ORIGINAL ARTICLE

Which sociodemographic factors are important on smoking behaviour of high school students? The contribution of classification and regression tree methodology in a broad epidemiological survey

C Özge, F Toros, E Bayramkaya, H Çamdeviren, T Şaşmaz

Postgrad Med J 2006;82:532–541. doi: 10.1136/pgmj.2005.040360

See end of article for authors' affiliations

Correspondence to:
Dr C Özge, Fatih Mah,
İstanbul Evleri C Blok D:5,
33170, Mezitli, Mersin,
Turkey; cengizozge@
mersin.edu.tr

Submitted 9 August 2005
Accepted
27 February 2006

Background: The purpose of this study is to evaluate the most important sociodemographic factors on smoking status of high school students using a broad randomised epidemiological survey.

Methods: Using in-class, self administered questionnaire about their sociodemographic variables and smoking behaviour, a representative sample of total 3304 students of preparatory, 9th, 10th, and 11th grades, from 22 randomly selected schools of Mersin, were evaluated and discriminative factors have been determined using appropriate statistics. In addition to binary logistic regression analysis, the study evaluated combined effects of these factors using classification and regression tree methodology, as a new statistical method.

Results: The data showed that 38% of the students reported lifetime smoking and 16.9% of them reported current smoking with a male predominancy and increasing prevalence by age. Second hand smoking was reported at a 74.3% frequency with father predominance (56.6%). The significantly important factors that affect current smoking in these age groups were increased by household size, late birth rank, certain school types, low academic performance, increased second hand smoking, and stress (especially reported as separation from a close friend or because of violence at home). Classification and regression tree methodology showed the importance of some neglected sociodemographic factors with a good classification capacity.

Conclusions: It was concluded that, as closely related with sociocultural factors, smoking was a common problem in this young population, generating important academic and social burden in youth life and with increasing data about this behaviour and using new statistical methods, effective coping strategies could be composed.

Tobacco smoking is the single leading cause of death in developed countries, causing more than 1 000 000 deaths annually. An average of 22 years of life is lost. About 80% of adult smokers begin their tobacco use before 18 years of age.¹ Tobacco use is an important public health issue also in Turkey where about 62.8% of boys and 24.3% of girls and overall 43.6% of the population over 15 years of age smoke tobacco.²

Numerous studies have investigated tobacco use among school children in Turkey and they showed that an increasing proportion of Turkish children and adolescents are exposed to smoke tobacco because smoking prevalence in Turkey has increased rapidly, especially among men, during past decades.^{3–6} Among adolescents nationwide, about 40% are currently exposed to tobacco smoke because at least one parent smokes.^{7–8} Results of a recent national survey further show that 35% of children younger than age 18 reside in homes where family members or visitors smoke in the home on a regular basis.⁹

Both social and commercial tobacco sources play significant parts in youth smoking. Regular smokers, boys, older smokers, and white people are more likely to use commercial sources.¹⁰ However, adolescent smoking is a complex behaviour. Factors that can affect the smoking habit include the sex of the child,¹¹ parental smoking patterns and attitudes,¹² socioeconomic status,¹³ peer influence,¹⁴ ethnic belonging,¹⁵ presence of other adolescent health risk behaviour,¹⁶ and psychological distress.¹⁷

The rapid mobility and the relative instability of living and employment conditions have contributed to the labelling of the high school population as “easy vulnerable” and a “target” population in Turkey like other countries of the world.¹⁸ At this point we must estimate the most susceptible population of youth and the prevention programmes should target this population. Which method should be used for this estimation?

The classification and regression tree method (CART) is a tree based multifactorial complex classification and regression model. When compared with traditional statistical methods, this method has both advantages and disadvantages. The CART method in particular is more useful than traditional methods when a dataset is large, and the number of variables is high. Moreover, the CART does not ignore interactions among factors, and it is not affected by high correlations between risk factors. In addition, the CART remains unaffected by missing values. This method uses a surrogate variable to replace any variable that has a missing value. It should be noted that the results obtained by the CART are based on visually presented data; this approach facilitates the interpretation of the results of an analysis. The authors suggest the use of the CART in studies involving numerical based outcome variables and for the investigation of a large number of variables and it may be more effective than traditional statistical methods in epidemiological studies that determine risk factor.¹⁹

Primary aims of this study are;

- To determine the prevalence of smoking, second hand smoking, and other sociodemographic variables of high school students in a representative sample drawn from different socioeconomic strata in Mersin,
- To examine multiple correlates and predictors of tobacco smoking,
- To determine risk factors affecting the smoking status of students by using the CART as an additional statistical method.

METHODS

This study was conducted in Mersin located on the Mediterranean coast of Turkey. Its population is roughly 800 000 and it is the 10th largest city in the country. Commercially it is an important harbour city and it is economically well developed. Mersin receives a considerable number of immigrants from different parts of Turkey, particularly from more underdeveloped eastern and south eastern regions. Nearly half of the population is less than 25 years of age.

Survey sample

This study was cross sectional in nature and multi-step, stratified, cluster sampling was used. The sample size randomisation was made with EPI 6 INFO program. Randomisation was also weighted according to sex and to each stratum (good, satisfactory, poor) of study population. In the first phase, all schools were classified into three groups according to their socioeconomic levels. In the second phase, classes were selected randomly according to the number of students in that particular school. Data were obtained from a self completed, group administered survey with sampling representative of all preparatory class, 9th, 10th, and 11th grade students (details of the randomisation samples can be seen in table 1). The study population consisted of 3352 students from 22 of 55 schools in the city. In data control process, we excluded 48 subjects with missing data of the outcome or exposure variables, and 3304 subjects were finally analysed.

Demographic variables

Students were asked their total average monthly income (both earned and other income), and the average educational durations (average education time of both the mothers and fathers) of their parent(s) in addition to questions about their life, behaviour, struggles, their parents' behaviour, etc. Detailed past personal and family history has also been

added. To assess environmental tobacco smoke exposure, we asked the student to list all household members and indicate whether each person currently smoked.

Measure of smoking status

Smoking behaviour was defined as follows: never smokers, past smokers (those who smoked in the past, but not in the past 30 days), current (past 30 days) non-daily smokers, and current (past 30 days) daily smokers similar to previous reports.⁷⁻¹⁰ Current smokers were asked to give the numbers of daily consumption. A final category, current smokers, combined both current non-daily and daily smokers. We also evaluated the number of subjects who reported lifetime smoking in addition to the factors affecting the status of current smoking or past smoking.

Statistical analysis

All of the samples have been inserted in a specific database by a professional with a supervision of biostatistics specialist (HC). After the randomised control process, a public health specialist (TS) made data quality control process meticulously. All of the confusing or unpaired data were excluded from the dataset. All statistics have been made after these processes. According to our statistics supervisor, the data quality control methods that we applied and the sampling size reduced our risk of uncontrolled α also. General statistical analyses were performed with SPSS 11.0 software (SPSS, Chicago, IL, USA). The results of descriptive analyses were tested and found to show normal distribution, thus data are given as the means and standard deviations. Difference of sociodemographic factors among the groups according to smoking behaviour was compared using appropriate statistical methods.

In this study, two regression methods have been used. The first one is binary logistic regression analysis and the second is classification and regression tree method. Binary logistic regression analyses using the enter method were performed to examine the association between smoking and the sociodemographic variables of students. We can calculate the affecting coefficient (OR) only by using regression analysis. To evaluate the important factors affecting smoking behaviour in each step, we used the CART methodology.

In this aspect, we think that the two methods can be used together as complementary to each other.

We used Statistica 6.0 (Statistica, 1984–2002, Tulsa, OK) software for CART, which was found to be a new useful method other than more traditional statistical methods for the determination of risk factors for smoking. At the initial stage, the building of a CART began with a root node, which contained all of the subjects; then, a series of yes/no

Table 1 Randomisation details of our samples

School type	Location	Sex	Population size	Sample size
Government school	Urban	Girls	11625	521
		Boys	12202	546
	Rural	Girls	1685	75
Private school	Urban	Boys	2023	91
		Girls	560	25
	Rural	Boys	582	26
		Girls	66	3
		Boys	65	3
Occupation High School	Urban	Girls	14499	649
		Boys	18287	819
	Rural	Girls	2419	108
		Boys	2996	134
Total			67000	3000

This figure was minimum sample size, but we calculated the sample size as 3352 subjects to avoid the effects of missing data.

Table 2 Sociodemographic characteristics of study sample

Variables	Girls (n = 1391)	Boys (n = 1913)	Total (n = 3304)
Average age (years) (SD)	15.8 (1.1)	16.1 (1.2)**	16.0 (1.2)
Stratified distribution of classes (%)			
Preparatory class	195 (14.0)	185 (9.6)	380 (11.5)
9th grade	497 (35.7)	773 (40.4)	1270 (38.4)
10th grade	367 (26.3)	476 (24.8)	843 (25.5)
11th grade	332 (23.8)	479 (25.0)	811 (24.5)
Number (SD) of family members	5.1 (2.0)	5.3 (2.2)**	5.2 (2.1)
Number of siblings (%)	n = 1358 (42.2)	n = 1857 (57.8)**	n = 3215
None	243 (17.9)	291 (15.7)	534 (16.6)
1-3	709 (52.2)	923 (49.7)	1632 (50.8)
>4	406 (29.9)	643 (34.7)	1049 (32.6)
Birth order (SD)	2.3 (1.6)	2.5 (1.8)**	2.4 (1.7)
Socioeconomic level of family (%)	n = 1361 (42)	n = 1880 (58)**	n = 3241
Poor	310 (22.8)	586 (31.2)	896 (27.6)
Satisfied	962 (70.7)	1217 (64.7)	2179 (67.2)
Good	89 (6.5)	77 (4.1)	166 (5.1)
Number (%) immigrated past year	156 (11.4)	319 (16.9)*	475 (14.6)
Not living with families (%)	58 (4.2)	121 (6.4)*	179 (5.4)
Mean (SD) age of mothers	40.5 (5.3)	40.9 (6.1)*	40.7 (5.8)
Number (%) separated from mothers	17 (1.2)	28 (1.4)	45 (1.4)
Number (%) of illiterate mothers	218 (15.8)	405 (21.4)**	623 (18.9)
Number (%) of unemployed mothers	1123 (81.3)	1614 (85.3)*	2737 (83.6)
Mean (SD) age of fathers	45.1 (6.0)	45.3 (6.3)	45.3 (6.2)
Number (%) separated from fathers	67 (4.9)	110 (5.7)	177 (5.4)
Number (%) of illiterate fathers	45 (3.4)	115 (6.4)	160 (5.1)
Number (%) of unemployed fathers	194 (14.5)	311 (17.2)*	505 (16)

Differences between two sexes *p<0.05; **p<0.001. The % figures represent the ratio of columns.

questions generated descendant nodes. These nodes were more homogenous than the root node and these terminal homogenous nodes were evaluated.¹⁹

In total 3304 students were included in this analysis and named as "learning sample". As a dependent variable, the question of current smoking was coded as 1 (yes) and 0 (no). The variables that possibly affect the smoking status were accepted as independent variables. It was composed of 8 parametric and 14 non-parametric, in total 22 sociodemographic variables, which can be seen in table 2. At this point because of missing variables, 2556 subjects were analysed.

After the maximum tree value was obtained, the pruning process began to obtain the optimum tree set using the cost-complexity pruning method. In pruning, the cost-complexity pruning method providing the balance between probability of misclassification calculating from risk matrix and tree complexity was used. During the application of this method, two estimation values (resubstitution estimate and 10-fold cross validation estimate) being used as estimation of misclassification ratio were examined as comparative. After obtaining the optimum tree, the results of this method have been evaluated.

Table 3 Tobacco smoking rates according to sex

Variables	Girls (n = 1391)	Boys (n = 1913)	Total (n = 3304)
Number of never smokers	980 (70.5)	1067 (55.8)	2047 (62)
Number of once in a lifetime smoking	411 (29.5)	846 (44.2)**	1257 (38)
Number of past smokers	289 (70.3)	435 (51.4)**	724 (57.6)†
Number of current non-daily smokers	38 (9.2)	89 (10.5)**	127 (10.1)†
Number of current daily smokers	84 (20.5)	322 (38.1)**	406 (32.3)†
Current smokers distribution	n = 122 (22.9)	n = 411 (77.1)**	n = 533 (16.1)
Preparatory class	8 (6.6)	9 (2.2)	17 (3.2)
9th grade	36 (29.5)	149 (36.3)	185 (34.7)
10th grade	45 (36.9)	105 (25.5)	150 (28.1)
11th grade	33 (27)	148 (36)	181 (34)
Tobacco rates distribution	n = 84 (20.7)	n = 322 (79.3)	n = 466 (14.1)
1-5/daily	45 (53.6)	162 (50.3)	207 (51)
6-10/daily	23 (27.4)	76 (23.6)	99 (24.4)
11-20/daily	12 (14.3)	55 (17.1)	67 (16.5)
> 20/daily	4 (4.8)	29 (9)	33 (8.1)
School distribution	n = 122 (22.9)	n = 411 (77.1)**	n = 533 (16.1)
Government school	22 (18)	112 (27.4)	134 (25.1)
Special government school	38 (31.1)	63 (15.3)	101 (18.9)
Trade school	53 (43.4)	213 (51.8)	266 (49.9)
Private school	9 (7.4)	23 (5.6)	32 (6)
Number of smoking mothers	56 (46.3)**	110 (27.2)	166 (31.6)
Number of smoking fathers	71 (61.2)	212 (55.2)	283 (56.6)
Number of smoking siblings	38 (31.1)	136 (33.1)	174 (32.6)
Number of second hand smoking	96 (78.7)	300 (73)	396 (74.3)
Number of relatives with smoking related disorders	47 (38.5)*	115 (28)	162 (30.4)

Percentages are shown in parentheses. Differences between the two sexes *p<0.05; **p<0.001. †The figures represented the ratios according to lifetime smoking of subjects. The: % figures represent the ratio of columns.

Table 4 Comparison of some family variables the of study groups

Variables	Smokers (n = 526)	Non-smokers (n = 2748)	p Value
Mean (SD) education time of mothers (year)	6.04 (4.48)	6.29 (4.21)	0.210
Education levels of the mothers (%)			0.009
Non-literate	117 (22.2)	506 (18.4)	
Primary school	222 (42.2)	1171 (42.6)	
Intermediate school	62 (11.8)	360 (13.1)	
High school	76 (14.4)	526 (19.1)	
University and after	49 (9.3)	185 (6.7)	
Mean (SD) education time of fathers (year)	7.85 (3.96)	8.44 (3.98)	0.002
Education levels of the fathers (%)			0.020
Non-literate	36 (7.2)	124 (4.7)	
Primary school	185 (37.0)	933 (35.2)	
Intermediate school	98 (19.6)	471 (17.8)	
High school	125 (25.0)	718 (27.1)	
University and after	56 (11.2)	402 (15.2)	
Mean (SD) income of the families (Turkish lira)	896.18 (1275.48)	852.84 (1447.16)	0.575

RESULTS

Descriptive statistics

Table 2 gives the descriptive statistics. After removing cases with missing data on control variables, sample sizes were in total 3304 students and their mean (SD) age was 16.04 (1.2) years. The composition of this sample was as follows; 380 students (11.6%) were preparatory class for high school, 1270 (38.4%) 9th grade, 843 (25.5%) 10th grade, and 811 (24.5%) 11th grade. Sex distributions were 1391 (42.1%) girls and 1913 (57.9%) boys. General academic performances of youths were 3.8 (0.78) (ranged from 1 to 5) according to personal reports of students. They commonly lived in moderate size households and satisfactory socioeconomic status, especially girls. Their mothers were commonly unemployed (83.6%) and younger than fathers (40.7 (5.8) years compared with 45.3 (6.2) years).

Smoking status of students

Table 3 shows tobacco smoking rates and trends according to the sex distribution. Significant increases in smoking prevalence and daily smoking rates for all grades were seen over time, especially in 11th grade students with a male predominance ($p = 0.000$). These students were generally non-smokers (62.7%) but a moderate rate of lifetime smoking prevailed (38%). The current smokers (16.1%) were commonly boys of higher grades of certain school types, having moderate tobacco consumption. Table 4 gives details of the some demographic family variables of the study groups (smoker compared with non-smokers). As shown in this table there are no significant difference between the groups

according to their economic status. On the other hand parents of smokers were lower educated than non-smokers.

Smoking status of the families

Among the current smokers, fathers (56.6%) were much more likely to smoke than mothers (31.6%). They have a high ratio of second hand smoking (74.3%) and a relative who complained of smoking related disorders (30.4%). Table 5 shows details of the burden of smoking. The findings pointed out the significantly increased ratio of suicide behaviour, sleep disturbances, psychiatric intervention, and both disturbed relationships with family and problems with friends, predominantly in girls. The high ratio of students who want to stop smoking (63.8%), predominantly in boys, is promising.

Risk factors affecting smoking and stopping smoking

Details of the factors that have significant effect on lifetime smoking, current smoking and quitting smoking are shown in table 6, table 7, and table 8, respectively. Various family characteristics were found to be prominently important on lifetime tobacco smoking and current smoking. On the other hand, the effects of cigarette smoking on social behaviour of students are also shown. Similarly, the students, who have good academic performance or good social environment showed the probable results on stopping smoking.

Classification and regression tree results

Sociodemographic variables, taken into consideration in RTM, are shown in tables 9A and 9B in detail. After the

Table 5 Burden of tobacco smoking on students' life events according to sex

Variables	Girls (n = 122)	Boys (n = 411)	Total (n = 533)
Decreased academic performances	21 (17.5)	99 (24)	120 (22.5)
Increased violence of students	52 (43.3)	226 (56.1)	278 (53.2)
Increased violence of parents	43 (36.1)	114 (28.1)	157 (30)
Disciplinary punishment	24 (20.3)	73 (19.5)	97 (19.7)
Idea/intent of suicide	47 (38.5)**	76 (18.5)	123 (23.1)
Failed a grade	10 (8.7)	116 (28.9)	126 (24.3)
Other substance use	66 (56.4)	221 (54.7)	287 (55.1)
Sleep disturbances	65 (54.2)*	170 (41.8)	235 (44.6)
Disturbed relationship with parents	74 (62.2)*	191 (46.8)	265 (50.3)
Disturbed relationship with friends	56 (47.1)*	138 (34.1)	194 (37)
Psychiatric intervention	37 (31.1)**	50 (12.4)	87 (16.7)
Idea as smoking is hazardous	108 (88.5)	349 (84.9)	457 (85.7)
Wanting to stop smoking	57 (50)	265 (67.8)**	322 (63.8)

Data shown as numbers with percentages in parentheses. Differences between sexes * $p < 0.05$; ** $p < 0.001$.

Table 6 Factors significantly affecting the lifetime smoking of students

Variables	Odds ratio (OR)	p Value
Age of students	46.6	0.000
Male sex	4.04	0.044
Private school education of students	13.98	0.032
Low academic performance	12.60	0.000
Having a current smoker sibling	32.24	0.000
Living together with their families	8.37	0.004
Associating with smoking students	13.60	0.000
Having a problem with a near friend	37.29	0.000
Having a problem with parents	30.35	0.000
The idea/intent of suicide	20.10	0.000
School absence	18.29	0.000
Substance use other than tobacco smoking	76.50	0.000
To earn own pocket money	6.68	0.010

splitting statistics, terminal nodes have been obtained. The structure of the optimum tree is given in figure 1. In this figure, nodes shown with a bold black square line represent terminal nodes, and the other nodes are root and child nodes. For these nodes, the value that is located in the upper right hand corner is the number of observations in each group (nodes), and the mean and variance of smoking status of these groups, respectively. The number located in the upper left hand corner represents the identification number (ID) of these groups. Detailed characteristics of some terminal nodes are as follows;

Smoking groups

Group 1 (ID:10)

The subjects whose ages are lower than 16.5, are going to public school, having low academic performances (the mean grades lower than 2.95), their mothers' age was lower than 43.5 years old and fathers' age was lower than 39.5 years old.

Group 2 (ID:9)

The subjects, aged lower than 16.5, are going to public schools, having low academic performances (the mean grades lower than 2.95), their mothers' age was lower than 43.5 years old.

Group 3 (ID:398)

The subjects, whose age ranged between 15.5 and 16.5, going to private school or public school, and have a smoker sibling.

Group 4 (ID:424)

The subjects, whose age ranged between 15.5 and 16.5, going to private school or public school, and have fathers younger than 40.5 years old, but non-smoker sibling.

Group 5 (ID:498)

The subjects older than 16.5 years old and having a smoker sibling.

Group 6 (ID:679)

These subjects are older than 16.5 years old, having a non-smoker sibling, low academic performance (the mean grades lower than 3.15), their fathers older than 44.5 years old, and mothers are non-smokers. The birth rank of these subjects was after the second child in the family.

Group 7 (ID:700)

The subjects are older than 16.5 years old, having a non-smoker sibling, moderate academic performance (the mean grades ranged from 3.15 to 4.05), mothers are non-smoker, and fathers are employed.

Group 8 (ID:675)

The subjects are older than 16.5 years old, having a non-smoker sibling, moderate academic performance (the mean grades lower than 4.05), their fathers are older than 44.5 years old, and mothers are smokers.

Table 7 Factors significantly affecting current smoking of students

Variables	Odds ratio (OR)	p Value
Increased age of students	18.1	0.000
Male sex	23.05	0.000
10th class of school	4.7	0.030
Over than 2nd birth rank	5.3	0.021
Trade school education	20.3	0.000
Special government school	15.2	0.000
Low academic performance	25.5	0.000
Increased mother age	4.1	0.041
Decreased father education level	3.8	0.049
Smoking of parents	19.3	0.022
Smoking of mothers	6.01	0.014
Smoking of sibling	9.75	0.002
Severe medical illness of relatives relating to smoking	14.05	0.000
Relationship problem with friends	4.8	0.027
Relationship problem with parents	10.2	0.001
Relationship problem between parents	21.02	0.000
Increased violence of students	21.02	0.000
Discipline punishment in school	11.6	0.001
To be hit in the home	6.2	0.012

General significance of this model is $p=0.000$.

Table 8 Factors significantly affecting past smoking of students

Variables	Odds ratio (OR)	p Value
Increased age of students	26.4	0.000
Male sex	32.9	0.000
Trade school	21.05	0.000
Special government school	21.48	0.000
Low academic performance	34.50	0.000
Smoking of parents	26.40	0.003
Smoking of mothers	5.92	0.015
Smoking of sibling	13.94	0.000
Poor or moderate socioeconomic status of family	6.17	0.046
Severe medical illness of relatives relating to smoking	6.69	0.010
Idea that smoking is hazardous	25.57	0.000
Relationship problem between parents	31.50	0.000
Relationship problem with parents	13.20	0.000
To be hit in the home	6.44	0.000

General significance of this model is $p=0.000$.

Group 9 (ID:746)

The subjects are older than 16.5 years old, having a non-smoker sibling, high academic performance (the mean grades higher than 4.05), and the average education time of mothers is lower than 2.5 years.

According to this model we calculated validation ratios of this CART (details are shown in table 10). This model can define the non-smoker subjects with a high specificity (88%), and moderate level of sensitivity for smokers (34%). General classification capacity of this tree is 69 % and this ratio can be accepted as satisfactory.

DISCUSSION

This study showed that smoking is an important health problem among high school students in Mersin using a

broad, stratified epidemiological survey. We also obtained a high ratio of lifetime smoking, but only the students who had proper socioeconomic conditions maintained smoking. The smoker students commonly showed prominent family characteristics and important behavioural problems with both their parents and friends.

Tobacco prevalence is typically estimated by self reported information collected during survey interviews. Studies of the validity of these self reports suggest that they are generally valid indicators of actual behaviour.²⁰ A recent meta-analysis of 51 comparisons of self reported behaviour and various biochemical assessments, for example, concluded that self reports are both sensitive (mean 87.5%) and specific (mean 89.2%).²¹⁻²³ In our study design, we used the stratified sample survey, as it is accepted as the best model for epidemiological

Figure 1 Optimum tree model of this sample. NS, non-smokers; S, smokers.

Table 9A Parametric variables used in classification and regression tree statistics (mean (SD))

Risk factors	Smokers	Non-smokers
Age*	16.4 (1.2) (n = 1257)†	15.8 (1.2) (n = 2047)
Number of siblings	3.5 (2.3) (n = 1222)	3.1 (2.1) (n = 1993)
Mean grade of academic performance*	3.6 (0.8) (n = 1088)	3.9 (0.8) (n = 1768)
Mother's age*	41.5 (6.1) (n = 1245)	40.3 (5.6) (n = 2029)
Education time of mother*	5.9 (4.4) (n = 1245)	6.4 (4.2) (n = 2029)
Father's age*	45.9 (6.3) (n = 1245)	44.9 (6.2) (n = 1956)
Education time of father	8.1 (4.0) (n = 1192)	8.5 (3.9) (n = 1956)
Number of family members	5.4 (2.4) (n = 1239)	5.1 (1.9) (n = 2018)

*Variables that are statistically significant splitters of this method. †The numbers in parentheses represented the numbers of subject analysed for each category of variables.

surveys, and we gave a detailed explanation to the students about the scope and the outcome measures of the study to obtain the best response rate, as soon as possible. We also achieved the full support of the teachers and principals.

The prevalence of tobacco smoking among school pupils has been worldwide researched in a detailed manner and various figures have been reported, with some methodological differences taken into consideration.^{6, 24-27} In an important recent report from France, Bogui *et al* reported a high frequency (12.9%) of tobacco smoking among students aged between 13 and 17 and reported increasing rate by age and they also reported an important ratio of second hand smoking (44.3% for home and 66.7% for public places) among non-smokers. In another population based study Valdivia *et al*²⁵ reported that 64% of students had smoked at least once in their lifetime. The ratio of our past smokers was 57.6% and current smokers were 16.9% with a boy predominancy. Similar to the mentioned study, second hand smoking was a high ratio among our study population (74.3%) and they presented especially among 11th grade of students in favour of boys with a statistical significance. Numerous studies have investigated tobacco use among school children in Turkey and they showed that an increasing

proportion of Turkish children and adolescents are exposed to tobacco smoke because smoking prevalence in Turkey has increased rapidly, especially among men, during past decades.³⁻⁶ Some studies reported similar rates of tobacco smoking between boys and girls.²⁸⁻³⁰ However, the lower rates for girls in our study may be explained on the basis of cultural differences in social roles related to countries and regions. Tobacco smoking is not as much approved and tolerated in both Turkish society and Mediterranean regions of Turkey, as mentioned previously.⁶

The relation between substance use and youth violence is not clear. In an important study from Israel, Molcho *et al* found that across all grades, sexes, and ethnicities, daily smoking, use of hard drugs, history of drunkenness and binge drinking were the best predictors of violent behaviour.³¹ We saw a high degree of violent behaviour among students against both friends and family members with a male predominancy. We also showed the important negative effect of smoking on both academic performances and social relationships. As a result of these burdens of smoking, students reported a high ratio of psychiatric intervention with a girl predominancy. This study also showed the negative effect of smoking on suicide attempt or behaviour. Similarly,

Table 9B Parametric variables used in classification and regression tree statistics (figures represent numbers and % of each column separately)

Risk factors	Categories	Smokers (%)	Non-smokers (%)
Sex	Girl (1)	411 (32.7)	980 (47.9)
	Boy (2)	846 (67.1)	1067 (52.1)
Categorised rank of born*	First (1)	381 (31.2)	751 (37.7)
	2nd and 3rd (2)	555 (45.4)	894 (44.9)
	≥4th (3)	286 (23.4)	348 (17.5)
High school type*	Public school (1)	343 (27.3)	606 (29.2)
	Occupation school (2)	521 (41.4)	597 (29.2)
	Special public school (3)	316 (25.1)	754 (36.8)
	Private school (4)	77 (6.1)	90 (4.4)
Mother's status	Alive (1)	1245 (99)	2029 (99.1)
	Dead (2)	12 (1)	18 (0.9)
	Stepmother (3)	10 (0.8)	5 (0.2)
Mother's employment*	Employed (1)	214 (17.2)	323 (15.9)
	Unemployed (2)	1031 (82.8)	1706 (84.1)
Father's status	Alive (1)	1192 (94.8)	1956 (95.6)
	Dead (2)	65 (5.2)	91 (4.4)
	Stepfather (3)	13 (1.1)	8 (0.4)
Father's employment	Employed (1)	982 (82.4)	1661 (84.9)
	Unemployed (2)	210 (17.6)	295 (15.1)
Smoking of mothers*	Smoker (1)	363 (29.2)	519 (25.6)
	Non-smoker (2)	882 (70.8)	1510 (74.4)
Smoking of fathers	Smoker (1)	643 (53.9)	969 (49.5)
	Non-smoker (2)	549 (46.1)	987 (50.5)
Smoking of siblings*	Smoker (1)	341 (27.1)	297 (14.5)
	Non-smoker (2)	916 (72.9)	1750 (85.5)
Socioeconomic levels of family	Poor (1)	389 (31.5)	507 (25.2)
	Satisfied (2)	779 (63.2)	1400 (69.7)
	Good (3)	65 (5.3)	101 (5)

*Variables that are statistically significant splitters of this method.

Table 10 Optimum classification and regression tree classification matrix

		Observed		
		Smokers	Non-smokers	Total
Expected	Smokers	372	192	564
	Non-smokers	595	1397	1992
	Total	967	1589	2556

Sensitivity of tree: 34% (372 of 967). Specificity of tree: 88% (1397 of 1589). Misclassification error of tree: 31% ((192+595)/2556). Ratio of true classification of tree: 69% ((1397+372)/2556).

smoking affected failure in classes as well. In another study from our regions, Tot *et al*⁶ showed a significant correlation between tobacco smoking and high ratio of associated alcohol consumption, not other drugs. Our study groups showed totally 844 students (25.5% of responders), but 287 of currents smokers (55.1%) reported associated substance use.

There are different predictors that have been reported to be associated with smoking onset.^{2 6 24} One important longitudinal survey analysis study showed that the prevalence of regular smokers increased from 1.7% to 22% among boys and from 1.6% to 38.2% among girls.²⁵ Although it was not a longitudinal survey, our study results suggested an important increasing trend of smoking rate with increasing age, especially in boys. These observations were suggested by some other reports from our country and some other countries as well.^{6 24 28-30}

The impact of parental smoking has been studied in a wide range of contexts in a large number of studies with a variety of outcomes. Studies examining the effect of paternal and maternal smoking separately have reported both to be significant,^{32 33} non-significant,^{34 35} or each one significant while the other was not.^{36 37} Some of the inconsistencies may reflect sex specific differences: several studies reported a significant effect only for girls^{38 39} whereas none found the reverse. It is unclear whether parental smoking has a stronger influence when it occurs in the same sex parent also.^{40 41} A dose-response effect may also be present, with a stronger influence if both parents smoke.^{42 43-45} Our study showed that only sibling smoking has an important effect on lifetime smoking but both parents and sibling smoking have important effects on current smoking of students. However, as rare reported data we showed a high important effect of severe medical illness of close relatives related to smoking on current smoking students. We also saw the important effect

of relationship problems of students on current smoking (for details see table 5).

In contrast with some studies that found that smoking increases with decreasing socioeconomic status, others showed a high prevalence of tobacco smoking among students coming from educated backgrounds.^{46 47} The results are also in agreement with a previous study, which found that 60% of the young Lebanese women who smoked belonged to the upper middle class while 22.5% belonged to the lower class.⁴⁷ We speculate whether educated parents who probably belong to a higher socioeconomic class adopt more open attitudes towards youth smoking, and thus do not actively oppose this habit.⁴⁸ A limitation of this study was that peer influences, which have been shown to be important for adolescent smoking, were not assessed.⁴⁹ Our study showed no important effect of socioeconomic status of family on either lifetime, or current smoking of high school students in Mersin, in contrast with one previous report of our region.⁶ We also saw an important effect of decreased father and mother education levels on increased current smoking of students (details can be seen in table 4). As it is frequently shown in the literature, information regarding socioeconomic level are subjective data with little statistical reliability. We did not use assessment methods such as the socioeconomic status score in this study because they exceed the scope of our study and these methods still do not have valid sensitivity and reliability tests that are confirmed by Turkish Ministry of Health.

Parent-child or family based interventions have been shown to be effective in preventing or reducing adolescent health behaviours including smoking, obesity, dietary changes, and alcohol use.⁵⁰⁻⁵² On the other hand, the personal income of adolescents have been associated with adolescent smoking: young people with more spending money showed higher levels of smoking presumably because money is needed for the purchase of tobacco.³² This study pointed out the important effect of parents' and siblings' smoking behaviour on current smoking, besides the significant effects of bad relationships and violence of subjects. We also found that having own money has a risk for lifetime smoking but has no important effect on current smoking in our sample.

The effect of household size on risk of smoking is unclear: studies have noted larger families to be associated with lower or higher levels of smoking, or have reported no significant relation. Higher levels of parental socioeconomic variables, such as education and social class, have often been found to be inversely related to smoking status in adolescents.³² This study showed that increased household size and later birth rank have important effect on current smoking with different scores.

Factors in the environment that potentially influence initiation and maintenance of smoking by adolescents have been the focus of many investigations as early studies showed the importance of peer and parental smoking as risk factors.^{32 53} We found some personal (that is, age, male sex) and environmental factors (that is, school type, smoking of family members, poor or satisfactory socioeconomic status of

What this paper adds

- The smoking students commonly showed prominent family characteristics and important behavioural problems with both their parents and friends.
- The significantly important factors that affect current smoking in high school students were increased household size, late birth rank, certain school types, low academic performance, increased ratio of second hand smoking, and stress (especially reported as separation from a close friend or violent behaviour in home).
- CART methodology showed the importance of some neglected sociodemographic factors with a good classification capacity. In the process of the estimation of target population for the development of main health care policies for smoking, CART methodology can be recommended as a new method.

families, bad relationship between family members, and violent behaviour in home) have important effects on past smoking.

Despite public health efforts that have produced dramatic declines in the prevalence of tobacco smoking, it still remains the worldwide leading cause of preventable disease and death.⁵⁴ The results of these studies should be enlightening on the future health care policies about smoking. At this point, the estimation of risk factors and primary target population are highly important. The classification and regression tree methodology, besides the logistic regression analysis answered this question. This analysis also highlighted the effects of some factors (for example, parents' age) on smoking of students.

Higher smoking rates in these subjects, who perceived their academic performance as poor is also an important finding that was also reported in some previous studies.⁵⁵⁻⁵⁶ This study also reported using more objective measures of evaluating academic performance that showed that smokers were less successful at school life and had accompanying behavioural problems.⁶ Despite some contradictory reports, we showed low academic performance is correlated with high rates of tobacco smoking of students. We also showed that academic performance score of students was an important splitting criteria in classification and regression tree methodology.

In addition to traditional regression analysis methods, classification and regression tree methodology achieved the estimation for the most probable population for smoking with a good classification ratio (69%). The splitter parameters of the tree also showed the importance of some neglected factors (for example, school type, parents' age, mother's education, smoking of sibling, birth rank, employment of father, etc) on current smoking. As the main message of the study, we can say that the CART method is useful for health policy makers. We tried to evaluate the application area of the CART in a population based study on smoking taking into consideration that it is an important health problem worldwide. We are of the opinion that statistical details, including the application of the CART, have not weakened the manuscript; in contrast this approach might increase the impact of the paper. Many of the social factors emerging as main contributors to the smoking habit can be the target of campaigns that make use of the results of this method. Also regional and cultural differences can be seen more easily using this method. Also this method can be considered for the other healthcare studies; for example, epidemiology of headache, depression, movement disorders.

Authors' affiliations

C Özge, Department of Chest Disease, Mersin University School of Medicine, Mersin, Turkey

F Toros, Child and Adolescent Psychiatry, Mersin University School of Medicine

E Bayramkaya, Department of Psychology, Mersin University School of Science and Letters

H Çamdeviren, Biostatistics, Mersin University School of Medicine

T Şaşmaz, Public Health, Mersin University School of Medicine

Funding: none.

Conflicts of interest: none.

REFERENCES

- 1 World Health Organisation. *The world health report. Life in the 21st century: a vision for all*. Geneva: World Health Organisation, 1998.
- 2 Çelik P, Esen A, Yorgancıoğlu A, et al. Attitudes of high school students towards smoking in Manisa. *Toraks Dergisi* 2000;1:61-6.
- 3 Pekşen Y. Sigara içiminin nedenleri, epidemiyolojisi, pasif içicilik. In: Tür A, ed. *Sigaranın Sağlığa Etkileri ve Bırakma yöntemleri*. İstanbul: Logos Yayıncılık, 1995:29-53.
- 4 Önder R, Egemen A. Lise çağı gençliğinin sigara içme durumu. *Türk Hij Den Biyol Derg* 1987;44:121-6.
- 5 Tümerdem Y, Ayhan B, Emekli Uve ark. İstanbul kentinde öğrenim gençliğinde sigara içme olayı etkinliklerinin araştırılması. *Solunum* 1986;6:412-16.
- 6 Tot S, Yazici K, Yazici A, et al. Psychosocial correlates of substance use among adolescents in Mersin, Turkey. *Public Health* 2004;118:588-93.
- 7 Patten CA, Lopez K, Thomas JL, et al. Reported willingness among adolescent nonsmokers to help parents, peers, and other to stop smoking. *Prev Med* 2004;39:1099-106.
- 8 Farkas AJ, Gilpin EA, White MM, et al. Association between household and workplace smoking restrictions and adolescent smoking. *JAMA* 2000;284:717-22.
- 9 Schuster MA, Franke T, Pham CB. Smoking patterns of household members and visitors in homes with children in the United States. *Arch Pediatr Adolesc Med* 2002;156:1094-100.
- 10 Johnston LD, O'Malley PM, Terry-McElrath YM. Methods, locations, and ease of tobacco access for American youth, 1997-2002. *Am J Prev Med* 2004;27:267-76.
- 11 Van Teigkingen E, Engs R. Correlates of alcohol, tobacco and marijuana use among Scottish post-secondary helping-profession students. *J Stud Alcohol* 1997;58:435-44.
- 12 Newman I, Ward J. The influence of parental attitude and behavior on early adolescent tobacco smoking. *J Sch Health* 1989;59:150-3.
- 13 Emmons K. Predictors of smoking among US college students. *Am J Public Health* 1998;88:104-7.
- 14 Maziak W, Mzayek F. Characterization of the smoking habit among high school students in Syria. *Eur J Epidemiol* 2000;16:1169-76.
- 15 McDermott RJ, Sarvelo PD, Hoatt PN, et al. Multiple correlates of tobacco use among high school students. *J Sch Health* 1992;62:146-51.
- 16 Escobedo LG, Peddicord JP. Smoking prevalence in US birth cohorts: the influence of gender and education. *Am J Public Health* 1996;86:231-6.
- 17 Acierno RA, Kilpatrick DG, Resnick HS, et al. Violent assault, post-traumatic stress disorder and depression: risk factors for tobacco use among adult women. *Behav Modif* 1996;20:368-84.
- 18 Xinggang C, Xiaoming L, Stanton B, et al. Tobacco smoking among rural-to-urban migrants in Beijing, China. *Prev Med* 2004;39:666-73.
- 19 Breiman L, Friedman JH, Olshen RA, et al. *Classification and classification and regression trees*. Belmont, CA: Wadsworth, 1984:358.
- 20 Patrick DL, Cheadle A, Thompson DC, et al. The validity of self-reported smoking: a review and meta-analysis. *Am J Public Health* 1994;84:1086-93.
- 21 Hatziandreu EJ, Pierce JP, Fiore MC, et al. The reliability of self-reported tobacco consumption in the United States. *Am J Public Health* 1989;79:1020-3.
- 22 Strecher VJ, Becker MH, Clark NM, et al. Using patients' descriptions of alcohol consumption, diet, medication compliance, and tobacco smoking: the validity of self-reports in research and practice. *J Gen Intern Med* 1989;4:160-6.
- 23 Velicer WF, Prochaska JO, Rossie JS, et al. Assessing outcome in smoking cessation studies. *Psychol Bull* 1992;111:23-41.
- 24 Valdivia G, Simonetti F, Cumsille P, et al. Smoking habit in school age children, in Chile. *Rev Med Chil* 2004;132:171-82.
- 25 Nebot M, Tomas Z, Ariza C, et al. Factors associated with smoking onset: 3 year cohort study of schoolchildren. *Arch Bronconeumol* 2004;40:495-501.
- 26 Kraus B, Wodarz N, Kuchlmaier K, et al. Use of tobaccos, alcohol and drugs by adolescents—a questionnaire survey among 9th grade of students. *Psychiatr Prax* 2004;31(suppl 1):99-101.
- 27 Bogui P, Yessoh M, Tuo N, et al. Tobacco consumption of school pupils and students between 8 and 22 years of age in Abidjan, 2002. *Rev Mal Respir* 2004;21:693-703.
- 28 Hughes JR. Nicotine-related disorders. In: Sadock BJ, Sadock VA, eds. *Kaplan and Sadock's comprehensive textbook of psychiatry*. Philadelphia: Lippincott Williams and Wilkins, 2000;1:1033-8.
- 29 Thomas M, Goddard E, Hickman M, et al. *General household survey: 1992*. London: HMSO, 1993.
- 30 Engs RC, Van Teijlingen E. Correlates of alcohol, tobacco and marijuana use among Scottish postsecondary helping-profession students. *J Stud Alcohol* 1997;58:435-44.
- 31 Molcho M, Harel Y, Dina LO. Substance use and youth violence. A study among 6th to 10th grade Israeli school children. *Int J Adolesc Med Health* 2004;16:239-51.
- 32 Tyas SL, Pederson LL. Psychosocial factors related to adolescent smoking: a critical review of the literature. *Tab Control* 1998;7:409-20.
- 33 Tuakli N, Smith MA, Heaton C. Smoking in adolescence: methods for health education and smoking cessation. A MIRNET study. *J Fam Pract* 1990;31:369-74.
- 34 Bolvin GJ, Baker E, Goldberg CJ, et al. Correlates and predictors of smoking among black adolescents. *Addict Behav* 1992;17:97-103.
- 35 Jensen EJ, Overgaard E. Investigation of smoking habits among 14-17-year-old boarding school pupils: factors which influence smoking status. *Public Health* 1993;107:117-23.
- 36 Hoyer SJ, Gaffney LR. Factors associated with smoking behavior in adolescent girls. *Addict Behav* 1988;13:139-45.
- 37 Hops H, Tildesley E, Lichtenstein E, et al. Parent-adolescent problem-solving interactions and drug use. *Am J Drug Alcohol Abuse* 1990;16:239-58.
- 38 Charlton A, Blair V. Predicting the onset of smoking in boys and girls. *Soc Sci Med* 1989;29:813-18.
- 39 Swan AV, Creeser R, Murray M. When and why children first start to smoke. *Int J Epidemiol* 1990;19:323-30.
- 40 Burchfiel CM, Higgins MW, Keller JB, et al. Initiation of tobacco smoking in children and adolescents of Tecumseh, Michigan. *Am J Epidemiol* 1989;130:410-15.

- 41 **Green G**, Macintyre S, West P, *et al.* Like parent like child? Associations between drinking and smoking behaviour of parents and their children. *Br J Addict* 1991;**86**:745–58.
- 42 **Ogawa H**, Tominaga S, Gellert G, *et al.* Smoking among junior high school students in Nagoya, Japan. *Int J Epidemiol* 1988;**17**:814–20.
- 43 **Shibata A**, Fukuda K, Hirohata T. Smoking habits among senior high school students and related factors. *Kurume Med J* 1990;**37**:129–40.
- 44 **Dusenbury L**, Kerner JF, Baker E, *et al.* Predictors of smoking prevalence among New York Latino youth. *Am J Public Health* 1992;**82**:55–8.
- 45 **Quine S**, Stephenson JA. Predicting smoking and drinking intentions and behavior of pre-adolescents: the influence of parents, sibling, and peers. *Fam Sys Med* 1990;**8**:191–200.
- 46 **Malmstadt JR**, Nordstrom DL, Carty DC, *et al.* Tobacco smoking in Wisconsin: the influence of race, ethnicity and socioeconomic. *Wis Med J* 2001;**100**:29–33.
- 47 **Tamim H**, Terro A, Kassem H, *et al.* Tobacco use by university students, Lebanon, 2001. *Addiction* 2003;**98**:933.
- 48 **Kandela P**. Nargile smoking keeps Arabs in the wonderland. *Lancet* 2000;**356**:1177.
- 49 **Beal AC**, Ausiello J, Perrin JM. Social influences on health-risk behaviors among minority middle school students. *J Adolesc Health* 2001;**28**:474–80.
- 50 **Epstein LH**, Valoski A, Wing RR, *et al.* Ten-year follow-up of behavioral, family-based treatment for obese children. *JAMA* 1990;**264**:2519–23.
- 51 **Baranowski MD**. Adolescents' attempted influence on parental behaviors. *Adolescence* 1978;**52**:585–604.
- 52 **Salo PM**, Xia J, Johnson CA, Li Y, *et al.* Respiratory symptoms in relation to residential to residential coal burning and environmental tobacco smoke among early adolescents in Wuhan, China: a cross-sectional study. *Environmental Health: A Global Access Science Source* 2004;**3**:14.
- 53 **Newman IM**, Ward JM. The influence of parental attitude and behavior on early adolescent tobacco smoking. *J Sch Health* 1989;**59**:150–2.
- 54 **Wetter DW**, Cofta-Gunn L, Irvin JE, *et al.* What accounts for the association of education and smoking cessation? *Prev Med* 2005;**40**:452–60.
- 55 **Hops H**, Tildesley E, Lichtenstein E, *et al.* Parent-adolescent problem-solving interactions and drug use. *Am J Drug Alcohol Abuse* 1990;**16**:239–58.
- 56 **Glynn TJ**. Adolescent drug use and the family environment: a review. *J Drug Issues* 1985;**2**:271–95.