

Smoking in young adolescents: an approach with multilevel discrete choice models

J Pinilla, B González, P Barber, Y Santana

J Epidemiol Community Health 2002;**56**:227–232

Study objective: To understand the context for tobacco smoking in young adolescents, estimating the effects of individual, family, social, and school related factors.

Design: Cross sectional analysis performed by multilevel logistic regression with pupils at the first level and schools at the second level. The data came from a stratified sample of students surveyed on their own, their families' and their friends' smoking habits, their schools, and their awareness of cigarette prices and advertising.

Setting: The study was performed in the Island of Gran Canaria, Spain.

Participants: 1877 students from 30 secondary schools in spring of 2000 (model's effective sample sizes 1697 and 1738).

Main results: 14.2% of the young teenagers surveyed use tobacco, almost half of them (6.3% of the total surveyed) on a daily basis. According to the ordered logistic regression model, to have a smoker as the best friend increases significantly the probability of smoking (odds ratio: 6.96, 95% confidence intervals (CI) (4.93 to 9.84), and the same stands for one smoker living at home compared with a smoking free home (odds ratio: 2.03, 95% CI 1.22 to 3.36). Girls smoke more (odds ratio: 1.85, 95% CI 1.33 to 2.59). Experience with alcohol, and lack of interest in studies are also significant factors affecting smoking. Multilevel models of logistic regression showed that factors related to the school affect the smoking behaviour of young teenagers. More specifically, whether a school complies with antismoking rules or not is the main factor to predict smoking prevalence in schools. The remainder of the differences can be attributed to individual and family characteristics, tobacco consumption by parents or other close relatives, and peer group.

Conclusions: A great deal of the individual differences in smoking are explained by factors at the school level, therefore the context is very relevant in this case. The most relevant predictors for smoking in young adolescents include some factors related to the schools they attend. One variable stood out in accounting for the school to school differences: how well they enforced the no smoking rule. Therefore we can prevent or delay tobacco smoking in adolescents not only by publicising health risks, but also by better enforcing no smoking rules in schools.

See end of article for authors' affiliations

Correspondence to:
Professor B González
López-Valcárcel, University
of Las Palmas de Gran
Canaria, Department of
Quantitative Methods,
Campus de Tafira, 35017
Las Palmas, Spain;
bea@empresariales.ulpgc.es

Accepted for publication
24 July 2001

Precisely what leads adolescents to take up smoking has been the subject of exhaustive study in the recent epidemiological literature.^{1–5} These studies have revealed the multifactorial nature of the phenomenon, and the methodological and empirical problems to be faced when trying to model and predict smoking patterns.⁶ Peer^{7–8} and parental influences are well documented,^{9–12} as well as personality variables.¹³ Differences in prevalence between girls and boys have been found.^{14–15} The wish to “experiment” at this age incites some teenagers to try cigarettes¹⁶; the same stands for alcohol consumption, which has been proved to be complementary with smoking for the teenagers group.¹⁷

It has been clearly and empirically stated in the literature that smokers aged 12–14 tend to have a circle of friends who are smokers. In addition, they often have at least one parent who smokes, little interest in school, come from broken homes, and may exhibit a precocious physical maturity. Experience with alcohol tends to be present, sometimes combined with other unhealthy habits.

In addition to these individual predictors, there are others specific to the environment at the school the young teenager attends. One such factor, having a particular importance, is the awareness of and compliance with antitobacco school regulations pertaining both to the sale and consumption of cigarettes on the part of teachers and students.¹⁸ Enforced smoking bans at school may reduce teenage smoking.¹⁹ Besides this, schools also can be an important source of infor-

mation about the health risks of smoking^{20–21} and an optimal place to implement preventive programmes.^{22–25}

This paper shows some results and analysis of the first wave of a three years longitudinal survey. It is part of a research project about the formation of the smoking habit in young adolescents. Its main objective is to estimate the relative influence of those factors based on the school setting as compared with the factors and variables related to the teenager's personal characteristics and to their family. The most important contribution of this study is that it estimates the effect of compliance with antismoking rules in the schools on the prevalence of smoking among young adolescents. To the best of our knowledge, this is the first time that this kind of results are reported in the public health literature.

METHODS

The data used in the empirical application derive from a cross section survey made in the spring of 2000, part of the first stage of a broader, three year, longitudinal study. The broader study was begun with secondary school students in their second year (aged 13–14).*

The sample system used was that of clusters with probabilities proportional to the number of groups of second year students in each school, doubly stratified: by municipality and by kind of school—public, subsidised, and private. The reason for stratification by municipality was that schools on Gran Canaria are highly concentrated (59%) in 2 of the 21 municipalities. For this reason we defined three strata, one each for

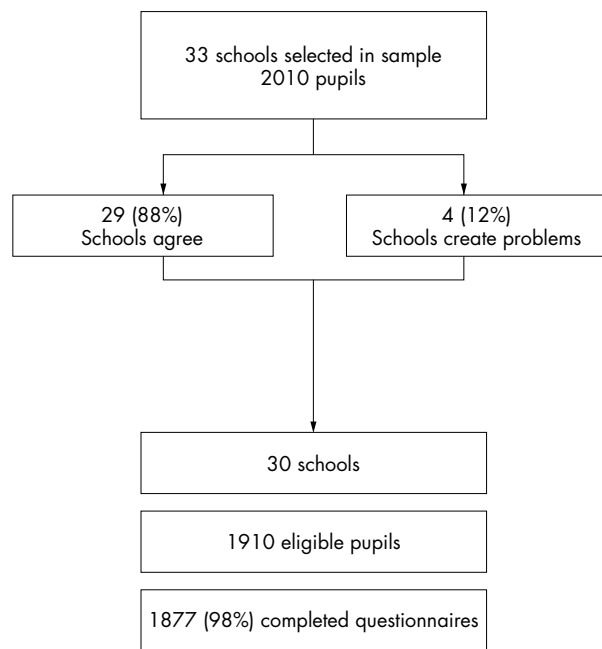


Figure 1 Survey sampling strategy.

the two cities with most of the schools, Las Palmas de Gran Canaria and Telde, and the third for the other 19 municipalities.

Stratification by type of school similarly was attributable to uneven distribution: 72% of the schools are public, 19% subsidised, and just 9% private.

The initial sample included a total of 2010 students in 33 schools, out of a total of 10 839 students and 178 schools, which yields an estimation error of 1.3% for 95% level of confidence for $\hat{p}=0.1$, and an error of 6.76% for $\hat{p}=0.5$, also at the 95% level of confidence. Figure 1 shows our sampling strategy, in which 88% of the schools initially selected (30 of 33) participated in the experiment. Two schools no longer existed, and two schools had merged in the year since our list of schools had been drawn up. So including this merged school we had a total of 1910 survey responses, of which we used 1877, rejecting 23 that were self contradictory or obviously joking.

The questionnaire, which was self administered, had four sections. The first included sociodemographic variables: age, sex, number of people in the home, and spending money, followed by variables related to in school and out of school activities. Then it asked about consumption of cigarettes and variables that might be related: interest in studies, alcohol consumption, tobacco consumption among peer group, by parents, and by other family members. We also asked about cigarette brands, prices, and the attitudes towards advertising, and concluded with questions about the students' knowledge about the negative effects of tobacco, attitudes toward tobacco, and their own prediction about whether they would become smokers. The questionnaire was approved by a committee of independent researchers with experience in surveys of teenage smoking, and tested with an initial pilot sample of 200 children. When students were given the questionnaires, they were told their answers would be confidential; teachers

were not allowed to be present in order to guarantee free responses. The survey was administered in the school classrooms by the same, single researcher, who provided instructions and answered students' doubts.

We estimate two types of models. Firstly, we estimate an ordered logistic regression model to explain the question "Do you smoke?", with four ordered categories from No smoking to Yes, smokes daily. The explanatory variables, at the individual level, are sex, interest in school (ordered, with four categories), alcohol consumption (never; occasionally; more than once a week); whether the best friend smokes or not; and the number of smokers living at home. The second type of model was a binary logistic multilevel regression (smokes daily or not). It has as explanatory variable at the school level (context) the compliance with antismoking rules in the school. This is a dichotomous variable, based on the replies of school directors about smoking in various school areas where smoking is forbidden by the antismoking rules: cafeteria, courtyard, classrooms, bathrooms, corridors or halls. If the director says that smoking is allowed in one or more of these places, the dummy variable for compliance is set to zero, otherwise it is set to one.

Multilevel models

In this study we specify and estimate a multilevel model^{26 27} adjusted for the traits of the adolescent smoker. It combines effects related to the individual with those for the school where the young teenager studies. Multilevel models, designed as they are for data grouped in hierarchies or levels, are particularly appropriate for this kind of analysis. With these models we can evaluate how much of the variability of the dependent variable—tobacco use in the adolescent—is attributable to individual circumstances, and how much to the effect of the group, the school attended.

Multilevel regression models are indicated when there is a hierarchical structure in levels of data, with a single dependent variable measured at the lowest level and a set of explanatory variables on each of the levels. The advantage of these models lies in their capacity to define and explore variations at each level of the hierarchy after controlling for relevant explanatory variables. The application of multilevel models in health economics has become quite common, in health administration,²⁸ as well as in other areas.^{29 30}

To better understand multilevel models, we suggest starting with a model of components of variance. When applying it, let us suppose that we have data from successive repetitions of experiments with J groups of different individuals, in each of which there are n_j persons in the sample. Taking Y_{ij} as the value of a given variable or property of individual i of group j in the sample, we can represent this variable as

$$Y_{ij} = \beta_j + e_{ij} \quad (1)$$

where β_j represents the mean of the variable Y of all the individuals of group j and e_{ij} the deviation of the value of the variable Y for the individual i from the mean of that variable in the other members of group j .

Deviating the β_j from its mean $\bar{\beta}$, equation (1) can be expressed as:

$$Y_{ij} = \bar{\beta} + u_j + e_{ij} \quad \text{with } u_j = \beta_j - \bar{\beta} \quad (2)$$

The model formulated in this equation coincides with the components of variances with fixed and random effects. The component $\bar{\beta}$ of the equation represents the so called fixed effects, while $u_j + e_{ij}$ contains the random effects. This model assumes that the random effects u_j are distributed normally

*Secondary education in Spain is obligatory for students aged 12–16. It is intended to prepare students for integration in the active population, to enter technical school, or to study bachillerato, the equivalent of high school. The students surveyed were in what is known in Spain as the "segundo año de ESO".

with mean 0 and variance $\sigma_{u_j}^2 = \sigma_{\beta_j}^2$, which stands for differences in the variable Y attributable to the group. It also assumes that the error component e_{ij} is also distributed normally with mean 0 and variance σ^2 . And finally it assumes that the random effects u_j and the error component e_{ij} are independent, $r_{ue} = 0$, and that the e_{ij} are all independent one from another.

The objective, therefore, consists in estimating from the available data the variances σ_u^2 and σ^2 which represent, respectively, the variability of the variable Y attributable to the group the individual belongs to, and also the variability of the variable Y attributable to differences between individuals, defining ρ the intraclass correlation as the percentage of the total variance among the J groups.

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma^2} \times 100 \quad (3)$$

Let us suppose that we also have information about $K-1$ personal characteristics for each individual in our sample. These characteristics could be included in a vector X'_{ij} that when introduced in the model would involve the appearance of K new fixed effects, $K-1$ variables, and a constant. Equation (2) would then become:

$$Y_{ij} = X'_{ij}\beta_j + u_j + e_{ij} \quad (4)$$

in which X includes K regressors and $e_{ij} \approx N(0, \sigma^2)$.

The next step in hierarchical models is to parametrise the coefficients β_j of equation (4) by adding L explanatory variables at the group level, leaving the model as:

$$Y_{ij} = X'_{ij}\bar{\beta}_j + u_j + e_{ij} \quad (5)$$

in which the fixed effects are now dependent on L possible variables of the groups.

The models we have seen so far represent general specifications of multilevel models in which the responses to a continuous variable are related in a linear function to a set of explanatory variables and to a simple hierarchical structure. Many of the applications to health economics are not linear, and so work better with generalised linear models with different functions: logit, probit, Poisson, negative binomial, etc. In our case we applied multilevel analysis with a logistic function.

For a binary response, unlike in the Normal case, even if the only random coefficient is the intercept, the level 1 variance ($\text{Var}(Y_{ij}) = \pi_{ij}(1-\pi_{ij})$, where π_{ij} is the probability of smoking of the i -student attending to the j -school) depends on the expected value of the dependent variable. Furthermore, the level 2 variance, σ_u^2 , is measured on the logistic scale, so is not directly comparable to the level 1 variance. Therefore in multilevel models for binary response, intraclass correlation cannot be calculated with equation (3). Goldstein proposes four different procedures to provide at least approximate estimates³¹ of the intraclass correlations in this models, but their interpretation is unclear.

The sequential process of modelling

To ensure the convergence of the estimations, we collapsed the dependent variable Y_{ij} which we initially specified with four possible values (smokes daily, on weekends, less than once a week, and not at all) as a simple binary: smokes daily/does not smokes daily. Then we estimated the following sequence of four models: Model 1, unconditional means model; Model 2,

including a school level (level 2) predictor, the compliance with antismoking rules in each school; Model 3, including pupils level (level 1) predictors (fixed effects); finally Model 4, including both level 1 and level 2 predictors. In models 3 and 4, the random errors at the school level involve only intercepts. So, there is only one explicit random effect, and there is not interaction assumed between the context level and the individual level.

Statistical analysis

We analysed the data with the program MIWin, version 1.0.³² We used RGLS (restricted generalised least squares), with extra-binomial variation and without it, to compare both results. The estimation without extra-binomial variation constrains to 1 the variance of the level 1 error. In the other case (extra-binomial variation is allowed), the level 1 variance is estimated.³³ The reported results are those with extra-binomial variation. There is some discussion about this point. It was said that, as a "rule of thumb", an estimate close to 1.0 indicates conformity to the binomial assumption, but a recent simulation³⁴ to investigate the effects of sample size and model misspecification on estimating the level 1 variance for binary responses has shown that "an estimate very close to 1.0 seems to be an indication for the incorrect specification rather than for a correct binomial assumption" (page 12). And, furthermore, "in small samples, estimates around 0.8 can appear although the model is correctly specified under the binomial assumption (page 13)".

RESULTS

Table 1 shows the descriptive analysis of the variables that appear in the models, including some characteristics of the schools that had been included at first, but that were excluded from the final specifications. There is not significant association between the school classification and antismoking rules compliance. Private schools have higher rates of academic success. We have no information about the prevalence of smoking in the teachers group, but more than 90% of 12 years old students know that at least one of their teachers smoke, and this is the case for both private and public schools. We observe that 14.3% of the young teenagers surveyed use tobacco, almost half of them (6.3% of the total surveyed) on a daily basis. As for the no smoking rules, we see that in 33% of the schools there is compliance with them.

Smoking prevalence was higher (9.53%) in schools with weak compliance with smoking rules as compared with schools with medium/strong compliance (5.45%) The difference in means is significant ($t=2.017$; $p=0.053$).

Firstly, we estimated an ordered logit model to find out the relation between the young teenager's smoking habits and other individual characteristics. This model is designed for ordered categorical dependent variables, to take into account their ordinal nature. In our case, the dependent variable has four ordered categories, no smoke, less than once a week, on weekends and smokes daily. Table 2 presents the results of this estimation, with the odds ratio of each explanatory variable and the corresponding confidence levels at 95% as well as the univariate (unadjusted) odds ratios. These odds ratios are calculated to compare the pupils smoking daily with those that never smoke. The multicollinearity among the explanatory variables causes the discrepancy between the unadjusted and the adjusted odds ratios (through the multivariate ordered logistic regression model). The pupils that do not have any interest in school tend to drink alcohol and they tend also to have a smoker as their best friend. We find that on the whole, girls are more inclined to smoke than boys, with an odds ratio of 1.85. Young teenagers who are less interested in school are more likely to smoke, and alcohol consumption is strongly related to smoking. Tobacco use by relatives and in the peer

Table 1 Descriptive statistics of the variables in the models

Type of variable	Definition	Number of observations (%)		
Dependent variable	Do you smoke?			
	No smoke	1590	(85.8)	
	Yes, less than once a week	92	(5.0)	
	Yes, on weekends	55	(3.0)	
	Yes, smokes daily	116	(6.3)	
Individual explanatory covariates	Sex:			
	Male	996	(52.3)	
	Female	907	(47.7)	
	Interest in school			
	Very little	329	(17.3)	
	Little	693	(36.5)	
	Quite	657	(34.6)	
	High	222	(11.7)	
	Alcohol consumption			
	Never	1263	(66.6)	
	Occasionally or on weekends	607	(32.0)	
	Some times during the week or every day	27	(1.4)	
	Number of smokers living at home			
Discrete variable	Mean 1.25; SD 1.20			
School characteristics ("context" variables)	Best friend smokes?			
	Yes	442	(23.5)	
	No	1439	(76.5)	
	Antismoking rules in the schools (number of schools)			
	Weak	8	1	1
Medium/strong	15	3	2	
% of academic failure in the school (Grades 1 and 2 of the ESO)				
Mean	Public 50.9	Private 97.9	Subsidised 90.0	
SD	18.1	2.3	8.7	
Median	46	100	97	
Min	21	95	75	
Max	100	100	97	

group has a strong positive effect on the probability of being a smoker. The peer group has the strongest effect, with an odds ratio of 6.96.

Table 3 presents the estimations obtained from the multilevel models. The sample size is 1.738. The dependent variable is dichotomous (1 if he/she smokes daily, 0 if he/she not smokes daily). According to the first multilevel logistic regression model (Model 1), the school level variance, σ^2_{ur} , is significant, and we may attempt to explain it by the differences in school related factors. As the computation and the interpretation of the intraclass correlation in this model

has some problems (see discussion above), we leave it out of all models in table 3. Model 2 introduced a fixed coefficient at level 2, for compliance with no smoking rules. Its coefficient is significant. We tested other school related variables such as the number of students per classroom, the kind of school (public, subsidised, private), and the percentage of students who pass their courses, by subject matter, but eliminated them from the final model for lack of statistical significance. In Model 2, the school level variance (significant too), is a 28% lower than in the previous model.

In Model 3 we included in the unconditional model, as fixed effects, the set of those explanatory variables related to the

Table 2 Smoking by young teenagers. Ordered logistic regression model estimation results

	Categories	Frequency (%)	
<i>Dependent variable</i> "Do you smoke?"	No smoke	1456 (85.8)	
	Yes, less than once a week	83 (4.9)	
	Yes, on weekends	48 (2.8)	
	Yes, smokes daily	110 (6.5)	
<i>Explanatory variables</i>	<i>Unadjusted odds ratios</i>	<i>Adjusted odds ratios (95% CI)</i>	
	Sex (female/male*)	1.75	1.85 (1.33 to 2.59)
	Interest in school (high/very little*)	0.17	0.23 (0.19 to 0.27)
	Alcohol consumption (occasionally/never*)	21.74	5.88 (3.95 to 8.75)
	Alcohol consumption (more than once a week/never*)	58.30	9.81 (6.12 to 15.75)
	Number of smokers at home (one/none*)	1.77	2.03 (1.22 to 3.36)
	Best friend smokes (yes/no*)	24.74	6.96 (4.93 to 9.84)
Log likelihood	-696.0774098		
χ^2 for covariates	p=0.0001		

Effective sample size: 1697. The odds ratios are calculated to compare "smokes daily" with "does not smoke". Dependent variable, 0 for no smoke, 1 for smoke less than once a week, 2 for smoke on weekends, 3 for smokes daily. Variables contents: sex (0 male; 1 female), interest in school (ordinal for interest in school; 1 very little; 2 little; 3 quite; 4 high), alcohol consumption occasionally (1 yes; 0 no), alcohol consumption frequently (1 yes; 0 no), Smokers at home (percentage of smokers at home), best friend smokes (1 yes; 0 no). Odds ratios 95% confidence intervals in parentheses. *Odds ratios for categories extremes, smokes daily versus no smoke.

Table 3 Differences in smoking habits explained by factors at the school level and individual characteristics (n=1738). Logistic multilevel regression results. Dependent variable: Smokes daily (yes=1/no=0)

	Regression coefficients (SE)			
	Model 1	Model 2	Model 3	Model 4
<i>Fixed effects</i>				
Intercept	-2.655 (0.159)	-2.174 (0.233)	-5.228 (0.475)	-4.74 (0.192)
Sex			0.645 (0.200)	0.650 (0.193)
Interest in school			-0.504 (0.113)	-0.521 (0.111)
Alcohol consumption occasionally			1.586 (0.264)	1.598 (0.256)
Alcohol consumption frequently			2.402 (0.281)	2.426 (0.272)
Smokers at home			1.181 (0.291)	1.199 (0.282)
Best friend smokes			1.832 (0.216)	1.817 (0.209)
Antismoking rules in the schools		-0.815 (0.300)		-0.770 (0.332)
<i>Random effects</i>				
σ^2_u	0.433 (0.189)	0.312 (0.159)	0.518 (0.212)	0.447 (0.192)
σ^2	0.939 (0.032)	0.938 (0.032)	0.636 (0.022)	0.605 (0.021)

Dependent variable, 0 for no daily smokers (1 for daily smokers). Variable contents: sex (0 male; 1 female), Interest in school (ordinal for interest in school; 1 very little; 2 little; 3 quite; 4 high), alcohol consumption occasionally (1 yes; 0 no), alcohol consumption frequently (1 yes; 0 no), smokers at home (percentage of smokers at home), best friend smokes (1 yes; 0 no), antismoking rules in the schools (0 weak; 1 medium/strong). Extra-binomial variation is allowed. Standard errors in parentheses. All variables are 95% significance.

individual that proved significant in the initial logit ordered model. Model 4 adds the antismoking enforcement. The level 2 component of variance (for the differences between schools) is significant. It has been reduced from 0.518 (in Model 3) to 0.447. Again, as in the shift from Model 1 to Model 2, the addition of the explanatory variable at the school level reduces the unexplained variability among schools. There are not significant cross level fixed or random effects. As, furthermore, it would have impeded convergence, they were not included in any of the final models.

DISCUSSION

Our study coincides with those in other European countries as to the main risk factors for tobacco consumption in adolescents,³⁵⁻³⁸ principally, smoking by close friends, tobacco consumption by parents or other close relatives, experience with alcohol, and lack of interest in school.

According to our estimates, part of the variability among teenagers' probabilities of smoking is related to individual and family characteristics, and the remaining belongs to the "context" level, due to school variables. The school variable that significantly contributes to explain the probability of smoking is the compliance with no smoking rules. The significance of this variable in our models shows how important a strict enforcement of no smoking health laws could be for the pupils. Other characteristics of the schools have no significant effect. There is no difference among public, private and subsidised schools; the rate of academic success in each school is not significant, either. We have tried with some other school variables, but none of them is significant in our model.

The school variable that significantly contributes to explain the probability of smoking is the compliance with no smoking rules.

The period of secondary schooling coincides with adolescence, in which students undergo an intense process of physical, intellectual, and emotional change. Students undergo tension, not only because of these internal changes, but also because of pressure from the family and the school. Though counselling and guidance schools should help their students form self identity, build self esteem, and resist group pressure in order to promote values and habits against the use of tobacco and other drugs.

According to the latest statistics on unhealthy consumption in the school population, tobacco use among students has

increased at an alarming rate.³⁹⁻⁴⁰ According to our survey, 14.3% of students aged 13-14 smoke, and almost half of these do so daily.⁴¹ These young teenager smokers may well be leaders or especially popular figures in peer groups, which would explain why there is a higher percentage, 23.5%, of affirmative replies to the question whether your best friend smokes, than the 14.3% overall rate of smokers.

Girls do so more than boys, reflecting the feminisation of tobacco consumption in Spain in the past years.

Lax enforcement of no smoking rules in schools has a double impact on the young teenagers. It makes it easier for them to begin to smoke themselves, and it also allows them to see older students, even teachers, smoke. Habits brought by teachers and by students from "smoking homes" become visible, available, and attractive to young students from "non-smoking homes."

Tobacco is one of the most consumed drugs; it causes the most health problems. Preventing its consumption is thus one of the main objectives of public health policy.⁴²⁻⁴⁵ We must attempt to neutralise the relation that is made in the social arena (induced by advertising) between the values of masculinity (competitiveness, initiative, power) and the consumption of tobacco. This relation is attractive now especially for girls, as they too prepare for positions of social power. Furthermore, as we know that initiation in the habit takes place in the circle of friends, and is associated with the symbolic perception of passage to adult life, our intervention should be focused on strengthening resistance to group pressure. For this reason it is the first years of secondary school (ESO in Spain) in which we must increase preventive measures in regard to tobacco, if we want to delay as long as possible teenager smoking, if not avoid it altogether. The school itself is a key arena for this preventive action, both in what it does, and in what it allows to be done.

The main limitation of our study lies in the multilevel model's assumption that teenager and school effects are uncorrelated. If some individual variables, omitted from the model because of lack of data (for instance, the family income), turned out to be relevant and correlated with the school effects (better compliance of the antismoking rules in the "rich" schools, for example), our hypothesis would not be justified.

This study is to be continued with the second (and third) wave of the survey. Longitudinal data on our schoolchildren will provide us with new material to pursue our research. New information will emerge because some of our sample teenagers are moving to a new school.

ACKNOWLEDGEMENTS

The authors are grateful to Bill Christian, PhD for his helpful comments of this manuscript and for his assistance in the translation, and to the anonymous reviewers for many helpful and constructive comments. The authors are also grateful to Jose Ramon Calvo, PhD, to Anselmo Lopez PhD, and to Araceli Caballero for their comments and ideas about the survey.

Funding: this paper is a product of the research project FUNCIS PI17/99 funded by FUNCIS (Fundación Canaria de Investigación y Salud).

Authors' affiliations

J Pinilla, B González, P Barber, Y Santana, Department of Quantitative Methods, University of Las Palmas de Gran Canaria, Canary Islands, Spain

REFERENCES

- 1 **Goddsard E**. *Why children start smoking*. London: HMSO, 1990.
- 2 **Conrad KM, Flay BR, Hill D**. Why children start smoking cigarettes: predictors of onset. *Br J Addict* 1992;**87**:1711–24.
- 3 **McNeill AD, Jarvis MJ, Stapleton JA, et al**. Prospective study of factors predicting uptake of smoking in adolescent. *J Epidemiol Community Health* 1988;**43**:72–8.
- 4 **McGee R, Stanton WR**. A longitudinal study of reasons for smoking in adolescence. *Addiction* 1993;**88**:265–71.
- 5 **Vries H de, Dijkstra M, Kok G**. *Young people and their smoking behaviour*. In: Heller T, Bailey L, Pattison S, eds. *Preventing cancers*. London: The Open University, 1992:154–61.
- 6 **Chassin L, Presson CC, Sherman SJ, et al**. The natural history of cigarette smoking: predicting young-adult smoking outcomes from adolescent smoking patterns. *Health Psychol* 1990; **9**:701–16.
- 7 **Aloise-Young PA, Graham JW, Hansen WB**. Peer influence on smoking initiation during early adolescence: a comparison of group members and group outsiders. *J Appl Psychol* 1994;**79**:281–7.
- 8 **Urberg KA, Shyu S-J, Liang J**. Peer influence in adolescent cigarette smoking. *Addict Behav* 1990;**15**:247–55.
- 9 **Biglan A, Duncan TE, Ary DV, et al**. Peer and parental influences on adolescent tobacco use. *J Behav Med* 1995;**18**:315–30.
- 10 **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.
- 11 **Newman IM, Ward JM**. The influence of parental attitude and behaviour on early adolescent cigarette smoking. *J Sch Health* 1989;**59**:150–2.
- 12 **Fergusson DM, Lynskey MT, Horwood UJ**. The role of peer affiliations, social, family and individual factors in continuities in cigarette smoking between childhood and adolescence. *Addiction* 1995;**90**:647–59.
- 13 **Burt RD, Dinh KT, Peterson AV Jr, et al**. Predicting adolescent smoking: a prospective study of personality variables. *Prev Med* 2000;**30**:115–25.
- 14 **Charlton A, Blair V**. Predicting the onset of smoking in boys and girls. *Soc Sci Med* 1989;**29**:813–18.
- 15 **Swan AV, Melia RJW, Fitzsimmons B, et al**. Why do more girls than boys smoke cigarettes? *Health Ed J* 1989;**48**:59–64.
- 16 **Brynin M**. Smoking behaviour: predisposition or adaptation? *Journal of Adolescence* 1999;**22**:635–46.
- 17 **Dee TS**. The complementarity of teen smoking and drinking. *J Health Econ* 1999;**18**:769–93.
- 18 **Wakefield M, Chaloupka F, Kaufman N, et al**. *Do smoking restrictions at home, at school and public places, influence youth smoking?* Chicago IL: University of Illinois, 2000. Impact Teen Working Papers Series no 3.
- 19 **Wakefield MA, Chaloupka FJ, Kaufman NJ, et al**. Effect of restrictions on smoking at home, at school, and in public places on teenage smoking: cross sectional study. *BMJ* 2000;**321**:333–7.
- 20 **Stead M, Hastings G, Tudor-Smith C**. Preventing adolescent smoking: a review of options. *Health Ed J* 1991;**50**:3–8.
- 21 **Perry C, Killen J, Telch M, et al**. Modifying smoking behaviour of teenagers: a school-based intervention. *Am J Public Health* 1980;**70**:722–5.
- 22 **Bellew B, Wayne D**. Prevention of smoking among schoolchildren: a review of research and recommendations. *Health Ed J* 1991;**50**:3–8.
- 23 **CDC**. *Guidelines for school health programs to prevent tobacco use and addiction*. *MMWR* 1994;**43**:1–18.
- 24 **Flay BR, Koepke D, Thomson SJ, et al**. Six-year follow-up of the first Waterloo school smoking prevention trial. *Am J Public Health* 1989;**79**:1371–6.
- 25 **Aubá J, Villalbí JR**. Prevención desde la escuela del uso de tabaco y de otras sustancias adictivas. *Gac Sanit* 1990;**4**:70–5.
- 26 **Goldstein, H**. *Multilevel statistical models*. London: Edward Arnold, 1995.
- 27 **Kreft I, de Leeuw J**. *Introducing multilevel modeling*. London: Sage, 1998.
- 28 **Carey K**. A Multilevel modeling approach to analysis of patient costs under managed care. *Health Econ* 2000;**9**:435–46.
- 29 **Scott A, Shiell A**. Do fee descriptors influence treatment choices in general practice? A multilevel discrete choice model. *J Health Econ* 1997;**16**:323–42.
- 30 **Rice N, Jones AM**. Multilevel models and health economics. *Health Econ* 1997;**6**:561–75.
- 31 **Goldstein H, Rasbash J**. Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society A* 1996;**159**:505–13.
- 32 **MLWIN version 1.0**. *Multilevel models project*. London: Institute of Education, University of London.
- 33 **Wright DB**. Extra-binomial variation in multilevel logistic models with sparse structures. *Br J Math Stat Psychol* 1997;**50**:21–9.
- 34 **Jacob M**. Extra-binomial variation in logistic multilevel models – a simulation. *Multilevel Modelling Newsletter* 2000;**12**:8–14.
- 35 **Stephens T, Morin M, eds** (with Health Canada). *Youth Smoking Survey, 1994: technical report*. Ottawa, Canada: Minister of Supply and Services, 1996.
- 36 **Wang S-Q, Yu J-J, Zhu B-P, et al**. Cigarette smoking and its risk factors among senior high school students in Beijing, China, 1988. *Tobacco Control* 1994;**3**:107–14.
- 37 **Smith C, Nutbeam D, Moore L, et al**. Current changes in smoking attitudes and behaviours among adolescents in Wales, 1986–92. *J Public Health Med* 1994;**16**:165–71.
- 38 **US Department of Health and Human Services**. *Preventing tobacco use among young people; a report of the Surgeon General, 1994*. Atlanta, GA: Public Health Service, Center for Disease Control and Prevention, Office on Smoking and Health, 1994. [US Government Printing Office Publication No S/N 017-001-00491-0].
- 39 **Villalbí JR, Nebot M, Comin E, et al**. Consumo precoz de tabaco en escolares. *Rev Sanid Hig Publica (Madr)* 1990; **64**:613–23.
- 40 **Mendoza R, Batista JM, Sánchez M, et al**. El consumo de tabaco, alcohol y otras drogas en los adolescentes escolarizados españoles. *Gac Sanit* 1998;**12**:263–71.
- 41 **Redmond WH**. Trends in adolescent cigarette use: the diffusion of daily smoking. *J Behav Med* 1999;**22**:379–95.
- 42 **Amos A**. Young people, tobacco and 1992. *Health Ed J* 1991;**50**:26–30.
- 43 **Law M, Tang JL**. An analysis of effectiveness of interventions intended to help people stop smoking. *Arch Intern Med* 1995;**155**:1933–41.
- 44 **Aveyard P, Cheng KK, Almond J, et al**. Cluster randomised controlled trial of expert system based on the transtheoretical (“stages of change”) model for smoking prevention and cessation in schools. *BMJ* 1999;**319**:948–53.
- 45 **Prochaska JO, Aveyard P, Cheng KK, et al**. Stages of change model for smoking prevention and cessation in schools. *BMJ* 2000;**320**:447.