



Published in final edited form as:

Addiction. 2007 November ; 102(11): 1831–1839. doi:10.1111/j.1360-0443.2007.01974.x.

Consistency of self-reported smoking over a 6-year interval from adolescence to young adulthood

Cassandra A. Stanton, George Papandonatos, Elizabeth E. Lloyd-Richardson, and Raymond Niaura

Brown Medical School, Providence, RI, USA

Abstract

Aims—To examine the reliability of self-report cigarette smoking questions by describing recanting (denial of previous smoking reports) in a nationally representative sample of US adolescents followed throughout young adulthood. Predictors of recanting across stages of smoking uptake/progression are examined.

Design—A total of 12 985 respondents to cigarette smoking questions during in-home interviews at waves I and III (6 years apart) of the National Longitudinal Study of Adolescent Health (Add Health). The sample survey procedures of Stata 9.0 were used to produce nationally representative estimates, with standard errors adjusted for both clustering at the school level and stratification by geographical region.

Measurements—Recanting probabilities determined by reports of stages of smoking uptake/progression at each time-point were predicted by race/ethnicity, parental education, household income, poverty level, depression and peer daily smoking.

Findings—Stage-specific results indicated that recanting is higher when the earlier smoking was less frequent/intense. Recanters were older, from lower-income households and had higher baseline depression levels. Non-Hispanic black youth were significantly more likely to recant previous smoking compared to non-Hispanic white youth, even in multivariate models controlling for socio-demographic variables. Predictors of recanting differed by level of tobacco involvement. The greater likelihood of non-Hispanic black respondents to deny previous smoking may be a reflection of less intense or more intermittent use of tobacco that leads to recall differences over time.

Conclusions—Racial/ethnic subgroups and/or respondents endorsing depressive symptoms may be more vulnerable to misclassification during interpretation of national survey data and subsequently not identified properly for prevention/intervention programs.

Keywords

Adolescents; African Americans; depression; ethnic groups; longitudinal studies; recanting; reproducibility of results; socioeconomic factors; smoking; young adults

INTRODUCTION

National smoking prevalence rates in the United States are often derived from large-scale surveys that use self-report and retrospective questions to assess smoking behavior. While

self-report measures of smoking behaviors are used frequently and generally accepted, the increasing availability of longitudinal survey data has brought into question the reliability of retrospective reports of substance use, particularly among racial/ethnic subgroups [1–7]. Longitudinal studies have found that reports of life-time smoking status are more accurate than retrospective questions regarding starting and quitting ages [3,4,7,8]. Reliability of self-reported age of first substance use has been found to be more consistent when reporters are adults, and in response to questions posed over no more than a 2-year time span [6]. Moreover, subgroup differences have been reported in which non-Hispanic whites, females, respondents from households at or above poverty level and respondents with higher education levels were found to provide more reliable self-reports of substance use over time [1,6,8,9].

The reproducibility of self-reported behaviors in longitudinal surveys can be gauged by the extent of logically inconsistent responses of life-time drug use, also referred to as recanting. Recanting is a specific form of inconsistent reporting, in which a respondent denies previous life-time substance use at a follow-up interview, despite reporting some use in an earlier interview. Recanting is a lower-prevalence and conservative form of inconsistent reporting, in which two reports taken at separate times directly contradict each other. For example, participants reporting regular smoking at follow-up who described themselves as occasional smokers at baseline may well have misreported their prior smoking status, but cannot be classified as ‘recanters’. The use of the term does not imply intentional misreporting, but rather captures irrefutable inconsistencies that bring into question the reliability of the response at one of the time-points. Recanting of smoking and drug use in longitudinal studies has been reported in school-based surveys [9], household surveys [10] and cross-cultural surveys of youth outside the United States [1].

Empirical studies reporting rates of drug use recanting in longitudinal studies are growing, and these studies consistently demonstrate subgroup differences. For example, minority respondents to the National Longitudinal Survey of Youth in the United States have been found in multiple studies to be more likely than whites to recant their previous reports of life-time marijuana use, even after controlling for educational status and other demographic variables [11,12]. More recently, it has been reported that black respondents have higher rates of recanting drug use than white subjects [9]. In general, recanting substance use in survey research in the United States and Northern Ireland appears to be higher among respondents for whom disclosure may have a higher social cost [1,12,13].

Inconsistent reporting—as manifested by recanting—affects prevalence estimates in longitudinal general population surveys, and may account partially for racial/ethnic differences found consistently in adolescent smoking rates. This study seeks to further our understanding of the reliability of survey-based self-report smoking questions by describing recanting of adolescent smoking reports after a 6-year time interval in a nationally representative sample. We will add uniquely to the literature in this area by examining differences in recanting patterns according to socio-demographic subgroups and psychological covariates at different stages of smoking uptake and progression. Previous studies have shown life-time smoking reports to be more reliable than specific questions regarding smoking patterns [3,4,7,8]; it is therefore hypothesized that recall of retrospective data may differ according to the frequency and intensity of the smoking behavior being reported. Additionally, empirical studies to date have not examined misreporting of smoking history according to strongly associated covariates, such as depression status. To the extent that depressive symptoms are associated with other risk factors related to misreporting, such as race/ethnicity, living status and/or measures of socio-economic status [14–16], it can be hypothesized that recall/report of smoking history from depressed patients may be less

reliable. We test this hypothesis by examining recanting according to both baseline and follow-up depressive symptoms, as well as change in depressive symptoms over time.

METHODS

Study sample and survey procedure

The sample was drawn from the National Longitudinal Study of Adolescent Health (Add Health [17]), a school-based study of the health-related behaviors of a United States' national representative sample of 90 118 adolescents in grades 7–12 and approved by the Institutional Review Board of the University of North Carolina. In-home interviews were conducted on a subset of these wave I participants in 1995 ($n = 20\,745$; median age was 16 years 2 months, with 95% of the sample falling between 13 and 19 years). Approximately 1 year later, a second in-home interview was conducted ($n = 14\,738$) for wave II, but cigarette smoking questions differed slightly from wave I. Therefore, wave II data are not examined in this study. The in-home wave III sample consisted of wave I respondents re-interviewed 6 years later in 2001–02, when the majority of them were aged between 18 and 26 years ($n = 15\,170$). Data were recorded on laptop computers to maintain confidentiality. For less sensitive material, the interviewer read the questions and entered the respondent's answers. For more sensitive material, the respondent entered his or her own answers in privacy. Additional details are available elsewhere [18].

Measures

Smoking stage—Previous reports utilizing Add Health data [19] have proposed stage-specific categorizations of cigarette smoking. Based on parallel tobacco questions at each wave ('Have you ever tried cigarette smoking, even just one or two puffs?', 'How old were you when you smoked a whole cigarette for the first time?' (wave I) or 'Have you ever smoked an entire cigarette?' (wave III), 'Have you ever smoked cigarettes regularly—that is, at least one cigarette every day for 30 days?', 'During the past 30 days, on how many days did you smoke cigarettes?'), we employed a similar classification procedure [19] based on smoking frequency and recency, as follows: never smokers (denied ever trying a puff or two of cigarettes), puffers (endorsed trying cigarettes, but denied ever smoking a whole cigarette), light smokers (endorsed trying cigarettes and ever smoking a whole cigarette, but denied ever smoking regularly and denied any smoking within the past 30 days), occasional smokers (reported smoking between 1 and 29 in the past 30 days), regular smokers (reported smoking on a daily basis within the past 30 days) and ex-regular smokers (endorsed regular past smoking, but denied any smoking within the past 30 days).

Covariates—Socio-demographic variables and social-psychological variables found to be highly associated with adolescent smoking [19] were assessed. Race and ethnicity were categorized as non-Hispanic white, non-Hispanic black, Hispanic (irrespective of race), non-Hispanic Asian, mixed, and other. Parental education was assessed via wave I parent report of highest education level. Household income was also assessed via wave I parent report of total household income before taxes in 1994. Poverty level was then defined as total household income below 1.5 times the US Census Bureau 1994 poverty thresholds, adjusted for age of the primary household member listed in a family roster, family size and number of related children under 18 years of age [20]. Depression reflected past week depressive symptomatology based on nine items asked at each wave derived from the Center for Epidemiologic Studies Depression Scale [21]. Principal components analysis yielded a single factor solution, with good internal consistency (wave I $\alpha = 0.80$, wave III $\alpha = 0.81$). Peer daily smoking was assessed with a wave I question: 'Of your three best friends, how many smoke more than one cigarette per day?'

Data reduction and statistical analyses

The sample survey procedures of Stata 9.0 [22] were used to produce nationally representative estimates, with standard errors adjusted for both clustering at the school level and stratification by geographical region [18]. Of 15 170 wave I respondents who participated in wave III, 265 had missing school identification, while 374 had missing region of origin, leaving 14 322 adolescents from 132 schools in four regions (North-east, Midwest, South and West) available for analysis. For these subjects, Add Health calculated weights that adjust simultaneously for unequal probability of selection at wave I, non-response at wave III and chance fluctuations in the composition of the sample from the composition of the population as a whole; trimming to the range 18.9–7547 ensured that no single individual exerted undue influence on the results.

In modeling recanting rates, we estimated initially the joint distribution of smoking stages across survey waves, using the svytabulate procedure [22] to convert sample proportions to nationally representative population estimates (Table 1). Subjects who were never-smokers at baseline were uninformative, as none of their wave III responses could be taken as contradicting reports of their wave I smoking status. The remaining subjects were classified as ‘recanters’ if: (i) they reported never having smoked even a puff in their life-time at wave III, despite having reported smoking at wave I, or (ii) they denied ever smoking a whole cigarette at wave III, contradicting wave I reports. Population characteristics were compared separately between recanters and non-recanters by smoking stage at baseline (Table 2). For continuous variables, the svymean procedure [22] was used to calculate both first and second moments. For discrete variables, the svytabulate procedure [22] was used to calculate prevalences. Mean comparisons employed a linear Taylor expansion to approximate the standard error of their difference, whereas prevalences were compared using the Rao–Scott correction to the Pearson χ^2 statistic [23]. Variables showing statistically significant differences across recanting strata for at least one baseline smoking stage were included as covariates in a multivariate logistic regression model for predicting recanting rates.

Regression models were fitted subsequently for each smoking stage at baseline, and simplified via a joint backward elimination procedure, in which a particular variable was excluded only if it failed to attain statistical significance for all subpopulations of interest. By retaining covariates significant in predicting recanting for at least one baseline smoking stage, we ensured that a common set of covariates would appear across all stage-specific models, so that the adjusted odds ratios (OR) for each covariate could be compared and contrasted across smoking stages.

RESULTS

Smoking stages across waves

Table 1 depicts the distribution of smoking stages across survey waves among the 12 985 adolescents with appropriate sample weights (GSWGT3-2) and smoking data available at both time-points—a sample representative of $n = 19\,895\,953$ adolescents. After excluding never-smokers at baseline, only 7021 individuals remained, representing $n = 11\,107\,426$ smoking-exposed adolescents (55.83% of the original population). Recanting rates were obtained for each baseline smoking stage by comparing the prevalence in the cells shown in bold type to the respective row totals: 16.95% among puffers, 26.94% among light smokers, 13.92% among occasional smokers, 5% among regular smokers and 9.46% among ex-smokers.

Population characteristics by smoking stage

Table 2 characterizes this population—both overall and separately by recanting status—in terms of continuous covariates. Recanters were older than non-recanters by about 6 months among puffers ($P = 0.004$) and by about 4 months among light smokers ($P = 0.033$). Household income was lower among recanters for all but the ex-smokers, with standardized mean differences in the 0.15–0.25 range. Baseline depression appears to have been 0.18–0.30 standard deviations higher among recanters for all but regular smokers, with highly significant differences for puffers ($P = 0.020$), light smokers ($P = 0.012$) and occasional smokers ($P = 0.004$). Further, although a larger drop in depression from baseline to follow-up was observed for recanters among puffers ($P = 0.032$) and light smokers ($P = 0.010$), this result appears to have been driven entirely by wave I differences. Consequently, only baseline depression was considered for inclusion in the multivariate regression models.

Table 3 characterizes this population further in terms of discrete covariates. Although recanters were mainly female, the only significant gender differences emerged for regular smokers ($P < 0.001$), with a preponderance of males observed among recanters. Further, recanters were consistently more likely to be members of minority groups than non-recanters. Blacks were over-represented among recanters across smoking stages, whereas Hispanics were more prevalent among light and occasional smokers, and particularly common among ex-smokers (8% versus 30%). Both poverty and parental educational attainment showed similar patterns, with recanters uniformly more likely to have been poor, and from homes in which parents received no more than a high school education. No differences emerged with respect to the number of peers who smoked daily for any subpopulation of interest (not shown).

Regression modeling of recanting status by smoking stage

The backward elimination procedure was initiated with a model that included age, gender, race/ethnicity, household income, poverty, educational level and baseline depression. Continuous variables were standardized by subtracting the overall population mean and dividing by the population standard deviation. Of this initial set of covariates, household income, parental education and baseline depression were not significant for any subpopulation of interest, given the other variables in the model. As a result, Table 4 includes only a reduced set of predictors made up of baseline age (population mean = 16.17, population SD = 1.72), gender, race/ethnicity and poverty level. Missing data for household income reduced the sample available for multivariate analyses to 78% of the original population of interest. Respondents whose parents reported household income were more likely to be younger, white and to have completed high school, resulting in a slight selection bias in Table 4.

The intercepts in Table 4 correspond to the odds of recanting among 16-year-old white male adolescents with annual income exceeding 1.5 times the poverty threshold, a group chosen deliberately as the referent for its generally low recanting rates. Converting odds to proportions, these recanting rates and associated confidence intervals (CIs) were estimated as follows: puffers = 0.12 (95% CI: 0.08, 0.17), light smokers = 0.18 (95% CI: 0.14, 0.24), occasional smokers = 0.08 (95% CI: 0.06, 0.12), regular smokers = 0.07 (95% CI: 0.03, 0.13) and ex-smokers = 0.06 (95% CI: 0.02, 0.14). Relative to this reference group, the odds of recanting among puffers rose by almost half for each standard unit increase in baseline age above its mean (OR = 1.44, 95% CI: 1.14, 1.82); no significant age effects on recanting were detected for the remaining smoking stages. Although female gender appeared to increase recanting for all but regular smokers, it attained significance only in this latter group, where it led instead to a fourfold reduction in the odds of recanting (OR = 0.26, 95% CI: 0.13, 0.55). Race/ethnicity differences were highly significant, with P -values of multivariate Wald

tests given by: puffers, $P = 0.044$; light, $P < 0.001$; occasional, $P < 0.001$; regular, $P < 0.001$; ex-smokers, $P = 0.009$. Compared with the univariate tests shown in Table 3, it appears that adjusting for age, gender and poverty rates diluted the significance of race/ethnicity among puffers—within the subsample of respondents for which household income is available—concentrating it on a black–white contrast, with black adolescents having significantly higher odds of recanting than whites (OR = 1.91, 95% CI: 1.18, 3.10). For light smokers, differences between minority and white youth were observed across the spectrum, with black (OR = 3.43, 95% CI: 2.27, 5.20), Hispanic (OR = 1.92, 95% CI: 1.11, 3.31) and Asian (OR = 4.42, 95% CI: 1.70, 11.52) adolescents all reporting higher recanting rates than whites. A similar picture was observed for occasional smokers, with significant OR for both blacks (OR = 3.60, 95% CI: 1.97, 6.57) and Hispanics (OR = 1.95, 95% CI: 1.02, 3.73), although not for Asians. Among regular smokers, blacks had higher—albeit not significant—odds of recanting, whereas Hispanics (OR = 0.07, 95% CI: 0.02, 0.29) and Asians (OR = 0.19, 95% CI: 0.03, 1.00) had sharply lower odds. Finally, for ex-smokers, the significant race/ethnicity effect could be attributed primarily to Hispanic–white differences (OR = 8.05, 95% CI: 2.29, 28.27). In summary, recanting rates for blacks tended to be higher than whites across all smoking stages, whereas Asian rates were mostly lower, with the exception of light smokers at baseline. For Hispanics the picture appeared mixed, with results largely stage-dependent. Finally, poverty led to higher recanting rates across all smoking stages, but attained significance only among light (OR = 1.58, 95% CI: 1.02, 4.15) and occasional smokers (OR = 1.93, 95% CI: 1.36, 2.73).

DISCUSSION

Overall, results indicated that the likelihood of denying life-time smoking behaviors in a follow-up interview 6 years after reporting smoking in an initial adolescent interview was higher when the earlier smoking behavior was less frequent/intense (e.g. puffers, light smokers). These findings are consistent with other studies that have found lighter smoking behaviors to be more susceptible to recall and report discrepancies [4]. For example, analyses of various longitudinal and cross-sectional surveys of adults in the United States demonstrated that retrospective reports of life-time smoking that did not agree over time tended to be among respondents who were classified as light smokers based on the number of cigarettes per day reported initially [4].

In our study, socio-demographic and psychosocial predictors of recanting were found to vary by smoking stage. For example, older age at baseline was predictive of recanting for only the lowest smoking intensity stage of puffing. Thus, older baseline participants who reported puffing later in adolescence were more likely to take it back in young adulthood. This could suggest that older adolescents who have only ever puffed were less likely to progress to higher smoking stages and thus more likely to minimize a smoking history later, whereas younger puffers may have progressed to greater smoking frequency and in turn been less likely to recant. Consistent with previous research [24], males were more likely to report smoking inconsistently than females. Our findings—that account uniquely for patterns of smoking—indicated gender differences only for reports of regular smoking. Further studies are needed to explain the source of these differences, e.g. by testing the hypothesis that males are more susceptible to over-reporting or exaggerating regular smoking in adolescence, but not in young adulthood.

Across smoking stages, adolescent respondents from low-income households or who reported depression were more likely to recant previous smoking behaviors as young adults, regardless of smoking stage. One can speculate that socio-demographic and psychological stressors may be associated with problem behaviors and polysubstance use and thus, combined, may contribute to response bias at either time-point. Studies using the Add

Health data have found that adolescent smokers had higher depression scores than non-smokers [15], including a temporal effect for females in which depression increases around the onset of smoking, and decreases around the time of quitting [25]. Fluctuating depression symptoms may impact smoking patterns and interfere subsequently with later recall. Whereas depression during adolescence was found to be predictive of recanting, depression in young adulthood was not. Mean depression levels of the students dropped from wave I to wave III, as the respondents grew out of adolescence (mean age rises from around 16 years to approximately 22 years), suggesting perhaps that depressive symptoms need to be above a certain threshold to affect reporting accuracy.

Although inconsistent reporting could be attributed to socio-demographic factors, our results indicated that minority youth were more likely to recant than white youth, even after these factors had been controlled for via regression adjustment. These results replicate the findings of previous longitudinal studies of other substance use [7,9] that appear to suggest that racial/ethnic differences in recanting rates go beyond recall distortions and other forms of measurement error, and may reflect honesty in reporting or social/cultural influences [5].

Studies that have used biochemical measures to validate self-reports of cigarette smoking among minority and white adolescents have found higher misreporting rates among minority teens than whites [26,27]. Additionally, black research participants have been found to be more reluctant than other participants to report tobacco use when they tested positive for cotinine [28]. Although little research has examined specifically why race/ethnicity may be related to under-reported smoking, researchers have suggested that reporting differences may be related to distrust of research in general or perceived differences between the respondent and interviewer regarding the stigmatization of reporting tobacco use [5,28]. In an examination of self-reported response distortion in adolescent smokers, authors attributed findings of minority and younger adolescents' greater likelihood of dishonest smoking reports to social desirability and fear of repercussion [5]. Despite the impact of social/cultural determinants on intentional response distortions, our findings suggest that we must also take into account initial levels of tobacco use. Previous Add Health studies [19,29] have established that minorities report lower levels of ever trying cigarettes or progression to higher levels of smoking (i.e. daily smoking). Recanting analyses indicated that black respondents were less likely to inconsistently report smoking, if they had reported being a daily smoker or ex-daily smoker 6 years earlier. Thus, the greater likelihood of minority respondents to inconsistently or under-report previous smoking behavior may simply be a reflection of less intense or more intermittent tobacco use among minority subgroups that leads to recall differences over time. Moreover, it has been speculated that there may be cultural differences in understanding survey questions. For example, the meaning of 'ever smoked' may differ across ethnic or socio-economic status groups such that experimental or light smokers may interpret the items differently, and this classification may vary systematically by ethnic background [24].

Results of this study do not allow us to speculate as to why inconsistencies existed or to determine at which time point smoking behavior was misreported. Biochemical verification data were not available at any time-point in the Add Health sample, prohibiting us from examining the validity of self-reports [30]. We did not examine cognitive/recall ability or capacity to understand the survey directions, and therefore cannot control for cognitive limitations. It is also noteworthy that recanting in this paper was defined by irrefutable contradictions in smoking reports across the waves and is therefore a conservative measure of inconsistent reporting. Although Add Health is a nationally representative sample in the United States, analyses revealed that respondents with missing income data were more likely to fit the profile of a respondent who may recant previous smoking (e.g. older, less educated,

minority), and therefore results of the multivariate analyses controlling for income may under-represent the recanting phenomenon.

Large-scale school-based data collection often depends upon the use of survey methodology and therefore studies that bring us closer to understanding more clearly the extent and correlates of survey measurement error are critical [2]. This is the first study, to our knowledge, to focus on stages of smoking uptake and employ stage-specific multivariate analyses to demonstrate that recanting is higher when the earlier smoking behavior was less frequent/intense (e.g. puffers, light smokers). Moreover, predictors identified subgroups more susceptible to these patterns of inconsistent reporting, such as racial/ethnic subgroups or respondents who endorse depressive symptoms. These subgroups may be more vulnerable to misclassification during interpretation of survey data and subsequently not identified correctly for prevention/intervention programs.

The implications of these results suggest that survey researchers should use caution in calculating prevalence rates of substance use over time and should account for patterns of inconsistent reporting that may bias estimates. Moreover, they should remain cognizant of factors likely to minimize inconsistent reporting, such as providing recall prompts, minimizing skip patterns, shortening survey instruments and the time-intervals between surveys and addressing social desirability factors within testing conditions [2]. Particular to longitudinal surveys, consistency checks using computerized or tailored surveys can be used to prompt respondents to resolve inconsistent information about life-time substance use [9]. Moreover, follow-up interviews with selected sub-samples of respondents [1] may allow for more in-depth understanding of factors that impact youths' honesty in reporting substance use and sources of response/recall distortions.

Acknowledgments

Work on this article was supported partially by the National Cancer Institute (K07CA95623, C. Stanton, PI), the National Heart, Lung and Blood Institute (K23HL069987, E. Lloyd-Richardson, PI) the USPHS, NIH Transdisciplinary Tobacco Use Research Center (TTURC) Award (P50 CA084719) and by the Robert Wood Johnson Foundation. 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 a grant P01-HD31921 from the National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Special acknowledgement is due to Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. People interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524 (addhealth@unc.edu). Additionally, the authors wish to thank Lan Jiang and Yunxia Sui for assistance with data management and statistical analyses.

References

1. Percy A, McAlister S, Higgins K, McCrystal P, Thornton M. Response consistency in young adolescents' drug use self-reports: a recanting rate analysis. *Addiction*. 2005; 100:189–196. [PubMed: 15679748]
2. Fendrich M. The undeniable problem of recanting. *Addiction*. 2005; 100:143–144. [PubMed: 15679741]
3. Parra GR, O'Neill SE, Sher KJ. Reliability of self-reported age of substance involvement onset. *Psychol Addict Behav*. 2003; 17:211–218. [PubMed: 14498815]
4. Kenkel D, Lillard DR, Mathios A. Smoke or fog? The usefulness of retrospectively reported information about smoking. *Addiction*. 2003; 98:1307–1313. [PubMed: 12930218]
5. Stein LA, Colby SM, O'Leary TA, Monti PM, Rohsenow DJ, Spirito A, et al. Response distortion in adolescents who smoke: a pilot study. *J Drug Educ*. 2002; 32:271–286. [PubMed: 12556133]
6. Johnson TP, Mott JA. The reliability of self-reported age of onset of tobacco, alcohol and illicit drug use. *Addiction*. 2001; 96:1187–1198. [PubMed: 11487424]

7. Johnston LD, O'Malley PM. The recanting of earlier reported drug use by young adults. *NIDA Res Monogr.* 1997; 167:59–80. [PubMed: 9243557]
8. Shillington AM, Clapp JD. Self-report stability of adolescent substance use: are there differences for gender, ethnicity and age? *Drug Alcohol Depend.* 2000; 60:19–27. [PubMed: 10821986]
9. Fendrich M, Rosenbaum DP. Recanting of substance use reports in a longitudinal prevention study. *Drug Alcohol Depend.* 2003; 70:241–253. [PubMed: 12757962]
10. Brittingham A, Tourangeau R, Kay W. Reports of smoking in a national survey: data from screening and detailed interviews, and from self- and interviewer-administered questions. *Ann Epidemiol.* 1998; 8:393–401. [PubMed: 9708875]
11. Mensch BS, Kandel DB. Underreporting of substance use in a national longitudinal youth cohort: individual and interviewer effects. *Public Opin Q.* 1988; 52:100–124.
12. Fendrich M, Vaughn CM. Diminished lifetime substance use over time: an inquiry into differential underreporting. *Public Opin Q.* 1994; 58:96–123.
13. Huerta M, Chodick G, Balicer RD, Davidovitch N, Grotto I. Reliability of self-reported smoking history and age at initial tobacco use. *Prev Med.* 2005; 41:646–650. [PubMed: 15917064]
14. Wight RG, Aneshensel CS, Botticello AL, Sepulveda JE. A multilevel analysis of ethnic variation in depressive symptoms among adolescents in the United States. *Soc Sci Med.* 2005; 60:2073–2084. [PubMed: 15743655]
15. Duncan B, Rees DI. Effect of smoking on depressive symptomatology: a reexamination of data from the National Longitudinal Study of Adolescent Health. *Am J Epidemiol.* 2005; 162:461–470. [PubMed: 16076832]
16. Goodman E. The role of socioeconomic status gradients in explaining differences in US adolescents' health. *Am J Public Health.* 1999; 89:1522–1528. [PubMed: 10511834]
17. Udry, JR. The National Longitudinal Study of Adolescent Health (Add Health), Waves I & II, 1994–1996; Wave III, 2001–2002. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill; 2003. [machine-readable data file and documentation]
18. Harris KM, Florey F, Tabor J, Bearman PS, Jones J, Udry JR. The National Longitudinal Study of Adolescent Health: Research Design. accessed 6 February 2007 Available at: <http://www.cpc.unc.edu/projects/addhealth/design>
19. Lloyd-Richardson EE, Papandonatos G, Kazura A, Stanton C, Niaura R. Differentiating stages of smoking intensity among adolescents: stage-specific psychological and social influences. *J Consult Clin Psychol.* 2002; 70:998–1009. [PubMed: 12182283]
20. U.S. Census Bureau. [assessed 20 February 2006] Poverty Thresholds in 1994, by Size of Family and Number of Related Children Under 18 Years. 2002. Available at: <http://www.Census.Gov/Hhes/Poverty/Threshld/Thresh94.Html>
21. Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. *Appl Psychol Meas.* 1977; 41:385–401.
22. StataCorp. Stata Statistical Software: Release 9. College Station, TX: StataCorp LP; 2005.
23. Rao, JNK.; Thomas, DR. Chi-squared tests for contingency tables. In: Skinner, CJ.; Holt, D.; Smith, TMF., editors. *Analysis of Complex Surveys.* New York: Wiley; 1989. p. 89-114. Ch. 4
24. Siddiqui O, Mott JA, Anderson TL, Flay BR. Characteristics of inconsistent respondents who have 'ever used' drugs in a school-based sample. *Subst Use Misuse.* 1999; 34:269–295. [PubMed: 10069753]
25. Steuber TL, Danner F. Adolescent smoking and depression: which comes first? *Addict Behav.* 2006; 31:133–136. [PubMed: 15922514]
26. Bauman KE, Ennett ST. Validity of adolescent self-reports of cigarette smoking. *Am J Public Health.* 1998; 88:309–310. [PubMed: 9491034]
27. Wills TA, Cleary SD. The validity of self-reports of smoking: analyses by race/ethnicity in a school sample of urban adolescents. *Am J Public Health.* 1997; 87:56–61. [PubMed: 9065227]
28. Fendrich M, Mackesy-Amiti ME, Johnson TP, Hubbell A, Wislar JS. Tobacco-reporting validity in an epidemiological drug-use survey. *Addict Behav.* 2005; 30:175–181. [PubMed: 15561458]

29. Hu MC, Davies M, Kandel DB. Epidemiology and correlates of daily smoking and nicotine dependence among young adults in the United States. *Am J Public Health*. 2006; 96:299–308. [PubMed: 16380569]
30. Caraballo RS, Giovino GA, Pechacek TF. Self-reported cigarette smoking vs. serum cotinine among U.S. adolescents. *Nicotine Tob Res*. 2004; 6:19–25. [PubMed: 14982684]

\$watermark-text

\$watermark-text

\$watermark-text

Table 1

Joint frequency of smoking stages (percent) across waves I and III among original population of size $n = 19\,895\,953$.*

Wave I smoking	Wave III smoking					Row total	
	Never smoked	Puffer	Light	Occasional	Regular		Ex-smoker
Never smoked	20.84	6.48	8.66	2.97	5.16	0.06	44.17
Puffer	2.20	3.54	3.66	1.21	2.34	0.02	12.98
Light	1.43	2.28	5.17	1.81	3.01	0.08	13.77
Occasional	1.41	0.86	3.83	3.35	6.78	0.08	16.31
Regular	0.44	0.05	0.19	1.49	7.62	0.02	9.81
Ex-smoker	0.19	0.09	0.64	0.59	1.43	0.02	2.96
Column total	26.51	13.29	22.15	11.43	26.34	0.28	100.00

* Figures in bold represent wave III responses incompatible with wave I self-reports.

Table 2

Mean (SD) of continuous variables by wave I smoking stage among original population of size $n = 11\ 107\ 426$.[†]

Covariates	Wave I smoking stage				
	Puffer	Light	Occasional	Regular	Ex-smoker
Age (years)					
Overall	15.81 (1.76)	15.98 (1.72)	16.14 (1.65)	16.93 (1.47)	16.34 (1.59)
Recant = no	15.72 (1.77)	15.89 (1.77)	16.13 (1.64)	16.93 (1.46)	16.33 (1.68)
Recant = yes	16.24 (1.78)	16.21 (1.71)	16.21 (1.63)	16.93 (1.58)	16.37 (1.37)
<i>P</i> -value	0.004**	0.033*	0.600	0.991	0.897
Income (thousands of dollars)					
Overall	42.56 (38.50)	44.68 (34.03)	48.26 (49.54)	40.18 (36.17)	44.56 (35.76)
Recant = no	43.49 (39.91)	46.06 (32.18)	49.88 (50.87)	40.64 (36.77)	44.50 (35.38)
Recant = yes	37.64 (29.50)	40.85 (38.45)	38.08 (38.71)	31.72 (20.68)	45.39 (40.96)
<i>P</i> -value	0.047*	0.143	0.004**	0.096	0.937
Depression (wave I)					
Overall	5.36 (3.89)	6.00 (4.22)	6.84 (4.51)	7.24 (4.71)	6.20 (4.56)
Recant = no	5.24 (3.92)	5.79 (4.06)	6.70 (4.50)	7.24 (4.75)	6.08 (4.59)
Recant = yes	5.96 (3.68)	6.59 (4.55)	7.75 (4.41)	7.22 (3.78)	7.43 (3.85)
<i>P</i> -value	0.020*	0.012*	0.004**	0.970	0.143
Depression (wave III)					
Overall	4.54 (4.08)	4.45 (3.92)	5.00 (4.27)	4.57 (4.34)	5.06 (4.77)
Recant = no	4.56 (4.08)	4.47 (3.92)	4.90 (4.24)	4.62 (4.36)	5.06 (4.60)
Recant = yes	4.45 (4.07)	4.40 (3.93)	5.63 (4.33)	3.52 (3.95)	5.10 (6.15)
<i>P</i> -value	0.754	0.803	0.058	0.171	0.974
Depression (wave III-I)					
Overall	-0.82 (4.49)	-1.55 (4.76)	-1.84 (5.10)	-2.67 (5.34)	-1.14 (5.21)
Recant = no	-0.68 (4.43)	-1.31 (4.70)	-1.79 (5.03)	-2.62 (5.39)	-1.02 (5.18)
Recant = yes	-1.51 (4.71)	-2.19 (4.88)	-2.12 (5.51)	-3.70 (4.27)	-2.33 (5.33)
<i>P</i> -value	0.032*	0.010*	0.535	0.201	0.271

* $P < 0.05$;

\$watermark-text

\$watermark-text

\$watermark-text

** $P < 0.01$

† P -values for mean differences between recanters and non-recanters calculated using survey estimation commands in Stata 9.0 (StataCorp. [22]).

Table 3

Prevalence of categorical variables by wave I smoking stage among original population of size $n = 11\ 107\ 426$.[†]

<i>Covariates</i>	<i>Wave I smoking stage</i>					
	<i>Puffer</i>	<i>Light</i>	<i>Occasional</i>	<i>Regular</i>	<i>Ex-smoker</i>	
Percentage female	Overall	0.50	0.48	0.48	0.49	0.48
	Recant = no	0.48	0.47	0.48	0.50	0.48
	Recant = yes	0.57	0.51	0.49	0.21	0.52
	<i>P</i> -value	0.047*	0.451	0.872	< 0.001***	0.723
Race/ethnicity	Overall	0.56	0.66	0.70	0.85	0.75
White		0.23	0.14	0.12	0.03	0.09
Black		0.13	0.12	0.11	0.07	0.10
Hispanic		0.04	0.03	0.02	0.01	0.02
Asian		0.05	0.05	0.04	0.05	0.03
Other		0.59	0.73	0.74	0.86	0.78
White	Recant = no	0.20	0.09	0.09	0.03	0.08
Black		0.13	0.11	0.10	0.07	0.08
Hispanic		0.04	0.02	0.02	0.01	0.02
Asian		0.05	0.04	0.05	0.04	0.04
Other		0.44	0.48	0.50	0.71	0.46
White	Recant = yes	0.37	0.26	0.28	0.08	0.22
Black		0.11	0.15	0.17	0.08	0.30
Hispanic		0.03	0.04	0.03	0.00	0.00
Asian		0.05	0.07	0.02	0.13	0.02
Other	<i>P</i> -value	< 0.001***	< 0.001***	< 0.001***	0.122	< 0.001***
Poverty rate	Overall	0.33	0.26	0.25	0.30	0.26
	Recant = no	0.31	0.23	0.22	0.29	0.25
	Recant = yes	0.40	0.36	0.43	0.46	0.39
	<i>P</i> -value	0.080	< 0.001***	< 0.001***	0.233	0.310
Education (percentage beyond high school)	Overall	0.60	0.62	0.62	0.56	0.65

\$watermark-text

\$watermark-text

\$watermark-text

<i>Covariates</i>	<i>Wave 1 smoking stage</i>				
	<i>Puffer</i>	<i>Light</i>	<i>Occasional</i>	<i>Regular</i>	<i>Ex-smoker</i>
Recant = no	0.62	0.65	0.64	0.56	0.68
Recant = yes	0.51	0.54	0.52	0.41	0.39
<i>P</i> -value	0.033 *	0.011 *	0.017 *	0.248	0.011 *

* $P < 0.05$;*** $P < 0.001$.† P -values for prevalence differences between recanters and non-recanters calculated using survey estimation commands in Stata 9.0 (StataCorp. [22]).

Table 4

Logistic regression modeling of recanting status by wave I smoking stage among original population of size $n = 8\ 678\ 962$.

Covariates	Wave I smoking stage											
	Puffer		Light		Occasional		Regular		Ex-smoker			
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI		
Intercept*	0.14	0.09, 0.21	0.22	0.16, 0.32	0.09	0.06, 0.14	0.06	0.03, 0.12	0.02	0.01, 0.10		
Age [†]	1.44	1.14, 1.82	1.17	0.97, 1.41	1.09	0.88, 1.35	0.98	0.53, 0.81	0.95	0.53, 1.72		
Female	1.27	0.84, 1.93	1.04	0.72, 1.49	1.15	0.77, 1.71	0.26	0.13, 0.55	2.14	0.52, 8.78		
Race												
Black	1.91	1.18, 3.10	3.43	2.27, 5.20	3.60	1.97, 6.57	2.54	0.38, 16.75	4.55	0.84, 24.61		
Hispanic	0.77	0.36, 1.62	1.92	1.11, 3.31	1.95	1.02, 3.73	0.07	0.02, 0.29	8.05	2.29, 28.27		
Asian	0.76	0.23, 2.48	4.42	1.70, 11.52	0.77	0.22, 2.62	0.19	0.03, 1.00	0.62	0.16, 2.39		
Other	1.81	0.83, 3.93	2.04	1.00, 4.15	0.73	0.30, 1.76	3.78	0.86, 16.55	‡	‡		
Poverty	1.28	0.80, 2.07	1.58	1.10, 2.28	1.93	1.36, 2.73	2.28	0.57, 9.12	1.57	0.40, 6.15		

OR = odds ratio; CI = confidence interval.

* Intercept denotes baseline odds.

[†] Age has been standardized to zero mean and unit variance.

[‡] Race = Other is a perfect predictor of non-recanting among ex-smokers; therefore, these observations were dropped and the model was fitted on the remaining ex-smokers.