SOME RELEVANT FACTORS IN THE CONSUMPTION AND NON CONSUMPTION OF NICOTINE IN ADOLESCENCE

Elena Gervilla, Berta Cajal, and Alfonso Palmer

Department of Psychology, University of the Balearic Islands (Spain)

Abstract

Tobacco is the second most consumed substance among Spanish adolescents and its effects on health are well known. The aim of this study is to analyze the predictor value of different personal, family and environmental factors on tobacco consumption in the adolescent population using appropriate modelling techniques and strategies included in *Data Mining*. The total sample is made up of 9,300 students aged between 14 and 18 years. The adolescents anonymously answered a questionnaire which asked about the frequency of use of addictive substances as well as a series of psychosocial variables. Tobacco consumption by friends, frequency of nights out, age, production of prohibited, impulsive behaviour and the control exerted by parents bear a statistically significant influence on the number of cigarettes consumed per week and constitute risk factors for initiating consumption. Future studies should analyze, using the techniques proposed in this work and longitudinal studies, whether the tougher anti-smoking laws have been transformed into a decrease in tobacco consumption in adolescents.

Keywords: Nicotine, adolescence, Data Mining, Decision Tree

Resumen

El tabaco es la segunda sustancia más consumida entre los adolescentes españoles y sus efectos sobre la salud son bien conocidos. El objetivo de este estudio es analizar el valor predictivo de diferentes variables personales, familiares y ambientales sobre el consumo de tabaco en la población adolescente mediante técnicas de modelado adecuadas y estrategias incluidas en *Data Mining*. La muestra total está formada por 9.300 estudiantes con edades comprendidas entre los 14 y los 18 años. Los adolescentes contestaron de forma anónima un cuestionario que preguntaba por la frecuencia de uso de diferentes sustancias adictivas así como una serie de variables psicosociales. El consumo de tabaco por parte de los amigos, la frecuencia de salidas nocturnas, la edad, la emisión de conductas prohibidas e impulsivas y el control ejercido por parte de los padres, influyen de forma estadísticamente significativa en el número de cigarrillos consumidos a la semana y constituyen factores de riesgo para iniciar el consumo. Futuros estudios deberían analizar mediante las técnicas propuestas en este trabajo y estudios longitudinales si el endurecimiento de las leyes antitabaco se ha traducido en una disminución del consumo de tabaco de los adolescentes.

Palabras clave: Nicotina, adolescencia, Minería de datos, Árboles de decisión.

Correspondence: Elena Gervilla. Facultad de Psicología. Ctra. Valldemossa, km 7.5, 07122 Palma de Mallorca, Spain. E-mail: elena.gervilla@uib.es

Introduction

Tobacco is the second most consumed substance amongst Spanish adolescents (Delegación del Gobierno para el Plan Nacional Sobre Drogas, 2008). The proportion of smokers increases with age and is at its greatest among girls at all ages. This consumption is found to be influenced by a series of psychosocial variables out of which pressure by the group of friends, some patterns of paternal action and certain personality variables stand out.

Actually, one of the more outstanding explanatory variables due to their influence on drug consumption in adolescence is use of the same by the peer group and/or ease of access to the same (Bricker, Andersen, Rajan, Sarason, & Peterson, 2007; Carvajal & Granillo, 2006; Ciairano, Bosma, Miceli, & Settani, 2008; De Vries, Engels, Kremers, Wetzels, & Mudde, 2003; Dick et al., 2007; Kobus, 2003). However, Mercken, Snijders, Steglich, Vertiainen, and de Vries (2010) have found that this influence would be mediated according to gender, in the sense that only girls would be affected by the group pressure to smoke.

On the other hand, recent studies continue to insist on the existence of different variables related with the family environment as risk and protection factors involved in drug consumption by adolescents (Avenoli & Merikangas, 2003; Becoña, 2002; Darling & Cumsille, 2003; Fernández, Secades, Vallejo, & Errasti, 2003; Jiménez, Musitu, & Murgui, 2008; Martínez, Fuertes, Ramos, & Hernández, 2003; Muñoz & Graña, 2001). However, it seems that the relationship between parental practices and consumption behaviour could be mediated by the number of friends who consume substances (Simons-Morton, 2007).

Moreover, certain personality traits, such as antisocial behaviour, thrill seeking or impulsivity have been related in different research studies with the use of addictive substances in adolescence (Doran, McCharge, & Cohen, 2007; Farrell, Sullivan, Esposito, Meyer, & Valois, 2005; Martínez & Alonso, 2003; Nadal, 2008; Otten, Wanner, Vitaro, & Engels, 2009; Peña, Andreu, & Graña, 2009; Timmermans, Van Lier, & Koot, 2008).

Tobacco consumption has well known health risks, especially when its initiation takes place at an early age. Therefore, knowing the factors associated with its consumption is of special interest. Many studies have approached research into the

factors related to the use of nicotine in adolescence using traditional data analysis techniques. Nevertheless, there is still scarce bibliography taking advantage of the benefits of *Data Mining* methodology to discover interesting relationships in the data (Buscema, 2002; Gervilla et al., 2009; Gervilla & Palmer, 2009; Gervilla & Palmer, 2010; Gervilla, Cajal, Roca, & Palmer, 2010; Kitsantas, Moore, & Sly, 2007; Larsson, Lilja, Borg, Buscema, & Hamilton, 2001; Palmer, Montaño, & Calafat, 2000; Sears & Anthony, 2004).

The aim of this study is to analyze the predictive value of different personal, family and environmental variables on tobacco consumption in the adolescent population using appropriate modelling techniques and strategies included in *Data Mining*.

Method

Participants

A random sample by conglomerates of schools in the island of Mallorca was conducted and 47 schools were selected out of a total of 122. The total sample was made up of 9,300 students aged between 14 and 18 years. The sample size represented 41.16% of the population size it was extracted from ($N = 22,593$).

After eliminating the very unreliable answers of 6 adolescents; the useful sample was made up of 9,284 adolescents. In Table 1 you can consult a summary of the most important socio-demographic characteristics of the sample.

Procedure

The adolescents answered a questionnaire which asked about the frequency of use of different addictive substances as well as a series of psychosocial variables.

In this study the data are analyzed using statistical modelling of count data (*Stata* 10.0 program) and using a decision tree (SPSS program15.0).

E. Gervilla et al.

Table 1. Demographic information concerning the sample.

Gender	Masculine	47.1%
	Feminine	52.9%
Academic year	2º ESO	1.5%
	3º ESO	35.8%
	4º ESO	26.3%
	1º Baccalaureate	24.1%
	2º Baccalaureate	9.8%
	Middle Grade Vocational Training	2.5%
Population	Urban	47.5%
	Rural	52.5%
Economic level	High	4.3%
	Medium-high	19.4%
	Medium	59.5%
	Medium-low	5.9%
	Low	0.9%
Who you live with	Parents	83.7%
	Grandparents	0.6%
	Friends	0.1%
	Alone	0.2%
	With the mother	12.7%
	With the father	1.8%
	Others	0.6%

Variables

We took into account environmental (tobacco consumption by friends, ease of access, frequency of nights out midweek and at the weekend and alcohol consumption), family (parents' patterns of upbringing) and personal (gender, age and personality factors) variables.

Tobacco consumption by the peer group was a categorical variable, represented in the model by the variables codified as *friend1* ("some of my friends consume tobacco"), *friend2* ("half of my friends consume tobacco"), *friend3* ("most of my friends consume tobacco") and *friend4* ("all my friends consume tobacco") with "none of my friends consume tobacco" as the reference category.

The parents' style of upbringing was introduced in the model through a series of variables referring to the parents that the adolescents had to assess using the words "Never", "Sometimes" or "Always", independently for the father and for the mother. Thus, the participants informed as to whether their parents were controlling, whether

they spent time with them, whether they did things with them and whether they spent quite a lot of time at home.

The information concerning the personality was included in the model through twenty dichotomous variables which referred to impulsivity, thrill seeking, antisocial behaviour and self-concept.

Ease of access, alcohol consumption and gender were introduced in the model through binary variables (0/1). The gender was codified as 0 = boy and 1 = girl. Age and frequency of nights out midweek and at weekends were introduced in the models using quantitative variables.

The dependent variable was the number of cigarettes consumed a week.

With the aim of organizing the variables that predict tobacco consumption in adolescence, a decision tree was calculated, which, based on the aforementioned variables, classified the subject as a smoker or non smoker.

Results

The prevalence of tobacco consumption in the whole sample is 25%. Table 2 shows the prevalence of tobacco consumers in the sample according to age and gender. It can be seen how the prevalence of consumption increases with age, with a greater percentage of girl smokers than boys at all ages.

Table 2. Prevalence of adolescent smokers in the sample ($n = 9,300$).

	14	15	16	17	18	Total
Boys	9.6%	16.4%	23.9%	24.2%	29.4%	19.2%
Girls	16.3%	27.1%	36.8%	37.1%	44.3%	30.1%
Total	13.2%	22.0%	30.6%	31.2%	37.0%	25.0%

Table 3 shows the mean number of cigarettes consumed per week among the adolescents who reported they were smokers (see Table 3). In general, we can observe a tendency for the mean weekly cigarette consumption to increase with age, in both genders.

Table 3. Consumption of cigarettes per week among adolescent smokers (mean, standard deviation and sample size).

	14	15	16	17	18	Total
Boys	13.9	17.2	21.1	28.9	22.9	21.3
	(SE = 26.5)	(SE = 32.9)	(SE = 35.5)	(SE = 40.2)	(SE = 42.0)	(SE = 35.9)
	N = 84	N = 193	N = 277	N = 177	N = 73	N = 804
Girls	11.5	18.1	25.7	21.3	30.1	21.5
	(SE = 24.2)	(SE = 32.7)	(SE = 39.5)	(SE = 34.6)	(SE = 42.8)	(SE = 35.9)
	N = 172	N = 339	N = 469	N = 325	N = 116	N = 1,421
Total	12.3	17.7	23.9	24.1	27.3	21.4
	(SE = 24.9)	(SE = 32.4)	(SE = 38.1)	(SE = 36.8)	(SE = 42.5)	(SE = 35.9)
	N = 256	N = 536	N = 748	N = 503	N = 189	N = 2,232

One of the basic assumptions of the Poisson regression model is that of equidispersion, which can be evaluated through the Cameron and Trivedi regression test (1990). After confirming the failure of this assumption, we compared the models called *Poisson Regression Model* (PRM), *Negative Binomial Regression Model* (NBRM), *Zero Inflated Poisson* (ZIP) and *Zero Inflated Negative Binomial* (ZINB) and it could be seen that the ZINB model was the one that best fitted the data (Mullahy, 1986; Greene, 1994; Lambert, 1992).

In Table 4 (Percentage change in the expected count for smokers) it can be seen that there are environmental (consumption by friends and frequency of nights out), personal (age, antisocial behaviour and impulsivity) and family (control exerted by the parents) variables which bear a statistically significant influence on cigarette consumption. Thus, it can be seen that if all the friends consume tobacco the weekly number of cigarettes consumed rises 374.7%, and 301.4% if most of the friends smoke, with respect to if no friend is a smoker, so long as the rest of the factors remain constant. Likewise, for each night an adolescent goes out at the weekend the weekly number of cigarettes smoked increases 32.4%; whereas for each year age is increased, 28.9% more cigarettes are smoked a week, so long as the other factors remain constant. Finally, if the adolescents report they like a wild life and lack of inhibition and/or carry out forbidden, illegal behaviour, the number of cigarettes consumed a week also increases (25.5% and 27.8%, respectively).

It is worth noting that, of the family variables analyzed, the control exerted by the parents is the factor that diminishes the weekly consumption of cigarettes in a statistically significant way (23.6% for paternal control and 35.9% for maternal

—

control). Likewise, doing sports/extreme activities constitutes a factor that decreases nicotine consumption.

Table 4. Percentage change in the expected count for smokers.

Variable	p	%	Variable	p	%
Gender (girl)	.638	-4.9	Nights out midweek	.106	8.5
Age	.000*	28.9	Ease of access	.291	-21.4
Father always controlling	.259	-15.9	I do many things well	.455	-7.9
Father sometimes controlling	.038	-23.6	Satisfied with myself	.779	3.3
Mother always controlling	.040	-35.9	I like myself physically	.706	-3.9
Mother sometimes controlling	.056	-32.8	I have good qualities	.631	6.3
Father always spends time with me	.139	-26.0	I'm a failure as a person	.433	-12.3
Father sometimes spends time with me	.137	-22.0	I do things impulsively	.885	-1.6
Mother always spends time with me	.468	-15.3	Difficulty keeping still	.186	14.7
Mother sometimes spends time with me	.792	-5.4	I say things without thinking	.237	13.2
Father always does things with me	.251	-20.8	Impatience to get something	.551	6.8
Father sometimes does things with me	.867	2.3	I normally plunge into things	.182	14.9
Mother always does things with me	.236	26.1	I do sports/extreme activities	.005	-24.8
Mother sometimes does things with me	.179	21.1	I like exciting, new experiences	.103	26.5
Father is always at home quite a lot	.588	9.9	I like activities that involve danger	.552	6.5
Father is sometimes at home quite a lot	.276	-14.6	Boredom at always doing the same things	.330	-10.5
Mother is always at home quite a lot	.916	-2.3	I like a wild life and lack of inhibition	.029	25.5
Mother is sometimes at home quite a lot	.392	-17.0	I cause disturbances and make rackets	.320	10.8
All my friends smoke	.002	374.7	I do forbidden, illegal things	.032	27.8
Most of my friends smoke	.005	301.4	I break, burn or deteriorate other people's property	.073	-29.6
Half of my friends smoke	.078	141.4	I fight with and insult others	.571	7.4
Few of my friends smoke	.216	90.0	I answer older people back	.761	-4.0
Nights out at the weekend	.000*	32.4	Alcohol consumption	.907	1.6

* p < .001.

The ZINB model - as well as informing of the factors that increase and decrease the number of cigarettes consumed on a weekly basis - calculates the risk and protection factors associated with cigarette consumption. The risk factors that turned out to be

statistically significant are shown in Table 5 (Change in the *odds* of continuing as a non smoker).

The percentages that appear in Table 5 show how much the odds of continuing being a non smoker increase (or decrease). Thus, it is worth noting that as the number of smoker friends increases, the odds of continuing as a non smoker decrease (if the other factors remain constant). Besides, it can be seen that if the adolescent is a girl she has 46.2% lower odds of continuing without consuming tobacco. Likewise, for each night adolescents go out at the weekend, the odds of continuing without smoking decrease 18.4%. Lastly, impulsive behaviour, committing forbidden behaviour and alcohol consumption also reduce the odds of maintaining a non smoking state in a statistically significant way. On the other hand, if the adolescent is self-satisfied, the odds of continuing without consuming tobacco rises 36.1%.

Table 5. Change in the *odds* of continuing as a non smoker.

Variable	p	%	Variable	p	%
Gender (girl)	.000*	-46.2	Nights out midweek	.004	19.3
Age	.867	-0.9	Ease of access	.779	-6.8
Father always controlling	.251	-20.8	I do many things well	.780	3.7
Father sometimes controlling	.468	-11.6	Satisfied with myself	.040	36.1
Mother always controlling	.603	15.5	I like myself physically	.447	10.2
Mother sometimes controlling	.624	14.0	I have good qualities	.898	2.1
Father always spends time with me	.811	6.6	I'm a failure as a person	.246	-22.8
Father sometimes spends time with me	.725	-7.4	I do things impulsively	.001	-33.9
Mother always spends time with me	.288	37.9	Difficulty keeping still	.681	5.3
Mother sometimes spends time with me	.282	34.3	I say things without thinking	.156	20.2
Father always does things with me	.146	47.0	Impatience to get something	.595	7.3
Father sometimes does things with me	.093	35.7	I normally plunge into things	.238	-14.2
Mother always does things with me	.284	-24.0	I do sports/extreme activities	.793	3.4
Mother sometimes does things with me	.168	-22.9	I like exciting, new experiences	.389	-12.9
Father is always at home quite a lot	.534	14.3	I like activities that involve danger	.090	-20.5
Father is sometimes at home quite a lot	.433	-13.9	Boredom at always doing the same things	.609	-6.3
Mother is always at home quite a lot	.407	-20.8	I like a wild life and lack of inhibition	.050	-23.9
Mother is always at home quite a lot	.407	-20.8	I like a wild life and lack of inhibition	.050	-23.9

Table 5 (Continued). Change in the *odds* of continuing as a non smoker.

Variable	P	%	Variable	P	%
Mother is always at home quite a lot	.407	-20.8	I like a wild life and lack of inhibition	.050	-23.9
Mother is sometimes at home quite a lot	.817	6.6	I cause disturbances and make rackets	.877	-2.1
All my friends smoke	.000*	-87.2	I do forbidden, illegal things	.005	-35.8
Most of my friends smoke	.000*	-84.6	I break, burn or deteriorate other people's property	.501	-17.8
Half of my friends smoke	.000*	-79.5	I fight with and insult others	.934	-1.4
Few of my friends smoke	.035	-57.5	I answer older people back	.786	4.8
Nights out at the weekend	.005	-18.4	Alcohol consumption	.000*	-69.3

* $p < .001$.

Among the most popular *Data Mining* techniques we find decision trees (Han & Kamber, 2006; Kantardzic, 2003; Witten & Frank, 2005; Ye, 2003), which offer a concise way of defining groups that are consistent in their attributes but which vary in terms of the dependent variable. Using this technique a set of rules concerning the decisions to be taken into account in order to assign a certain element to a class is represented graphically. One of the most outstanding advantages in the decision tree models is their descriptive nature, which makes it possible to understand and easily interpret the decisions made by the model, as we have access to the rules that are used in the predictive task.

Since the prevalence of tobacco consumption in the sample was 25% and with the aim that the trees calculated should be able to predict both abstinence and the conduct of tobacco consumption, in order to carry out the decision tree model, a balanced sample of smokers and non smokers was built. That is, all the adolescent smokers were selected ($n = 2,220$) and, out of the total non smokers, a random sample was chosen ($n = 2,220$).

Figure 1 shows a decision tree for tobacco consumption ($n = 4,440$) calculated using the *Classification And Regression Trees* (CART) algorithm (Breiman, Friedman, Olshen, & Stone, 1984) which makes it possible to predict with a risk of 0.273 whether or not an adolescent will smoke based on the number of consumers in the peer group, alcohol consumption, carrying out forbidden behaviour, gender and the frequency of nights out.
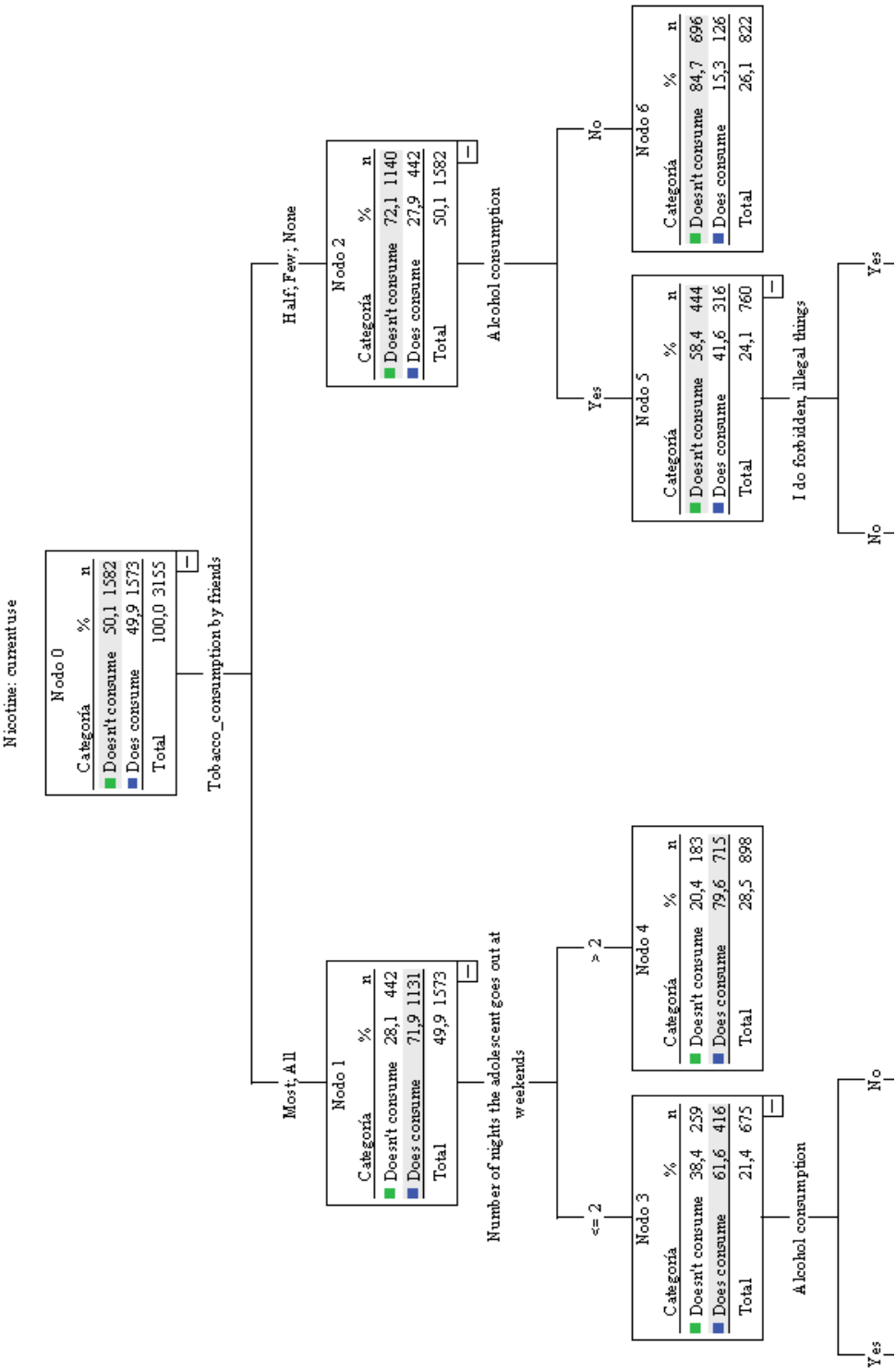
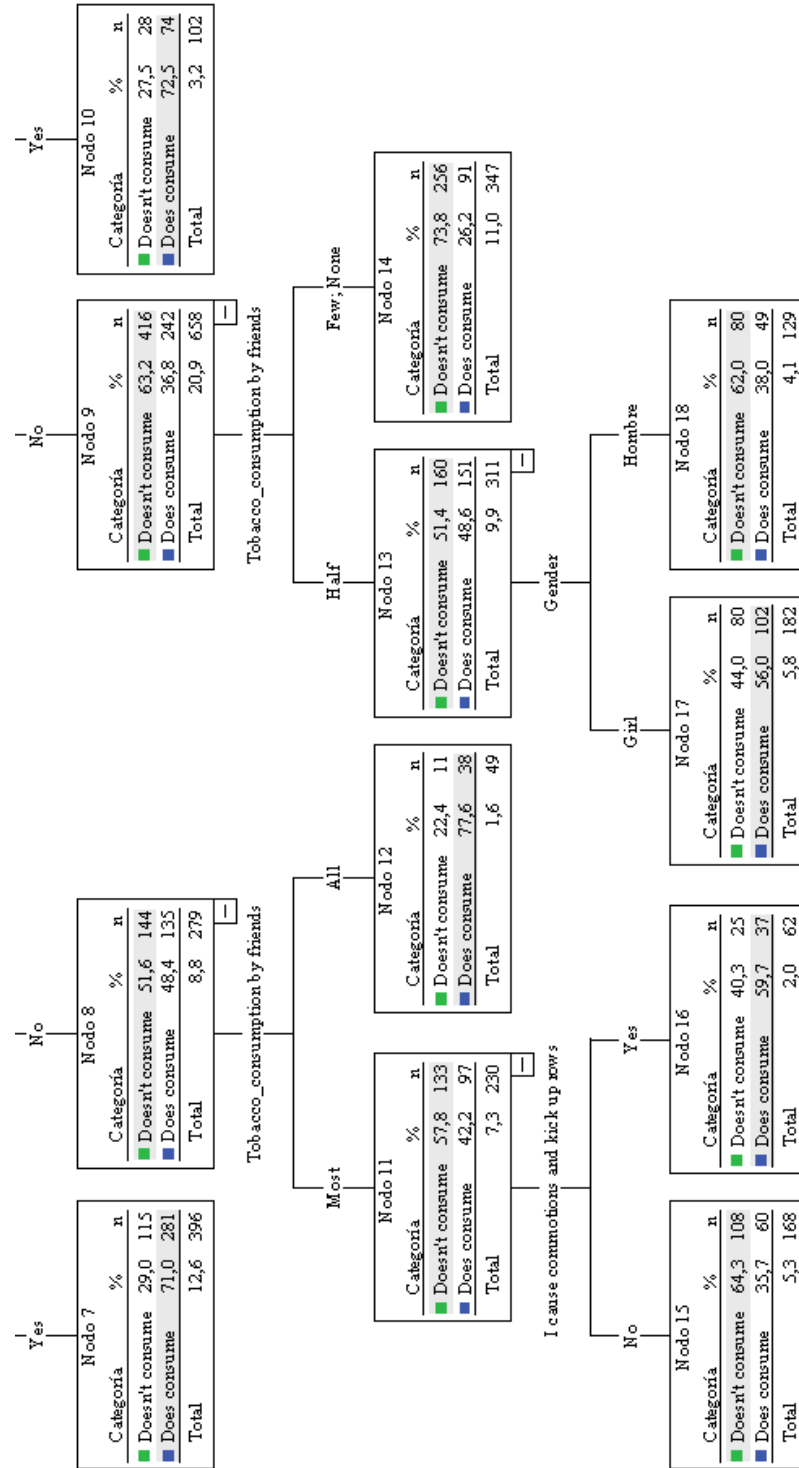**Figure 1:** Decision tree for tobacco consumption (n = 4,440)

Figure 1 (Continued): Decision tree for tobacco consumption (n = 4,440)

Discussion

The aim of this study is to analyze the influence of psychosocial risk factors on tobacco consumption in adolescence using appropriate *Data Mining* modelling techniques and tools. Along the lines of other research, the results highlight the influence of environmental variables such as tobacco consumption by the peer group, ease of access or the frequency of nights out and the relationship between alcohol and tobacco consumption (Geckova et al., 2005; Hoffman, Welte, & Barnes, 2001; Molyneux et al*., 2004*). Moreover, a relationship, albeit more moderate, between tobacco consumption and the production of socially deviant behaviour can be observed. By way of protection factors, paternal action, doing sport and having a good self-concept stand out.

Concerning the influence of friends on the specific consumption of tobacco, many theories have been proposed to explain it (Hoffman, Sussman, Unger, & Valente, 2006). Recently, Otten, Engels, and Prinstein (2009), in a longitudinal study, found that having a predominantly smoking peer group significantly increases the risk of overestimating the prevalence of the smoking habit and that this overestimation is associated with a greater risk of smoking in the future, which was also found by Molyneux et al. (2004) in a longitudinal study, confirming the influence of friends who smoke on the risk of smoking.

Different studies have found evidence of the so-called "false consensus effect" among smokers, showing that people who smoke have greater odds of overestimating the prevalence of smoking, including adolescents (Castrucci, Gerlach, Kaufman, & Orleans, 2002, Prinstein & Wang, 2005). One theory points to this overestimation being the result of selective exposure by people; that is, adolescents who smoke would predominantly choose to join up with other smokers; whereas others suggest that people who are exposed to a high prevalence of smokers among their closest people have greater odds of assuming that these perceptions are representative of the population as a whole, leading to an overestimation of the prevalence of smoking.

In this sense, most of the research conducted on the relationship between peers has focused on the similarity of behaviour between friends, which can be explained from two theories: the effect of socialization (adolescents become like their friends because of their influences) or the selection effect (adolescents choose friends who

behave and think in the same way as they do). Even though both tendencies were considered alternatives in the past, it is now recognized that both are produced in the processes that take place in peer groups (Hall & Valente, 2007; Mercken, Candel, Willems, & De Vries, 2007; Simons-Morton, 2007).

The interest of this study lies in the availability of a large sample which makes up nearly 50% of the population studied, and in the contribution of the use of decision tree models, included in *Data Mining* methodology, which are capable of handling large volumes of data and which constitute a good method for classifying and predicting (see Larose, 2005 and Shmueli, Patel, & Bruce, 2007).

Nevertheless, the results of this work must be interpreted bearing in mind certain limitations. First of all, the data must be analyzed from a transversal perspective. In this sense, the appropriate design in order to identify predictor factors for the use of a substance is a longitudinal study. Secondly, the data analyzed are self-reported and do not include the view of different family members. Some authors have found that self reports of drug consumption may be biased, in such a way that adolescents report a lower consumption in self-reported protocols in comparison with an interview (Stone & Latimer, 2005), leading to a lower validity (Williams & Nowatzki, 2005).

Future studies should analyze the relationship between consumption by parents and friends and nicotine consumption using the techniques proposed in this work and longitudinal studies, since it has been observed that having parents and friends who are smokers exerts a variable influence throughout adolescence (Bricker et al., 2007; Peterson et al., 2006). Likewise, it would be worth evaluating whether the tougher antismoking laws have been transformed into a decrease in the places of exposure to tobacco in adolescence and whether this has an influence, as some studies point out (Seo, Torabi, & Weaver, 2008), in tobacco consumption in adolescents.

References

Avenoli, S., & Merikangas, K.R. (2003). Familial influences on adolescent smoking. *Addiction, 98*, 1-20.

Becoña, E. (2002). Factores de riesgo y protección familiar para el uso de drogas. In R. Secades & J. R. Fernández (Eds.), *Intervención familiar en la prevención de las drogodependencias*. Madrid: Plan Nacional sobre Drogas.

Breiman, L., Friedman, J.H., Olshen, R.A., & Stone, C.J. (1984). *Classification and regression trees*. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software.

Bricker, J.B., Peterson, A.V., Sarason, I. G., Andersen, M.R., & Rajan, K.B. (2007). Changes in the influence of parents' and close friends' smoking on adolescent smoking transitions. *Addictive Behaviors, 32*, 740-757.

Buscema, M. (2002). A brief overview and introduction to artificial neural networks. *Substance Use & Misuse, 37*, 1093-1148.

Cameron, A.C., & Trivedi, P.K. (1990). Regression-based tests for overdispersion in the Poisson model. *Journal of Econometrics, 46*, 347-364.

Carvajal, S.C., & Granillo, T. M. (2006). A prospective test of distal and proximal determinants of smoking initiation in early adolescents. *Addictive Behaviors, 31*, 649-660.

Castrucci B.C., Gerlach K.K., Kaufman N.J., & Orleans, C.T. (2002). The association among adolescents' tobacco use, their beliefs and attitudes, and friends' and parents' opinions of smoking. *Maternal and Child Health Journal, 6*, 159–67.

Ciairano, S., Bosma, H.A., Miceli, R., & Settani, M. (2008). Adolescent substance use in two European countries: Relationships with psychosocial adjustment, peers, and activities. *International Journal of Clinical and Health Psychology, 8*, 119-138.

Darling, N., & Cumsille, P. (2003). Theory, measurement, and methods in the study of family influences on adolescent smoking. *Addiction, 98*(Suppl. 1), 21-36.

De Vries, H., Engels, R., Kremers, S., Wetzels, J., & Mudde, A. (2003). Parents' and friends' smoking status as predictors of smoking onset: Findings from six European countries. *Health Education Research, 18,* 627-636.

Dick, D.M., Pagan, J.L., Viken, R., Purcell, S., Kaprio, J., Pulkkinen, L., & Rose, R.J. (2007). Changing environmental influences on substance use across development. *Twin Research and Human Genetics, 10*, 315-326.

Delegación del Gobierno para el Plan Nacional Sobre Drogas (2008). *Encuesta Estatal sobre Uso de Drogas en Enseñanzas Secundarias (ESTUDES), 1994-2008*. URL Retrieved February 10, 2009, from http://222.pnsd.msc.es/ca/Categoria2/observa/pdf/Estudes2008_Web.pdf.

Doran, N., McCharge, D., & Cohen, L. (2007). Impulsivity and the reinforcing value of cigarette smoking. *Addictive Behaviors, 32*, 90-98.

Farrell, A.D., Sullivan, T.N., Esposito, L.E., Meyer, A.L., & Valois, R.F. (2005). A latent growth curve analysis of the structure of aggression, drug use, and delinquent behaviors and their interrelations over time in urban and rural adolescents. *Journal of Research on Adolescence, 15*, 179-204.

Fernández, J.R., Secades, R., Vallejo, G., & Errasti, J.M. (2003). Evaluation of what parents know about their children's drug use and how they perceive the most common family risk factors. *Journal of Drug Education, 33*, 334-350.

Geckova, A.M., Stewart, R., Van Dijk, J.P., Orosova, O., Groothof, J.W., & Post, D. (2005). Influence of socio-economic status, parents and peers on smoking behaviour of adolescents. *European Addiction Research, 11*, 204-209.

Gervilla, E., & Palmer, A. (2009). Predicción del consumo de cocaína en adolescentes mediante árboles de decisión. *Revista de Investigación en Educación, 6*, 7-13.

Gervilla, E., & Palmer, A. (2010). Prediction of cannabis and cocaine use in adolescence using decision trees and logistic regression. *The European Journal of Psychology Applied to Legal Context, 2*, 19-35.

Gervilla, E., Cajal, B., Roca, J., & Palmer, A. (2010). Modelling alcohol consumption during adolescente using Zero Inflated Negative Binomial and Decision Trees. *The European Journal of Psychology Applied to Legal Context, 2*, 145-159.

Gervilla, E., Jiménez, R., Montaño, J.J., Sesé, A., Cajal, B., & Palmer, A. (2009). La metodología del data mining. Una aplicación al consumo de alcohol en adolescentes. *Adicciones, 21*, 65-80.

Greene, W. (1994). *Accounting for excess zeros and sample selection in poisson and negative binomial regression models* (working paper number, 94-10). New York: New York University, Stern School of Business.

Hall, J.A., & Valente, T.W. (2007). Adolescent smoking networks: The effects of influence and selection on future smoking. *Addictive Behaviors, 32*, 3054-3059.

Han, J., & Kamber, M. (2006). *Data mining: Concepts and techniques* (2nd ed.). San Francisco, CA: Morgan Kauffmann.

Hoffman, B.R., Sussman, S., Unger, J.B., & Valente, T.W. (2006). Peer influences on adolescent cigarette smoking. *Substance Use & Misuse, 41*, 103-155.

Hoffman, J.H., Welte, J.W., & Barnes, G.M. (2001). Co-ocurrence of alcohol and cigarette use among adolescents. *Addictive Behaviors, 26*, 63-78.

Jiménez, T.I., Musitu, G., & Murgui, S. (2008). Funcionamiento familiar y consumo de sustancias en adolescentes: El rol mediador de la autoestima. *International Journal of Clinical and Health Psychology, 8*, 139-151.

Kantardzik, M. (2003). *Data mining: Concepts, models, methods, and algorithms.* New York, NY: John Wiley and Sons.

Kitsantas, P., Moore, T.W., & Sly, D.F. (2007). Using classification trees to profile adolescent smoking behaviors. *Addictive Behaviors, 32*, 9-23.

Kobus, K. (2003). Peers and adolescent smoking. *Addiction, 98*(Suppl. 1), 37-55.

Lambert, D. (1992). Zeros-Inflated Poisson regression, with an application to defects in manufacturing. *Technometrics, 34*, 1-14.

Larose, D.T. (2005). *Discovering knowledge in data: An introduction to data mining.* Hoboken, NJ: John Wiley and Sons.

Larsson, S., Lilja, J., Borg, S., Buscema, M., & Hamilton, D. (2001). Toward an integrative approach in the analysis of dependency problems. *Substance Use & Misuse, 36*, 1323-1356.

Martínez, J.L., Fuertes, A., Ramos, M., & Hernández, A. (2003). Consumo de drogas en la adolescencia: Importancia del afecto y la supervisión parental. *Psicothema, 15*, 161-166.

Martínez, M., & Alonso, C. (2003). Búsqueda de sensaciones, autoconcepto, asertividad y consumo de drogas ¿Existe relación? *Adicciones, 15*, 145-158.

Mercken, L., Snijders, T.A. B., Steglich, C., Vertiainen, E., & De Vries, H. (2010). Smoking-based selection and influence in gender-segregated friendship networks: A social network analysis of adolescent smoking. *Addiction, 105*, 1280-1289.

Mercken, L., Candel, M., Willems, P., & De Vries, H. (2007). Disentangling social selection and social influence effects on adolescent smoking: The importance of reciprocity in friendship. *Addiction, 102*, 1483-1492.

Molyneux, A., Lewis, S., Antoniak, M., Browne, W., McNeill, A., Godfrey, C., et al. (2004). Prospective study of the effect of exposure to other smokers in high school tutor groups on the risk of incident smoking in adolescence. *American Journal of Epidemiology, 159*, 127-132.

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics, 33*, 341-65.

Muñoz, M., & Graña, J. L. (2001). Factores familiares de riesgo y de protección para el consumo de drogas en adolescentes. *Psicothema, 13*, 87-94.

Nadal, R. (2008). La búsqueda de sensaciones y su relación con la vulnerabilidad a la adicción y al estrés. *Adicciones, 20*, 59-72.

Otten, R., Engels, R.C.M.E., & Prinstein, M.J. (2009). A prospective study of perception in adolescent smoking. *Journal of Adolescent Health, 44*, 478-484.

Otten, R., Wanner, B., Vitaro, F., & Engels, R.C.M.E. (2009). Disruptiveness, peer experiences and adolescent smoking: A long-term longitudinal approach. *Addiction, 104*, 641-650.

Palmer, A., Montaño, J.J., & Calafat, A. (2000). Predicción del consumo de éxtasis a partir de redes neuronales artificiales. *Adicciones, 12*, 29-41.

Peña, M.E., Andreu, J.M., & Graña, J.L. (2009). Multivariate model of antisocial behavior and substance use in spanish adolescents. *Journal of Child & Adolescent Substance Abuse, 18*, 207-220.

Peterson, A.V., Leroux, B.G., Bricker, J., Kealey, K.A., Marek, P.M., Sarason, I.G., et al. (2006). Nine-year prediction of adolescent smoking by number of smoking parents. *Addictive Behaviors,31*, 788-801.

Prinstein, M.J., & Wang, S.S. (2005). False consensus and adolescent peer contagion: Examining discrepancies between perceptions and actual reported levels of friends' deviant and health risk behaviors. *Journal of Abnormal Child Psychology,33*, 293–306

Sears, E.S., & Anthony, J.C. (2004). Artificial neural networks for adolescent marijuana use and clinical features of marijuana dependence. *Substance Use & Misuse, 39*, 107-134.

Seo, D.C, Torabi, M.R., & Weaver, A.E. (2008). Factors influencing openness to future smoking among non-smoking adolescents. *Journal of School Health, 78*, 328-336.

Shmueli, G., Patel, N.R., & Bruce, P.C. (2007). *Data mining in excel: Lecture notes and cases*. Arlington, VA: Resampling Stats, Inc.

Simons-Morton, B. (2007). Social influences on adolescent substance use. *American Journal of Health Behavior, 31*, 672-684.

Stone, A.L., & Latimer, W.W. (2005). Adolescent substance use assessment: Concordance between tools using self-administered and interview formats. *Substance Use & Misuse, 40*, 1865-1874.

Timmermans, M., Van Lier, P.A.C., & Koot, H.M. (2008). Which forms of child/adolescent externalizing behaviors accounts for late adolescent risky sexual behaviour and substance use? *Journal of Child Psychology and Psychiatry, 49*, 386-394.

Williams, R.J., & Nowatzki, N. (2005). Validity of adolescent self-report of substance use. *Substance Use & Misuse, 40,* 299-311.

Witten, I.H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques* (2nd ed.). San Francisco, CA: Morgan Kaufmann.

Ye, N. (2003). *The handbook of data mining.* Mahwah, NJ: Lawrence Erlbaum Associated.