Drug Use Prevention Data, Missing Assessments and Survival Analysis

Jennifer M. Bacik[1], Susan A. Murphy[1], and James C. Anthony[2]

[1]*Department of Statistics, Pennsylvania State University,*
326 Thomas Building, University Park, PA 16802
&
Department of Epidemiology,
[2]*Johns Hopkins University School of Hygiene & Public Health*
624 North Broadway, Baltimore, MD 21205

Summary

In prevention studies, it is often of interest to investigate the incidence of initial drug experimentation or other drug use milestones and their relationship to individual attributes such as the level of parental monitoring or rebelliousness. Thus survival analysis is the methodology of choice. Survival analysis methods deal efficiently with data from individuals who leave the study prematurely and do not return. However often individuals do return to the study. The application of survival analysis to a situation in which individuals miss assessments and later return is nonstandard. This paper examines the use of multiple imputation as a methodology for utilizing information from assessments following a missed assessment.

Key words: Multiple Imputation, Censoring

1. Introduction

In many drug use prevention studies, researchers are interested in the relationship between attributes such as the level of parental monitoring or the rebelliousness of children and the incidence of initial drug experimentation. Survival analysis is well suited for understanding such relationships as it provides estimates of how the incidence of initial drug experimentation varies with both time and attributes of a child.

In general, survival analysis methods efficiently handle *right-censored* initiation times. Due to time and economic pressures, longitudinal prevention studies are often set to be a certain length; the study starts and finishes at prespecified calendar times. Thus the time to initial drug experimentation is known only for those subjects who begin use before the end of the study. For the remaining subjects, all that is known is that the time until initial drug experimentation is greater than the study period. These subjects' event times are said to be right-censored at the end of the study period. In addition, some subjects may be unable to continue participating in the study and providing follow-up information. In this case one may only know that the subject did not experiment prior to leaving the study. These subjects' event times are right-censored at the loss to follow-up time.

In contrast to missing data caused by right censoring, standard survival analysis does not deal effectively with intermittent missed assessments. Some students may be absent from school the day the questionnaire is administered or they may move away from the area one year only to return the following year. Efforts are often made by the researcher to administer the questionnaire to those students not present at a particular assessment, but some of the efforts may be futile. Therefore, instead of collecting the year of the initial drug experimentation, the researcher may only know that initial use occurred within a range of years.

This paper investigates the use of multiple imputation in discrete time survival analysis and evaluates its effectiveness in using the information from future assessments. The focus is on discrete time survival analysis because in many prevention studies, the time of the event of interest is measured up to a given time interval. For example, the time of initial drug experimentation may be known only up to the year or age

of initiation. In discrete time analyses, the incidence or hazard of the event for a particular time interval is the conditional probability that a subject experiences the event of interest in the time interval, given that the subject did not experience the event prior to this interval. The collection of estimates, each corresponding to a time interval in the study, is called the estimated hazard function.

The next section illustrates how intermittent missed assessments combined with *prototypical* questions lead to knowledge of the initiation times only up to a range of time intervals. In section three, two methods for dealing with the imprecision in timing caused by intermittent missed assessments are discussed. In the last section, both methods are applied to each of two data sets and a comparison is made.

2. Prototypical questions

There have been many prevention studies designed and administered with the purpose of investigating and deterring onset of substance use in youth. It is common for these studies to collect information on a number of endpoints, or events of interest, via a set of prototypical questions. The questions generally focus on lifetime use, use in the recent past, and use in the very recent past. For example, the AAPT study [1] used the following three questions to gather information about a subject's cigarette use: "How many cigarettes have you smoked in your whole life?", "How many cigarettes have you smoked in the past month (30 days)?" and "How many cigarettes have you had in the past week (7 days)?" Possible responses ranged from "none" and "only one puff" to "more than one pack" or "more than 5 packs", depending on the time interval that was being addressed. Similarly, the Monitoring the Future study [2] included a general question about overall lifetime cigarette use; possible responses included whether the subject had been a regular user and whether he or she had quit smoking. This general question was followed by a specific question on frequency of use in the past 30 days, where the responses ranged from "not at all" and "less than 1 cigarette per day" to "2 packs or more per day".

AAPT and Monitoring the Future are just two of the many studies in the prevention field that address their hypotheses of interest with these prototypical questions. This line of questioning has particular ramifications for survival analysis, creating difficulty

when assessments are missed. For example, consider a study pertaining to time until initial cigarette use by children in 7th, 8th, 9th, and 10th grades. Suppose that the study asks the following question at the end of each school year, "Have you ever smoked a cigarette?" Consider a student who is present for the 7th, 8th and 10th grade assessments, but who is absent for the 9th grade assessment. Suppose the student tries smoking for the first time in 9th grade, but because of the missed assessment and the type of question asked, this information is unknown to the researcher. All that is known is that this student began use sometime during the 9th or 10th grade. It is impossible to fill in the initiation time due to the formulation of the question.

An alternate mode of data collection is illustrated by a epidemiological study of urban-dwelling children initially aged 8-12 years in Baltimore, Maryland. One purpose of the study was to test the hypothesis that parental monitoring and supervision is associated with a reduced risk of drug use in the elementary school years. Detailed discussions of the data and analyses are given in Chilcoat, Dishion and Anthony [3], Kellam and Anthony [4] and Chilcoat and Anthony [5]. This study was designed with the intention of estimating the incidence of tobacco sampling and thus, in addition to asking prototypical questions, the children were asked to give the age at which they first sampled tobacco. To minimize recall error, the age at first tobacco use was designated to be the reported age on the first occasion of reporting such use. In this study, if a child who has reported no prior tobacco use at age 9, misses an assessment at age 10 but is present at age 11 and reports prior tobacco use, the researcher will know at the time of the age 11 assessment whether the child initiated use at age 10 or age 11.

Since many longitudinal drug use studies are not designed for the purpose of studying hypotheses concerning incidence or timing of drug use milestones, the prototypical questions listed above are common. Given this setting and the fact that after data collection it is often of interest to test hypotheses concerning timing of drug use milestones, it is important to provide methodology for dealing with the consequences of the combination of intermittent missed assessments and prototypical questions. The remainder of this paper compares the use of multiple imputation with an alternate simplistic method of dealing with intermittent missed assessments.

3. Two methods of dealing with missed assessments

Both methods make the assumption that the missing assessments are missing at random (MAR) [6]. In this setting this means that the chance that a subject misses an assessment can only depend on the measured covariate pattern and the subject's past responses. That is, given the measured covariates and the subject's past responses, the chance of the subject missing the assessment should not depend on whether the subject initiated drug use in the time interval immediately prior to the assessment. For example, consider Panel 2 of the AAPT study [1] in which students were given a questionnaire every year from seventh grade through tenth grade for a total of four assessments. Suppose the researcher wishes to determine whether the time until initiation of smoking differs by gender and plans to fit a regression model including gender as a covariate. Split the students into two groups according to their gender and consider only students who did not initiate prior to eighth grade. If within each gender, it can be assumed that the timing of initiation among those who miss the eighth grade assessment is approximately equal to timing of initiation among those who do not miss the assessment, then the MAR assumption holds for missed eighth grade assessments. If the timing of initiation in the two groups cannot be assumed to be approximately equal, then the MAR assumption is violated. An example of a MAR violation is if the boys who initiate smoking tend to miss assessments more frequently than boys who do not initiate smoking. Or stated in another way, the boys who miss assessments initiate smoking earlier than the boys who do not miss assessments. This issue is discussed in greater detail below.

3.1 Artificial Censoring

An effective but wasteful method used to handle intermittent missed assessments is to censor the event time prior to the first missed assessment. In this method, any information gained from later assessments is discarded. Assuming MAR, this approach does not introduce bias but it certainly does not make efficient use of the data. Bias is not introduced because one discards later assessments without regard to whether the child is later observed to initiate. However from power considerations, assessments after a missed assessment may be quite informative. For example, in a study comparing

the incidence of tobacco initiation between poorly monitored and well monitored children, it may be that most of the children missing assessments are precisely the poorly monitored children. In this case, censoring the event time prior to the first missed assessment will primarily discard all later tobacco use information on the poorly monitored children, thus reducing the power of a test to detect differences in initiation rates between the two groups of children. Thus in order to maximize power in a study with intermittent missed assessments, it is necessary to find an appropriate way of using the information from the subjects who leave the study but then later return.

Consider Panel 2 of the AAPT study as described above. Suppose a student was present for the 7th and 8th grade assessments, and by his responses to the question regarding lifetime cigarette use, it is determined that at neither of these time points had he previously used cigarettes. Therefore, information about his cigarette-use status is available for the first interval, the time between the 7th and 8th grade assessments. Now suppose that the student missed the 9th grade assessment but was present at the 10th grade assessment and that his response to the question indicates that prior to the 10th grade assessment, he had tried smoking. Even though the researcher may conclude that the student began using cigarettes either between the 8th and 9th grade assessments or between the 9th and 10th grade assessments, the researcher disregards this information and censors the student at the end of the first interval (at the 8th grade assessment). That is, the researcher acts as if he/she only knows that the student did not initiate up to the time of the 8th grade assessment. This approach is inefficient because it does not make use of all the information the researcher has about each student – in this case, the information collected at the 10th grade assessment.

In order to ensure that the MAR assumption holds, it is important that the researcher include in the regression model all covariates which may explain both the timing of initiation and why subjects miss assessments. For example, it may be that the boys who miss assessments tend to initiate smoking earlier than the boys who do not miss assessments. However, if the boys are further cross classified into well monitored and poorly monitored boys then it may be plausible to assume that poorly monitored boys who miss assessments tend to initiate smoking around the same time as the poorly monitored boys who do not miss assessments. Suppose a similar assumption

can be made for the well monitored boys. In this case both gender and monitoring level should be included as covariates in the survival analysis regression model.

3.2 Multiple imputation

Multiple imputation [7, 8] is a Bayesian procedure which solves a missing-data problem by replacing each missing observation in the data set by M plausible values. For example, consider again the student from the AAPT study who is present for the 7th and 8th grade assessments, misses the 9th grade assessment, but is present for the 10th grade assessment. Recall that the student reports no prior use of cigarettes at the 7th and 8th grade assessments, but at the 10th grade assessment, he does report prior use. Because of the missed assessment, it is not known if this student began use between the 8th and 9th grade assessments, or between the 9th and 10th grade assessments. Using an imputation model for the distribution of smoking initiation given the observed data, the time of initiation is randomly imputed as one of the two intervals. The imputation process transforms the data set with missing values into a complete-data version. Therefore, repeating the process M times results in M complete data sets. It may be the case that in the first of the M imputations, the time of initiation is imputed as the time between the 8th and 9th grade assessments (interval 2), whereas in the second imputation, the time of initiation may be imputed as the time between the 9th and 10th grade assessments (interval 3). Each of the M data sets is then analyzed using a survival analysis model and the results of the M analyses are subsequently combined to arrive at one overall estimate for each parameter of interest. The resulting estimates incorporate two sources of variability: the within-imputation variation, which is a measure of ordinary sampling variation; and the between-imputation variation, a measure of the uncertainty due to the missing data. In many cases, the number M of imputations needed for valid estimates may be as small as 5 [8]. When using multiple imputation, it is important to distinguish between the imputation model and the analysis model. The survival analysis model may not be the same as the imputation model but the validity of multiple imputation requires that the multiple imputation model be at least as complex as the survival analysis model. For example, the imputation model may contain more covariates than the survival analysis model but the survival analysis model may not contain more covariates than the imputation model.

After the imputation procedure, censoring occurs only if the subject did not initiate smoking prior to leaving the study. This, of course, is due to the fact that there are no intermittent missed assessments in the imputed data sets. Further nontechnical discussion of the technique, assumptions under which multiple imputation is valid, the rules for combining the estimates and standard errors, along with computer code, are given by Bacik [11]. For a more complete discussion see Schafer [8].

If the MAR assumption holds, the multiple imputation procedure uses all available information from all subjects without introducing bias into the estimation of the hazard function. As in the previous method, in order for the MAR assumption to hold, all covariates which may explain both the timing of initiation and why subjects miss assessments must be included in the imputation model. As before it may be that the boys who miss assessments tend to initiate smoking earlier than the boys who do not miss assessments. However, suppose the boys may be further cross classified into well monitored and poorly monitored boys. If it is plausible to assume that poorly monitored boys who miss assessments tend to initiate smoking around the same time as the poorly monitored boys who do not miss assessments, then the MAR assumption holds. In this case both gender and monitoring level should be included as covariates in the imputation model.

4. A comparison of the two methods

The two methods of handling intermittent missed assessments will be compared on data collected from the epidemiological study of urban-dwelling children described earlier. The sample for this study was drawn from a group of several thousand children originally recruited for a larger study of the first graders enrolled in 19 urban elementary schools in Baltimore between 1985 and 1986. 1610 children were first interviewed in spring 1989, when they were in grades 3 or 4, and then were reinterviewed each spring from 1990 through 1994 [3, 5].

At the first interview in spring 1989, the level of parental monitoring was assessed by use of a 10-item scale. Each child was asked questions about supervisory rules and the level of supervision provided by parents and other caretakers. The responses

to the items were summed to form a continuous measure and then the subjects were categorized according to quartiles of the resulting scale.

A subset of 622 girls had not initiated tobacco sampling at the first interview. For the purposes of comparing the two methods, the focus of this paper is on the girls and on relating the level of parental monitoring to initiation of tobacco sampling. Only two levels of parental monitoring are used; the lowest quartile (low parental monitoring) versus the upper three quartiles (high parental monitoring). Of the 622 girls, 142 girls belonged to the lowest quartile of parental monitoring. Recall that the children were asked to give the age at which they first sampled tobacco. Thus missing assessments do not cause imprecision in the age at first use. In addition, each year the absence/presence of the girl at the assessment is recorded.

The analyses to follow are simplified analyses that include only the level of parental monitoring. A complete analysis would control for other student attributes such as level of antisocial behavior. For more complete analyses see Chilcoat, Dishion and Anthony [3], Kellam and Anthony [4] and Chilcoat and Anthony [5].

Four data sets are created from the original subset of 622 girls. The first data set corresponds to the original subset (call this the complete-true data). Because of the way the girls were asked about their initiation of tobacco use, there are no imprecise initiation times due to intermittent missed assessments. The second data set corresponds to the data if instead of asking the girls for the age of first use, the girls were asked if they had ever used tobacco (a prototypical question). This second data set suffers from imprecise initiation times due to intermittent missed assessments. This is called the classical-true data set. To see how the complete-true data set corresponds to the classical-true data set, consider a girl who was present at ages 9, 10, 13 and 14 but was absent at ages 11 and 12. At ages 9 and 10, she reports that she has never tried tobacco. At age 13, she reports that she first tried tobacco at age 12. So the complete-true data set has that she initiated tobacco sampling at age 12. In contrast, the classical-true data set has that she initiated tobacco sampling at either age 11 or age 12.

The third data set is identical to the complete-true data set except that 8 girls who are in the low parental monitoring group and who had not initiated are recorded as initiating at age 12. This data set is called the complete-test data set. The purpose of

this data is to provide an comparison of the two methods when future assessments are informative. As before, a classical-test data set can be formed by acting as if the girls are only asked if they have ever initiated tobacco use rather than at what age they first used tobacco. This forms the fourth data set.

A discrete time survival analysis model may be fit to the two complete data sets using PROC LOGISTIC in SAS (see [9, 10]). Recall that the hazard is the conditional probability of a girl initiating tobacco sampling at a particular age, given that the girl had no prior use. Denote the hazard probability for girl i at age j as h_{ij}. Including a main effect of age, and an age-varying effect of parental monitoring, leads to a logistic discrete-time hazard model

$$log\frac{h_{ij}}{1-h_{ij}} = \alpha_j + \beta_j Z_i,$$

where α_j represents the baseline profile of risk at age j, Z_i is an indicator of girl i's level of parental monitoring at the first assessment(with $Z = 1$ for a low level and $Z = 0$ otherwise) and β_j is the regression coefficient for the level of parental monitoring. The effect of parental monitoring is allowed to vary by age, resulting in a parental monitoring regression coefficient for each age. Positive regression coefficients indicate that low parental monitoring level is associated with an increased incidence of tobacco initiation as compared to a high parental monitoring level. The following p-values are one-sided.

Tables 1 and 2 contain estimators of the regression coefficients for a survival analysis on the complete-true data and for a survival analysis on the artificially censored classical-true data, respectively. The estimators based on the complete data are unbiased. Theoretically the estimators based on the artificially censored classical data should also be unbiased. This means that the difference between the complete data estimator and the artificially censored classical data estimator of β_i should, on average, be zero. To check if this is indeed the case, a 95% confidence interval centered around the difference between the two estimators of β_i is constructed for each i. If there is no evidence of bias in the estimators based on the artificially censored classical data, then the confidence intervals should contain zero. Note that all of the confidence intervals in Table 2 contain zero. Additionally the standard errors did not substantially increase from Table 1 to Table 2. This means that in the true data, assessments occurring after

Table 1: Results from Complete-True Data

Parameter	Parameter Estimate	Standard Error	p-value
β_9	1.06	0.727	0.07
β_{10}	-0.06	0.479	0.55
β_{11}	0.60	0.422	0.08
β_{12}	-0.10	0.512	0.58
β_{13}	1.24	0.475	0.005

Table 2: Results using Artificial Censoring on the Classical-True Data

Parameter	Parameter Estimate	Standard Error	p-value	C.I. for 0
β_9	1.13	0.730	0.06	(-0.37,0.23)
β_{10}	-0.16	0.577	0.61	(-1.23,1.43)
β_{11}	0.64	0.455	0.08	(-0.40,0.31)
β_{12}	0.06	0.526	0.46	(-0.43,0.10)
β_{13}	1.56	0.623	0.006	(-1.28,0.63)

a missing assessment are not very informative. As a result, regression coefficients found to be significant in an analysis of the complete data are also found to be significant when all assessments following a missing assessment are discarded (i.e., the artificial censoring method).

In Table 3, ten multiple imputations are used to produce complete data sets which are then analyzed by discrete time survival analysis by use of the hazard model given previously. The imputations were carried out using software written by J.L. Schafer and available at **http//www.stat.psu.edu/~jls/misoftwa.html#top** on the world wide web. The multiple imputation model is a multinomial model (see [8, chapter 7]) which does not constrain the interactions between the two levels of parental monitoring, the two possible responses at age 9, the two possible responses at age 10,..., and the two possible responses at age 14 except to disallow impossible response patterns. For example the imputation model gives probability zero to an initiation at age 10 and

Table 3: Results using Ten Multiple Imputations on the Classical-True Data

Parameter	Parameter Estimate	Standard Error	p-value	C.I. for 0
β_9	1.02	0.727	0.08	(-0.47,0.56)
β_{10}	-0.32	0.565	0.72	(-0.46,0.99)
β_{11}	0.64	0.418	0.06	(-0.37,0.27)
β_{12}	0.02	0.505	0.48	(-0.47,0.23)
β_{13}	1.41	0.575	0.007	(-0.74,0.39)

then a subsequent response at age 11 of no previous initiation.

Because, as mentioned earlier, the assessments following the missing assessment are not very informative, the results using multiple imputation are essentially equivalent to those found by discarding the later assessments. Note also that the 95% confidence intervals for zero indicate no bias in the estimated regression coefficients.

In general, assessments occurring after a missed assessment will be informative if deleting the later assessments leads to a loss of power. Since MAR is assumed, deleting the later assessments should not lead to bias. It is difficult to precisely state when deletion of later assessments leads to power loss. Certainly it is necessary that the population regression coefficient for parental monitoring at a particular age is positive and girls must miss assessments at or prior to the age and be present later. In particular if the proportion of girls who are at risk of initiating tobacco sampling and who miss the assessment at the age is greater for the low parental monitoring than for the high parental monitoring group, one may expect to see a loss of power.

One case in which it is clear that a loss of power should result by artificial censoring is when a girl, who is absent at one or more assessments, returns, reports no prior tobacco use and then later is observed to initiate. The test data present just such a scenario.

Recall that the test data were obtained by randomly choosing eight girls in the low monitoring group who were present at age 12 but did not initiate tobacco sampling during the study duration, and by changing their records so as to indicate initiation at age 12. As a result, the survival analysis of the complete-test given in Table 4 indicates

Table 4: Results from Complete-Test Data

Parameter	Parameter Estimate	Standard Error	p-value
β_9	1.06	0.727	0.07
β_{10}	-0.06	0.479	0.55
β_{11}	0.60	0.422	0.08
β_{12}	0.93	0.375	0.006
β_{13}	1.33	0.476	0.003

a significant positive regression coefficient at age 12. Four of the girls had missed an assessment prior to age 12. Thus when artificial censoring is used to deal with the missed assessments, these girls are not seen to initiate. This causes the significance of the regression coefficient to decrease (the p-value of the regression coefficient increases from .006 in Table 4 to .053 in Table 5).

The artificial censoring method must delete ALL information after a missed assessment so as to not produce bias in the initiation rates. For example suppose one girl reports no initiation at age 9, misses an assessment at age 10, reports no initiation at age 11 and then initiates at age 12. If this girl is included in the analysis by filling in a response of no initiation at age 10, then one would be differentially censoring girls based on whether they report prior initiation at age 11. That is, girls who report prior initiation at age 11 are not included but girls who report no prior initiation are included. This would produce bias in the estimation of the initiation rates at age 10.

By using multiple imputation all information from the eight girls may be used and as a result the p-value for the regression coefficient at age 12 indicates a highly significant effect as can be seen in Table 6. As in the analysis of the classical-true data, the imputation model is a general multinomial model which is not subject to any constraints beyond the specification of impossible response patterns.

Table 5: Results using Artificial Censoring on the Classical-Test Data

Parameter	Parameter Estimate	Standard Error	p-value	C.I. for 0
β_9	1.13	0.733	0.06	(-0.36,0.24)
β_{10}	-0.16	0.577	0.61	(-1.20,1.42)
β_{11}	0.64	0.455	0.08	(-0.40,0.30)
β_{12}	0.70	0.434	0.053	(-0.25,0.72)
β_{13}	1.63	0.624	0.005	(-1.25,0.66)

Table 6: Results using Ten Multiple Imputations on the Classical-Test Data

Parameter	Parameter Estimate	Standard Error	p-value	C.I. for 0
β_9	0.95	0.721	0.09	(-0.27,0.50)
β_{10}	-0.24	0.574	0.66	(-0.46,0.83)
β_{11}	0.62	0.435	0.08	(-0.39,0.34)
β_{12}	0.91	0.369	0.007	(-0.23,0.28)
β_{13}	1.54	0.564	0.003	(-0.78,0.37)

5. Discussion

Intermittent missed assessments are unavoidable in longitudinal drug use prevention research. This problem necessitates finding an appropriate way of using the information gained from the subjects who leave the study but then later return. Multiple imputation is an appropriate methodology to consider for this situation since it uses all of the information gained from future assessments.

In many situations, the usual tools survival analysis provides for intermittent missed assessments, i.e. censoring prior to the first missed assessment, may be adequate. This depends on the amount of information contained in the later assessments. In the true data set there was little information added by assessments following a missed assessment but in the test data later assessments contained much information. It is clear that the test data could have been constructed so as to make the later assessments even more informative. For example, the test data could have contained even more girls who miss an early assessment and return for later assessments. In general, in any given setting it is difficult to know when one is at risk of losing power by using the artificial censoring method. Certainly when the future assessments are not informative, proper use of multiple imputation does not hurt the analysis. Yet if a data set contains many subjects who miss one or more assessments early in the study and later return, one is at risk of losing power by using the artificial censoring method. Thus in studies in which power to detect alternatives is of paramount importance, multiple imputation is indeed promising.

Another advantage of multiple imputation is that it may be used when an individual's covariates are missing. In the example treated here, the covariate, parental monitoring was always known but this was not necessary. Both continuous and discrete covariates are handled by J.L. Schafer's software.

Multiple imputation is also useful in randomized trials. Consider a comparison of two programs aimed at delaying initiation of smoking. Subjects are randomized to the two intervention programs. As is usually the case with longitudinal studies, there are missing responses for some of the subjects at some of the assessments. Now level of commitment at the end of program administration may help explain the missingness patterns and thus ensure that the MAR assumption holds. For example, it may be

that those who were less committed to the program are more likely to miss assessments than those who were more committed. However, level of commitment may be a venue via which the intervention program causally delays smoking initiation. Level of commitment may be a considered a response to the intervention programs. Therefore, the researcher should not include level of commitment in a regression model intended to assess intevention impact. If the artificial censoring method is to be used, this presents a dilemma. Level of commitment should be included in any regression model so as to ensure MAR holds but to make a causal inference concerning intervention, level of commitment should not be included in the regression of smoking initiation on intervention program.

Use of multiple imputation allows both of the above goals (ensuring that missing at random holds and determining the effect of intervention program) to be met because it differentiates between the *imputation model* and the *analysis model*. Including the additional variables, such as level of commitment, in the imputation model helps to ensure that the MAR assumption holds. Then, when the imputed data are analyzed to determine whether the intervention program has an effect on delaying time until initiation of smoking, level of commitment does not need to be included as a covariate in the analysis model. In this way, using multiple imputation in a randomized trial setting helps to account for problems that may develop if it is thought that whether a subject misses an assessment is related to post-treatment variables.

References

[1] W. B. Hansen and J. W. Graham. (1991) Preventing alcohol, marijuana, and cigarette use among adolescents: peer pressure resistance training versus establishing conservative norms. Preventive Medicine **20** pg. 414–430.

[2] J.G. Bachman, K.N. Wadsworth, P.M. O'Malley, L.D. Johnston and J.E. Schulenberg. (1997) *Smoking, Drinking, and Drug Use in Young Adulthood: The Impacts of New Freedoms and New Responsibilities.* Research Monographs in Adolescence, Lawrence Erlbaum Associates, Inc. Mahwah, New Jersey.

[3] H.D. Chilcoat, T.J. Dishion, and J.C. Anthony. (1995) Parent monitoring and the incidence of drug sampling in urban elementary school children. Am J Epidemiol **141** pg. 25–31.

[4] S.G. Kellam and J.C. Anthony. (1995) Impact of the good behavior game on the risk of initiation tobacco use. *under revision*

[5] H.D. Chilcoat and J.C. Anthony. (1996) Impact of parent monitoring on initiation of drug use through late childhood. J Am Acad Child Adolesc Psychiatry **35** pg. 91–100.

[6] D.B. Rubin. (1976) Inference and missing data. Biometrika **63** pg. 581–592.

[7] D.B. Rubin. (1987) *Multiple Imputation for Nonresponse in Surveys* John Wiley & Sons, New York.

[8] J.L. Schafer. (1997) *Analysis of Incomplete Multivariate Data.* Chapman & Hall, London.

[9] J.D. Singer and J.B. Willett. (1993) It's about time: using discrete-time survival analysis to study duration and the timing of events. J Educ Statist **18** pg. 155–195.

[10] SAS Institute Inc. (1990) *SAS/STAT User's Guide, Version 6.* fourth edition.

[11] J.M. Bacik (1997) Missing response times in survival analysis. Technical Report #97-16, Center for the Study of Prevention through Innovative Methodology, The Pennsylvania State University.

Acknowledgements

This research was supported by Grant #1 P50 DA10075 from the National Institute on Drug Abuse.