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

J Pinilla, B González, P Barber and Y Santana

J. Epidemiol. Community Health 2002;56;227-232
doi:10.1136/jech.56.3.227

Updated information and services can be found at:
http://jech.bmjjournals.com/cgi/content/full/56/3/227

These include:

References
This article cites 37 articles, 9 of which can be accessed free at:
http://jech.bmjjournals.com/cgi/content/full/56/3/227#BIBL

8 online articles that cite this article can be accessed at:
http://jech.bmjjournals.com/cgi/content/full/56/3/227#otherarticles

Rapid responses
You can respond to this article at:
http://jech.bmjjournals.com/cgi/eletter-submit/56/3/227

Email alerting service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article

Topic collections
Articles on similar topics can be found in the following collections

Smoking (1072 articles)
Other Statistics and Research Methods: descriptions (587 articles)
Adolescent medicine (491 articles)

Notes

To order reprints of this article go to:
http://www.bmjjournals.com/cgi/reprintform

To subscribe to Journal of Epidemiology and Community Health go to:
http://www.bmjjournals.com/subscriptions/
PUBLIC HEALTH POLICY AND PRACTICE

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

J Pinilla, B González, P Barber, Y Santana

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.
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 \( p = 0.1 \), and an error of 6.76% for \( p = 0.5 \), 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,\(^{26} \) as well as in other areas.\(^{25-26} \)

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

\[
Y_{ij} = \beta_j + e_{ij}
\]

where \( \beta_j \) 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 \( \beta_j \) from its mean \( \bar{\beta} \), equation (1) can be expressed as:

\[
Y_{ij} = \bar{\beta} + u_j + e_{ij} \quad \text{with} \quad u_j = \beta_j - \bar{\beta}
\]

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...
with mean $\theta$ and variance $\sigma^2_{ij}=\sigma^2_{\epsilon}$, which stands for differences in the variable $Y$ attributable to the group. It also assumes that the error component $\epsilon_i$ is also distributed normally with mean $0$ and variance $\sigma^2$. And finally it assumes that the random effects $u_i$ and the error component $\epsilon_i$ are independent, $u_i=0$, and that the $\epsilon_i$ are all independent one from another.

The objective, therefore, consists in estimating from the available data the variances $\sigma^2_{ij}$ and $\sigma^2$ 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 $p$ the intraclass correlation as the percentage of the total variance among the $J$ groups.

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

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 $X_i$ 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:

$$Y_{ij} = X_{ij}'\beta + u_i + \epsilon_{ij} \quad (4)$$

in which $X$ includes $K$ regressors and $\epsilon_i \sim N(0, \sigma^2)$.

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

$$Y_{ij} = X_{ij}'\beta' + u_i + \epsilon_{ij} \quad (5)$$

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 $\text{Var}(Y_{ij})=\pi(1-\pi)_ij$, where $\pi$ 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, $\sigma^2_{ij}$, 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 estimates of the intraclass correlations in these models, but their interpretation is unclear.

The sequential process of modelling

To ensure the convergence of the estimations, we collapsed the dependent variable $Y$, 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 MLWin, version 1.0. We have 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. 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 simulation investigates the effects of sample size and model mis specification 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
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, $\sigma^2_{ui}$, 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>
<thead>
<tr>
<th>Table 1</th>
<th>Descriptive statistics of the variables in the models</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of variable</strong></td>
<td><strong>Definition</strong></td>
</tr>
<tr>
<td>Dependent variable</td>
<td>Do you smoke?</td>
</tr>
<tr>
<td>No smoke</td>
<td>1590 (85.8)</td>
</tr>
<tr>
<td>Yes, less than once a week</td>
<td>97 (5.6)</td>
</tr>
<tr>
<td>Yes, on weekends</td>
<td>55 (3.0)</td>
</tr>
<tr>
<td>Yes, smokes daily</td>
<td>116 (6.3)</td>
</tr>
<tr>
<td>Individual explanatory covariates</td>
<td>Sex: Male</td>
</tr>
<tr>
<td>Female</td>
<td>907 (47.7)</td>
</tr>
<tr>
<td>Interest in school Very little</td>
<td>329 (17.3)</td>
</tr>
<tr>
<td>Little</td>
<td>693 (36.5)</td>
</tr>
<tr>
<td>Quite</td>
<td>657 (34.6)</td>
</tr>
<tr>
<td>High</td>
<td>222 (11.7)</td>
</tr>
<tr>
<td>Alcohol consumption Never</td>
<td>1263 (66.6)</td>
</tr>
<tr>
<td>Occasionally or on weekends</td>
<td>607 (32.0)</td>
</tr>
<tr>
<td>Same times during the week or every day</td>
<td>27 (1.4)</td>
</tr>
<tr>
<td>Number of smokers living at home</td>
<td>Discrete variable</td>
</tr>
<tr>
<td>Mean</td>
<td>1.25, SD 1.20</td>
</tr>
<tr>
<td>Best friend smokes? Yes</td>
<td>442 (23.5)</td>
</tr>
<tr>
<td>No</td>
<td>1439 (76.5)</td>
</tr>
<tr>
<td>School characteristics (context variables)</td>
<td>Antismoking rules in the schools (number of schools)</td>
</tr>
<tr>
<td>Public</td>
<td>8</td>
</tr>
<tr>
<td>Private</td>
<td>1</td>
</tr>
<tr>
<td>Subsidised</td>
<td>1</td>
</tr>
<tr>
<td>Medium/strong</td>
<td>15</td>
</tr>
<tr>
<td>% of academic failure in the school (Grades 1 and 2 of the ESO)</td>
<td>Mean</td>
</tr>
<tr>
<td>SD</td>
<td>18.1</td>
</tr>
<tr>
<td>Median</td>
<td>46</td>
</tr>
<tr>
<td>Min</td>
<td>21</td>
</tr>
<tr>
<td>Max</td>
<td>100</td>
</tr>
</tbody>
</table>

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

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

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.
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 no 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, 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 subsidized schools; the rate of academic success in each school 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.

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. According to our survey, 14.3% of students aged 13–14 smoke, and almost half of these do so daily. 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. 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 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 P17/99 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

23 CDC. Guidelines for school health programs to prevent tobacco use and addiction. MMWR 1994; 43:1–18.