

A Tool to Identify Adolescents at Risk of Cigarette Smoking Initiation

Marie-Pierre Sylvestre, PhD,^{a,b} Nancy Hanusaik, PhD,^a David Berger, MSc,^a Erika Dugas, MSc,^a Lori Pbert, PhD,^c Jonathan Winickoff, MD, MPH,^{d,e,f} Jennifer L. O'Loughlin, PhD^{a,b}

abstract

OBJECTIVES: To describe the development of a prognostic tool to identify adolescents at risk for transitioning from never to ever smoking in the next year.

METHODS: Data were drawn from the Nicotine Dependence in Teens study, a longitudinal investigation of adolescents (1999 to present). A total of 1294 students initially age 12 to 13 years were recruited from seventh-grade classes in 10 high schools in Montreal. Self-report questionnaire data were collected every 3 months during the 10-month school year over 5 years (1999–2005) until participants completed high school ($n = 20$ cycles). Prognostic variables for inclusion in the multivariable analyses were selected from 58 candidate predictors describing sociodemographic characteristics, smoking habits of family and friends, lifestyle factors, personality traits, and mental health. Cigarette smoking initiation was defined as taking even 1 puff on a cigarette for the first time, as measured in a 3-month recall of cigarette use completed in each cycle.

RESULTS: The cumulative incidence of cigarette smoking initiation was 16.3%. Data were partitioned into a training set for model-building and a testing set to evaluate the performance of the model. The final model included 12 variables (age, 4 worry or stress-related items, 1 depression-related item, 2 self-esteem items, and 4 alcohol- or tobacco-related variables). The model yielded a c-statistic of 0.77 and had good calibration.

CONCLUSIONS: This short prognostic tool, which can be incorporated into busy clinical practice, was used to accurately identify adolescents at risk for cigarette smoking initiation.



^aCentre de Recherche du Centre Hospitalier de l'Université de Montréal, Montréal, Canada; ^bDepartment of Social and Preventive Medicine, School of Public Health, Université de Montréal, Montréal, Canada; ^cDivision of Preventive and Behavioral Medicine, Center for Tobacco Treatment Research and Training, Medical School, University of Massachusetts, Worcester, Massachusetts; ^dDivision of General Academic Pediatrics, Department of Pediatrics, and ^eTobacco Research and Treatment Center, Massachusetts General Hospital, Boston, Massachusetts; and ^fDepartment of Pediatrics, Harvard Medical School, Harvard University, Boston, Massachusetts

Dr Sylvestre conceived and designed the study, interpreted the data, drafted the initial manuscript, and reviewed and revised the manuscript; Dr Hanusaik contributed to the interpretation of the data, drafted the initial manuscript, and reviewed and revised the manuscript; Mr Berger conducted the analysis and contributed to the interpretation of the data; Ms Dugas coordinated the study and reviewed and revised the manuscript; Drs Winickoff and Pbert critically reviewed the manuscript for important intellectual content; Dr O'Loughlin conceived of and designed the study, obtained the funding, participated in its coordination, contributed to interpretation of the data, and participated in the drafting, review, and revision of the manuscript; and all authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

DOI: <https://doi.org/10.1542/peds.2017-3701>

Accepted for publication Aug 20, 2018

WHAT'S KNOWN ON THIS SUBJECT: Pediatricians and family practitioners are important sources of smoking preventive counseling. However, the lack of a prognostic tool to assist clinicians rapidly identify youth at risk of transitioning from never to ever smoking is a major barrier to counseling.

WHAT THIS STUDY ADDS: Using data from a longitudinal investigation of adolescents, we developed a 12-item prognostic tool for use in clinical practice to identify adolescents at risk for initiating cigarette smoking. This tool has good predictive ability.

To cite: Sylvestre M-P, Hanusaik N, Berger D, et al. A Tool to Identify Adolescents at Risk of Cigarette Smoking Initiation. *Pediatrics*. 2018;142(5):e20173701

Cigarette smoking typically begins during adolescence, and **the younger the age of initiation, the greater the risk of daily smoking**,^{1,2} heavy cigarette consumption,^{3,4} nicotine dependence (ND), and difficulty quitting.⁵ The prevalence of “tried smoking” has declined markedly in North American youth (from 20% of US middle school students in 2013⁶ to 7% in 2016⁷ and from 45% of sixth- through ninth-grade Canadian students in 1994⁸ to 8% in 2014–2015⁹). However, 25% to 30% of never-smokers lack firm commitment to never smoke and are classified as “susceptible to smoking.”^{8,10} These individuals represent a key target group for prevention¹¹ because the transition from never to ever smoking can lead to rapidly increasing cigarette use.¹²

Pediatricians and family practitioners are important sources of preventive counseling.¹³ It was recently recommended that education and brief counseling aimed at preventing school-aged youth from trying a first cigarette, be integrated into counseling.^{14,15} However, because of competing priorities,¹⁶ time and/or resource constraints, and low provider self-efficacy,^{17,18} routine counseling remains infrequent.^{16,19} In addition, the lack of a prognostic tool to assist clinicians rapidly identify at-risk youth is a major barrier to delivering counseling.²⁰ The susceptibility to smoking cigarettes (SSC) index is widely used in research to identify individuals at risk for becoming a smoker, but it has not been tested in clinical settings and is focused solely on smoking intentions, disregarding other factors known to affect initiation.^{11,21,22} With an accurate prognostic tool, at-risk youth could be selectively targeted for counseling, rendering counseling more efficient.

We describe the development of a prognostic tool for use by clinicians, which identifies adolescents at risk for transitioning from never to ever

smoking in the next year. It was developed on the basis of a large literature identifying a wide range of factors that predict initiation among never smokers.²² It incorporates **12 questions**, most with **yes or no responses**.

METHODS

This current study is an extension of the Nicotine Dependence in Teens (NDIT) study, a longitudinal investigation of **1294** seventh-grade students ages **12 to 13 years** recruited in **10 high schools** in **Montreal, Canada**.²³ The NDIT study aimed to describe the natural course of cigarette smoking and ND. A total of 55.4% of eligible students participated (some teachers refused to collect consent forms because of a labor dispute). Parents and/or guardians provided informed consent, and participants assented. Questionnaire data (1999–2005) were collected **every 3 months** during the 10-month school year **over 5 years**, for a total of 20 cycles. Because this current study was focused on initiation, prevalent smokers at inception were excluded. Ethics approval was obtained from ethics committees at the Montreal Department of Public Health, McGill University, and the University of Montreal.

Study Variables

Cigarette smoking was assessed in each cycle in a 3-month recall,²⁴ which measured the number of days in each of the 3 preceding months in which participants had smoked and number of cigarettes smoked per day on average during that month. Test-retest reliability for these 2 items was good.²⁵ Initiation was considered to have occurred during the cycle in which participants smoked for the first time.

A total of 58 prognostic variables were selected on the basis of a review of cigarette smoking predictors in

youth,²² as well as the feasibility of collecting data from youth in clinical settings. Selected variables pertained to sociodemographics, smoking habits of family and friends, and lifestyle factors (Supplemental Table 3, Supplemental Information). In addition, personality traits and mental health were measured by using validated scales (Supplemental Table 3), the items of which we considered as variables.

Data Analysis

The prognostic tool was developed in 6 steps (Supplemental Fig 4).

Create Data Sets

To enable predicting the 1-year risk of initiation, the data set was divided into 4 consecutive 5-cycle waves with baselines in cycles 1, 5, 9, and 13. Predictor variable values were drawn from the baseline cycle, and the initiation indicator was based on the subsequent 4 cycles corresponding to a 1-year follow-up. Initiators in a given wave were excluded from all subsequent waves. Never-smokers could be included in up to 4 waves. The four 1-year waves were pooled. Continuous variables were standardized to ensure a common unitless scale.²⁶ Prognostic models produce better prediction in the data sets in which they are built than in independent data sets, a phenomenon known as overoptimism.²⁷ Thus, the analytical sample was randomly divided into a training sample (80% of observations) in which the models were developed, and a test sample (20% of observations) was used to estimate model performance. Both samples had the same proportion of initiators.

Imputation

We used a nonparametric, computationally efficient multiple imputation (MI) method on the basis of random forests, to impute missing values of the predictors.²⁸

Select Prognostic Variables

Prognostic variables were selected by using the bootstrap-enhanced least absolute shrinkage operator (Bolasso)²⁹ algorithm in the training data set. Bolasso improves on variable selection in least absolute shrinkage and selection operator (Lasso) by combining it with bootstrapping.²⁹ Lasso is a penalized regression in which a penalty parameter λ is selected to control the number of variables that enter a model, with large values of λ leading to sparse models.³⁰ In effect, coefficients of less influential variables are penalized to exactly 0, which is how variable selection is performed.^{30,31} When predictors are strongly correlated, Lasso is not consistent, such that a given penalty λ can lead to different sets of variables.²⁹ Bolasso relies on bootstrapping to stabilize the variable selection process. For a given λ , variable selection is performed by choosing the variables selected by Lasso in the vast majority of bootstrapped copies.²⁹ In our implementation of Bolasso, we considered 150 bootstraps and selected variables that appeared in 95% of bootstraps. We incorporated MI by estimating Lasso in each MI data set and by selecting the set of variables for which the averaged coefficient over the 15 imputation data sets was >0.001 as an absolute value. We considered model sizes ranging from 1 to 20 variables. Supplemental Figure 4 is used to summarize the variable selection process.

Estimate Coefficients

We used 10-fold cross-validation to estimate coefficients and validate each model. For each model of size 1 to 20, we divided the training data set into 10 partitions of equal size and initiation prevalence, imposing the same partition on each imputation data set. We then estimated the model on nine-tenths of the data and

averaged the regression coefficients over the 15 imputation data sets. We repeated the procedure 10 times, each time excluding a different tenth of data. Supplemental Figure 5 is used to summarize the estimation and validation process.

Model Performance

For the prognostic tool, we chose from among all models with c -statistics >0.70 , the model with 5 to 20 variables that minimized optimism and had good calibration performances in the validation data set (Supplemental Table 4).

The c -statistic measured model performance in discriminating participants who did and did not initiate smoking. We used calibration plots to assess the level of agreement between predicted initiation probabilities and the observed outcome.³² Using calibration curves in the training data set allows for internal calibration (ie, checking whether the model was missing important predictors). Using calibration curves in the test data set allowed for external validation, which assessed whether the model over- or underpredicted initiation for a given range of observed probabilities resulting in poor performance in external data.³³ Calibration plots are used to depict the smoothed estimated relationship between observed outcomes and the predicted probability of the outcome.³² Perfect calibration is indicated by a diagonal line with unit slope, and large discrepancies from the diagonal indicate segments in the range of the predictions in which the model under- or overpredicts the outcome.³²

Compute Scores

We defined thresholds to identify adolescents at risk of initiation from the probabilities estimated in the model using a utility-based approach that emphasizes sensitivity over specificity by constraining the

sensitivity to be ≥ 0.80 .³⁴ This is warranted when the intervention (ie, smoking prevention counseling) is not invasive so that false-positives are less problematic than false-negatives (ie, counseling low-risk adolescents is a less important problem than not counseling adolescents at risk).³⁵ To facilitate the model interpretation, we describe a scoring system to evaluate the initiation risk for 8 scenarios.

Data analysis was conducted in R version 3.3.2 (R Core Team, Vienna, Austria) by using the *missForest*,²⁸ *caret*,³⁶ *glmnet*,³⁷ and *rms*³³ packages.

RESULTS

Of 1294 participants, 461 participants who reported cigarette smoking at inception or joined the NDIIT study after inception were excluded. Wave 1 included 833 never smokers; 22.8% initiated smoking. Waves 2 to 4 comprised 584, 457, and 388 never smokers of whom 16.4%, 10.3%, and 16.7% initiated. Together, the 4 waves included 2266 observations contributed by 842 adolescents. The training data set comprised 1813 observations, and the test data set included 453 observations. Overall, 370 adolescents initiated smoking in both the training and test data sets, representing 16.3% of all observations. Missing value patterns are described in Supplemental Table 5. Table 1 is used to present baseline statistics.

With Supplemental Table 4, we report performance measures for models with 1 to 20 variables in the training data set. The final model was selected from among models with 5 to 20 variables because its c -statistic was the highest, and its measure of optimism was similar to that of models of similar size.

Twelve variables were associated with initiation, including age, 4

TABLE 1 Characteristics of Participants ($n = 2266$ Observations of 842 Participants; NDIT Study, 1999–2005)

Characteristic	Mean (SD) or %
Male sex, %	47.8
Age, y, mean (SD)	13.8 (1.1)
University-educated mother, %	48.0
Single-parent family, %	7.5
Parent(s) smoke, %	24.3
Sibling(s) smoke, %	10.8
Friend(s) smoke, %	38.1
Consumes alcohol, %	35.6
Participates in team sports, %	63.2
Feels the need for a cigarette, %	3.2
Hard not to smoke when others are smoking, %	9.3
Worry or stress about loneliness, %	23.8
Worry or stress about wt, %	29.1
Feels hopeless about the future, %	27.0
Worry or stress about relationship with siblings, %	26.8
Worry or stress about a health problem, %	23.1
Has something valuable to offer, %	94.0
Has a positive attitude toward oneself, %	91.4

Includes up to 4 observations per participant.

stress items (ie, worried or stressed about loneliness, weight, a health problem, or relationship with siblings), 1 depression-related item (ie, feeling hopeless about the future), 2 self-esteem items (ie, have something valuable to offer; have a positive attitude toward oneself), and 4 alcohol- or tobacco-related variables (ie, consumes alcohol, feels the need for a cigarette, finds it hard not to smoke when others are smoking, friend[s] smoke). With Fig 1, we report the coefficients for the 12 variables; the size of the bar corresponds to each coefficient proportional to its relative effect on the estimated probability of initiation. The risk decreased with age. Alcohol consumption and having friends who smoke increased the risk. Adolescents with positive self-esteem were at reduced risk. The estimated coefficients with 95% confidence intervals are shown in Supplemental Table 6.

After cross-validation in the training data set, the model had a mean c-statistic of 0.72 (SD = 0.07), a mean sensitivity of 0.80 (SD = 0.08), and a low optimism value (0.011). In the test data set, the c-statistic and sensitivity were 0.77 and 0.80, respectively, which is consistent

with good predictive ability. In the training data set, the cutoff indicating whether an adolescent was at risk of initiation was 0.11. In the test data set, the cutoff yielded a sensitivity and specificity of 0.80 and 0.55. The calibration curves being associated with the validation of the training data set (Fig 2A) suggest excellent calibration of the model with slight overprediction of probabilities beyond 0.40, which is well above the cutoff indicating whether an adolescent is at risk for initiation and therefore has no practical impact. The behavior of the locally weighted scatter-plot smoother curve in the test data set (Fig 2B) appears bumpy suggesting slight underprediction of probabilities >0.20.

The estimated regression coefficients (Fig 1) were used to construct the prognostic tool (Fig 3).³⁸ An online version of the tool can be used to automatically compute the 1-year risk of initiation (<http://www.mapageweb.umontreal.ca/sylvesma/logiciels-en.html>). Table 2 describes the application of the scoring system to 8 scenarios. Scenario 1 suggests that being drawn to smoking is not enough to place an adolescent at risk if the adolescent has high self-esteem, does not consume alcohol,

and does not have friends who smoke. However, the combination of being drawn to smoking and having low self-esteem (scenarios 3–5) or consuming alcohol and having friends who smoke (scenario 2) does place the adolescent at risk. Scenario 8 suggests that adolescents who are not drawn to cigarettes are at high risk if they have low self-esteem, consume alcohol, and have friends who smoke.

DISCUSSION

Given the burden of smoking, counseling to prevent initiation should be a top priority in pediatric practice. Growing evidence on how quickly ND symptoms can manifest after the first puff^{12,39,40} supports treating the first puff as a clinical emergency necessitating intervention to prevent long-term smoking. However, busy clinicians need to quickly and accurately identify youth who would benefit most, because counseling all patients is not feasible or necessary.

We developed a 12-item prognostic tool to identify at-risk youth, with the following 7 important attributes: (1) it identifies at-risk youth, implicitly acknowledging that the first puff is a sentinel event that can rapidly lead to ND symptoms and sustained smoking^{39,40}; (2) it capitalizes on a recent review on initiation predictors identified in high-quality longitudinal studies²²; (3) it was based on a broad socioecological understanding of risk²² and therefore includes diverse predictors; (4) it was designed for use in clinical settings; (5) it can be easily self-administered using a computer or smart phone application before a clinical visit⁴¹; (6) it is short and easily interpretable; and (7) its predictive validity was established by using cutting-edge analytic approaches.

We could not locate an independent data set in which to assess performance of the tool with

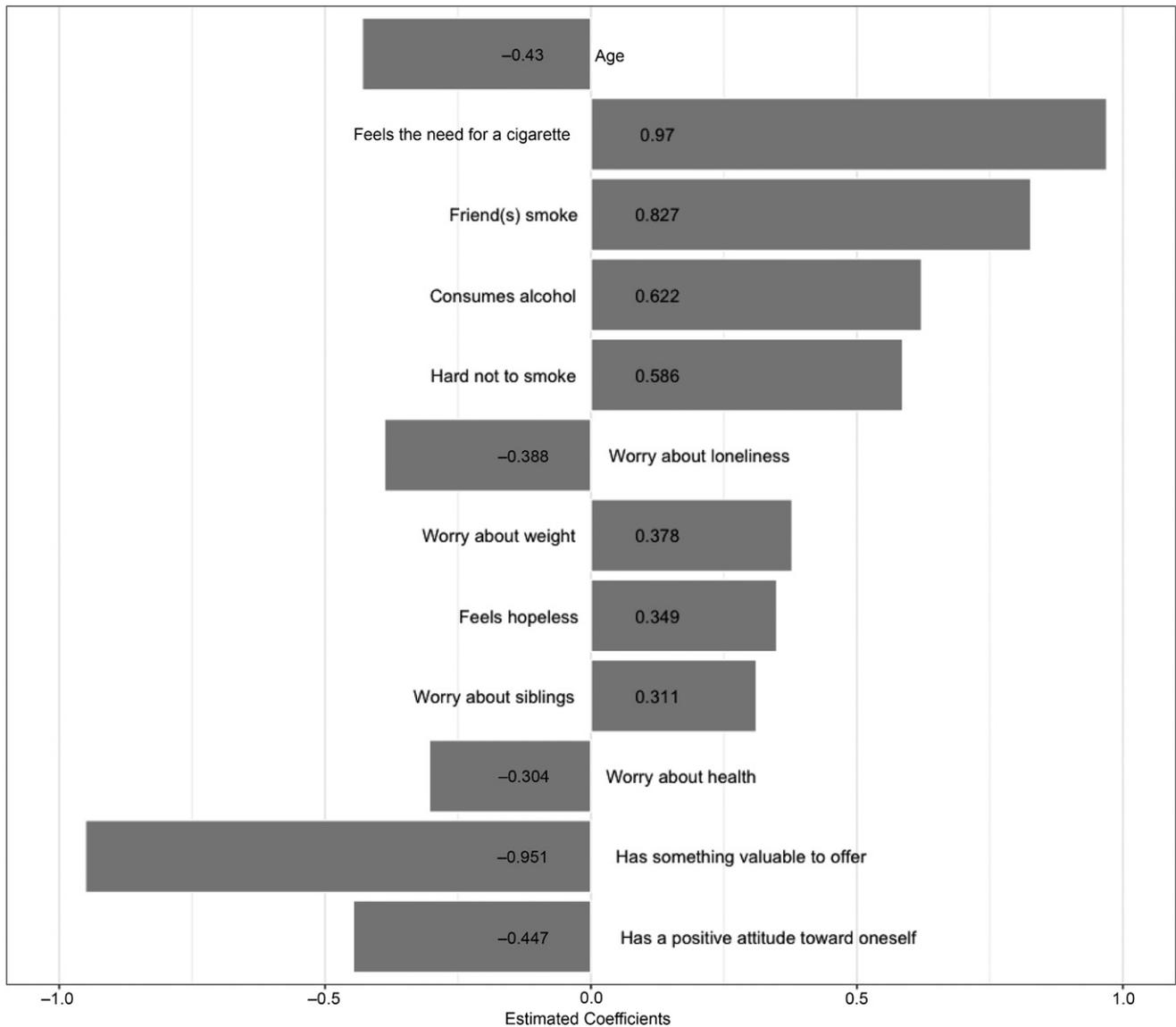


FIGURE 1
 β coefficients for variables included in the model ($n = 2266$ observations of 842 participants; NDIT study, 1999–2005).

the same variables and a 1-year follow-up for initiation after the measurement of the predictors. Therefore, we divided the data into a training and test data set. Our validation suggests that the model performed satisfactorily outside the training data set, but its predictive validity remains to be established in external populations. By presenting the tool in this forum, we hope to lay the groundwork for its use and validation in the many clinical settings in which it could be deployed. We view this tool not as a

static entity but as a way to address a gap in current clinical practice that can be iteratively improved over time. Future work should attempt to replicate the findings in an independent data set that measures the same variables in an adolescent population.

Our risk model shares similarities with that of Talluri et al⁴² who developed a 13-item model used to predict the 1-year risk of initiation using data from a prospective population-based sample of 1179 adolescents of

Mexican descent. Their items tapped individual characteristics, the social environment, and broader social-environment factors. This model has not been tested in other youth populations. Our model places less emphasis on the broader environment and taps more into adolescent characteristics and behaviors that “promote participation in social situations in which access to and availability of cigarettes is increased”.⁴³ Eight of the 12 predictors relate to stress, self-esteem, depression, and alcohol

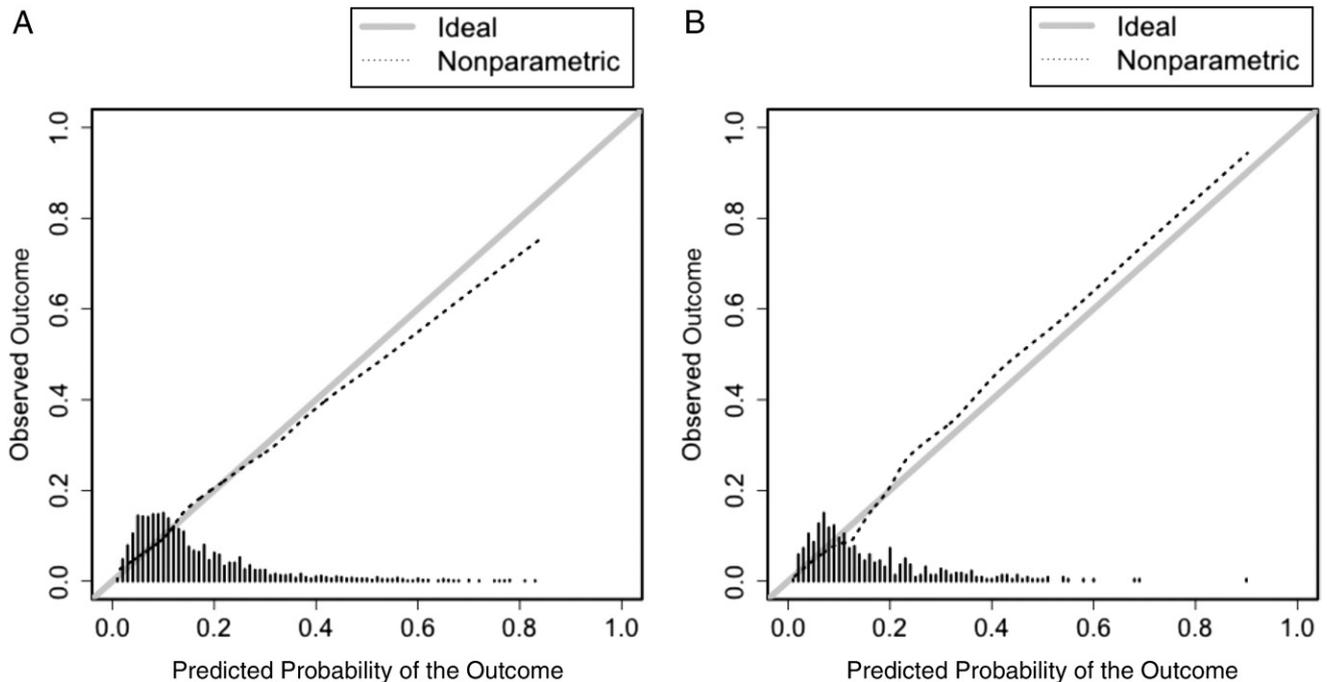


FIGURE 2

A, Nonparametric calibration plots using the training data set. B, Nonparametric calibration plots using the test data set.

use, all of which are amenable to prevention. Self-esteem training,⁴⁴ sensitization to the influence of tobacco advertising, rehearsal of refusal skills,⁴¹ watching 10 truth campaign ads,⁴⁵ and using commitment contracts to delay smoking⁴⁶ are strategies that may increase resilience to tobacco smoking.

The 3-item SSC index^{11,21} is widely used in research to identify youth at risk of becoming a smoker. Strong et al²¹ added a curiosity question to the original index¹¹ improving sensitivity (79% from 62%) and reducing specificity (36% from 50%). However, this index was not developed for, nor has it been tested in clinical settings or include diverse factors reflective of the many causes of smoking. Indeed, our scenarios suggest that adolescents who are not drawn to smoking but live in a high-risk environment are at risk of initiation. This would likely not be captured by the SSC index because it relies solely on feelings about cigarettes. It may be prudent

for researchers of future work to assess the predictive validity of these screening tools in head-on comparisons in the same setting.

Our prognostic tool may not generalize to other jurisdictions, especially if the prevalence of the items tapped differ importantly. A cutoff used to designate high or low risk depends on smoking prevalence. Our cutoff may only be meaningful if the adolescent smoking prevalence is ~16% (as it is currently in Canada,⁴⁷ which is slightly higher than in the United States).⁴⁸ If smoking prevalence differs substantially, our model can still be used to provide guidance on the relative importance of each predictor and allow clinicians to flag adolescents with several risk factors in the model. Similarly, because the legal drinking age is 18, alcohol use is relatively frequent among adolescents in Quebec.⁴⁹ The ability to discriminate high versus low risk using our cutoff may be compromised if adolescents do not drink to the same extent. However, as our scenarios illustrate, factors

including self-esteem and having friends who smoke have higher impacts on the predicted probability than alcohol use. Thus, the tool can still be used to identify at-risk adolescents, even in populations with lower alcohol consumption.

Although these data were collected almost a decade ago, our systematic review on longitudinal studies²² suggested no changes over time in the prognostic value of these predictors. In addition, we are not aware of a more recent data set with as comprehensive a set of measured predictors of initiation,²² which is required to meet the latest recommendations for constructing a prognostic tool with acceptable performance.^{50,51} Increased understanding of youth at risk could impel the development of therapeutic toolkits to prevent or delay initiation. For example, older age had a strong protective effect (the longer the delay in initiation, the lower the probability of initiation). A no-smoking contract for the next year might hold promise, as would strategies used to navigate

CIRCLE POINTS CORRESPONDING TO AGE	
AGE, years (round up)	POINTS
12	100
13	80
14	60
15	40
16	20
17	0
IF THE ANSWER IS “YES” TO THE QUESTIONS BELOW, CIRCLE CORRESPONDING POINTS	
TOBACCO & ALCOHOL	
Do any of the people whom you usually hang out with smoke cigarettes?	38
When you see other kids your age smoking cigarettes, is it difficult for you not to smoke?	27
Do you sometimes feel like you really need a cigarette?	45
Did you drink any alcohol (beer, wine or hard liquor) in the past 3 months?	29
STRESS & DEPRESSION	
In the past 3 months, have you.....	
been worried or stressed about loneliness?	-18
been worried or stressed about your weight?	18
been worried or stressed about your relationship with your brother(s) or sister(s)?	14
been worried or stressed about a health problem?	-14
felt hopeless about the future?	16
SELF-ESTEEM	
Do you.....	
think of yourself as someone who has something valuable to offer, at least as much as other people do?	-14
have a positive attitude towards yourself?	-44
ADD UP POINTS CIRCLED	
SUBTRACT	
FINAL SCORE	
- 118	

Score Converter

FINAL SCORE	1-year risk (%)
<-40	<5%
-40 to 0	5 to 11%
1 to 40	11 to 23%
41 to 80	23 to 41%
81 to 120	41 to 62%
121 to 160	62% to 78%
>160	>78%

FIGURE 3

Prognostic tool and score converter to assess 1-year risk of cigarette smoking initiation in adolescents. Negative scores correspond to low risk if using the suggested cutoff of 11% to identify adolescents at high risk of smoking in our population.

or avoid situations when cigarettes are present.

Electronic cigarette use was not measured and could not be incorporated in the prognostic tool, although evidence suggests that it is

associated with an increased risk of cigarette smoking initiation among adolescents.⁵² Physicians choosing to use the prognostic tool to identify at-risk adolescents should, as part of a comprehensive clinical assessment, also inquire about

electronic cigarettes and other forms of combustible and noncombustible tobacco. Further limitations included that subitems in psychological scales were likely correlated, which can adversely affect the performance of conventional

TABLE 2 Assessment of the 1-Year Risk of Cigarette Smoking Initiation in 8 Scenarios (NDIT Study)

	Scenarios							
	Drawn to Cigarette Smoking ^a				Not Drawn to Cigarette Smoking ^b			
	1	2	3	4	5	6	7	8
	High Self-Esteem	High Self-Esteem, Friends Who Smoke and Consume Alcohol	Low Self-Esteem	Low Self-Esteem and Friends Who Smoke	Low Self-Esteem and Consume Alcohol	Low Self-Esteem and Friends Who Smoke	Low Self-Esteem and Consume Alcohol	Low Self-Esteem, Friends Who Smoke and Consume Alcohol
Age, y	14	14	14	14	14	14	14	14
Friend(s) smoke	No	Yes	No	Yes	No	Yes	No	Yes
Consumes alcohol	No	Yes	No	No	Yes	No	Yes	Yes
Hard not to smoke when others are smoking	Yes	Yes	Yes	Yes	Yes	No	No	No
Feels the need for a cigarette	Yes	Yes	Yes	Yes	Yes	No	No	No
Worry or stress about loneliness	No	No	No	No	No	No	No	No
Worry or stress about wt	No	No	No	No	No	No	No	No
Feels hopeless about the future	No	No	No	No	No	No	No	No
Worry or stress about relationship with siblings	No	No	No	No	No	No	No	No
Worry or stress about a health problem	No	No	No	No	No	No	No	No
Has something valuable to offer	Yes	Yes	No	No	No	No	No	No
Has a positive attitude toward oneself	Yes	Yes	No	No	No	No	No	No
1-y predicted probability of initiating cigarette smoking (SE)	0.04 (0.02)	0.15 (0.05)	0.14 (0.08)	0.27 (0.13)	0.23 (0.11)	0.07 (0.05)	0.06 (0.03)	0.13 (0.08)
Score (increasing positive score implies higher risk)	-51	16	14	52	43	-20	-29	9
Assessment of risk of cigarette smoking initiation	Low	High	High	High	High	Low	Low	High

^a Drawn to cigarette smoking is defined by the following positive responses to 2 items: “hard not to smoke when others are smoking” and “feels the need for a cigarette.”

^b Not drawn to cigarette smoking is defined by negative responses to both items.

variable selection techniques such as stepwise regression.³³ We used the Bolasso algorithm,²⁹ which combines Lasso and bootstrapping and addresses correlation between predictors. Because the exact time of initiation was not measured, we used pooled logistic regression rather than survival analysis, although both methods lead to similar estimates.⁵³ It is unclear whether the correlation between intraindividual observations affected the performance of Bolasso, which assumes that observations are independent. However, ignoring the correlation between repeated measures usually affects the estimation of variances with a

negligible impact on regression coefficients.⁵⁴

CONCLUSIONS

We developed a 12-item prognostic tool that can be used to identify adolescents likely to initiate smoking in the next year. If the predictive ability is replicated in other settings, this tool can be used to help clinicians select who should be counseled and, because several items in the tool are amenable to prevention, how they should be counseled. The sensitivity of the tool combined with the potentially

lethal consequences of smoking initiation underscore an urgent need for such tools in pediatric settings.

ABBREVIATIONS

Bolasso: bootstrap-enhanced least absolute shrinkage operator
 Lasso: least absolute shrinkage and selection operator
 MI: multiple imputation
 ND: nicotine dependence
 NDIT: Nicotine Dependence in Teens
 SSC: susceptibility to smoking cigarettes index

Address correspondence to Marie-Pierre Sylvestre, PhD, Centre de Recherche du Centre Hospitalier de l'Université de Montréal, S03.458, 850 Rue St Denis, Montréal, Québec, Canada H2X 0A9. E-mail: marie-pierre.sylvestre@umontreal.ca

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

Copyright © 2018 by the American Academy of Pediatrics

FINANCIAL DISCLOSURE: The authors have indicated they have no financial relationships relevant to this article to disclose.

FUNDING: Supported by the Canadian Cancer Society (grant 010271, 017435). Dr Sylvestre is supported by a Chercheur–Boursier career award from the Fonds de Recherche du Québec–Santé. Dr O’Loughlin holds a Canada Research Chair in the Early Determinants of Adult Chronic Disease. The funders were not involved in the design or conduct of the study; collection, management, analysis, or interpretation of the data; or preparation, review, or approval of the manuscript.

POTENTIAL CONFLICT OF INTEREST: The authors have indicated they have no conflicts of interest to disclose.

COMPANION PAPER: A companion to this article can be found online at www.pediatrics.org/cdi/doi/10.1542/peds.2018-2298.

REFERENCES

1. Reidpath DD, Davey TM, Kadirvelu A, Soyiri IN, Allotey P. Does one cigarette make an adolescent smoker, and is it influenced by age and age of smoking initiation? Evidence of association from the U.S. Youth Risk Behavior Surveillance System (2011). *Prev Med*. 2014;59:37–41
2. Reidpath DD, Ling ML, Wellington E, Al-Sadat N, Yasin S. The relationship between age of smoking initiation and current smoking: an analysis of school surveys in three European countries. *Nicotine Tob Res*. 2013;15(3):729–733
3. Hwang JH, Park SW. Age at smoking initiation and subsequent smoking among Korean adolescent smokers. *J Prev Med Public Health*. 2014;47(5):266–272
4. Everett SA, Warren CW, Sharp D, Kann L, Husten CG, Crossett LS. Initiation of cigarette smoking and subsequent smoking behavior among U.S. high school students. *Prev Med*. 1999;29(5):327–333
5. Wilkinson AV, Schabath MB, Prokhorov AV, Spitz MR. Age-related differences in factors associated with smoking initiation. *Cancer Causes Control*. 2007;18(6):635–644
6. Arrazola RA, Neff LJ, Kennedy SM, Holder-Hayes E, Jones CD; Centers for Disease Control and Prevention (CDC). Tobacco use among middle and high school students—United States, 2013 [published correction appears in *MMWR Morb Mortal Wkly Rep*. 2015;64(33):924]. *MMWR Morb Mortal Wkly Rep*. 2014;63(45):1021–1026
7. Jamal A, Gentzke A, Hu SS, et al. Tobacco use among middle and high school students - United States, 2011–2016. *MMWR Morb Mortal Wkly Rep*. 2017;66(23):597–603
8. Reid JL, Hammond D, Rynard VL, Burkhalter R. *Tobacco Use in Canada: Patterns and Trends, 2015 Edition*. Waterloo, Canada: Propel Centre for Population Health Impact, University of Waterloo; 2015. Available at: https://uwaterloo.ca/tobacco-use-canada/sites/ca.tobacco-use-canada/files/uploads/files/tobaccouseincanada_2015_accessible_final-s.pdf. Accessed September 29, 2016
9. Reid JL, Hammond D, Rynard VL, Madill CL, Burkhalter R. *Tobacco Use in Canada: Patterns and Trends*. Waterloo, Canada: Propel Centre for Population Health Impact, University of Waterloo; 2017
10. Veeranki SP, Mamudu HM, Anderson JL, Zheng S. Worldwide never-smoking youth susceptibility to smoking. *J Adolesc Health*. 2014;54(2):144–150
11. 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(5):355–361
12. Gervais A, O’Loughlin J, Meshfedjian G, Bancej C, Tremblay M. Milestones in the natural course of onset of cigarette use among adolescents. *CMAJ*. 2006;175(3):255–261
13. Pbert L, Farber H, Horn K; Julius B. Richmond Center of Excellence Tobacco Consortium, et al. State-of-the-art office-based interventions to eliminate youth tobacco use: The past decade. *Pediatrics*. 2015;135(4):734–747
14. Farber HJ, Walley SC, Groner JA, Nelson KE; Section on Tobacco Control. Clinical practice policy to protect children from tobacco, nicotine, and tobacco smoke. *Pediatrics*. 2015;136(5):1008–1017
15. Harvey J, Chadi N; Canadian Paediatric Society, Adolescent Health Committee. Preventing smoking in children and adolescents: recommendations for practice and policy. *Paediatr Child Health*. 2016;21(4):209–221
16. Schauer GL, Agaku IT, King BA, Malarcher AM. Health care provider advice for adolescent tobacco use: results from the 2011 National Youth Tobacco Survey. *Pediatrics*. 2014;134(3):446–455
17. O’Loughlin J, Makni H, Tremblay M, et al. Smoking cessation counseling practices of general practitioners in Montreal. *Prev Med*. 2001;33(6):627–638
18. Makni H, O’Loughlin JL, Tremblay M, et al. Smoking prevention counseling practices of Montreal general practitioners. *Arch Pediatr Adolesc Med*. 2002;156(12):1263–1267
19. Shelley D, Cantrell J, Faulkner D, Haviland L, Heaton C, Messeri P. Physician and dentist tobacco use counseling and adolescent smoking behavior: results from the 2000 National Youth Tobacco Survey. *Pediatrics*. 2005;115(3):719–725
20. Ozer EM, Adams SH, Lustig JL, et al. Increasing the screening and counseling of adolescents for risky health behaviors: a primary care intervention. *Pediatrics*. 2005;115(4):960–968
21. Strong DR, Hartman SJ, Nodora J, et al. Predictive validity of the expanded susceptibility to smoke index. *Nicotine Tob Res*. 2015;17(7):862–869
22. Wellman RJ, Dugas EN, Dutczak H, et al. Predictors of the onset of cigarette smoking: a systematic review of longitudinal population-based studies in youth. *Am J Prev Med*. 2016;51(5):767–778
23. O’Loughlin J, Dugas EN, Brunet J, et al. Cohort profile: the Nicotine Dependence in Teens (NDIT) study. *Int J Epidemiol*. 2015;44(5):1537–1546
24. Centers for Disease Control and Prevention (CDC). Selected cigarette smoking initiation and quitting

- behaviors among high school students—United States, 1997. *MMWR Morb Mortal Wkly Rep.* 1998;47(19):386–389
25. Eppel A, O’Loughlin J, Paradis G, Platt R. Reliability of self-reports of cigarette use in novice smokers. *Addict Behav.* 2006;31(9):1700–1704
 26. Gelman A. Scaling regression inputs by dividing by two standard deviations. *Stat Med.* 2008;27(15):2865–2873
 27. Steyerberg E. Overfitting and optimism in prediction models. In: *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating.* New York, NY: Springer-Verlag; 2009:83–100
 28. Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics.* 2012;28(1):112–118
 29. Bach FR. Bolasso: model consistent Lasso estimation through the bootstrap. In: *Proceedings of the 25th International Conference on Machine Learning.* New York, NY: Association for Computing Machinery; 2008:33–40
 30. Rasmussen MA, Bro R. A tutorial on the Lasso approach to sparse modeling. *Chemom Intell Lab Syst.* 2012;119:21–31
 31. Tibshirani R. Regression shrinkage and selection via the Lasso. *J R Stat Soc Series B Methodol.* 1996;58(1):267–288
 32. Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Stat Med.* 2014;33(3):517–535
 33. Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis.* 1st ed. New York, NY: Springer-Verlag; 2001
 34. Steyerberg EW, Van Calster B, Pencina MJ. Performance measures for prediction models and markers: evaluation of predictions and classifications [in Spanish]. *Rev Esp Cardiol.* 2011;64(9):788–794
 35. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology.* Philadelphia, PA: Lippincott Williams & Wilkins; 2008
 36. Kairalla JA, Coffey CS, Muller KE. GLUMIP 2.0: SAS/IML software for planning internal pilots. *J Stat Softw.* 2008;28(7):1
 37. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* 2010;33(1):1–22
 38. Yang D. Build prognostic nomograms for risk assessment using SAS. In: *Proceedings of SAS Global Forum, Paper 264-2013; April 28–May 1, 2013; San Francisco, CA*
 39. DiFranza JR, Savageau JA, Fletcher K, et al. Symptoms of tobacco dependence after brief intermittent use: the Development and Assessment of Nicotine Dependence in Youth-2 study. *Arch Pediatr Adolesc Med.* 2007;161(7):704–710
 40. DiFranza JR, Rigotti NA, McNeill AD, et al. Initial symptoms of nicotine dependence in adolescents. *Tob Control.* 2000;9(3):313–319
 41. Pbert L, Farber H, Horn K, et al; American Academy of Pediatrics, Julius B. Richmond Center of Excellence Tobacco Consortium. State-of-the-art office-based interventions to eliminate youth tobacco use: the past decade. *Pediatrics.* 2015;135(4):734–747
 42. Talluri R, Wilkinson AV, Spitz MR, Shete S. A risk prediction model for smoking experimentation in Mexican American youth. *Cancer Epidemiol Biomarkers Prev.* 2014;23(10):2165–2174
 43. O’Loughlin J, Karp I, Koulis T, Paradis G, DiFranza J. Determinants of first puff and daily cigarette smoking in adolescents. *Am J Epidemiol.* 2009;170(5):585–597
 44. Onrust SA, Otten R, Lammers J, Smit F. School-based programmes to reduce and prevent substance use in different age groups: what works for whom? Systematic review and meta-regression analysis. *Clin Psychol Rev.* 2016;44:45–59
 45. Farrelly MC, Nonnemaker J, Davis KC, Hussin A. The influence of the national truth campaign on smoking initiation. *Am J Prev Med.* 2009;36(5):379–384
 46. Andersen A, Krølner R, Bast LS, Thygesen LC, Due P. Effects of the X:IT smoking intervention: a school-based cluster randomized trial. *Int J Epidemiol.* 2015;44(6):1900–1908
 47. Burkhalter R, Cumming T, Rynard V, Manske S. *2012/2013 Youth Smoking Survey Microdata User Guide.* 5th ed. Waterloo, Canada: Propel Centre for Population Health Impact, University of Waterloo; 2013:1–47
 48. US Department of Health Human Services. *Preventing Tobacco Use Among Youth and Young Adults: A Report of the Surgeon General.* Atlanta, GA: US 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
 49. Nanhou V, Ducharme A, Eid H. *Initiation to Tobacco, Alcohol and Drugs: An Overview of the Situation During the Transition From Grade 6 to First Year of High School [in French].* Institut de la Statistique du Québec; 2013
 50. Riley RD, Hayden JA, Steyerberg EW, et al; PROGRESS Group. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med.* 2013;10(2):e1001380
 51. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology.* 2010;21(1):128–138
 52. Leventhal AM, Strong DR, Kirkpatrick MG, et al. Association of electronic cigarette use with initiation of combustible tobacco product smoking in early adolescence. *JAMA.* 2015;314(7):700–707
 53. D’Agostino RB, Lee ML, Belanger AJ, Cupples LA, Anderson K, Kannel WB. Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study. *Stat Med.* 1990;9(12):1501–1515
 54. Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis.* 2nd ed. Hoboken, NJ: John Wiley & Sons; 2012

A Tool to Identify Adolescents at Risk of Cigarette Smoking Initiation
Marie-Pierre Sylvestre, Nancy Hanusaik, David Berger, Erika Dugas, Lori Pbert,
Jonathan Winickoff and Jennifer L. O'Loughlin

Pediatrics 2018;142;

DOI: 10.1542/peds.2017-3701 originally published online October 1, 2018;

Updated Information & Services

including high resolution figures, can be found at:
<http://pediatrics.aappublications.org/content/142/5/e20173701>

References

This article cites 43 articles, 9 of which you can access for free at:
<http://pediatrics.aappublications.org/content/142/5/e20173701#BIBL>

Subspecialty Collections

This article, along with others on similar topics, appears in the following collection(s):
Substance Use
http://www.aappublications.org/cgi/collection/substance_abuse_sub
Smoking
http://www.aappublications.org/cgi/collection/smoking_sub
Preventive Medicine
http://www.aappublications.org/cgi/collection/preventative_medicine_sub

Permissions & Licensing

Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at:
<http://www.aappublications.org/site/misc/Permissions.xhtml>

Reprints

Information about ordering reprints can be found online:
<http://www.aappublications.org/site/misc/reprints.xhtml>

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN™



PEDIATRICS®

OFFICIAL JOURNAL OF THE AMERICAN ACADEMY OF PEDIATRICS

A Tool to Identify Adolescents at Risk of Cigarette Smoking Initiation

Marie-Pierre Sylvestre, Nancy Hanusaik, David Berger, Erika Dugas, Lori Pbert,
Jonathan Winickoff and Jennifer L. O'Loughlin

Pediatrics 2018;142;

DOI: 10.1542/peds.2017-3701 originally published online October 1, 2018;

The online version of this article, along with updated information and services, is
located on the World Wide Web at:

<http://pediatrics.aappublications.org/content/142/5/e20173701>

Data Supplement at:

<http://pediatrics.aappublications.org/content/suppl/2018/10/24/peds.2017-3701.DCSupplemental>

Pediatrics is the official journal of the American Academy of Pediatrics. A monthly publication, it has been published continuously since 1948. Pediatrics is owned, published, and trademarked by the American Academy of Pediatrics, 141 Northwest Point Boulevard, Elk Grove Village, Illinois, 60007. Copyright © 2018 by the American Academy of Pediatrics. All rights reserved. Print ISSN: 1073-0397.

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN™

