Nicotine & Tobacco Research, Volume 11, Number 4 (April 2009) 455–462

Original Investigation

Use of propensity score matching in evaluating smokeless tobacco as a gateway to smoking

David S. Timberlake, Jimi Huh, & Cynthia M. Lakon

Abstract

Introduction: The contentious debate over the promotion of Swedish snus, a form of moist snuff, as an alternative to cigarettes has often focused on the product's potential as a gateway to smoking. Findings from prospective studies in the United States have suggested that smokeless tobacco (SLT) is a significant predictor of smoking onset, before and after adjustment for baseline covariates. Use of conventional regression methods in these studies may have resulted in biased parameter estimates, arising from imbalanced covariate distributions in the users and nonusers of SLT. An alternative approach, which has been used widely in the econometric literature, matches exposure or treatment levels on the basis of the propensity score distribution.

Methods: Using this approach, we matched current SLT users from the National Longitudinal Study of Adolescent Health with nonusers (496 pairs) and followed them from adolescence into young adulthood for determination of smoking status.

Results: Prior to matching, the unadjusted risk of becoming a daily smoker was significantly greater for the SLT users compared with nonusers ($n = 10{,}820$; range of relative risk = 1.3–2.0, $p < .001$). However, after pairing individuals on propensity score, we found no evidence for an increased risk of smoking among the SLT users.

Discussion: Baseline differences in the risk factors for smoking likely account for the association between the two tobacco products.

Introduction

Despite stable rates of tobacco use over the past two decades (Rodu, Stegmayr, Nasic, & Asplund, 2002), the incidence of smoking-related morbidity among Swedish males has declined (Foulds et al., 2003). This trend has been attributed to the transition from use of cigarettes to snus, a form of smokeless tobacco (SLT) that is low in tobacco-specific nitrosamines (Daniel Roth, Roth, & Liu, 2005). Estimates suggest that 40% of the annual one-half million smoking-related deaths in European males would be prevented if smoking rates throughout the continent were comparable with that of Swedish males (Rodu & Cole, 2004). Some groups, such as Action on Smoking and Health, petitioned the European Union to reconsider its ban on snus (Fagerström & Schildt, 2003). Although Sweden's experience is promising, debate continues over whether the promotion of snus in other countries can be implemented effectively in a campaign for tobacco harm reduction (Gartner, Hall, Chapman, & Freeman, 2007; Martinet, Bohadana, & Fagerström, 2006; Savitz, Meyer, Tanzer, Mirvish, & Lewin, 2006).

Another concern about tobacco harm reduction relates to the potential of SLT as a gateway to cigarette smoking. Researchers have argued convincingly that even if SLT use can lead to cigarette smoking, it would affect only a minority of SLT users (Kozlowski, O'Connor, Edwards, & Flaherty, 2003; O'Connor, Kozlowski, Flaherty, & Edwards, 2005). Using cross-sectional data from the 1987 National Health Interview Survey and the 2000 National Household Survey on Drug Abuse, Kozlowski et al. (2003) and O'Connor et al. (2005) reported that most users of SLT in the United States either never initiated smoking or initiated smoking prior to their use of SLT. In the 1987 and 2000 national surveys, 24% and 29% of current SLT users, respectively, had initiated their use prior to smoking, thus qualifying as potential gateway users. The increase in these users over time was not statistically significant (O'Connor et al., 2005), strengthening the authors' argument that potential gateway effects should not be at the forefront of discussions about harm reduction. Yet, a statistically significant increase could occur if nonsmoking adolescents were inadvertently influenced by SLT marketing, leading to a larger pool of those who first use SLT and then experiment with cigarettes. Thus, before SLT is promoted in a campaign for

David S. Timberlake, Ph.D., *Department of Population Health and Disease Prevention, Program in Public Health, University of California, Irvine, CA*
Jimi Huh, M.A., *Department of Psychology and Social Behavior, University of California, Irvine, CA*
Cynthia M. Lakon, Ph.D., *Department of Population Health and Disease Prevention, Program in Public Health, University of California, Irvine, CA*

Corresponding Author:
David S. Timberlake, Ph.D., *Program in Public Health, University of California, Irvine, 100 Theory, Suite 100, Irvine, CA 92697-7555, USA. Telephone: 949-824-3552; Fax: 949-824-1343; E-mail: dtimberl@uci.edu*

doi: 10.1093/ntr/ntp008
Advance Access publication on March 23, 2009
Received May 7, 2008; accepted September 30, 2008

tobacco harm reduction, epidemiologists must confirm that it is not causally associated with the uptake of cigarette smoking.

Evidence suggestive of a gateway effect originates from three U.S.-based cohorts: Air Force recruits (Haddock et al., 2001), adolescents from the Teenage Attitudes and Practices Survey (TAPS; Tomar, 2003), and adolescents from the Project Sixteen community intervention (Severson, Forrester, & Biglan, 2007). Among the TAPS adolescents who were nonsmokers at the baseline survey, regular users of SLT compared with never-users were 3.45 times more likely to be current smokers at the follow-up survey (Tomar, 2003). An additional analysis of the TAPS data suggested that baseline differences (e.g., poor school performance, fighting) between SLT users and nonusers accounted for much of the variance in subsequent smoking status (O'Connor, Flaherty, Quinio Edwards, & Kozlowski, 2003). This point is illustrated further by a survey of correlates of SLT use in military trainees (Lando, Haddock, Klesges, Talcott, & Jensen, 1999). SLT users in this study exhibited greater binge drinking, rebelliousness, and risk-taking behavior compared with individuals who never used tobacco. Having controlled for similar covariates, Severson et al. (2007) reported that among adolescent boys, SLT users were 2.5 times more likely than nonusers to initiate weekly smoking.

A concern about the aforementioned studies is the use of regression methods in accounting for the imbalanced covariate distributions in SLT users and nonusers. Conventional methods do not provide regression diagnostics for examining the joint distributions of explanatory variables (Rubin, 1997). Thus, for example, it cannot be determined whether a regression analysis fully accounts for the higher proportion of risk takers among users of SLT. One means of correcting for baseline differences is the use of the propensity score, defined as the probability of an event conditional on a series of predictor variables (Little & Rubin, 2000; Newgard, Hedges, Arthur, & Mullins, 2004). Differences between groups can be controlled for by matching individuals on the basis of similar propensity scores. Rubin (1997) has described this technique as a means of "reducing the entire collection of background characteristics to a single composite characteristic" (p. 757). The matching of SLT users and nonusers on this single composite characteristic facilitates an examination of the overlap of confounding factors. Furthermore, this approach does not rely on linearity, or log linearity, between the outcome (e.g., smoking) and covariate (e.g., age) within each exposure group (SLT use and nonuse). For regression, this is particularly problematic when the groups differ greatly in their covariate distributions, potentially leading to biased parameter estimates. Although propensity score matching has existed for some time, its widespread application to large epidemiological studies has occurred only recently. Examples of its application in medicine, for reducing bias associated with nonrandom assignment, include the allocation of medical treatments (e.g., diabetic patients receiving stents; Ko et al., 2008) and services (e.g., transfer of trauma patients; Newgard et al., 2004).

The application of propensity score matching in the present study is extended to adolescents who chose a particular behavior (SLT use and nonuse), which is likely influenced by predisposing characteristics (e.g., impulsivity). These characteristics are highly predictive of other forms of substance use, including regular cigarette smoking, by young adulthood. The present study is intended to complement existing longitudinal analyses of the effects of SLT use on smoking. It is hypothesized that matched pairs of adolescent SLT users and nonusers from the National Longitudinal Study of Adolescent Health (Add Health) do not differ in their likelihood of becoming daily cigarette smokers by either later adolescence (wave II) or young adulthood (wave III).

Methods

Selection of participants

Description of Add Health and its sampling frame have been reported previously (Alexander, Piazza, Mekos, & Valente, 2001). An in-school questionnaire was administered between 1994 and 1995 to a nationally representative sample of participants ($n = 90{,}118$) who attended either middle or high school. Samples of students were then selected to participate in lengthier in-home surveys, which included the wave I survey in 1995 ($n = 18{,}924$), the wave II survey in 1996 ($n = 13{,}570$), and the wave III survey ($n = 14{,}322$), which was administered between 2001 and 2002.

The sample at the wave I survey was chosen with consideration of two factors, statistical power and comparability with prior studies (O'Connor et al., 2003; Tomar, 2003). One of the critiques of the original analysis of the TAPS data was inclusion of experimental smokers in the baseline sample. These individuals were more likely to be SLT users, which confounded the association between SLT use and subsequent smoking status. In the present study, individuals were similarly defined as nonsmokers if they took at least a puff from a cigarette in their lifetime but had not smoked daily by the wave I survey. These individuals were initially excluded from the baseline sample as a conservative measure, yielding 5,909 never-smokers who were surveyed again at wave II. Only 189 of these individuals were current users of SLT at wave I, requiring that the nonsmokers be added to the sample to increase the number of SLT users ($n = 498$). Thus, for the primary analyses, which included the never- and nonsmokers at wave I ($n = 10{,}820$), smoking status was accounted for in development of the propensity score. Among the adolescent females in this sample, 1.7% reported past-month use of SLT. In contrast to the other U.S.-based cohorts (Severson et al., 2007; Tomar, 2003), females comprised a moderate proportion of total past-month users of SLT, ranging from 0.13 to 0.18 at all three waves. Thus, they were included in all analyses.

Measures

The primary outcome for participants at waves II and III was ever-daily smoking, defined as having smoked at least one cigarette on 30 or more consecutive days since the baseline survey. A secondary outcome, current smoking, was defined as having smoked a cigarette in the past-month, in addition to having smoked daily for at least 1 month since the wave I survey. Current smoking in other gateway studies, such as TAPS (O'Connor et al., 2003; Tomar, 2003), was defined as smoking at least one cigarette in the past-month and at least 100 cigarettes in a lifetime. The latter definition was not used in the Add Health survey. The primary predictor in the wave I survey, past-month use of SLT, was recoded into a binary variable from a question assessing the number of days of SLT use (0–30 days, chewing tobacco or snuff).

The covariates used in deriving the propensity scores were selected on the basis of their measures of risk for smoking (Kandel, Kiros, Schaffran, & Hu, 2004) and their prior use in

gateway studies of SLT (Haddock et al., 2001; O'Connor et al., 2003; Severson et al., 2007). All covariates were limited to measurements at the baseline survey and were categorized into one of three groups: (a) demographics, which included age, gender, race, and U.S. region of residency; (b) smoking-related exposures, which included number of best friends who smoke cigarettes (0–3), number of smoking parents (0–2), availability of cigarettes in the home (yes or no), smoking prevention taught in school (yes or no), and lifetime smoking (≥1 entire cigarette, yes or no); and (c) behavioral risk, which included past-year binge drinking (≥5 drinks in a row, 0 [never] to 6 [every day], past-month use of cannabis (yes or no), having repeated a grade in school (yes or no), past-year motorcycle riding (0 [never] to 4 [almost every day]), and use of a seat belt when riding in or driving a car (0 [never] to 4 [always]). In addition, two scales from the wave I survey included a measure of delinquency, the sum of 14 delinquent behaviors in the past year (Kandel et al., 2004), and a measure of depressive symptoms (CES-D), the sum of 18 symptoms (e.g., too tired to do things) experienced in the prior week (Radloff, 1977).

Data analyses

An analysis using propensity scores entails three steps: (a) estimation of the score using logistic regression, (b) incorporation of the score as a covariate or a stratified or matching variable, and (c) analysis of the score with respect to the outcome (i.e., daily smoking; Newgard et al., 2004). Treating the propensity score as a matching variable has particular advantages over conventional regression, notably the greater emphasis on study design. For example, the technique enables an investigator to assess whether covariates are comparable across groups. In a paired case–control study, participants are matched on a series of confounding variables prior to analyses. Such matching, however, on individual predictors is often impractical because of the few individuals who are concordant for a large number of potential confounders. An alternative is use of the "single composite characteristic" (propensity score), whereby individuals may not be matched perfectly on all confounders but are matched closely enough to reduce bias associated with their nonrandom assignment to a given group.

Estimation of propensity score. In the first step of the analysis, the propensity score was estimated as the conditional probability (π) of using SLT in the past-month, where $\pi = e^{\beta+\beta 1X1+\cdots}/1 + e^{\beta+\beta 1X1+\cdots}$ in the logit model, $g(x) = \ln[\pi/1-\pi] = \beta_0 + \beta_0 X_1 + \ldots + \beta_q X_q$. In this model, $X_1, \ldots, X_q$ represent the covariates described previously. A number of factors were considered in the development of the logit model (Weitzen, Lapane, Toledano, Hume, & Mor, 2004), including collinearity and the number of events per variable (>10; Peduzzi, Concato, Kemper, Holford, & Feinstein, 1996).

Matching on propensity score. In the second step of the analysis, baseline users and nonusers of SLT were matched individually on propensity score using the psmatch2 procedure in Stata Statistical Software version 10 (Leuven & Sianesi, 2003). One-to-one matching without replacement was chosen over stratified matching due to the large ratio (~20:1) between the number of nonusers and the number of users of SLT in the sample. The matching algorithm, nearest neighbor matching within caliper (Guo, Barth, & Gibbons, 2006), first involved the random ordering of the SLT users and nonusers. Upon selecting the first SLT user, we chose a nonuser with the nearest propensity score from within the same caliper, a region predetermined as one-quarter of the average standard deviation of the propensity score (Rosenbaum & Rubin, 1984). This matched pair was then removed from the sample, followed by selection of the next SLT user. A series of bivariate analyses were then conducted in the paired and nonpaired samples in assessing the effectiveness of the matching procedure. This included calculation of the percent reduction in bias for each covariate, $100(1 - b_m/b_i)$ (Rosenbaum & Rubin, 1985), where b_i and b_m represent the difference in covariate means of the two groups, before (i) and after (m) matching on propensity score.

Analyses of smoking outcomes. For paired data, the conditional logit and McNemar's χ^2 test were used in evaluating differences in the number of discordant pairs for the two dichotomous outcomes, current and ever-daily smoking. The odds ratio (*OR*) in these analyses was calculated as the ratio of the number of matched pairs that were discordant for smoking outcomes in subsequent surveys. The ratio of discordant pairs corresponds to the equation: *OR* = (SLT [smoker], non-SLT [nonsmoker])/(SLT [nonsmoker], non-SLT [smoker]). For all paired analyses, weights and design effects in the Add Health sample were not accounted for, given the difficulty of matching the SLT users and nonusers within each of the 132 school clusters or primary sampling units. For comparative purposes, the sampling considerations were implemented in a set of nonpaired analyses, in which two baseline variables (SLT use and propensity score) were modeled as predictors of the outcomes for smoking.

Power analysis. In comparing binomial proportions from paired samples, we estimated statistical power ($1 - \beta$) as a function of the number of matched pairs, the proportion of discordant pairs among all matched pairs (projected range = .33–.50), and the proportions of the two types of discordant pairs (effect size; Rosner, 1990). All estimates of power exceeded .85 in detecting significant *OR*s of 3.4 and 2.5 reported by Tomar (2003) and Severson et al. (2007), respectively. Power also was estimated on the basis of an effect (*OR* = 1.5) intermediate to the adjusted *OR*s (*OR*s = 1.41, 1.68), which were reported but not found to be statistically significant by O'Connor et al. (2003). Sufficient power existed for detecting statistical significance of a conditional *OR* of 1.5 in our sample of the never- and nonsmokers (n = 496 matched pairs, range of $1 - \beta$ = .73–.89) but was not adequate in the sample of never-smokers (n = 189 matched pairs, range of $1 - \beta$ = .35–.49).

Results

Users and nonusers of SLT significantly differed in their unadjusted risk of becoming smokers by the subsequent surveys. As indicated in Table 1, past-month SLT users had approximately a twofold greater risk of having smoked daily for at least 1 month by the wave II survey compared with nonusers. The incidence ratios, or relative risks, for the two outcomes were comparable by wave for both samples of Add Health participants, the never-smokers and the never- and nonsmokers. Yet, the test statistics corresponding to the relative risks were much greater among the latter, a likely function of greater statistical power. Among the never- and nonsmokers, the relative risks diminished over

Table 1. Relative risk of initiating daily smoking by past-month use of smokeless tobacco (SLT) at the baseline survey

Wave I sample		Wave II (~1 year)		Wave III (~6 years)	
Outcome/smoking status	SLT use (%)	Number of subjects	Relative risk (95% *CI*)[a]	Number of subjects	Relative risk (95% *CI*)
Ever-daily smoking					
Never-smokers	3.6	5,909	2.32 (1.19–4.50)*	4,766	1.21 (0.84–1.75)
Never- and nonsmokers	5.2	10,820	2.00 (1.60–2.49)†	8,685	1.34 (1.14–1.56)†
Current smoking[b]					
Never-smokers	3.6	5,879	1.81 (0.86–3.80)	4,576	1.16 (0.76–1.78)
Never and nonsmokers	5.2	10,706	2.09 (1.65–2.64)†	8,075	1.33 (1.11–1.61)**

Note. [a]Estimates account for sampling weights and design effects. Relative risk = (incidence of smoking outcome among SLT users)/(incidence of smoking outcome among nonusers).

[b]Analyses of the outcome for current smoking (ever-daily and ≥1 cigarette in past-month) exclude the former smokers.

*$p < .05$; **$p < .01$; †$p < .001$.

time from 2.1 (wave II) to 1.3 (wave III) but remained highly significant.

The distributions of propensity scores in the full baseline sample ($n = 14{,}985$) of the never- and nonsmokers varied substantially by past-month SLT use. As illustrated in Figure 1, the SLT users exhibited a greater variability in conditional probabilities compared with the nonusers.

Although most nonusers had conditional probabilities in the lower ranges, a sufficient number had probabilities in the right tail of the distribution for matches with SLT users. This was not readily apparent given the scaling properties of the figure (i.e., larger number of nonusers compared with users). An examination of quintiles of the propensity scores by SLT use confirmed speculations that stratification was not a viable option for analyses. The proportions of nonusers of SLT in the five quintiles (21.8%, 20.8%, 20.9%, 19.7%, and 16.7%) approximated those observed in the full sample (20%), yielding samples of sufficient size in each stratum. But the disproportions of SLT users in the first three of the five quintiles (3.4%, 5.2%, 9.0%, 19.3%, and 63.0%) did not yield sufficient samples for stratified matching.

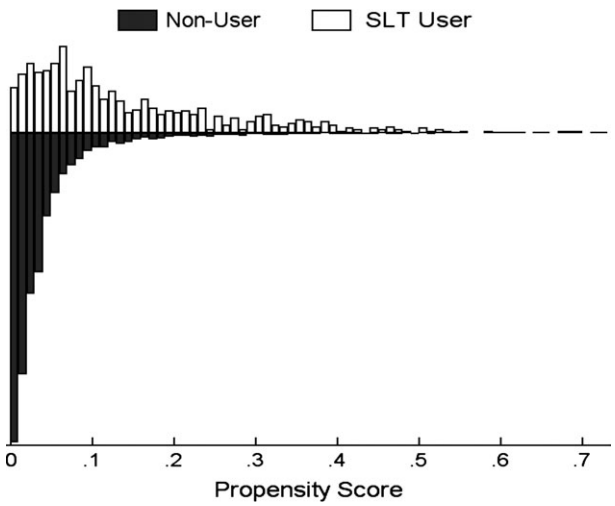

Figure 1. Histogram of propensity score by use of smokeless tobacco at the wave I survey.

Prior to matching on propensity score, we observed significant differences between the SLT users and nonusers for every covariate in the nonpaired sample (Table 2). Notable differences occurred by gender, race, having friends/parents who smoke, having smoked an entire cigarette, binging on alcohol, using cannabis, having repeated a grade in school, motorcycle riding, seat belt use, and delinquency. Compared with nonusers, SLT users had a much greater likelihood of being exposed to cigarette smoking and exhibiting harmful behaviors. In sharp contrast, the means and proportions of covariates were quite comparable between SLT users and nonusers in the paired sample. Not a single test statistic exceeded the threshold for statistical significance ($\alpha = .05$). Individuals were perfectly matched on cannabis use and were matched very closely on race, having smoked a cigarette, and binge drinking. Close matches occurred for a majority of the covariates, reflected by a reduction in bias exceeding 80%. In a separate analysis of the never-smokers, the matching yielded a comparable balance of covariate distributions.

Results from the two sets of propensity score analyses, conducted in the paired and nonpaired samples, are summarized in Table 3. Among the never-smokers, the numbers of discordant pairs at wave II were small ($n = 25$, 28 pairs), increasing by almost threefold by the wave III survey. An increase of this magnitude was reflective of the greater prevalence of ever-daily smoking among young adults compared with adolescents. By the third survey, the Add Health participants who used SLT were aged 18–27 years and appeared to be at greater risk ($OR \sim 1.5$, $p = ns$) of becoming daily smokers. But among pairs from the larger sample of the never- and nonsmokers, we found no evidence for an association between SLT use and daily smoking. This null finding also was observed in all the nonpaired analyses, conducted separately to account for the complex sampling of the Add Health survey.

Discussion

An exposure in an individual is considered a risk factor if it occurs prior to the onset of disease and is associated with the individual's increased probability of developing disease (Mrazeck & Haggerty, 1994). Our findings indicate that SLT qualifies as a risk factor for daily cigarette use, not accounting for other predictors for smoking. Consistent with findings from three U.S.-based

Table 2. Comparisons of baseline covariates in the sample of the never- and nonsmokers, before and after matching on propensity score

		Nonpaired sample			Paired sample		
Baseline variable	Scale or range	SLT users % or mean (*SD*)	Nonusers % or mean (*SD*)	Test statistic[a]	SLT users % or mean (*SD*)	Nonusers % or mean (*SD*)[b]	Reduction in bias (%)[c]
Sample size[d]		498	10,322		496 pairs		
Demographics							
Age (years)	11–21	16.1 (1.6)	15.7 (1.6)	5.6†	16.1 (1.6)	16.1 (1.6)	92.6
Male gender	y/n	81.1%	46.2%	232†	81.0%	78.6%	93.1
Race							
White	y/n	77.3%	64.2%	36.5†	77.2%	77.0%	98.5
Black	y/n	17.5%	25.7%		17.5%	18.1%	92.6
Asian or Native	y/n	5.2%	10.1%		5.2%	4.8%	91.7
American							
U.S. region							
West	y/n	17.3%	24.9%	20.7†	17.3%	16.5%	89.4
Midwest	y/n	27.1%	23.8%		27.2%	24.2%	8.8
South	y/n	44.2%	37.6%		43.9%	45.4%	78.6
Northeast	y/n	11.4%	13.7%		11.5%	13.9%	−5.6
Smoking exposures							
Smoking friends	0–3	1.05 (0.98)	.49 (0.84)	15.7†	1.04 (0.97)	1.10 (1.12)	88.8
Smoking parents	0–2	1.12 (0.77)	.97 (0.78)	4.2†	1.12 (0.77)	1.09 (0.77)	81.0
Cigarettes in home	y/n	33.3%	26.0%	13.1**	33.3%	31.8%	80.7
Smoked cigarette	y/n	50.4%	26.5%	136†	50.4%	50.0%	98.3
Prevention taught	y/n	90.4%	92.9%	4.4*	90.3%	90.5%	92.0
Behavioral risks							
Binge drink (log)	0–1.9	0.56 (0.70)	0.16 (0.40)	17.8†	0.55 (0.69)	0.55 (0.67)	98.2
Cannabis use	y/n	22.7%	6.9%	169†	22.4%	22.4%	100
Repeated grade	y/n	32.3%	18.5%	58.2†	32.2%	35.9%	73.7
Motorcycle riding	0–4	0.67 (1.17)	0.26 (0.74)	10.7†	0.66 (1.16)	0.65 (1.16)	98.0
Seat belt use	0–4	2.77 (1.27)	3.19 (1.10)	−8.0†	2.77 (1.27)	2.70 (1.32)	83.7
Scaled scores							
Delinquency (log)	0–3.8	1.37 (0.91)	0.98 (0.83)	9.4†	1.36 (0.91)	1.41 (0.90)	89.0
Depression (log)	0–3.9	2.32 (0.68)	2.21 (0.73)	3.3**	2.32 (0.68)	2.36 (0.68)	60.3

Note. y/n = yes or no; SLT, smokeless tobacco. All estimates are unweighted.

[a]Pearson's chi-square statistic corresponds to categorical variables (y/n); *z* statistic from the Wilcoxon rank-sum test corresponds to ordinal variables; *t* statistic applies to mean comparisons of age, delinquency, and depressive symptoms.

[b]None of the analogous paired tests (McNemar's chi-square, Wilcoxon's sign-rank, *t* tests) was statistically significant.

[c]Rounding of sample means accounts for differences.

[d]Corresponds to participants with complete data at waves I and II.

*$p < .05$; **$p < .005$; †$p < .0001$.

cohorts (Haddock et al., 2001; Severson et al., 2007; Tomar, 2003), the incidence rates for smoking were significantly greater among SLT users compared with nonusers in Add Health, both later in adolescence and in young adulthood. However, as suggested from the propensity score analyses, baseline differences between SLT users and nonusers accounted for most of the variation in smoking outcomes. Results from the paired analyses support the conclusions reached by O'Connor et al. (2003) and researchers in Sweden (Foulds, Ramstrom, Burke, & Fagerström, 2003), reinforcing the importance of matching levels of SLT use on the risk factors for cigarette smoking.

The baseline differences in substance use and risky behaviors, which were observed between the SLT users and nonusers in Add Health, were comparable to the differences reported in samples of adolescents (Coogan, Geller, & Adams, 2000) and young adults entering military training (Lando et al., 1999). Coogan et al. (2000) reported that among male high school students from Connecticut, current SLT users were much more likely than nonusers to smoke cigarettes, smoke cannabis, drink heavily, engage in fighting, receive failing grades, and rarely or never use seat belts. Air Force recruits who used SLT exhibited a similar pattern of risk-taking behavior, which was particularly prevalent among the polytobacco users. An unexpected finding in the present study was the higher prevalence of SLT use among female participants (1.7%). Comparable rates were observed at waves II and III, confirming the reliability of the estimate and the necessity of including the female users in the analyses. Although considered rare, SLT use does occur in women and has been reported in civilian and military samples (Vander Weg et al., 2005). Coinciding with the wave I Add Health survey (1995–1996), Lando et al. (1999) reported that 4% of female Air Force

Table 3. Odds of smoking outcomes by baseline use of smokeless tobacco (SLT), controlling for propensity scores in paired and nonpaired samples

Wave I sample		Wave II (~1 year)		Wave III (~6 years)	
Outcome/smoking	Data type[a]	Number of subjects[b]	Odds ratio (95% *CI*)	Number of subjects	Odds ratio (95% *CI*)
Ever-daily smoking					
Never-smokers	Paired	28/189	1.33 (0.63–2.82)	75/165	1.59 (0.99–2.52)
Never-smokers	Nonpaired	5,662	0.95 (0.30–3.05)	4,580	1.02 (0.57–1.83)
Never- and nonsmokers	Paired	149/496	0.93 (0.68–1.29)	203/446	1.03 (0.78–1.36)
Never- and nonsmokers	Nonpaired	10,270	1.14 (0.80–1.62)	8,282	1.03 (0.74–1.43)
Current smoking					
Never-smokers	Paired	25/185	1.08 (0.49–2.37)	60/145	1.50 (0.89–2.51)
Never-smokers	Nonpaired	5,633	0.53 (0.13–2.11)	4,399	0.99 (0.53–1.86)
Never- and nonsmokers	Paired	139/490	0.93 (0.67–1.30)	179/370	1.06 (0.79–1.42)
Never- and nonsmokers	Nonpaired	10,176	1.12 (0.78–1.61)	7,702	0.99 (0.71–1.41)

Note. [a]Analyses of nonpaired data account for sampling weights and design effects, in which the propensity score and SLT use were modeled as predictors of the smoking outcomes (weights account for discrepant sample sizes in Tables 2 and 3).

[b]Sample sizes for paired data (*n*/*N*) correspond to number of discordant pairs/number of total pairs.

recruits had experimented with SLT and approximately 0.5% had a history of regular use.

The studies conducted in Sweden have addressed the gateway question from a number of perspectives (for review, see Foulds et al., 2003). Some studies have challenged the hypothesis, citing a larger proportion of former smokers who are current snus users, as compared with the proportion of former snus users who are current smokers (Furberg, Lichtenstein, Pedersen, Bulik, & Sullivan, 2006; Rodu et al., 2002). Furberg et al. (2006) reported that most Swedish smokers had initiated use prior to age 25, in contrast to male snus users who exhibited a later age at onset. Given the older age at onset and high prevalence of snus use, it is questionable whether the Swedish and American users of SLT are comparable. For studying gateway effects, one might speculate that differences in risk-taking behaviors are not as evident between the Swedish users and nonusers of snus. Consequently, one might argue that this would diminish the likelihood of observing spurious associations between use of snus and cigarette smoking in the Swedish samples. Yet, risky behaviors, observed in adolescent SLT users in the United States, also have been reported among teenagers in Sweden (Galanti, Wickholm, & Gilljam, 2001). In a recent study, Swedish adolescents who initiated use of cigarettes and snus within the same time frame had the greatest likelihood of becoming current tobacco users (Galanti, Rosendahl, & Wickholm, 2008). Irrespective of culture, tobacco use in adolescence is probably an indicator of risk-taking behavior, having less to do with the use of any particular product.

Use of the Add Health sample was ideal in many respects for testing the association between SLT use and smoking. A number of risk factors for smoking, which were used in estimating propensity scores, were assessed at the wave I survey. The surveys at waves II and III provided outcome measurements in adolescence and young adulthood, which corresponded, respectively, with periods of observation in the Project Sixteen community intervention and TAPS. In addition, detailed questions on smoking history facilitated use of more than one sample. The unadjusted risks for the smoking outcomes were comparable by wave in the two Add Health samples, the never-smokers and the never- and nonsmokers. Yet, the unadjusted estimates were highly significant in the latter, but not the former, suggesting an increase in the power to detect statistical significance for effects of comparable magnitude. Statistical power was insufficient for modest effects ($OR = 1.5$) in matched pairs of the exclusive never-smokers. One might suggest that a larger paired sample of never-smokers would yield evidence in support of the gateway hypothesis. But, this argument is challenged by the null effects of SLT estimated from the weighted, nonpaired analyses. Inclusion of nonsmokers did not bias our estimates, given the near identical prevalence of lifetime smoking in the SLT users and nonusers in the paired sample.

Use of the Add Health sample and the decision to match individuals on propensity score entailed limitations. Lifetime use of SLT could not be determined among those who were not current users of the product. An effect of SLT use on smoking may have been underestimated by classifying lifetime, noncurrent users as nonusers in this study. The past-30-day measure of SLT use also was examined as a primary predictor by Severson et al. (2007). Similar to the TAPS sample (O'Connor et al., 2003), the Add Health sample included participants of variable age, which could have influenced the outcomes for smoking behavior. Furthermore, the lengthy interval between waves II and III was problematic for assessing events that typically occur in the latter part of adolescence. The use of propensity score matching is not a panacea for observational studies, particularly given its inability to adjust for unmeasured confounders (Rubin, 2004). Accounting for such effects, we considered the alternative approach of examining smoking outcomes among twin pairs that were discordant for SLT use in adolescence. This design has been used in assessing cannabis use as a gateway to hard drug use in Add Health siblings (Lessem et al., 2006). Given the low prevalence of SLT use in adolescence, however, the number of discordant pairs was not sufficient for a sibling-based analysis. An additional limitation of pairing individuals on propensity score was an inability to estimate the risk factors on

which individuals were matched, such as binge drinking. Nevertheless, the paired analyses provided sound methodology for testing and confirming our hypothesis.

Compared with conventional regression, propensity score matching is particularly useful when substantial differences in risk factors for the outcome exist between the groups being compared. The findings from the present study should contribute to the ongoing debate about tobacco harm reduction. A separate issue is whether smokers should be informed of the reduced harm associated with use of low-nitrosamine SLT. Although our results do not address this topic, they do suggest that SLT use in adolescence does not increase an individual's likelihood of becoming a daily smoker.

Funding

This research uses data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. No direct support was received from grant P01-HD31921 for this analysis.

Declaration of Interests

None declared.

Acknowledgments

The authors thank Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 West Franklin Street, Chapel Hill, NC 27516-2524 (addhealth@unc.edu). The authors also thank two anonymous reviewers for their helpful suggestions.

References

Alexander, C., Piazza, M., Mekos, D., & Valente, T. (2001). Peers, schools, and adolescent cigarette smoking. *Journal of Adolescent Health*, *29*, 22–30.

Coogan, P. F., Geller, A., & Adams, M. (2000). Prevalence and correlates of smokeless tobacco use in a sample of Connecticut students. *Journal of Adolescence*, *23*, 129–135.

Daniel Roth, H., Roth, A. B., & Liu, X. (2005). Health risks of smoking compared to Swedish snus. *Inhalation Toxicology*, *17*, 741–748.

Fagerström, K. O., & Schildt, E. B. (2003). Should the European Union lift the ban on snus? Evidence from the Swedish experience. *Addiction*, *98*, 1191–1195.

Foulds, J., Ramstrom, L., Burke, M., & Fagerström, K. (2003). Effect of smokeless tobacco (snus) on smoking and public health in Sweden. *Tobacco Control*, *12*, 349–359.

Furberg, H., Lichtenstein, P., Pedersen, N. L., Bulik, C., & Sullivan, P. F. (2006). Cigarettes and oral snuff use in Sweden: Prevalence and transitions. *Addiction*, *101*, 1509–1515.

Galanti, M. R., Rosendahl, I., & Wickholm, S. (2008). The development of tobacco use in adolescence among "snus starters" and "cigarette starters": An analysis of the Swedish "BROMS" cohort. *Nicotine & Tobacco Research*, *10*, 315–323.

Galanti, M. R., Wickholm, S., & Gilljam, H. (2001). Between harm and dangers. Oral snuff use, cigarette smoking and problem behaviours in a survey of Swedish male adolescents. *European Journal of Public Health*, *11*, 340–345.

Gartner, C., Hall, W., Chapman, S., & Freeman, B. (2007). Should the health community promote smokeless tobacco (Snus) as a harm reduction measure? *PLOS Medicine*, *4*, 1138–1141.

Guo, S., Barth, R. P., & Gibbons, C. (2006). Propensity score matching strategies for evaluating substance abuse services for child welfare clients. *Children and Youth Services Review*, *28*, 357–383.

Haddock, C. K., Weg, M. V., DeBon, M., Klesges, R. C., Talcott, G. W., Lando, H., et al. (2001). Evidence that smokeless tobacco use is a gateway for smoking initiation in young adult males. *Preventive Medicine*, *32*, 262–267.

Kandel, D. B., Kiros, G. E., Schaffran, C., & Hu, M. C. (2004). Racial/ethnic differences in cigarette smoking initiation and progression to daily smoking: A multilevel analysis. *American Journal of Public Health*, *94*, 128–135.

Ko, D. T., Chiu, M., Austin, P. C., Bowen, J., Cohen, E. A., & Tu, J. V. (2008). Safety and effectiveness of drug-eluting stents among diabetic patients: A propensity analysis. *American Heart Journal*, *156*, 125–134.

Kozlowski, L. T., O'Connor, R. J., Edwards, B. Q., & Flaherty, B. P. (2003). Most smokeless tobacco use is not a causal gateway to cigarettes: Using order of product use to evaluate causation in a national US sample. *Addiction*, *98*, 1077–1085.

Lando, H. A., Haddock, C. K., Klesges, R. C., Talcott, G. W., & Jensen, J. (1999). Smokeless tobacco use in a population of young adults. *Addictive Behaviors*, *24*, 431–437.

Lessem, J. M., Hopfer, C. J., Haberstick, B. C., Timberlake, D., Ehringer, M. A., Smolen, A., et al. (2006). Relationship between adolescent marijuana use and young adult illicit drug use. *Behavior Genetics*, *36*, 498–506.

Leuven, E., & Sianesi, B. (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. http://ideas.repec.org/c/boc/bocode/s432001.html (Version 3.00)

Little, R. J., & Rubin, D. B. (2000). Causal effects in clinical and epidemiological studies via potential outcomes: Concepts and analytical approaches. *Annual Review of Public Health*, *21*, 121–145.

Martinet, Y., Bohadana, A., & Fagerström, K. (2006). Would alternate tobacco products use be better than smoking?. *Lung Cancer*, *53*, 1–4.

Mrazeck, P. J., & Haggerty, R. J. (1994). *Reducing Risks for Mental Disorders: Frontiers for Preventive Intervention Research*. Washington, DC: Institute of Medicine, National Academy Press.

Newgard, C. D., Hedges, J. R., Arthur, M., & Mullins, R. J. (2004). Advanced statistics: The propensity score—a method for estimating treatment effect in observational research. *Academy of Emergency Medicine*, *11*, 953–961.

O'Connor, R. J., Flaherty, B. P., Quinio Edwards, B., & Kozlowski, L. T. (2003). Regular smokeless tobacco use is not a reliable predictor of smoking onset when psychosocial predictors are included in the model. *Nicotine & Tobacco Research*, *5*, 535–543.

O'Connor, R. J., Kozlowski, L. T., Flaherty, B. P., & Edwards, B. Q. (2005). Most smokeless tobacco use does not cause cigarette smoking: Results from the 2000 National Household Survey on Drug Abuse. *Addictive Behaviors*, *30*, 325–336.

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, *49*, 1373–1379.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measures*, *1*, 385–401.

Rodu, B., & Cole, P. (2004). The burden of mortality from smoking: Comparing Sweden with other countries in the European Union. *European Journal of Epidemiology*, *19*, 129–131.

Rodu, B., Stegmayr, B., Nasic, S., & Asplund, K. (2002). Impact of smokeless tobacco use on smoking in northern Sweden. *Journal of Internal Medicine*, *252*, 398–404.

Rosenbaum, P. R., & Rubin, D. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, *79*, 516–524.

Rosenbaum, P. R., & Rubin, D. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician*, *39*, 33–38.

Rosner, B. (1990). *Fundamentals of biostatistics* (3rd ed.). Boston: PWS-Kent Publishing Company.

Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine*, *127(Pt 2)*, 757–763.

Rubin, D. B. (2004). On principles for modeling propensity scores in medical research. *Pharmacoepidemiology and Drug Safety*, *13*, 855–857.

Savitz, D. A., Meyer, R. E., Tanzer, J. M., Mirvish, S. S., & Lewin, F. (2006). Public health implications of smokeless tobacco use as a harm reduction strategy. *American Journal of Public Health*, *96*, 1934–1939.

Severson, H. H., Forrester, K. K., & Biglan, A. (2007). Use of smokeless tobacco is a risk factor for cigarette smoking. *Nicotine & Tobacco Research*, *9*, 1331–1337.

Tomar, S. L. (2003). Is use of smokeless tobacco a risk factor for cigarette smoking? The U.S. experience. *Nicotine & Tobacco Research*, *5*, 561–569.

Vander Weg, M. W., DeBon, M., Peterson, A. L., Sherrill-Mittleman, D., Klesges, R. C., & Relyea, G. E. (2005). Prevalence and correlates of lifetime smokeless tobacco use in female military recruits. *Nicotine & Tobacco Research*, *7*, 431–441.

Weitzen, S., Lapane, K. L., Toledano, A. Y., Hume, A. L., & Mor, V. (2004). Principles for modeling propensity scores in medical research: A systematic literature review. *Pharmacoepidemiology and Drug Safety*, *13*, 841–853.