

Longitudinal Attrition in a RDD Survey of Adolescents

Thanh Lê¹, Charles Carusi¹, and Susan Swain¹
Westat¹

Abstract

The survey on the Visual Media Influences on Adolescent Smoking Behavior—sponsored by Dartmouth College—interviewed adolescents aged 10-14 between June and October, 2003, identified from households selected using random digit dialing. Three waves of follow-up interviews were conducted every eight months thereafter with the last wave occurring between June and October, 2005. Data from the survey will be used to study how movie exposure, combined with factors such as adolescent temperament, peer affiliation and parent involvement, affects smoking behavior in adolescents. Since sample attrition occurs after each wave, weighting adjustments are implemented to data from each wave to minimize nonresponse bias. This paper examines sampling attrition in this RDD survey after each of the first three waves of data collection, describes methods employed to reduce attrition, and determines demographic factors that differentiate nonrespondents from respondents even after weighting adjustments.

Keywords: Longitudinal, nonresponse bias, RDD, sample attrition

1. Introduction

The Visual Media Influences on Adolescent Smoking Behavior Study is designed to measure the relationship between smoking among U.S. adolescents and their viewing of tobacco use in popular contemporary movies. This longitudinal study will estimate the fraction of adolescent smoking initiation that is attributable to movies over a time period that coincides with the critical stage when youngsters make decisions about smoking initiation (10-14 years of age). The main effects are measured by combining adolescents' self-reports of movie viewing and smoking behaviors, attitudes, and intentions, with content analyses of tobacco use in hundreds of the top grossing box-office hits and video rentals during each year of the study.

This study is the first nationally representative study of the effects of movies on adolescent smoking and drinking. It is an extension of an earlier study by the Dartmouth University Medical School. The preliminary findings from that study show a strong relationship between movie smoking and adolescent smoking behavior in an earlier study conducted with a sample of students who were recruited in-person, in New England schools.

The survey component of the current study consists of a national random-digit-dialing CATI survey to empanel 6,400

adolescents aged 10 to 14 into a two-year survey in which they are being interviewed four times about smoking behavior and attitudes, movies viewed, and relevant personal, familial, and social environmental factors. In each interview, respondents are asked whether they have seen 50 movies randomly selected from a master list of movies that changes for each wave, based on recent box office or video rental sales. Researchers at Dartmouth have content-coded each movie for smoking occurrences, images, and context. Based on the coding, each movie receives a smoking exposure score. Respondents are given a smoking exposure score based on the cumulative sum of the exposure scores of the movies they have viewed over time. The self-reports of smoking behavior and attitudes are compared to the subjects' exposure scores to measure the differential effect of viewing smoking in movies on adolescent smoking behavior, after controlling for socio-demographics, friend/sibling/parent smoking, school performance, personality characteristics, and parenting style.

In June 2003, the first in a series of four telephone interviews was conducted with a national sample of 10 – 14 year old adolescents who were recruited via telephone. The fourth and final wave of interviews was completed in mid-October 2005. The data collection period for each wave continued for approximately four months; interviews were re-fielded at eight-month intervals, allowing for the total two-year span between the first and the final interviews.

2. Study Design

2.1 Longitudinal Sample Design

In the baseline survey, telephone numbers were randomly sampled. Screening interviews were conducted to enumerate and sample eligible adolescents aged 10 to 14. Only one age-eligible adolescent was sampled per household. Parents in screened households with sampled adolescents were asked for permission to interview their child. Finally, the adolescent was interviewed. If the household failed to complete the screener, no adolescents were sampled. If the household completed the screener but the parent refused consent for their child's interview, no adolescents were interviewed. After the baseline, three waves of followup interviews have been conducted at eight-month intervals. All baseline respondents are eligible for each of the subsequent followups. Table 1 shows the sample yield after three waves of data collection. A very small number of baseline adolescent respondents became ineligible in subsequent waves because they died or were sampled in error in the baseline.

Table 1. Sample yield before the last followup survey

	Sample size	Percent	Baseline nonsmokers
Households screened	69,516		
Screened households with 10-14 year olds	10,025	14	
Parent interviews completed	7,525	75	
Baseline child interviews completed	6,522	87	5,829
Followup 1 interviews completed ¹	5,503	85	4,717
Followup 2 interviews completed ²	5,019	77	4,116

¹6,514 adolescents remained eligible for the first followup.

²6,512 adolescents remained eligible for the second followup.

Source: Visual Media Influences on Adolescent Smoking Behavior Study, 2003 and 2004.

2.2 Coverage and Precision Goals

The goal of the study was to complete approximately 6,400 interviews with adolescents in the baseline, and retain enough of the adolescents to complete 3,000 at the end of the last followup wave. Core analyses will focus on nonsmokers in the baseline. They will be classified as having had high or low exposure to smoking depictions in films and as having transitioned or not transitioned to experimenting with smoking by the end of the study. At the end of the study, there should be at least 2,200 baseline nonsmokers to achieve a power of 90 percent to detect an adjusted odds ratio of 1.4 (that high exposure adolescents would experiment with smoking over low exposure) using a 5 percent two-sided test. Based on the results of the second followup wave, the end sample of baseline nonsmokers will exceed the target by a wide margin, most likely due to an underestimate of the followup response rates and an increase in the RDD sample size over what was thought to be needed to ensure that the target would be met.

3. Sample Attrition and Response Rates

3.1 Methods for Reducing Attrition

Several methods to reduce attrition were used to maintain the recruited cohort. Before each new wave of data collection, parents were sent a letter reminding them of their families' participation in the study, informing them of the forthcoming call, and providing them with a toll-free number to call to report a new telephone number or to set up an appointment for an interviewer to call. During the interview process, interviewers always spoke with the parent first to request permission to speak with the child again. While talking with the parent, interviewers confirmed the household mailing address, collected their email address if applicable, and collected the name and telephone number of someone else who would know how to find them in the future, in case they moved. If interviewers were unable to locate the parent or

child at the number last used to interview them, then standard tracing procedures were employed. Tracers were given all previously collected address, telephone and alternative contact information collected. Tracing interviewers made calls to all available numbers and may also have contacted directory assistance or performed web searches as needed. Email addresses were first collected in the third wave of data collection. Once the interview was complete, the respondent and his/her parent were each sent a thank you letter. The youth letter contained a monetary incentive as thanks for their participation and to encourage participation in subsequent waves.

Not all nonresponses were due to parents or adolescents refusing to cooperate. Table 2 shows the reasons for nonresponse after the baseline. In the first followup, parent and child refusals account for 30 percent of all nonresponse. The rest was due to the fact that we could not locate them or could locate them but could not reach them by phone after numerous attempts. For the second followup, the parent and adolescent refusal rate decreased but more cases could not be located. The total number of nonresponses is 1,011 in the first followup and 1,493 in the second followup. The pattern of nonresponse for each followup is also shown with respect to the other followup. For example, of the 15 percent of parents who refused in the first followup, 30 percent subsequently participated in the second followup, and of the 12 percent of parents who refused in the second followup, 60 percent had responded in the first followup.

Table 2. Unweighted percent distribution of reasons for nonresponse and pattern of nonresponse

	First followup (%)	Second followup (%)
Parent refused	15	12
<i>In the other followup</i>		
Responded	5	7
Refused	5	4
Maximum number of calls	2	0.5
Not located	3	0.5
Adolescent refused	15	9
<i>In the other followup</i>		
Responded	5	4
Refused	6	4
Maximum number of calls	2	0.5
Not located	2	0.5
Maximum number of calls reached	30	30
Not located	40	49

Source: Visual Media Influences on Adolescent Smoking Behavior Study, 2003 and 2004.

3.2 Response Rates and Completion Rates

Response rates and completion rates are two ways to describe the outcomes of data collection activities. A response rate is the ratio of the number of units with completed interviews (for example, the units could be households, parents, or adolescents) to the number of units sampled and eligible for the interview. The response rate indicates the percentage of possible interviews completed, taking all survey stages into account. On the other hand, the completion rate measures the percentage of interviews completed for a specific stage of the survey.

Response and completion rates can be either unweighted or weighted. The unweighted rate, computed using the raw number of cases, provides a useful description of the success of the operational aspects of the survey. The weighted rate, computed by summing the weights (usually the reciprocals of the probability of selecting the units) for both the numerator and denominator, gives a better description of the success of the survey with respect to the population sampled since the weights allow for inference of the sample data (including response status) to the population. Both rates are usually not very different unless the probabilities of selection and the response rates in the categories with different selection probabilities vary considerably.

Table 3 shows the weighted completion rates and overall response rates for three waves of the study.

Table 3. Weighted completion and response rates

	Completion rate (%)	Over response rate (%)
Screened households	48	48
Baseline/parent	76	37
Baseline/adolescent	87	32
Followup 1/adolescent	83	31
Followup 2/adolescent	76	28

Note: For each followup study, the denominator of the completion rate is the number of baseline respondents still eligible at the time of the followup study.

Source: Visual Media Influences on Adolescent Smoking Behavior Study, 2003 and 2004.

It is useful to examine the completion or response rates for important subgroups as a general indicator of the potential for nonresponse bias. Clearly, large differences in these rates for subgroups increase the potential for nonresponse bias in the estimates.

Figure 1 shows the completion rates of the adolescent interviews by selected subgroups for the baseline (BL). Of the adolescents whose parents gave permission to participate in the study, 87 percent completed the interview. The highest completion rate is for White and the lowest is for American Indians and Alaskan Natives (AI). By household income, the highest rate is for high income households and the lowest rate for low income households. Figure 2 shows the

completion rates of the baseline respondents for the two followup waves, by the same selected subgroups, with the addition of two characteristics collected in the baseline. It is consistent for all subgroups that the rates for the second followup are lower than the rates for the first followup. By household income, there is a very strong positive correlation between response rates and level of income in the two followups. We have seen earlier that nonresponses were due not only to refusals but also to sample units who could not be contacted or located. Households with lower income tend to move around more, and this is likely reflected in their response rate. By smoking status, the four subgroups are smokers (S) and nonsmokers (NS) with each group further divided into those who have family and friends who smoke (F) and those who do not have such family and friends (NoF). The last group of nonsmokers (NS/NoF) has the highest completion rates.

4. Nonresponse Bias Analysis

4.1 Methods for Reducing Nonresponse Bias

Weights were developed for each completed adolescent interview in each wave of the study. Weights are necessary to compensate for the sampling of telephone numbers and of one adolescent per eligible household (in the baseline), and to reduce potential bias due to nonresponse and coverage errors (in all waves). The use of these weights is essential to produce estimates that are representative of the entire population of adolescents aged 10 to 14 in U.S. households at the time of the baseline survey. Although weighting adjustments are aimed at reducing bias, these adjustments typically introduce variation in the weights and may increase the variances of survey estimates. Care was taken in the development and implementation of the weighting methodology to balance the bias reductions against the potential increases in variance.

In the baseline, we adjusted for nonresponse at three levels: (1) by allocating a portion of the telephone numbers with unknown residential status as nonresponding residential numbers, (2) by doing a screener nonresponse adjustment, and (3) by adjusting for parent/adolescent nonresponse by raking to control totals from the 2001 American Community Survey. This step should reduce undercoverage bias due to sampling only households with landline telephones, and reduce biases from nonresponse. For each of the followup waves, we adjusted for adolescent nonresponse by raking to baseline population estimates. Raking is a multivariate poststratification to reduce attrition bias and the variability due to sample attrition (Brick 2003). We used four raking dimensions for the baseline: geographic region, gender by age, race/ethnicity by household income and by whether the family owns or rents their house. For the followup waves, we used the same dimensions and added three characteristics of the adolescent: smoking status at baseline, rebellious scores at baseline, and number of movies seen at baseline. These dimensions were selected because they were important analytic variables and were correlated with response rates.

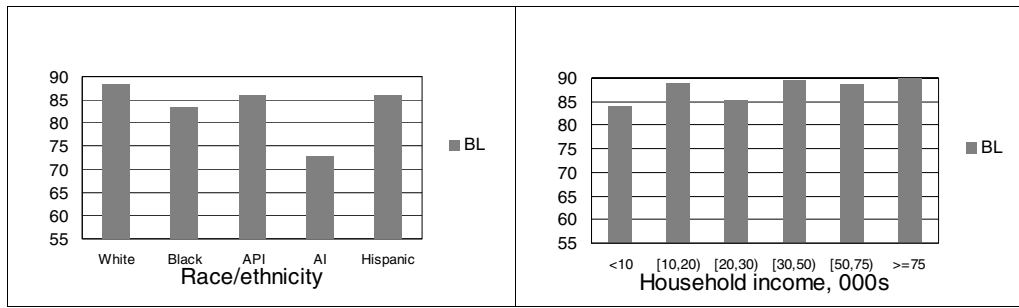


Figure 1. Weighted completion rates for baseline adolescent interviews

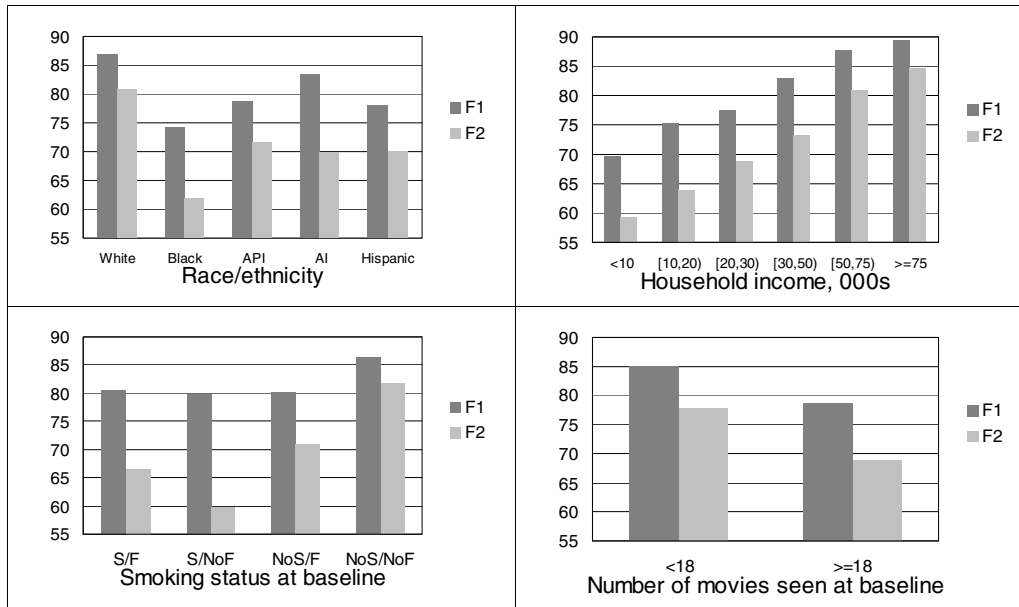


Figure 2. Weighted completion rates for baseline respondents in the followup rounds

Three procedures were used to evaluate potential nonresponse bias in the study: (1) comparing the estimates from the study with estimates from other sources, (2) comparing estimates from a later wave to estimates from the baseline, and (3) comparing estimates using adjusted weights to estimates using unadjusted weights.

4.2 Comparison of Estimates from the Study to Estimates from Other Data Sources

We compared the estimates from the second followup wave to estimates from three other sources: the most recent census (2000), the American Community Survey and Current Population Survey that are closest to the study (ACS 2003 and CPS 2003, respectively). As shown in Table 4, the estimates from the study compare reasonably well with the census data for most of the selected subgroups. They compared better with the ACS than with the CPS since the data were raked to the ACS.

4.3 Comparison of Second Followup Respondents to Baseline Respondents

This comparison examines the effect of nonresponse due to attrition. Estimates for 50 items from the baseline data (y_i) for the second followup respondents (using weights from the second followup not adjusted for nonresponse, w_i^*) were compared to estimates for the same items from the baseline respondents (using baseline final weights, w_i). This method gives a direct measure of nonresponse bias due to the additional nonresponse arising from the loss in the sample size since the baseline study. Bias was estimated as $\frac{\sum w_i^* y_i}{\sum w_i^*} - \frac{\sum w_i y_i}{\sum w_i}$. As shown in Figure 3 where data points are ordered by smallest (2%) to largest (99%) estimates, the relative bias, $\left(\frac{\sum w_i^* y_i}{\sum w_i^*} - \frac{\sum w_i y_i}{\sum w_i} \right) / \frac{\sum w_i y_i}{\sum w_i}$, varies between -7 and 17 percent.

Table 4. Percent distribution by selected subgroups

	Census 2000	ACS 2003	CPS 2003	Followup 2
Age at baseline				
10	21	20	20	21
11	20	20	20	20
12	20	20	21	20
13	20	20	20	19
14	20	20	19	19
Race/ethnicity				
Non-Hispanic				
White	63	61	62	62
Black	15	15	15	15
Asian/Pacific Islander	3	4	4	2
American Indian	1	1	1	0
Hispanic	15	17	16	16
More than one race	2	2	2	5
Highest education in HH				
Less than high school	20	8	8	15
High school	29	24	25	26
Some college	29	36	34	27
Bachelor or higher	22	32	33	31
Home tenure				
Own	71	71	72	70
Rent	29	29	28	30

Source: Visual Media Influences on Adolescent Smoking Behavior Study, 2003 and 2004.

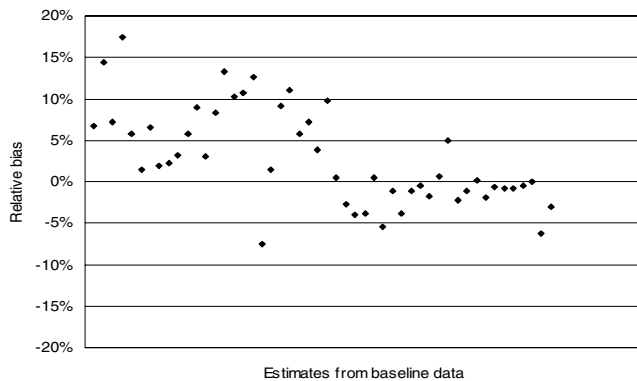


Figure 3. Relative bias if weights for second followup were not adjusted for followup nonresponse

4.4 Comparison of Unadjusted Estimates to Adjusted Estimates

To evaluate the raking method used for reducing attrition bias, we computed the same estimates above from the second followup respondents but this time using the final weights

from the second followup, i.e., weights that have been raked (w'_i). We compared these estimates to the baseline estimates using the baseline final weights (w_i as above). This time,

bias was estimated as $\frac{\sum w'_i y_i}{\sum w'_i} - \frac{\sum w_i y_i}{\sum w_i}$, and relative bias as

$$\left(\frac{\sum w'_i y_i}{\sum w'_i} - \frac{\sum w_i y_i}{\sum w_i} \right) / \frac{\sum w_i y_i}{\sum w_i}$$

Figure 4 shows that the relative bias has dropped significantly and the majority of the estimates center on 0. Note that the scale of the y-axis is different than in figure 3 and data points are still ordered by smallest (2%) to largest (99%) estimates. Relative bias now varies between -4 and 5 percent.

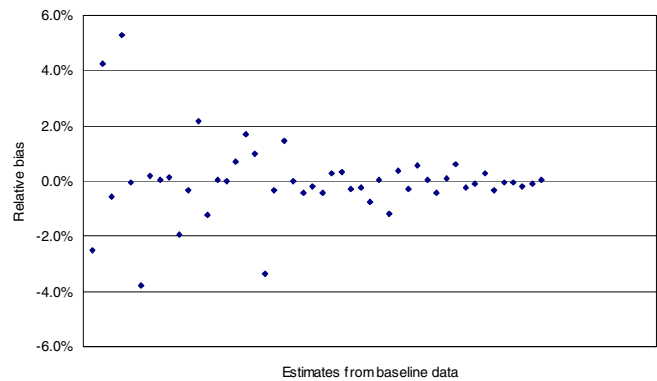


Figure 4. Relative bias after weights for second followup were adjusted for followup nonresponse

5. Summary and Conclusion

We presented in this paper the design of a study on whether viewing depictions of smoking scenes and images has an affect on adolescent smoking behavior, together with the results of sampling implementation. We presented the adjustment method used to deal with sample attrition. This method proved to be very effective in reducing attrition bias. We did not, however, examine potential bias in the domain estimates, and this could be pursued in the next research when the last wave of data collection of the study is completed.

6. Reference

Brick, J.M. Lê, T., and West, J. (2003). Dealing with movers in a longitudinal study of children. *Proceedings of the 2003 Statistics Canada Symposium on Challenges in Survey Taking for the Next Decade.*