Published in final edited form as:
Cancer Epidemiol Biomarkers Prev. 2014 October ; 23(10): 2165–2174. doi: 10.1158/1055-9965.EPI-14-0467.

A Risk Prediction Model for Smoking Experimentation in Mexican American Youth

Rajesh Talluri[1], **Anna V. Wilkinson**[2], **Margaret R. Spitz**[3], and **Sanjay Shete**[1,4]

[1]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

[2]The University of Texas School of Public Health, Austin Regional Campus, Austin, TX, USA

[3]Baylor College of Medicine, Houston, TX, USA

[4]Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Abstract

Background—Smoking experimentation in Mexican American youth is problematic. In light of the research showing that preventing smoking experimentation is a valid strategy for smoking prevention, there is a need to identify Mexican American youth at high risk for experimentation.

Methods—A prospective population-based cohort of 1179 adolescents of Mexican descent was followed for 5 years starting in 2005–06. Participants completed a baseline interview at a home visit followed by three telephone interviews at intervals of approximately 6 months and additional interviews at two home visits in 2008–09 and 2010–11. The primary end point of interest in this study was smoking experimentation. Information regarding social, cultural, and behavioral factors (e.g., acculturation, susceptibility to experimentation, home characteristics, household influences) was collected at baseline using validated questionnaires.

Results—Age, sex, cognitive susceptibility, household smoking behavior, peer influence, neighborhood influence, acculturation, work characteristics, positive outcome expectations, family cohesion, degree of tension, ability to concentrate, and school discipline were found to be associated with smoking experimentation. In a validation dataset, the proposed risk prediction model had an AUC of 0.719 (95% confidence interval, 0.637 to 0.801)for predicting absolute risk for smoking experimentation within 1 year.

Conclusions—The proposed risk prediction model is able to quantify the risk of smoking experimentation in Mexican American adolescents.

Keywords

Smoking; experimentation; risk prediction; cohort study; Mexican American youth

Corresponding author: Sanjay Shete, PhD, Department of Biostatistics, Unit 1411, The University of Texas MD Anderson Cancer Center, 1400 Pressler Dr, FCT4.6002, Houston, TX 77030, U.S.A., Phone: (713) 745-2483, Fax: (713) 563-4243, sshete@mdanderson.org.

Conflict of Interest statement: The authors declare no conflicts of interest.

Introduction

One out of three cancer deaths in the United States is caused by smoking(1), and longer duration and greater intensity of smoking increase the risk of lung cancer significantly(2–6). Early smoking experimentation is associated with a higher risk of habitual smoking(7–9) Also, individuals who experiment with smoking at an earlier age are less likely to successfully quit(7, 10). Several studies have shown that delaying or preventing experimentation is a valid strategy for smoking prevention(7, 10, 11). Because tobacco use results in diseases that cause the premature deaths of more than half a million Americans each year(12), even modest declines in smoking incidence could lead to remarkable public health benefits(13).

Few studies have focused on analyzing and preventing smoking experimentation. The risk factors associated with smoking experimentation are household smoking(14–18), cognitive susceptibility(15–17, 19–21), outcome expectations(14, 15, 20, 22–24), peer influence(15, 17, 24, 25), marketing/media influences(26–30), lower income(31), and lower education(31–33). However, most of the subjects in prior studies were Caucasian, and risk factors for other ethnic groups may vary owing to cultural and social differences between the populations. Hispanics are the most rapidly increasing ethnic group in the United States, and most US Hispanics are of Mexican origin. In 2011, nationally, 48.6% of Hispanics ever tried to smoke compared to 44.2% of non-Hispanic whites, and in Texas, 54.3% of Hispanics ever tried to smoke compared to 49.2% of non-Hispanic whites(34). Because of the higher incidence of cigarette experimentation in this population, there is a need to examine the risk factors associated with cigarette experimentation.

Risk prediction models are being developed to predict the risk of a variety of cancers(35–44), and cardiovascular diseases(45). A risk prediction model for smoking experimentation would be useful in identifying youth at high risk of becoming experimenters who may benefit from targeted interventions(46) to prevent progress along the smoking trajectory. Such risk prediction models can also be used to improve the design of intervention and prevention studies(47).

In this study, we developed a risk prediction model for smoking experimentation based on data from a prospective cohort of Mexican American youth. Our approach accounted for variability in the sampled cohort by resampling the data and aggregating the parameter estimates for the resampled datasets. We then used a resampling-based model selection algorithm to select the predictors to include in the final multivariable risk model. This approach guards against over-fitting the model and reduces the variance of the model parameters. The performance of the risk prediction model was evaluated using the area under the receiver operating characteristic curve (AUC). Using the risk prediction model, we computed the absolute risk of smoking experimentation in Mexican American youth.

Materials and Methods

Participants, Study Setting, and Population

The study participants were recruited from a population-based cohort of Mexican-American households that was part of a prospective study of smoking behavior involving adolescents of Mexican descent that was launched in 2001 at The University of Texas MD Anderson Cancer Center. The individuals recruited for the study were self-identified Mexican Americans who resided in Houston, Texas. A description of the cohort recruitment methodology has been published(48).

From this cohort, households with age-eligible potential participants (adolescents between the ages of 11 and 14 years) were identified. One child per household was enrolled in the study. In total, 3000 households had age-eligible potential participants. Of these 3000 households, 1425 households were contacted and over 90% of them agreed to participate in the study. Each participant enrolled in the study completed a 5-minute personal interview, during which they provided their sex, age, nativity status, and acculturation information. Each participant was then given a personal digital assistant (PDA) to complete the remainder of the interview so as to ensure privacy. The institutional review board at MD Anderson Cancer Center approved all aspects of this study.

Outcomes and Predictors

The primary end point of interest in this study was smoking experimentation. Participants completed a baseline interview at a home visit in 2005–06, followed by three telephone interviews at intervals of approximately 6 months and additional interviews at two home visits in 2008–09 and 2010–11. The participants' smoking experimentation status was assessed using two questions at each interview: "Have you ever smoked a cigarette?" and "Have you ever tried a cigarette, even a puff?". Individuals who answered "No" to these two questions were labeled as non-experimenters, but individuals who responded "Yes" to either of the questions were categorized as experimenters. The total number of individuals from which data were collected was 1328, out of which 149 individuals were previous experimenters or smokers at baseline and therefore, following the standard guidelines for a prospective cohort analysis(49), were excluded from our analyses.

Because only current information would be available for an individual for whom we want to predict the risk of smoking experimentation, we only used information collected at the baseline interview to model the risk of smoking experimentation. In total there were 146 continuous and categorical baseline predictors, including demographics (e.g., age, sex), cognitive susceptibility (e.g., "Would you smoke a cigarette if your best friend offers you one?"), household smoking behavior (e.g., "Does your father/mother/brother/sister smoke?"), peer influence (e.g., "How many of your friends smoke?"), family cohesion (e.g., "Does your family support each other?"), smoking messages, positive outcome expectations (e.g., "Do you think smoking would make you look more mature?"), work smoking (e.g., "Do people smoke where you work?"), school discipline (e.g., "How many detentions have you had in your school?"), and acculturation (e.g., "In which language do you generally think?").

Statistical Methods

The relationships between the predictors and smoking experimentation were assessed using the Cox proportional hazards regression model. The data were randomly split into 1000 training sets (constituting 67% of the individuals in the study), and 1000 test sets (constituting the remaining 33% of the individuals in the study). The training sets were used to develop the risk prediction model, and the corresponding test sets were used to validate the model.

Risk Model Building

The predictors were selected using a two-stage approach. In the first stage, survival regression was performed using the Cox proportional hazards model by regressing the time of smoking experimentation with each predictor individually. For an individual who experimented between two interviews, the midpoint of the interval between the two interviews was used as the time of smoking experimentation(50). The predictors that were significant at the 0.05 level individually were selected for the next step of the analysis. In the second stage, all the predictors that passed the first stage were included in a multivariable Cox proportional hazards regression model and regressed with the time of smoking experimentation. Backward selection was performed on the multivariable model to remove predictors that were not significant at the 0.05 level.

This two-stage approach is the standard used for developing risk prediction models (e.g. (51)). However, this approach does not account for the variability associated with the cohort being a random sample from the population. Hence, we applied a novel approach called Resampling-based Model Selection and Aggregation (RMSA) to account for this variability and improve the performance of the risk prediction model. The RMSA approach was accomplished using the following steps.

1. ***Resampling data:*** The data were randomly split into K (=1000) training sets and test sets. The K training sets were used to develop K multivariable models (one for each training set) using the standard two-stage approach mentioned above.
2. ***Importance of predictors:*** We computed the number of times each of the 146 predictors was selected in the K multivariable models. The higher the frequency the more likely the variable is important for predicting smoking experimentation.
3. ***Model building using a threshold:*** The final model included all predictor variables that were selected in at least C% of the K models. (Details about how the value of C was determined are presented in Step 6.)
4. ***Parameter estimates for resampled datasets:*** For each of the K training sets, the final model from step 3 was used to estimate the parameters of the Cox proportional hazards model.
5. ***Aggregation of estimates from resampled datasets:*** A random effect model, without assuming independence of the resampled datasets, was used to aggregate the parameter estimates and the associated variance-covariance matrix.

6. ***Assessment of model fit:*** The final model with the aggregated estimates of the model parameters was used to analyze the K test sets to assess the performance of the model. The receiver operating characteristic (ROC) curves for each of the models were constructed by computing the specificity and sensitivity of the model. The area under the ROC curve (AUC) was used to determine the model's ability to predict smoking experimentation. The process was repeated to obtain the optimal C % threshold that corresponded to the threshold of the model with the highest mean AUC value.

Absolute Risk Prediction

Our risk prediction model estimates the absolute risk of an individual experimenting with cigarettes in the next 1 to 5 years. The risk prediction model is based on the Cox proportional hazards model and developed using the RMSA approach, $h(t) = h_0(t)\exp(X\beta)$, where $h(t)$ is the hazard function, $h_0(t)$ is the baseline hazard function, X contains the predictors, and β contains the regression coefficients. Using the estimates of β and the variance-covariance matrix for β, M (=1000) random samples of the regression coefficients $[\beta^{(0)}, \beta^{(1)}, \ldots, \beta^{(M)}]$ were sampled from a multivariate normal distribution. For an individual with a set of predictors X, the hazard functions $[h(t)^{(0)}, h(t)^{(1)}, \ldots, h(t)^{(M)}]$ corresponding to $[\beta^{(0)}, \beta^{(1)}, \ldots, \beta^{(M)}]$ were computed. The probability of experimenting with smoking in the next T years was estimated using $\frac{1}{M}\sum_{i=1}^{M}\int_0^T h(t)^{(i)}$. This procedure quantifies uncertainty in the risk estimate for an individual, which can then be used to compute the 95% confidence interval for the risk of smoking experimentation.

It is cost efficient to provide interventions when individuals are classified as being at high risk for smoking experimentation. We developed two thresholds, P_1 and P_2, and individuals whose absolute risk was lower than P_1 were in the low-risk category and individuals whose absolute risk was higher than P_2 were in the high-risk category. We chose P_1 such that the negative predictive value was set to be 90%. P_2 was chosen to match the number of predicted experimenters with prevalence of experimentation in the population.

Results

Epidemiologic data from 1179 individuals enrolled in the prospective cohort (who were self-reported non-experimenters) were available for developing the risk prediction model for smoking experimentation. The mean age of the participants at baseline was 12.32 years (range, 11.01 to 14.69 years). The number of new experimenters identified over the course of the study was 380 (Table 1). The distribution of select predictors in experimenters and non-experimenters is presented in Table 2. The experimenters were more likely to be male than the non-experimenters (57.6% vs 42.6%;p<0.001). Non-experimenters were more likely than experimenters to say that they definitely would not try a cigarette soon (86.6% vs 74.2%; p<0.001). A higher proportion of experimenters had at least 1 detention in school over the past year (35.8% vs 21.0%; p<0.001), had friends who smoke (19.5% vs 5.8%; p<0.001), and knew whether one needed to show identification to buy cigarettes in their neighborhood (48.2% vs 34.5%;p<0.001).

Univariate analysis using the Cox proportional hazards model was first performed to identify the risk factors associated with smoking experimentation. Of the 146 predictors studied, 69 were significantly associated with smoking experimentation at the 0.05 level (see Supplementary Table S1 and Supplementary Table S2).

Multivariable Risk Model

The multivariable risk model constructed using the RMSA procedure included 18 predictors that were significantly associated with smoking experimentation at the 0.05 level (Table 3). The optimal threshold C (See Methods: Risk Model Building) for the RMSA procedure was estimated to be 22.5%. Work smoking had the highest impact on experimentation, with a hazard ratio (HR) of 2.32 (95% CI, 1.27 to 4.26). Sex was significantly associated with smoking experimentation, with adolescent girls having a lower risk of experimentation (HR=0.61, 95% confidence interval [CI], 0.51 to 0.72). Other predictors that were associated with smoking experimentation were having a mother who smoked (HR=2.22, 95%CI, 1.36 to 3.62), neighborhood characteristics (HR=0.65, 95% CI, 0.55 to 0.78), and having peers who smoke (HR=1.64, 95%CI, 1.39 to 1.93).

Model Validation and Predictive Power of the Model

We randomly sampled 1000 training sets constituting 66.67% of the individuals in the study and 1000 test sets constituting the remaining 33.33% of the individuals in the study. The model was built using the training set and validated using the corresponding test set. The AUC was calculated for each of the 1000 test sets, and the mean AUCs for 1, 2, 3, 4 and 5-year risk of smoking experimentation were 0.719, 0.714, 0.688, 0.671 and 0.666 respectively (Table 4). The ROC curves for 1, 2, 3, 4 and 5-year risk of smoking experimentation are presented in Supplementary Figure S1.

Estimation of Absolute Risk for Smoking Experimentation

We used the risk prediction model to estimate the absolute risk for smoking experimentation in a time interval. The final model was as follows:

$$\begin{aligned} h(t) = h_0(t)\exp\,(&0.341\,Age - 0.495\,Gender + 0.293\,CognitiveSusceptibility1 \\ &+ 0.376\,CognitiveSusceptibility2 + 0.103\,Tension + 0.082\,Concentration \\ &- 0.194\,FamilyCohesion1 - 0.193\,FamilyCohesion2 + 0.799\,MotherSmoking \\ &+ 0.762\,SisterSmoking + 0.479\,OtherSmoking + 0.494\,PeerInfluence1 \\ &+ 0.140\,PeerInfluence2 + 0.845\,WorkSmoking - 0.424\,Neigborhood \\ &+ 0.091\,ThinkingLanguage - 0.195\,POE + 0.036\,Detentions), \end{aligned}$$

where the predictors were as described in Table 3.

As an example, consider an adolescent boy who is 12-years-old, probably not susceptible to trying a cigarette, has a few friends who smoke, has no parents' friends who smoke, has a mother who smokes in the house, has no siblings or others in the household who smoke, is rarely tense and rarely has difficulty concentrating, agrees that smoking would give him bad breath, strongly agrees that they help each other and can do whatever they want in his family, doesn't work, has no detentions or suspensions in school in the past year, thinks

equally in English and Spanish, and knows whether one needs to show identification to buy cigarettes in his neighborhood. The absolute risk for this individual to experiment in the next 1 year is 32.3% (95%CI, 18.0% to 50.0%).

Discussion

Using data from a prospective cohort of Mexican American youth, we developed a multivariable model for predicting risk of smoking experimentation. We proposed an approach called RMSA that accounts for variability associated with the cohort being a random sample from the population, by resampling the data and then aggregating the parameter estimates of the resampled datasets to estimate the model parameters. RMSA also safeguards against over-fitting the model because the model is optimized over all of the resampled datasets. Age, sex, cognitive susceptibility, household smoking behavior, peer influence, neighborhood influence, acculturation, work characteristics, positive outcome expectations, family cohesion, degree of tension, ability to concentrate, and school discipline were found to be significantly associated with smoking experimentation.

Several other studies including our own have identified cognitive susceptibility(15–17, 19, 21), peer influence(14, 15, 17, 25), age(16, 20, 22), and male sex(14, 17, 20, 22) as important risk factors for smoking experimentation. Positive outcome expectations(14, 15, 20, 22, 23), household smoking(14–17), neighborhood characteristics(52), anxiety and depression(25, 53), and school suspensions(54)have also been shown to be associated with smoking experimentation in several studies. Our study also found that family cohesion, co-worker smoking status, and acculturation were associated with smoking experimentation.

Based on findings from this study, we developed an online risk calculator for smoking experimentation that is applicable to Mexican American youths (https://biostatistics.mdanderson.org/SmokingExperimentRisk/). This risk prediction model can be used to identify individuals at high risk of smoking experimentation and provide suitable interventions to reduce the risk. The ability of the risk model to distinguish between experimenters and non-experimenters was measured using the AUC. As a general rule, a prediction model with an AUC greater than 0.7 is considered to have acceptable discriminative ability(55). The AUC for 1-year risk of smoking experimentation in our model was 0.719, which is higher or comparable to risk prediction models for diseases such as breast (0.58), ovarian (0.59), and endometrial (0.68) cancer(56). Furthermore, the AUC for the dataset that included only low- and high-risk individuals (P_1=0.03 and P_2=0.215) was 0.901 for 1-year risk of experimentation. Any model with an AUC greater than 0.9 is considered to have outstanding predictive ability(55). According to the 2012 Surgeons General's Report, nearly 9 out of 10 smokers have experimented with cigarettes by age 18. The time (from baseline) at which an individual's predicted risk exceeds the high-risk threshold (P_2) is of public health relevance. Our risk prediction model estimates this time based on the individual's baseline information and can be used in determining the time at which interventions would be most beneficial.

Interventions are available for many of the modifiable risk factors identified in our risk prediction model (e.g., household smoking, co-worker smoking status, family cohesion,

positive outcome expectations). Interventions such as smoke-free homes(57) are available for individuals who have smokers in the household, and workplace interventions(58) for smoking cessation help reduce the risk of smoking experimentation among adolescents who work. A variety of family therapies (e.g., Bowenian family system(59)) can be administered to improve family cohesion. Similarly, susceptibility to smoking can be reduced using anti-smoking media campaigns(60). The risk of smoking experimentation due to anxiety or tension could be reduced by the use of cognitive-behavioral therapy(61).

Our prospective cohort study has various strengths and limitations. The cohort represents a homogeneous sample of low-income Mexican American youth, who are relatively understudied compared to other populations. The cohort was balanced with respect to sex. The privacy of the participating children was ensured because the data were collected using a PDA, which likely increased accuracy of the participants' self-reports. The study had a high retention rate: 87% of the participants participated in all five follow-up interviews.

The limitation of this risk prediction model is that it can only be used for Mexican American individuals between 11 and 14 years old and may not be applicable to other populations and age groups. In this study, internal validation using separate training and testing datasets was performed. Therefore, the findings are preliminary and need to be validated in external cohorts. Another limitation of the study is that the status of smoking experimentation in the cohort was self-reported, which may include bias. However, the bias was reduced by informing the participants that they may be selected to provide a saliva sample to test their smoking status(62).

This risk prediction model is able to quantify the risk of smoking experimentation in Mexican American adolescents. This model can be used by teachers, parents, and counselors to assess the risk of smoking experimentation in Mexican American youth. This information can then be used to provide suitable interventions to reduce that risk. In the future we plan to include genetic information in the risk model to improve its performance even more, as genetics play an important role in addictive behaviors.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the cohort staff for conducting all field interviews and maintaining the high participation rates. We thank the participants for providing the data and their parents for permitting their children to join the study. Without their support this research would not be possible.

Financial Support: This work was supported in part by National Institutes of Health [grants R01CA131324 (to S. Shete), R01DE022891 (to S. Shete), R25 DA026120 (to S. Shete), CA105203 (to M.R. Spitz) and CA126988 (to A.V. Wilkinson)]. This research was supported in part by Barnhart Family Distinguished Professorship in Targeted Therapy (to S. Shete). This research was supported in part by a cancer prevention fellowship for R. Talluri supported by a grant from the National Institute of Drug Abuse (NIH R25 DA026120). The Mexican American Cohort receives funds (1) collected pursuant to the Comprehensive Tobacco Settlement of 1998 and appropriated by the 76th legislature to The University of Texas MD Anderson Cancer Center, (2) from the Caroline W. Law Fund for Cancer Prevention, and (3) from the Dan Duncan Family Institute for Risk Assessment and Cancer Prevention. The funders did not contribute to the design and conduct of the study; the collection, analysis, or interpretation of the data; or the preparation, review, or approval of the manuscript.

References

1. American Cancer Society. Cancer facts & figures 2013. Atlanta: American Cancer Society; 2013.
2. Doll R, Hill AB. A study of the aetiology of carcinoma of the lung. Br Med J. 1952; 2:1271–86. [PubMed: 12997741]
3. Doll R, Peto R. Cigarette-smoking and bronchial-carcinoma – dose and time relationships among regular smokers and lifelong non-smokers. J Epidemiol Community Health. 1978; 32:303–13. [PubMed: 744822]
4. Hecht SS. Tobacco smoke carcinogens and lung cancer. J Natl Cancer Inst. 1999; 91:1194–210. [PubMed: 10413421]
5. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. CA Cancer J Clin. 2014; 64:9–29. [PubMed: 24399786]
6. Wynder EL, Graham EA. Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma – a study of 684 proved cases. Jama-Journal of the American Medical Association. 1950; 143:329–36.
7. Chassin L, Presson CC, Rose JS, Sherman SJ. The natural history of cigarette smoking from adolescence to adulthood: Demographic predictors of continuity and change. Health Psychol. 1996; 15:478–84. [PubMed: 8973929]
8. Chassin L, Presson CC, Sherman SJ, Edwards DA. The natural-history of cigarette-smoking – predicting young-adult smoking outcomes from adolescent smoking patterns. Health Psychol. 1990; 9:701–16. [PubMed: 2286181]
9. Taioli E, Wynder EL. Effect of the age at which smoking begins on frequency of smoking in adulthood. N Engl J Med. 1991; 325:968–9. [PubMed: 1881424]
10. Ershler J, Leventhal H, Fleming R, Glynn K. The quitting experience for smokers in 6th through 12th grades. Addict Behav. 1989; 14:365–78. [PubMed: 2782120]
11. Breslau N, Peterson EL. Smoking cessation in young adults: Age at initiation of cigarette smoking and other suspected influences. Am J Public Health. 1996; 86:214–20. [PubMed: 8633738]
12. Cole HM, Fiore MC. The war against tobacco: 50 years and counting. JAMA. 2014; 311:131–2. [PubMed: 24399546]
13. Frieden TR. Tobacco control progress and potential. JAMA. 2014; 311:133–4. [PubMed: 24399547]
14. Abroms L, Simons-Morton B, Haynie DL, Chen RS. Psychosocial predictors of smoking trajectories during middle and high school. Addiction. 2005; 100:852–61. [PubMed: 15918815]
15. Flay BR, Hu FB, Richardson J. Psychosocial predictors of different stages of cigarette smoking among high school students'. Prev Med. 1998; 27:A9–A18. [PubMed: 9808813]
16. Jackson C. Cognitive susceptibility to smoking and initiation of smoking during childhood: A longitudinal study. Prev Med. 1998; 27:129–34. [PubMed: 9465363]
17. Pierce JP, Choi WS, Gilpin EA, Farkas AJ, Merritt RK. Validation of susceptibility as a predictor of which adolescents take up smoking in the united states. Health Psychol. 1996; 15:355–61. [PubMed: 8891714]
18. Jackson C, Henriksen L, Dickinson D, Messer L, Robertson SB. A longitudinal study predicting patterns of cigarette smoking in late childhood. Health Educ Behav. 1998; 25:436–47. [PubMed: 9690102]
19. Huang M, Hollis J, Polen M, Lapidus J, Austin D. Stages of smoking acquisition versus susceptibility as predictors of smoking initiation in adolescents in primary care. Addict Behav. 2005; 30:1183–94. [PubMed: 15925127]
20. Wilkinson AV, Waters AJ, Vasudevan V, Bondy ML, Prokhorov AV, Spitz MR. Correlates of susceptibility to smoking among mexican origin youth residing in houston, texas: A cross-sectional analysis. BMC Public Health. 2008; 8
21. Spelman AR, Spitz MR, Kelder SH, Prokhorov AV, Bondy ML, Frankowski RF, et al. Cognitive susceptibility to smoking: Two paths to experimenting among mexican origin youth. Cancer Epidemiol Biomarkers Prev. 2009; 18:3459–67. [PubMed: 19959696]

22. Elder JP, Campbell NR, Litrownik AJ, Ayala GX, Slymen DJ, Parra-Medina D, et al. Predictors of cigarette and alcohol susceptibility and use among hispanic migrant adolescents. Prev Med. 2000; 31:115–23. [PubMed: 10938211]
23. Wilkinson AV, Shete S, Vasudevan V, Prokhorov AV, Bondy ML, Spitz MR. Influence of subjective social status on the relationship between positive outcome expectations and experimentation. J Adolesc Health. 2009; 44:342–8. [PubMed: 19306792]
24. U.S. Department of Health and Human Services. Preventing tobacco use among youth and young adults: A report of the surgeon general. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2012.
25. Patton GC, Carlin JB, Coffey C, Wolfe R, Hibbert M, Bowes G. Depression, anxiety, and smoking initiation: A prospective study over 3 years. Am J Public Health. 1998; 88:1518–22. [PubMed: 9772855]
26. Henriksen L, Schleicher NC, Feighery EC, Fortmann SP. A longitudinal study of exposure to retail cigarette advertising and smoking initiation. Pediatrics. 2010; 126:232–8. [PubMed: 20643725]
27. Lovato C, Watts A, Stead LF. Impact of tobacco advertising and promotion on increasing adolescent smoking behaviours. Cochrane Database Syst Rev. 2011:CD003439. [PubMed: 21975739]
28. Primack BA, Longacre MR, Beach ML, Adachi-Mejia AM, Titus LJ, Dalton MA. Association of established smoking among adolescents with timing of exposure to smoking depicted in movies. J Natl Cancer Inst. 2012; 104:549–55. [PubMed: 22423010]
29. Wilkinson AV, Spitz MR, Prokhorov AV, Bondy ML, Shete S, Sargent JD. Exposure to smoking imagery in the movies and experimenting with cigarettes among mexican heritage youth. Cancer Epidemiol Biomarkers Prev. 2009; 18:3435–43. [PubMed: 19959693]
30. Wilkinson AV, Vandewater EA, Carey FR, Spitz MR. Exposure to pro-tobacco messages and smoking status among mexican origin youth. J Immigr Minor Health. 2013
31. Emmons K, Li FP, Whitton J, Mertens AC, Hutchinson R, Diller L, et al. Predictors of smoking initiation and cessation among childhood cancer survivors: A report from the childhood cancer survivor study. J Clin Oncol. 2002; 20:1608–16. [PubMed: 11896111]
32. Emmons KM, Butterfield RM, Puleo E, Park ER, Mertens A, Gritz ER, et al. Smoking among participants in the childhood cancer survivors cohort: The partnership for health study. J Clin Oncol. 2003; 21:189–96. [PubMed: 12525509]
33. Tao ML, Guo MD, Weiss R, Byrne J, Mills JL, Robison LL, et al. Smoking in adult survivors of childhood acute lymphoblastic leukemia. J Natl Cancer Inst. 1998; 90:219–25. [PubMed: 9462679]
34. Centers for Disease Control and Prevention (CDC). Atlanta (GA): 1991–2013 High School Youth Risk Behavior Survey Data. [cited 2014 Jan 1]. Available from http://nccd.cdc.gov/youthonline/
35. Taplin SH, Thompson RS, Schnitzer F, Anderman C, Immanuel V. Revisions in the risk-based breast-cancer screening-program at group health cooperative. Cancer. 1990; 66:812–8. [PubMed: 2386908]
36. Tice JA, Cummings SR, Ziv E, Kerlikowske K. Mammographic breast density and the gail model for breast cancer risk prediction in a screening population. Breast Cancer Res Treat. 2005; 94:115–22. [PubMed: 16261410]
37. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors (vol 23, pg 1111, 2004). Stat Med. 2005; 24:156.
38. Imperiale TF, Wagner DR, Lin CY, Larkin GN, Rogge JD, Ransohoff DF. Using risk for advanced proximal colonic neoplasia to tailor endoscopic screening for colorectal cancer. Ann Intern Med. 2003; 139:959–65. [PubMed: 14678915]
39. Selvachandran SN, Hodder RJ, Ballal MS, Jones P, Cade D. Prediction of colorectal cancer by a patient consultation questionnaire and scoring system: A prospective study. Lancet. 2002; 360:278–83. [PubMed: 12147370]
40. Wijnen JT, Vasen HFA, Khan PM, Zwinderman AH, van der Klift H, Mulder A, et al. Clinical findings with implications for genetic testing in families with clustering of colorectal cancer. N Engl J Med. 1998; 339:511–8. [PubMed: 9709044]

41. Cho E, Rosner BA, Feskanich D, Colditz GA. Risk factors and individual probabilities of melanoma for whites. J Clin Oncol. 2005; 23:2669–75. [PubMed: 15837981]
42. Fears TR, Guerry D, Pfeiffer RM, Sagebiel RW, Elder DE, Halpern A, et al. Identifying individuals at high risk of melanoma: A practical predictor of absolute risk. J Clin Oncol. 2006; 24:3590–6. [PubMed: 16728488]
43. Hartge P, Whittemore AS, Itnyre J, Mcgowan L, Cramer D, Casagrande JT, et al. Rates and risks of ovarian-cancer in subgroups of white women in the united-states. Obstet Gynecol. 1994; 84:760–4. [PubMed: 7936508]
44. Eastham JA, May R, Robertson JL, Sartor O, Kattan MW. Development of a nomogram that predicts the probability of a positive prostate biopsy in men with an abnormal digital rectal examination and a prostate-specific antigen between 0 and 4 ng/ml. Urology. 1999; 54:709–13. [PubMed: 10510933]
45. Chambless LE, Dobson AJ, Patterson CC, Raines B. On the use of a logistic risk score in predicting risk of coronary heart-disease. Stat Med. 1990; 9:385–96. [PubMed: 2362977]
46. Freedman AN, Seminara D, Gail MH, Hartge P, Colditz GA, Ballard-Barbash R, et al. Cancer risk prediction models: A workshop on development, evaluation, and application. J Natl Cancer Inst. 2005; 97:715–23. [PubMed: 15900041]
47. Bach PB, Kattan MW, Thornquist MD, Kris MG, Tate RC, Barnett MJ, et al. Variations in lung cancer risk among smokers. J Natl Cancer Inst. 2003; 95:470–8. [PubMed: 12644540]
48. Wilkinson AV, Spitz MR, Strom SS, Prokhorov AV, Barcenas CH, Cao YM, et al. Effects of nativity, age at migration, and acculturation on smoking among adult houston residents of mexican descent. Am J Public Health. 2005; 95:1043–9. [PubMed: 15914831]
49. Ray WA. Evaluating medication effects outside of clinical trials: New-user designs. Am J Epidemiol. 2003; 158:915–20. [PubMed: 14585769]
50. Ismail BF, Craven T, Banerji MA, Grp AT. Effect of intensive treatment of hyperglycaemia on microvascular outcomes in type 2 diabetes: An analysis of the accord randomised trial (vol 376, pg 419, 2010). Lancet. 2010; 376:1466. [PubMed: 21036274]
51. Spitz MR, Hong WK, Amos CI, Wu XF, Schabath MB, Dong Q, et al. A risk model for prediction of lung cancer. J Natl Cancer Inst. 2007; 99:715–26. [PubMed: 17470739]
52. Xue Y, Zimmerman MA, Caldwell CH. Neighborhood residence and cigarette smoking among urban youths: The protective role of prosocial activities. Am J Public Health. 2007; 97:1865–72. [PubMed: 17761584]
53. Okeke NL, Spitz MR, Forman MR, Wilkinson AV. The associations of body image, anxiety, and smoking among mexican-origin youth. J Adolesc Health. 2013; 53:209–14. [PubMed: 23669646]
54. Breslau N, Fenn N, Peterson EL. Early smoking initiation and nicotine dependence in a cohort of young-adults. Drug Alcohol Depend. 1993; 33:129–37. [PubMed: 8261877]
55. Hosmer, DW.; Lemeshow, S. Applied logistic regression. 2. New York: Wiley; 2000.
56. Pfeiffer RM, Park Y, Kreimer AR, Lacey JV Jr, Pee D, Greenlee RT, et al. Risk prediction for breast, endometrial, and ovarian cancer in white women aged 50 y or older: Derivation and validation from population-based cohort studies. PLoS Med. 2013; 10:e1001492. [PubMed: 23935463]
57. Klepeis NE, Hughes SC, Edwards RD, Allen T, Johnson M, Chowdhury Z, et al. Promoting smoke-free homes: A novel behavioral intervention using real-time audio-visual feedback on airborne particle levels. PLoS One. 2013; 8
58. Cahill K, Moher M, Lancaster T. Workplace interventions for smoking cessation. Cochrane Database of Systematic Reviews. 2008
59. Bowen, M. Family therapy in clinical practice. New York: J. Aronson; 1978.
60. Meshack AF, Hu S, Pallonen UE, McAlister AL, Gottlieb N, Huang P. Texas tobacco prevention pilot initiative: Processes and effects. Health Educ Res. 2004; 19:657–68. [PubMed: 15199003]
61. Sawyer MC, Nunez DE. Cognitive-behavioral therapy for anxious children: From evidence to practice. Worldviews Evid Based Nurs. 2014
62. Murray DM, Oconnell CM, Schmid LA, Perry CL. The validity of smoking self-reports by adolescents – a reexamination of the bogus pipeline procedure. Addict Behav. 1987; 12:7–15. [PubMed: 3565116]

Impact

Accurately identifying Mexican American adolescents who are at higher risk for smoking experimentation who can be intervened will substantially reduce the incidence of smoking and thereby subsequent health risks.

Table 1

Progression of the cohort from non-experimenters to experimenters at various stages of the study.

Cohort (N=1179)	B	T1	T2	T3	H1	H2	Total
New Experimenters*	0	59	48	43	86	144	380
Non-Experimenters	1179	1120	1072	1029	943	799	799

* Individuals who reported experimentation in this interview but were non-experimenters before this interview.

B corresponds to baseline home visit interview. T1, T2, T3 correspond to three telephone interviews and H1, H2 correspond to interviews at two home visits in chronological order.

Table 2

Distribution of study population by select variables at baseline.

Variables	Experimenters (N=380)	Non Experimenters (N=799)	P-value*
Mean Age (SD+)	12.57 (0.92)	12.2 (0.85)	<.001
Sex, n(%)			
Males	219 (57.6)	340 (42.6)	
Females	161 (42.4)	459 (57.4)	<.001
Cognitive Susceptibility			
Do you think you will try a cigarette soon?			
Definitely Not	282 (74.2)	692 (86.6)	
Probably Not	79 (20.8)	101 (12.6)	<.001
Probably Yes	18 (4.7)	6 (0.8)	
Definitely Yes	1 (0.3)	0 (0)	
Do you think you will be smoking cigarettes 1 year from now?			
Definitely Not	319 (83.9)	743 (93.0)	
Probably Not	58 (15.3)	55 (6.9)	<.001
Probably Yes	3 (0.8)	1 (0.1)	
Definitely Yes	0 (0.0)	0 (0.0)	
Do you feel anxious or tense?			
Never	224 (58.9)	585 (73.2)	
Very Rarely	82 (21.6)	120 (15.0)	<.001
Rarely	42 (11.1)	47 (5.9)	
Sometimes	17 (4.5)	27 (3.4)	
Mostly	5 (1.3)	14 (1.7)	
Always	10 (2.6)	6 (0.8)	
Do you have difficulty concentrating?			
Never	216 (56.8)	529 (66.2)	
Very Rarely	57 (15.0)	130 (16.3)	<.001
Rarely	46 (12.1)	68 (8.5)	
Sometimes	31 (8.2)	34 (4.3)	
Mostly	19 (5.0)	19 (2.4)	
Always	11 (2.9)	19 (2.4)	
Family Cohesion			
In my family we really help and support one another.			
Strongly Disagree	8 (2.1)	12 (1.5)	
Disagree	14 (3.7)	18 (2.3)	
Agree	230 (60.5)	427 (53.5)	0.015
Strongly Agree	128 (33.7)	341 (42.7)	
We can do whatever we want in our family.			
Strongly Disagree	201 (52.9)	380 (47.6)	
Disagree	166 (43.7)	368 (46.0)	

Variables	Experimenters (N=380)	Non Experimenters (N=799)	P-value*
Agree	9 (2.4)	46 (5.8)	0.023
Strongly Agree	4 (1.1)	5 (0.6)	
Positive Outcome Expectations			
I think smoking would give me bad breath.			
Strongly Disagree	17 (4.5)	22 (2.8)	
Disagree	7 (1.8)	8 (1.0)	
Agree	103 (27.1)	182 (22.8)	0.054
Strongly Agree	253 (66.6)	587 (73.5)	
Peer Influence			
How many of your friends smoke?			
None	306 (80.5)	753 (94.2)	
Few	57 (15.0)	37 (4.6)	<.001
Some	13 (3.4)	8 (1.0)	
Most	3 (0.8)	1 (0.1)	
All	1 (0.3)	0 (0.0)	
How many of your parents friends smoke?			
None	158 (41.6)	449 (56.2)	
Few	152 (40.0)	279 (34.9)	
Some	52 (13.7)	60 (7.5)	<.001
Most	16 (4.2)	9 (1.1)	
All	2 (0.5)	2 (0.3)	
School Discipline			
During this school year how many detentions or suspensions have you had?			
0	244 (64.2)	631 (79.0)	
>0	136 (35.8)	168 (21.0)	<.001
Acculturation			
In which language do you generally think?			
Only Spanish	22 (5.8)	54 (6.8)	
More Spanish Than English	33 (8.7)	120 (15.0)	
Both Equally	105 (27.6)	225 (28.2)	
More English than Spanish	85 (22.4)	190 (23.8)	0.002
Only English	135 (35.5)	209 (26.2)	
Neighborhood Characteristics			
If you try to buy cigarettes will you be asked to show your ID?			
Yes/No	183 (48.2)	276 (34.5)	
I don’t know	197 (51.8)	523 (65.5)	<.001
Work Characteristics			
Do people smoke where you work?			
Yes	8 (2.1)	7 (0.9)	
No/I don’t work	372 (97.9)	792 (99.1)	0.096
Household Smoking Behavior			

Variables	Experimenters (N=380)	Non Experimenters (N=799)	P-value*
Does your mother/stepmother smoke?			
No	343 (90.26)	752 (94.12)	
Smokes in the house	14 (3.68)	7 (0.88)	
Smokes but not in the house	20 (5.26)	37 (4.63)	0.005
Smokes but doesn't live with me	3 (0.79)	3 (0.38)	
Does your sister smoke?			
Have no sisters	64 (16.8)	130 (16.3)	
No	304 (80)	654 (81.9)	
Smokes in the house	1 (0.3)	2 (0.3)	
Smokes but not in house	5 (1.3)	11 (1.4)	
Smokes but doesn't live with me	6 (1.6)	2 (0.3)	0.146
Does anybody else who lives in the house with you smoke?			
No	341 (89.7)	752 (94.1)	
Smoke in the house	4 (1.1)	5 (0.6)	
Smoke but not in house	35 (9.2)	42 (5.3)	0.023

*P value from the two-sided Fisher exact test (for categorical variables) and Student's t test (for continuous variables).

+SD = Standard deviation.

Table 3

Multivariable regression model for smoking experimentation based on RMSA framework.

Risk factor	Coefficient	SD	P-value
Age	0.341	0.051	<0.001
Sex	−0.495	0.088	<0.001
CognitiveSusceptibility1	0.293	0.095	0.002
CognitiveSusceptibility2	0.376	0.131	0.004
Tension	0.103	0.042	0.013
Concentration	0.082	0.036	0.021
FamilyCohesion1	−0.194	0.072	0.007
FamilyCohesion2	−0.193	0.072	0.007
MotherSmoking	0.799	0.249	0.001
SisterSmoking	0.762	0.354	0.031
OtherSmoking	0.479	0.151	0.002
PeerInfluence1	0.494	0.084	<0.001
PeerInfluence2	0.140	0.053	0.008
WorkSmoking	0.845	0.308	0.006
Neighborhood	−0.424	0.088	<0.001
ThinkingLanguage	0.091	0.038	0.016
POE	−0.195	0.061	0.001
Detentions	0.036	0.015	0.019

CognitiveSusceptibility1 – “Do you think that you will try a cigarette soon?”

CognitiveSusceptibility2 – “Do you think you will be smoking cigarettes in 1 year from now?”

Tension – “Do you feel anxious or tense?”

Concentration– “Do you have difficulty concentrating?”

FamilyCohesion1– “In my family we really help and support one another”

FamilyCohesion2– “We can do whatever we want in my family”

MotherSmoking– “Does your mother/stepmother smoke?”

SisterSmoking– “Do any of your sisters/stepsisters smoke?”

OtherSmoking – “Does anybody else who lives in the house with you smoke?”

PeerInfluence1– “How many of your friends smoke?”

PeerInfluence2– “How many of your parents’ friends smoke?”

WorkSmoking– “Do people smoke where you work?”

Neighborhood– “If you try to buy cigarettes will you be asked to show ID?”

ThinkingLanguage– “In what language do you usually think?”

POE– “I think smoking would make give me bad breath”

Detentions– “During this school year how many detentions and suspensions have you had?”

Table 4

Area under the curve for 1 to 5-year risk of smoking experimentation.

AUC	Mean	Median	SD
1 Year	0.719	0.720	0.042
2 Year	0.714	0.715	0.031
3 Year	0.688	0.689	0.028
4 Year	0.671	0.671	0.025
5 Year	0.666	0.666	0.024

SD = Standard deviation.