
10	Illit. + Lit., primary or secondary	0≤ HW ≤3	*	3<Veg.≤20	*	0≤TV≤3	*	Thin	*	*	25.00
20	High or university	*	Non-smoker	*	*	*	*	*	*	Secondary school	24.32

*: Non significant; Veg.: Vegetable consumption; HW: Homework; Illit.: Illiterate; Lit.: Literate; B: Boy; G: Girl.

Performance comparison of LR and decision tree.

A comparison of the sensitivity, specificity and predictive rate for the two models is shown in Table 5. The CHAID algorithm had a better sensitivity and predictive rate (61.19 and 67.70% respectively) than LR (50.00 and 64.29% respectively). However, CHAID had slightly lower specificity (74.25%) compared with LR (75.69%).

DISCUSSION

The future of mankind depends on the quality of the intellectual development of mankind. Over the years, the school has become the most important agent for intellectual training. Investments aimed at improving school performance of children are therefore a worthwhile venture; and the enhancement of school performance of children is necessary for a sustainable future development. For this reason, studies providing support in increasing educational success are of extreme importance. This study revealed those factors with a major effect on school performance among a series of clinical factors such as nutrition, health indicators, and risky behaviors, and other factors such as personal characteristics and family indicators.

We compared LR and a decision tree algorithm, CHAID, since LR has assumed the major position as a method for predicting or classifying outcomes based on the specific characteristics of each individual case.[16,19] Similar to the study by Chae et al.[19] and Ho et al.,[16] the CHAID algorithm (67.70%) performed better than LR (64.29%) in predicting low performance, and it provided a much higher sensitiv-

Table 4. Gain Chart by CHAID algorithm

	Node-by-node						Cumulative					
Node	Node (n)	Node (%)	Resp (n)	Resp (%)	Gain (%)	Index (%)	Node (n)	Node (%)	Resp (n)	Resp (%)	Gain (%)	Index (%)
14	18	2.06	16	3.65	88.89	177.17	18	2.06	16	3.65	88.89	177.17
6	70	8.02	61	13.93	87.15	173.69	88	10.08	77	17.58	87.50	174.40
17	16	1.83	12	2.74	75.00	149.49	104	11.91	89	20.32	85.58	170.57
5	130	14.89	91	20.78	70.00	139.52	234	26.80	180	41.10	76.92	153.32
22	59	6.76	38	8.68	64.41	128.37	293	33.56	218	49.77	74.40	148.30
13	18	2.06	11	2.51	61.11	121.80	311	35.62	229	52.28	73.63	146.76
11	69	7.90	39	8.90	56.52	112.66	380	43.53	268	61.19	70.53	140.57
7	34	3.89	16	3.65	47.06	93.80	414	47.42	284	64.84	68.60	136.73
21	168	19.24	66	15.07	39.29	78.30	582	66.67	350	79.91	60.14	119.86
16	115	13.17	45	10.27	39.13	77.99	697	79.84	395	90.18	56.67	112.95
10	28	3.21	7	1.60	25.00	49.83	725	83.05	402	91.78	55.45	110.52
20	148	16.95	36	8.22	24.32	48.48	873	100.00	438	100.00	50.17	100.00

Resp.: Respondents.

ity (61.19%) than LR (50.00%). While mother's education level was found to be the most important risk factor on low performance in LR, father's educational level was the most important determining factor in CHAID method. Smoking status, time reserved for homework, and nutrition were the next important factors predicting low school performance according to CHAID method.

In addition, we demonstrated how CHAID could be used in risk analysis and target segmentation for the pre-detection and management of low performance at school. While LR provides risk factors for low performers, it does not provide specific information about the segment characteristics of risk factors that may be useful for the management of potential low performance. The CHAID algorithm provided cumulative statistics demonstrating how well we found the low performance by taking the best segments of the sample. The gains chart also provided valuable information about which segments to target and which to avoid. In addition, we presented the rules that provided an occurrence relationship among the factors. Such information, which could not be obtained from the LR, can be used in examining the effects of individual factors on a specific segment of the target population.

Table 5. Comparison of the performance of LR and CHAID

Model	Sensitivity (%)	Specificity (%)	Predictive rate (%)
LR	50.00	75.69	64.29
CHAID	61.19	74.25	67.70

CHAID: Chi-squared automatic interaction detection; LR: Logistic regression.

There were some limitations in this study. The importance of social factors with regard to academic achievement may be effected from cultural, ethnic, or religious differences. The sufficient validation for the generalization of our findings to other populations or groups needs to be shown. Other limitations were the lack of input variables, such as the daily amount of reading, regularity of school attendence, motivation towards learning, potential factors associated with smoking (being involved in peer groups such as bands etc.), and relationships with teachers and friends. We believe that the results of this study will contribute to developing guidelines for educators, managers and parents.

REFERENCES

1. Lassiter KS, Bardos AN. The relationship between young children academic achievement and measures of intelligence. Psychology in the Schools

1995;32:170-7.
2. McDonald AS. The prevalence and effects of test anxiety in school children. Educational Psychology 2001;21:89-101.
3. Meijer AM, van den Wittenboer GLH. The joint contribution of sleep, intelligence and motivation to school performance. Personality and Individual Differences 2004;37:95-106.
4. Kurdek LA, Sinclair RJ. Relation of eight graders' family structure, gender, and family environment with academic performance and school behavior. Journal of Educational Psychology 1988;80:90-4.
5. Abidoye RO. Comparative school performance through better health and nutrition in nsukka, enugu, nigeria. Nutr Res 2000;20:609-20.
6. Goux D, Maurin E. The effect of overcrowded housing on children's performance at school. Journal of Public Economics 2005;89:797-819.
7. Zalilah MS, Khor GL, Sarina S, Mirnalini K. Factors contributing to academic achievement among a sample of Indian and Malay school children in Malaysia. Asia Pac J Clin Nutr 2004;13(Suppl):S125.
8. Dappen A, Schwartz RH, O'Donnell R. A survey of adolescent smoking patterns. J Am Board Fam Pract 1996;9:7-13.
9. Heaven P, Mak A, Barry J, Ciarrochi J. Personality and family influences on adolescent attitudes to school and self-rated academic performance. Personality and Individual Differences 2002;32:453-62.
10. Kramer RA, Allen L, Gergen PJ. Health and social characteristics and children's cognitive functioning: results from a national cohort. Am J Public Health 1995;85:312-8.
11. Abidoye RO, George IA, Akitoye CO. Effect of nutritional status on intellectual performance of Nigerian children. Early Child Dev and Care 1991;65:87-94.
12. Hu TW, Lin Z, Keeler TE. Teenage smoking, attempts to quit, and school performance. Am J Public Health 1998;88:940-3.
13. Yorulmaz F, Akturk Z, Dagdeviren N, Dalkilic A. Smoking among adolescents: relation to school success, socioeconomic status nutrition and self-esteem. Swiss Med Wkly 2002;132:449-54.
14. Michael JA, Gordon SL. Data mining technique for marketing, sales and customer support. New York: John Wiley; 1997.
15. Hair JF, Anderson RE, Tatham RL, Black WC. Multivariate data analysis. Englewood cliffs, NJ: Prentice-Hall; 1998.
16. Ho SH, Jee SH, Lee JE, Park JS. Analysis on risk factors for cervical cancer using induction technique. Expert Systems with Applications 2004;27:97-105.
17. Ture M, Kurt I, Kurum AT, Ozdamar K. Comparing classification techniques for predicting essential hypertension. Expert Systems with Applications 2005;29:583-8.
18. Magidson J. SPSS for Windows. CHAID. Release 6.0. SPSS, Chicago: 1993.
19. Chae YM, Ho SH, Cho KW, Lee DH, Ji SH. Data mining approach to policy analysis in a health insurance domain. Int J Med Inform 2001;62:103-11.