Patterns of Nonresponse for Key Questions in NSDUH and Implications for Imputation

Peter Frechtel[1], Elizabeth Copello[2]

[1] Statistics and Epidemiology, RTI International, 701 13th Street, NW, Suite 750, Washington, DC 20005

[1] Statistics and Epidemiology, RTI International, Research Triangle Park, NC 27705

Abstract

The idea of using "soft nonrespondents" to represen t "hard nonrespondents" is not new to survey research. Callbacks are often used to adjust for nonresponse in surveys. The goal is to control nonresponse bias by assuming that the hard nonrespondents are more similar to the callback respondents than they are to the original respondents.

The National Survey on Drug Use and Health (NSDUH), an annual nationwide survey involving approximately 70,000 subjects per year, does not make use of callbacks. However, for several key questions in the NSDUH, follow-up questions, or "probes," are presented to subjects who entered a response of "don't know" or "refused" to the original questions. The probes are intended to increase item response rates by simulating an actual interviewer. The probe respondents can be considered soft nonrespondents, and the subjects that answer neither the original question nor the probe can be viewed as hard nonrespondents.

In NSDUH imputation procedures, subjects who responded to the original question are treated exactly the same as subjects who responded to the probes. This may not be the best approach. An earlier study which used data from the 2000 NSDUH showed some evidence that "original respondents" differ from "probe respondents," especially those probe respondents who refused to answer the original question.

The analysis from the earlier study will be expanded to include data pooled from the 2000-2005 surveys. The values of auxiliary variables will be compared between original respondents, probe respondents, and nonrespondents, to see whether the nonrespondents resemble the probe respondents more than the original respondents. This comparison will help estimate the potential bias caused by failing to distinguish between original respondents and probe respondents in imputation. Finally, ways to incorporate this information into the imputation procedure currently used in NSDUH, called Predictive Mean Neighborhoods, will be discussed.

KEY WORDS: nonresponse bias, item nonresponse, imputation

1. Introduction

The National Survey on Drug Use and Health (NSDUH) is an annual nationwide survey involving about 67,500 completed interviews per year. In the NSDUH, unit nonresponse has historically been handled by weighting, and item nonresponse has historically been handled using imputation. This approach is common for large-scale surveys (Lohr, 1999, p. 272).

As in most large-scale surveys, various methods are used to mitigate nonresponse bias in the NSDUH. Both the weighting procedures and the imputation procedures attempt to mitigate nonresponse bias by using information from auxiliary variables. In the weighting procedures, the nonresponse adjustment attempts to compensate for differential propensities to respond among different demographic groups, for example (Chen et al., 2006). In the imputation procedures, often responses to other questions in the questionnaire are used as auxiliary information. For example, numerous person-level demographic variables are used in the imputation of variables related to drug use; and often information about other drugs is used (Grau et al., 2006; see esp. Appendix F).

So both the weighting and imputation procedures attempt to mitigate nonresponse bias using auxiliary variables. However, sometimes the nonresponse is not only related to the auxiliary variables: it is also related to the variable of interest itself. For example, if those who used marijuana in the past month are less likely to respond to the marijuana questions than those who did not use marijuana in the past month, and past month marijuana use itself cannot be perfectly described by all the auxiliary information, then not all of the nonresponse bias will be eliminated by the auxiliary information. It is not easy to determine whether this sort of nonresponse pattern exists in the data, since the actual responses of the item nonrespondents are unknown!

There are several common ways to assess nonresponse bias at the unit level. One common approach is double sampling, or two-phase sampling (Lohr, 1999; Thompson, 1992). This method involves the use of callbacks; a subsample of the nonrespondents are re-contacted, usually using a different (and usually more expensive) mode, and this subsample is used to represent all the unit nonrespondents. NSDUH does not ordinarily use callbacks. One study which used callbacks for the 1990 NSDUH found no evidence of serious nonresponse bias in Washington, DC (Caspar, 1992). Though the double sampling method is a good way to adjust for nonresponse bias, it is often costly.

The unequal weighting effect might also more than cancel out the benefits derived from the correction for nonresponse bias, with respect to the mean square errors of the statistics of interest (Singh, Iannacchione, & Dever, 2003).

Methods other than double sampling have been attempted to assess nonresponse bias at the unit level, including the use of record-of-call data (Wang, Murphy, Baxter, & Aldworth, 2005; Biemer & Wang, in process). Wang, Murphy, Baxter, & Aldworth (2005) found some evidence that, in the 2004 NSDUH, subjects who were interviewed on the first day they were contacted were lifetime users of drugs less often than subjects who required additional call days. Biemer & Wang (in process) are attempting to model outcome variables as a function of the level of effort required to get a completed interview, for use in the weight adjustment for unit nonresponse.

The common goal of the studies mentioned in the above paragraphs is to somehow use "soft nonrespondents" to represent "hard nonrespondents". That is, the subjects who respond reluctantly are used to represent the subjects who do not respond at all.

The same ideas could be applied at the item level. Although the mechanism for item nonresponse may be different from the mechanism for unit nonresponse, soft nonrespondents can still be used to represent hard nonrespondents. In order to mitigate nonresponse bias at the item level, and in order to increase the item response rate, probes were added to key questions in the 2000 NSDUH questionnaire. The success of the probes in increasing the item response rate in the 2000 NSDUH was examined closely by Caspar, Penne, & Dean (2005). The authors also found some evidence that, for certain drugs, the subjects who responded to the original questions were less likely to be lifetime users than the subjects who responded to the probes. This latter point suggests that perhaps the imputation method should be modified so that the soft nonrespondents (i.e., the probe respondents) are used to represent the hard nonrespondents (i.e., those subjects who declined to answer both the original question and the probe). In the rest of this paper, the nonresponse bias analysis begun by Caspar, Penne, and Dean (2005) will be expanded to include data from the 2000-2005 NSDUHs, and the use of the probes to mitigate bias due to item nonresponse will be examined more thoroughly.

2. Types of Item Nonresponse

Little and Rubin (1987) describe three types of item nonresponse:

- Missing Completely at Random (MCAR): the set of item nonrespondents is a simple random subsample of the set of all subjects in the sample. The missingness is not related to any auxiliary variables, nor is it related to the outcome variable. In this case, there is no nonresponse bias and imputation is not considered necessary as a correction for it.
- Missing at Random (MAR): the nonresponse is related to auxiliary variables, but not to the outcome variable itself. For example, one question in the NSDUH asks the subject whether he or she has ever used marijuana in his or her lifetime. If the nonresponse for this question depends only on auxiliary variables such as age, race, gender, etc., then the nonresponse would be considered MAR. Imputation is considered necessary in this case. Ignoring the item nonrespondents would cause nonresponse bias.
- Not Missing Completely at Random (NMAR): the nonresponse is related to the outcome variable itself. In this case, imputation involving only the auxiliary variables would not completely correct for nonresponse bias.

Most imputation methods, including the one used for the drug outcome variables in the NSDUH, are designed to handle the MAR case. It is difficult to determine whether the data are truly NMAR, since the actual responses of the item nonrespondents are unknown. However, one way to assess whether the data are NMAR is by using some of the methods mentioned above, including double sampling. If the probe respondents differ from the original respondents, the data may be NMAR.

If the data are NMAR, then perhaps the probes could be used to enhance the imputation methodology to reduce bias due to item nonresponse.

3. Predictive Mean Neighborhoods

The imputation method used in the NSDUH for the outcome variables discussed in this paper is called Predictive Mean Neighborhoods (PMN). The theoretical underpinnings of the method are described in Singh, Grau, & Folsom (2001). The application of PMN to the NSDUH is described in detail in Grau et al. (2006).

For the purposes of this paper, it suffices to say that PMN is a model-based hot-deck imputation method. A neighborhood of potential donors is formed for each recipient (i.e., item nonrespondent), and one donor is selected from the neighborhood to be the final donor. The neighborhood is formed via constraints. One of the constraints requires the donor to have predicted value(s) close to the recipient's predicted value(s), where the predicted value(s) are calculated from regression models. These predicted values will be used to evaluate whether the current imputation method is successfully assessing the differences between the probe respondents and the original respondents.

For all variables discussed in this paper, the imputation processing was done separately within three age groups: 12 to 17, 18 to 25, and 26 and older.

4. The NSDUH Probes

The NSDUH questionnaire allows subjects to decline to answer any question by entering "Don't Know" (DK) or "Refused" (REF) as a response. Probes were added to some of the questions on drug use in the 2000 NSDUH. For example, the following question appeared in the 2000 NSDUH for subjects who entered a "REF" response to the question, "Have you **ever** smoked part or all of a cigarette?"

> The information respondents provide about their cigarette smoking is very important to the success of this study. We recognize that this information is personal. Please remember that the answers you give will be kept strictly confidential and they will never be linked to your name.
>
> Please reconsider answering this question: Have you **ever** smoked part or all of a cigarette? [yes, no]

Table 1: Questions for Which Probes Existed in 2000-2005 NSDUHs
("DK" means that the subject is probed if the response to the original question was "DK". "REF" means that the subject is probed if the response to the original question was "REF". "N/A" means that the questionnaire does not include a question related to this quantity.)

Drug	Lifetime	Recency	30-day Frequency
Cigarettes	REF	DK/REF	DK/REF
Chewing Tobacco	REF	DK/REF	DK/REF
Snuff	REF	DK/REF	DK/REF
Cigars	REF	DK/REF	DK/REF
Pipes	REF	No probe	n/a
Alcohol	REF	DK/REF	DK/REF
Marijuana	REF	DK/REF	DK/REF
Cocaine	REF	DK/REF	DK/REF
Crack	REF	DK/REF	DK/REF
Heroin	REF	DK/REF	DK/REF
Hallucinogens	n/a	DK/REF	DK/REF
LSD	REF	DK/REF	n/a
PCP	REF	DK/REF	n/a
Ecstasy (2001 onward)	REF	DK/REF[1]	n/a
Inhalants	REF	DK/REF	DK/REF
Pain Relievers	REF	DK/REF	n/a
Oxycontin	No probe	No probe	n/a
Tranquilizers	REF	DK/REF	n/a
Stimulants	REF	DK/REF	n/a
Methamphetamine	REF	DK/REF[1]	n/a
Sedatives	REF	DK/REF	n/a

[1]There was no probe in the 2000 questionnaire. These probes first appeared in the 2001 questionnaire.

No probe appeared if the subject entered a "DK" response to the original question. Similar probes appeared for many of the questions on lifetime use of drugs.

Probes also appeared for questions about recency of drug use. For example, the following question appeared in the 2000 NSDUH for subjects who entered a "DK" response to the question, "How long has it been since you **last** used any inhalant for kicks or to get high?"

> What is your **best guess** of how long it has been since you **last** used any inhalant for kicks or to get high? [past 30 days, more than 30 days ago but within the past 12 months, more than 12 months ago]

The following question appeared in the 2000 NSDUH for subjects who entered a "REF" response to the same question:

> The information respondents provide about their use of inhalants is very important to the success of this study. We recognize that this information is personal. Please remember that the answers you give will be kept strictly confidential and they will never be linked to your name.
>
> Please reconsider answering this question: How long has it been since you last used any inhalant for kicks or to get high? [past 30 days, more than 30 days ago but within the past 12 months, more than 12 months ago]

The only other questions that were probed were the questions asking past month users for the number of days in the past 30 during which they used the drug. For exmaple, the following question appeared if a subject entered a response of "DK" or "REF" to the question, "During the past 30 days, on how many days did you use cocaine?"

What is your **best estimate** of the number of days you used cocaine during the past 30 days? [1 or 2 days, 3 to 5 days, 6-9 days, 10-19 days, 20-29 days, all 30 days]

These probes have appeared in all NSDUH questionnaires for a similar set of questions, in every year since 2000. Table 1 shows the questions for which probes existed in the 2000-2005 NSDUHs. The table also shows whether DK, REF, or both responses were probed for each question. An entry of "n/a" means that there was no question in the NSDUH on this topic. For example, the questionnaire does not ask past month users of pipes for their 30-day frequency.

4. Comparison of Probe Responses to Original Responses

4.1 Lifetime Drug Use

First, it is important to note that all of the questions about lifetime drug use have very high response rates.

Table 2 shows the level of nonresponse for the lifetime drug use questions in the 2005 NSDUH. It also shows logical bounds for the actual prevalence rates after imputation: the lower bound is simply the prevalence estimate assuming that all the item nonrespondents are lifetime nonusers; and the upper bound is the prevalence estimate assuming that all the item nonrespondents are lifetime users.

Table 2: Response Rates and Lifetime Use Bounds for Key Drugs in the 2005 NSDUH
(Data from other survey years are likely to be similar)

Drug	*Item Response Rate*	*Lower Bound on Prevalence*	*Upper Bound on Prevalence*
Cigarettes[1]	100%	66.55%	66.55%
Chewing Tobacco	99.99%	13.66%	13.67%
Snuff	99.98%	13.19%	13.22%
Cigars	99.98%	36.28%	36.30%
Pipes	100.00%	15.81%	15.82%
Alcohol	99.98%	82.90%	82.92%
Marijuana	99.90%	40.08%	40.18%
Cocaine	99.97%	13.84%	13.87%
Crack	99.96%	3.25%	3.29%
Heroin	99.98%	1.45%	1.48%
Hallucinogens	99.81%	13.86%	14.05%
LSD	99.94%	9.22%	9.28%
PCP	99.94%	2.71%	2.77%
Ecstasy	99.94%	4.72%	4.79%
Inhalants	99.88%	9.33%	9.45%
Pain Relievers	99.81%	13.44%	13.62%
Oxycontin	99.83%	1.43%	1.60%
Tranquilizers	99.87%	8.63%	8.76%
Stimulants	99.88%	7.83%	7.95%
Methamphetamines	99.93%	4.24%	4.31%
Sedatives	99.87%	3.67%	3.80%

[1]Subjects who declined to answer the question about lifetime use of cigarettes were treated as unit nonrespondents. So the lower bound and upper bound are equal. See the description of the "usable case rule" in Kroutil, Handley, and Smarrella (2006).

Table 3 shows the sample sizes and the (unweighted) proportion of lifetime users for each drug lifetime use question, for both probe responses and original responses.

Even when the data is combined across six survey years, sample sizes are small. Only marijuana, cigarettes, alcohol, and cocaine appear to have enough of a sample size to support any kind of further analysis. For marijuana and cocaine, the probe respondents showed a much higher prevalence rate than the original respondents. For alcohol, the probe respondents showed a much lower prevalence rate. The prevalence rates were about the same for cigarettes. Perhaps subjects who refuse to answer questions about illegal drugs like marijuana and cocaine are hiding sensitive information (Caspar, Penne, & Dean, 2005). Practically all of the other illicit drugs show the same pattern, but sample sizes are very small.

Table 3: Sample Sizes and Prevalence Rates for Original Responses and Probe Responses, for 2000-2005 NSDUHs Combined
(*Results are not presented for certain drugs because the lifetime use variable is actually a summary of information contained in several questions. For example, the sedatives module contains five lifetime use questions about types of sedatives. The probe only appears if the subject refuses to answer all five of them.*)

	Percent of Lifetime Users		*Sample Size*	
Drug	*Original*	*Probe*	*Original*	*Probe*
Cigarettes	58.3%	56.4%	412,570	101
Snuff	13.4%	18.2%	412,433	22
Chewing Tobacco	12.9%	26.7%	412,588	15
Cigars	32.9%	33.3%	412,577	21
Pipes	9.3%	21.1%	412,589	19
Alcohol	72.6%	42.9%	412,445	98
Marijuana	38.7%	79.3%	412,005	285
Cocaine	11.1%	57.7%	412,391	78
Crack[1]	25.0%	73.3%	45,650	15
Heroin	1.2%	16.7%	412,447	12
LSD	9.2%	41.9%	412,292	31
PCP	2.3%	11.8%	412,311	17
Ecstasy[2]	6.6%	28.6%	340,570	14
Methamphetamine[2]	4.2%	4.0%	340,586	25

[1]Subjects who did not respond affirmatively to the question about cocaine lifetime use were automatically skipped out of the crack module of the questionnaire. The reported percentages and sample sizes cover only those who reported lifetime use of cocaine.
[2]The questions about lifetime use of ecstasy and methamphetamines were not probed in the 2000 questionnaire. They were first probed in the 2001 questionnaire.

Also, for all survey years from 2000 to 2005, the proportion of users among the probe respondents exceeds the proportion of users among the original respondents for both marijuana and cocaine. If there truly is no difference in prevalence rates between the probe respondents and the original respondents, then the probability of seeing the same result for all six years is equal to the probability of six coin flips coming up as either all heads or all tails. The P-value from such a "sign test" is .03125 (Hollander & Wolfe, 1999). Weighted X^2 tests were also done to examine whether respondent type (original vs. probe) and lifetime use were independent, using SUDAAN software. The tests took the design of the NSDUH sample into account. The

tests showed similar results to the sign tests: for marijuana, the X^2 test statistic was 41.46 ($p < .0001$), and for cocaine, it was 12.45 ($p = .0004$). The pattern also holds across year and age group: out of the 18 combinations of years and age groups, the proportion of lifetime users among the probe respondents was higher than the proportion of lifetime users among the original respondents 17 times for marijuana ($p < .0001$) and 16 times for cocaine ($p = .0007$).

4.2 Recency of Drug Use

Table 4 shows the number of subjects who responded to the original, "DK" probe, and "REF" probe questions for recency for each tobacco-based drug. Data is pooled across the 2000-2005 NSDUHs as before. Note that only subjects who respond affirmatively to the lifetime use questions are presented with the recency questions. For tobacco-based drugs, only the second recency question is probed. That is, the first recency question asks subjects whether they are past month users. If they respond negatively, they are asked whether they last used more than 30 days ago but within the past 12 months, more than 12 months ago but within the past 3 years, or more than 3 years ago.

Table 4: Sample Sizes for Original and Probe Respondents for Recency of Drug Use, Tobacco-Based Drugs Only
(*Data is pooled across 2000-2005 NSDUHs*)

Drug	*Original*	*DK Probe*	*REF Probe*
Cigarettes	129,513	226	65
Snuff	42,414	56	16
Chewing Tobacco	46,027	70	19
Cigars	107,019	165	19

Sample sizes are relatively small since only lifetime users who did not respond affirmatively to the past month use question are presented with this question.

The proportions of responses in each recency category for each drug are presented in Table 5.

Table 5: Distribution of Responses of Original and Probe Respondents for Recency of Drug Use, Tobacco-Based Drugs Only
(*Data is pooled across 2000-2005 NSDUHs. 1 = past year but not past month, 2 = past three years but not past year, 3 = lifetime but not past three years*)

	Original Respondents			*DK Probe Respondents*			*REF Probe Respondents*		
Drug	*1*	*2*	*3*	*1*	*2*	*3*	*1*	*2*	*3*
Cigarettes	19.1	19.5	61.4	19.9	19.9	60.2	32.3	16.9	50.8
Snuff	15.5	18.5	66.0	17.9	26.8	55.4	0.0	6.3	93.8
Chewing Tobacco	13.1	18.0	68.9	22.9	21.4	55.7	26.3	10.5	63.2
Cigars	27.6	29.0	43.4	32.1	29.7	38.2	52.6	15.8	31.6

The REF probe respondents tend to be more likely to be past month users than the original respondents and the probe respondents, for all drugs except snuff.

Since the recency probes operate slightly differently for non-tobacco-based drugs, the results for these drugs are presented in separate tables. Namely, there was only one original recency question instead of two; and the three levels were:

1) Within the past 30 days
2) More than 30 days ago but within the past 12 months
3) More than 12 months ago

Table 6 shows sample sizes for non-tobacco-based drugs. Sample sizes are still small, but generally not quite as small as for tobacco-based drugs.

Table 6: Sample Sizes for Original and Probe Respondents for Recency of Drug Use, Non-Tobacco-Based Drugs Only
(*Data is pooled across 2000-2005 NSDUHs*)

Drug	*Original*	*DK Probe*	*REF Probe*
Alcohol	298,257	379	156
Marijuana	158,962	161	218
Cocaine	45,492	37	38
Crack	11,380	16	6
Heroin	4,888	8	6
Hallucinogens	58,406	94	60
Inhalants	45,810	244	140
Pain Relievers	58,290	354	163
Tranquilizers	29,628	93	35
Stimulants	23,069	94	41
Sedatives	9,694	59	8

Table 7 shows the distribution of recency responses for each drug, for the original question, the DK probe, and the REF probe.

Table 7: Distribution of Responses of Original and Probe Respondents for Recency of Drug Use, Non-Tobacco-Based Drugs Only
(*Data is pooled across 2000-2005 NSDUHs. 1 = past year but not past month, 2 = past three years but not past year, 3 = lifetime but not past three years*)

	Original Respondents			*DK Probe Respondents*			*REF Probe Respondents*		
Drug	*1*	*2*	*3*	*1*	*2*	*3*	*1*	*2*	*3*
Alcohol	61.1	22.0	16.9	20.8	30.1	49.1	35.9	26.9	37.2
Marijuana	24.4	18.5	57.1	16.8	21.7	61.5	52.3	16.1	31.7
Cocaine	9.5	18.3	72.3	5.4	16.2	78.4	34.2	21.1	44.7
Crack	7.3	14.5	78.1	6.3	31.3	62.5	0.0	16.7	83.3
Heroin	7.1	13.5	79.4	0.0	12.5	87.5	0.0	16.7	83.3
Hallucinogens	6.0	17.9	76.1	8.5	24.5	67.0	16.7	15.0	68.3
Inhalants	5.1	13.8	81.0	12.3	20.5	67.2	11.4	11.4	77.1
Pain Relievers	18.4	28.8	52.8	15.0	33.3	51.7	16.0	16.0	68.1
Tranquilizers	11.7	23.8	64.4	11.8	30.1	58.1	8.6	20.0	71.4
Stimulants	9.9	18.9	71.2	7.4	37.2	55.3	12.2	12.2	75.6
Sedatives	6.2	12.5	81.3	8.5	22.0	69.5	0.0	37.5	62.5

For marijuana and cocaine, the REF probe respondents tend to be more recent users than the original respondents. As was the case for lifetime, the pattern was consistent across all six survey years for both marijuana and cocaine: REF probe respondents reported past month use more often than original respondents.

4.3 30-day Frequency of Use

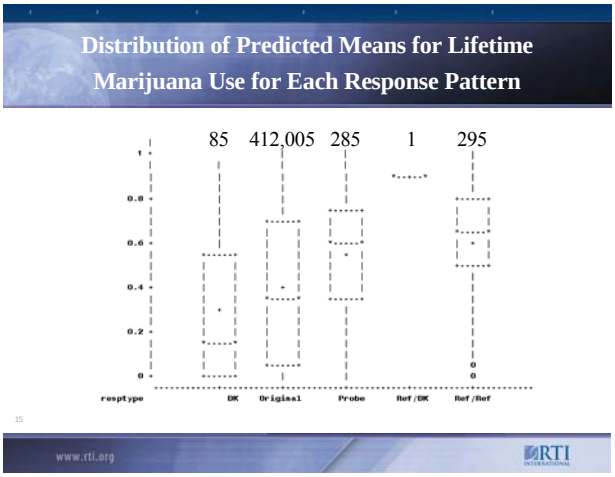
Because the set of possible responses to the 30-day frequency question were different in the probe than they were in the original question, and because the same probe appeared whether the subject responded to the original question with "DK" or "REF", no further investigations were done on 30-day frequency of drug use.

5. Comparison of Predicted Means, Marijuana and Cocaine Lifetime Use

Up to this point, only comparisons of actual responses have been made. What about the item nonrespondents—i.e., those subjects who responded to neither the probe nor the original question? Their responses are unknown, but auxiliary variables can be examined to assess whether the item nonrespondents are similar to the original respondents, the probe respondents, or neither.

Given that the PMN method is model-based, it makes sense to compare the predicted means of each type of subject. With respect to the lifetime use questions, there are only five types of subjects: those who responded to the original question; those who entered DK for the original question; those who entered REF for the original question but answered the probe; those who entered REF for the original question and DK for the probe; and those who entered REF for the original question and REF for the probe. Figure 1 displays side-by-side box plots of the predicted means for each of the five types of subjects for the marijuana lifetime use question. Sample sizes are displayed at the top of each box plot.

Figure 1: Distribution of predicted means for lifetime marijuana use for each response pattern



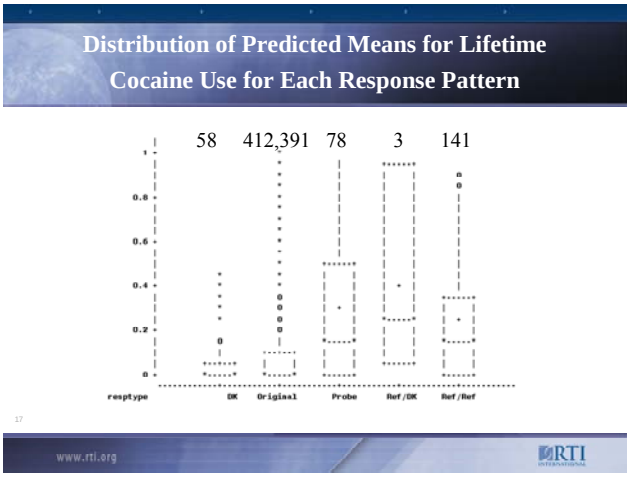
The subjects who refused to respond to the original question (i.e., the last three columns) tended to have higher predicted means than the other subjects. The subjects who refused both the original question and the probe had even higher predicted means than the subjects who responded to the probes. This suggests that perhaps it is reasonable to use the "soft nonrespondents" to represent the "hard nonrespondents" in imputation. It also suggests that the pattern is partly being accounted for by the auxiliary variables: the regression models are accounting for some of the difference between the probe respondents and the original respondents.

Perhaps the models are not accounting for all of the difference, though. As seen in Table 3 above, the (unweighted) prevalence rate among the probe respondents was approximately 79%. The median of the predicted means for the probe respondents was only 58.3%, and the mean was only 54.2%.

Figure 2 is the same as Figure 1, but for cocaine lifetime use instead of marijuana lifetime use. The pattern is similar. The models seem to be accounting for some, but not all, of the differences between the probe respondents and the original respondents. Table 3 shows that the

cocaine prevalence rate among probe respondents was 57.7%, but the median of the predicted means of the probe respondents was only 15.4%, and the mean was only 28.0%. These are larger than the values for the original respondents, but not as large as the actual value.

Figure 2: Distribution of predicted means for lifetime cocaine use for each response pattern



6. Conclusions

The probes offer a cheap alternative to callbacks, for the mitigation of nonresponse bias. There is a balance to be obtained, however: the NSDUH is already a very long survey, and probes may serve to lengthen the interview and annoy the subjects (Caspar, Penne, & Dean, 2005). The response patterns suggest that, for the illicit drugs, subjects who refuse to respond to the original question but respond to the probes are more often lifetime users, and more often recent users, than subjects who respond to the original question.

The presence of the probes seems to be correcting for some of the bias, simply by adjusting the estimates relative to what they would be if the probes did not exist. However, the probes could be further used in imputation, which would enhance the adjustment for the nonresponse bias. A comparison of the predicted means of the different response patterns suggests that, at least for marijuana and cocaine lifetime use, the imputation method is able to pick up some, but not all, of the difference between the original respondents and the probe respondents.

The imputation method could be enhanced either by adding an indicator of the source of the response (i.e., original or probe) as a covariate to the models, or by adding a stricter constraint to the hot deck programs, increasing the likelihood that the donor will be a probe respondent as opposed to an original respondent.

However, for all questions examined in this paper, item nonresponse is low. The imputation methodology certainly does not have much of an effect on the final estimates for the overall US population. Still, given that 1) the lifetime use questions are used in the usable case rule; 2) analysts often subset the data into subgroups, which might magnify the impact of a few cases; and 3) these are probably the most important and visible variables in the NSDUH, a modification to the imputation methodology may still be warranted.

Acknowledgements

The authors would, first and foremost, like to thank the other members of the NSDUH Imputation Team at RTI for their support and for discussing all sorts of imputation research topics at length, including this one. At the time of writing, the other team members were Jeremy Aldworth and Kim Ault as co-team leaders, and Kortnee Barnett-Walker, Lisa Carpenter, Amy Licata, Bing Liu, and Peilan Martin as other team members. We would also like to thank our former team leader, Eric Grau, who played a major role in the conception of these ideas. Also, we'd like to thank the following people for reviewing drafts of this paper and the accompanying presentation, and/or discussing some of its ideas: Art Hughes and Dicy Painter of the Substance Abuse and Mental Health Services Administration; and Vince Iannacchione, Karol Krotki, Mike Penne, Jim Chromy, Mike Witt, Rachel Caspar, Kevin Wang, and Adam Safir of RTI.

References

Biemer, P., and K. Wang. In process. Using Callback Models To Adjust for Nonignorable Nonresponse in Face-to-Face Surveys.

Caspar, R.A. 1992. A follow-up study of nonrespondents to the 1990 National Household Survey on Drug Abuse. In *Proceedings of the 1992 American Statistical Association, Survey Research Section, Boston, MA* (pp. 476-481). Alexandria, VA: American Statistical Association.

Caspar, R.A., Penne, M.A., and E. Dean. 2005. Evaluation of follow-up probes to reduce item nonresponse in NSDUH. In J. Kennet & J. Gfroerer (Eds.), *Evaluating and improving methods used in the National Survey on Drug Use and Health* (DHHS Publication No. SMA 05-4044, Methodology Series M-5, pp. 121-148). Rockville, MD: Substance Abuse and Mental

Health Services Administration, Office of Applied Studies.

Chen, P., Dai, L., Folsom, R., Gordek, H., Grau, E., Shi, W., and M. Westlake. 2006. Person-level sampling weight calibration [2004]. In *2004 National Survey on Drug Use and Health: Methodological resource book* (Section 3, prepared for Office of Applied Studies, Substance Abuse and Mental Health Services Administration, under Contract No. 283-03-9028, Deliverable No. 28; RTI/0208726). Research Triangle Park, NC: RTI International. [Available as a PDF at http://www.oas.samhsa.gov/nhsda/methods.cfm#2k4]

Grau, E.A., Barnett-Walker, K., Copello, E., Frechtel, P., Licata, A., Liu, B., and P. Martin. 2006. Imputation Report [2004]. In *2004 National Survey on Drug Use and Health: Methodological resource book* (Section 5, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-03-9028, Deliverable No. 26, RTI/0208726. Research Triangle Park, NC: RTI International. [Available as a PDF at http://www.oas.samhsa.gov/nsduh/methods.cfm#2k4]

Hollander, M., and D.A. Wolfe. 1999. *Nonparametric Statistical Methods, Second Edition.* New York: Wiley.

Kroutil, L.A., Handley, W., Smarrella, D.J., Suresh, P., and S. Calvin. 2006. 2004 National Survey on Drug Use and Health: Procedures for editing supplementary self-administered data in the 2004 NSDUH computer-assisted interview. In *2004 National Survey on Drug Use and Health: Methodological resource book* (Prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-03-9028, Deliverable No. 28, RTI/0208726. Research Triangle Park, NC: RTI International. [Available as a PDF at http://www.oas.samhsa.gov/nsduh/methods.cfm#2k4]

Little, R.J.A. and D.B. Rubin. 1987. *Statistical analysis with missing data.* New York: Wiley.

Lohr, S.L. 1999. *Sampling: Design and Analysis.* New York: Duxbury Press.

Singh, A.C., Grau, E.A., and R.E. Folsom, Jr. (2001). Predictive mean neighborhood imputation with application to the person-pair data of the National Household Survey on Drug Abuse. In *Proceedings of the 2001 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Atlanta, GA* [CD-ROM]. Alexandria, VA: American Statistical Association.

Singh, A.C., Iannacchione, V.G., and J.A. Dever. 2003. Efficient estimation for surveys with nonresponse follow-up using dual-frame calibration. In *Proceedings of the 2003 Joint Statistical Methods, Survey Research Methods Section, San Francisco, CA* (pp. 3919-3930). Alexandria, VA: American Statistical Association.

Thompson, S.K. 1992. *Sampling.* Wiley: 1992.

Wang, K., Murphy, J., Baxter, R., and W.J. Aldworth. 2005. Are two feet in the door better than one? Using process data to examine interviewer effort and nonresponse bias. Presented at the Federal Committee on Statistical Methodology Research Conference, Washington, DC. [Available as a PDF at http://www.fcsm.gov/05papers/Wang_Aldworth_etal_VIB.pdf]