Modeling Nonresponse and Underreporting in Response in Surveys of Arrestees

Zhiwei Zhang, Ph.D.[1]
NORC at the University of Chicago

[1] Bethesda Office, NORC at the University of Chicago, 4350 East West Highway, Suite 800, Bethesda, Maryland, 20814, U.S.A. zhang-zhiwei@norc.org

Key Words: arrestee, bias, crime, nonresponse, survey, underreporting

Abstract

Policy makers in the major developed countries have used social surveys of human subjects to generate influential information and make informed decisions and plans. However, possible combined and unexamined biases caused by noncontact, refusal, and underreporting in responses may have the survey data quality compromised. Utilizing auxiliary administrative records, respondents' self-reports, paradata on interviewers and survey processes, as well as laboratory test results, this study aims to ascertain factors influencing survey procedural outcomes at various stages through logit models, understand the presence of biases using the propensity score approach, and investigate whether the propensities of participation and/or cooperation could be significantly associated with the likelihood of providing false information in surveys of arrestees. It is concluded that survey estimation adjustments need to be made not only for the biases caused by nonresponses but also for the measurement error such as underreporting. These adjustments should be made with a consideration of the full process involving multiple stages at which nonuniform and differentiated conditionality is embedded and interrelated.

1. Introduction

For decades, policy makers in the major developed countries have used social surveys of human subjects to generate influential information and make informed decisions and plans. Individuals being approached for interviews under different circumstances in various survey settings have the freedom to choose whether to give consent to be interviewed. Many surveys of the household population are experiencing increasingly higher refusal rates over the years (Groves, 2006). Factors and conceptual framework have been examined and proposed in the past to explain the household survey participation. At the same time when professional standards and agency responsible for budgets of survey funders continue to urge and value high response rates (AAPOR, 2001; OMB, 1999, 2006), special surveys, such as in business establishments or in institutional settings, have encountered a wide range of response rates ranging, i.e., from less than 30% in the Drug Abuse Warning Network (DAWN) data collection in hospital

emergency department setting to more than 99% for the Census of State and Local Law Enforcement Agencies (CSLLEA) data collection in criminal justice system setting. Detailed investigations of the field data collections in highly controlled institutionalized settings have been --relatively speaking -- rarely found in the mainstream methodological research domains.

Recent studies, literatures, and workshops on survey nonresponses in general have highlighted that there could be no fixed linkage between nonresponse rates and nonresponse biases, that multiple tools may be available to assess whether and to what extent the potential bias exist, and that there are ways to make posterior adjustments to have bias and variances under control. Obtaining survey measures on a sampling unit after the sampling step is probably one of most important components in the execution stage of all the federal government surveys and even the censuses. What are less known are how similar the factors influencing survey participation from individuals in non-household environment are, and how and to what extent different stages of survey processes, such as the likelihood of contact, survey participation, arrestee's willingness for further cooperation at the end of the survey, may be related with each other. Could it possible that the "quality" secured at one stage, say, acquiring high rates of survey responses, be at the cost of lousy outcome at another stage, say, having large proportion of "untrue" responses? Specifically, are the propensities of participation and/or cooperation significantly associated with the likelihood of providing false information in the surveys?

Through the integrated use of the auxiliary administrative records, the survey response data, the interviewers' profiles, as well as the accompanied data on the lab test results in multiple years large scale surveys conducted within the criminal justice system, this study aims to (1) test whether factors identified, or the associated theories developed, from the analyses of the household survey process predicting the survey participation can be generalized to other settings such as in highly controlled and institutionalized criminal justice system; (2) examine whether factors influencing initial willingness of participating would be similar to those influencing the willingness of providing biomarker specimens after the participation; and (3) explore how and to what extent the measurement errors would be associated with the nonresponse tendencies.

2. Survey Data Collection at Arrestee Booking Facilities

Survey respondents and nonrespondents in the general population, when being approached in household settings by interviewers at front doors or remotely over phones, can be in various non-mutually-exclusive everyday life situations, i.e., in relatively isolated mode and therefore wanting (or avoiding) to talk to someone, showing interests in the survey topic, enjoying helping others and thus participating altruistically anyway, being under tight time pressure to do something else on schedule or of more importance, being alerted to guard one's privacy, etc. In contrast, interviewing arrestees in highly-controlled institutional settings seems to be shadowed by the peculiar context framed by some special pre-interview events. Arrestees, who might have just lived in a usual

household or other domicile living place just a couple of days ago, had just been caught and charged for violations of law and had just apparently followed, even not reluctantly, the law enforcement orders and actions to be carried, likely through the police vehicles, to a booking or detention facility, and thereafter offered figure prints, had picture taken, and provided basic demographic information on records. Each sampled arrestee eligible for the survey had already carried at least one penal charge code against him/her which could lead to jail or prison terms with various lengths. With uncertain destiny looming larger than a survey, it seemed unclear whether they would choose to participate in a survey which can contribute to an understanding of the aggregated level of drug abuse by others (as they were told, the research organization). Working experiences at multiple urban areas in the United States told us that more than 80% of the arrestees being contacted for interviews would actually agree to participate, with unknown reasons that may or may not mimic what could motivate many household-based survey participants, such as due to ample "free" time, as a part of an arrest momentum to follow instructions and deferring to authority, because of their willingness to help, or being interested in getting a small incentive like a chocolate bar, etc.

The data used for this study here are from the national database of 164,037 male and 29,369 female arrestee collected during 2000-2003 from the Arrestee Drug Abuse Monitoring (ADAM) Program which was a National Institute of Justice (NIJ) program initiated in 2000. ADAM collected voluntary interviews and urine samples from real-time samples of booked arrestees 3 to 4 times per year in each of approximately 39 metropolitan areas. New arrestees were interviewed within 48 hours and asked a battery of questions on arrest history, drug use patterns, drug acquisition, and prior participation in treatment programs.

3. Noncontact and Refusal as Nonresponse

Eligible sampled arrestees might not have been contacted due to reasons such as physical illness, language problem, being transferred, having been released, physical and mental health problem, or violent behavior, etc. While there are merits to disentangle the detailed reasons in the analysis, two large categories were aggregated as two important components of the "nonresponse," namely, noncontact and refusal. The noncontact covers both those who were not contacted for interview and those who happened to be unavailable for interview.

Response rates can be calculated in a myriad of ways, for different purposes. The following interview codes are considered in calculating the noncontact and refusal rates:

1. Eligible, contacted and interviewed (I + P)
2. Eligible, contacted and refused (R)
3. Eligible, not-contacted (NC)
4. Eligible, not available (NC)

Here, the Non-contact Rate is defined as: NC / (I + P + R + NC), where I=completed interviews; P=partial interviews; NC= known eligible units not contacted or not

available; and R=refused eligible units. In ADAM 2000-2003, the Noncontact Rate = 62,447 / 193,400 = 32.29%.

Once contacted, the interviewer would read to the arrestee a verbal consent script which includes the following text:

> *"This is a Federally-funded project designed to collect information about drug use, illegal activities, and service needed among individuals who have been arrested. Your participation is voluntary. … You may find some questions embarrassing or distressing, and you can refuse to answer any question. At the end of the interview, I will ask you to provide a urine sample. If you listen to all my questions and provide the urine sample, you will be given a [incentive]. Can we begin now?"*

Therefore, each arrestee approached for interview knew before the interview that a urine sample would be needed at the end of the interview.

The Refusal Rate is defined as: R / (I + P + R) and the ADAM Refusal Rate = 21,717 / 130,953 = 16.58%. The Total Response Rate = (Contact Rate)x(Cooperation Rate) = (67.71%)x(83.42%) = 56.48%.

4. Measurement Error in the Form of Underreporting of Illicit Drug Use

All the arrestees in the ADAM program during 2000-2003 who completed the interviews were asked to provide, voluntarily, their urine samples at the end of the interviews at the booking facilities where interviews were conducted. The ADAM Urine Sample Refusal Rate was about 5.50% (=5,864 /106,701). The urine samples were then analyzed by a professional contractor in the lab tests. The results were submitted back to the ADAM national contractor and merged with the survey data. Comparisons were performed on each of the four major illicit drugs – Marijuana, Cocaine (including both powder and crack cocaine), Heroin, and Methamphetamine. If the urine test result for a specific drug is positive but the self-reported drug use for the same drug in the past three days was none, underreporting was recorded for the arrestee on this particular drug. Overall, the Rate of Under-Reporting Any of the Four Major Illicit Drugs was 34.78% (=34,138/98,164). Overreportings of drug use, in contrast, were rare in ADAM and were therefore not analyzed or modeled separately.

5. Logistic Regression Modeling Using Sampling Frame Variables from Administrative Records and Paradata

This study classifies nonrespondents by reasons (e.g., noncontact, survey refusal, and urine sample refusal) and performs modeling for these groups to identify the sources of bias. To estimate the impact of the arrestee characteristics, the survey design factor, and the interviewer attributes on the survey process outcomes at various stages –i.e., the arrestees' contactability, consent to respond to the interview, cooperation to provide a urine sample, and underreporting -- four types of logistic regression models were used. These include: a contact model; a survey cooperation model, conditional on contact; a

bio-specimen sample cooperation model, conditional on survey cooperation; and a measurement error model in which the deliberate disavowal of the drug use despite of the positive urine test result would be the model outcome.

Difference between statistics from the full sample and statistics from respondent-only is an indicator of nonresponse bias. Significant associations between the survey stage outcomes and the independent variables would suggest the presence of the nonresponse bias. Used as the independent variables are a series of basic demographic and arrest information which were drawn from the arresting facilities records, including information about the arrestees and the nature and the severity of the offenses with which they were charged. Interviewers play roles and contribute to the survey responses (Hanson and Marks, 1958; Bailey, Moore, and Bailar, 1978). Here, the anonymous and unique identifier of each ADAM interviewer was linked to all the sampled arrestees for whom the interviewer was assigned to conduct the in the interview. The analytic models performed here include the interviewers' age, gender, and race/ethnicity as part of the independent variables. For simplicity in the initial exploratory analysis, we did not control for the clustering of the arrestees within urban areas or the nesting of arrestees within the same interviewers. Estimations in the models predicting the noncontactability and the survey refusals were unweighted. Analytic weight was used in the model predicting the urine refusals among the survey participants, and special urine analytic weight was used in the model predicting the underreporting status among the survey participants who provided urine samples.

6. Results

Arrestee Characteristics as Factors

Age: As compared to younger arrestees (i.e., 34 or younger), older arrestees (age 45 or older) were more likely not to be contacted or not available, more likely to refuse to participate among those who were contacted, and more likely to underreport their illicit drug use behavior if they were interviewed. The youngest arrestees (age 20 or younger) were least likely not to be contacted/available, least likely to refuse to participate in the survey once they were contacted, least likely to refuse to provide urine sample once they were interviewed, and least likely to deny and underreport their drug use.

Gender: Female arrestees were less likely than male arrestees not to be contacted, less likely to refuse to participate, less likely to refuse to provide urine sample, but more likely to underreport drug use.

Race/Ethnicity: As compared to white arrestees, black arrestees were less likely not to be contacted, and Hispanic arrestees were more likely not to be contacted. Black arrestees were more likely to refuse to participate, and refused to provide urine sample if they did participate. Both black and Hispanic arrestees were more likely to underreport their drug use than white and other arrestees.

Table 1: Logistic Regression Coefficients in Models Predicting NonContact, Refusal, Urine-Sample Refusal, and Under-reporting

	Non-Contact	Refusal	Urinal-sample Refusal	Under-reporting
Age (ref: 21-25)				
≤ 20	-0.111***	-0.358***	-0.266*	-0.306***
26 – 34	-0.013	0.195	0.017	0.433***
35 – 44	-0.051*	0.276***	0.02	0.896***
45 or older	0.115***	0.205***	0.08	1.059***
Gender (ref: Male)				
Female	-0.228***	-0.144***	-0.401**	0.343***
Race/Ethnicity (ref: White)				
Black	-0.256***	0.076***	0.161*	0.833***
Hispanic	0.261***	-0.186***	-0.084	0.498***
Other	0.153	-0.107*	0.087	-0.196**
Offense Type (ref: other)				
Violence	-0.256***	0.078**	0.122	-0.274***
Property	-0.191***	-0.038	-0.055	0.105**
Drug	0.106***	-0.063*	-0.107	0.088**
Offense Severity (ref: Misdemeanor)				
Felony	-0.613***	-0.082***	-0.177**	0.017
Survey Situation (ref: Stock)				
Flow	-1.389***	0.158***	0.268***	0.078**
Interviewer Gender (ref: Male)				
Female	0.089***	-0.217***	-0.247***	0.070*
Interviewer Age (ref: 25 or younger)				
26-30	0.382***	-0.025	-0.003	-0.039
31-40	0.303***	-0.146***	-0.393***	0.092**
41 or older	0.570***	-0.212***	-0.466***	0.016
Interviewer Race/Ethnicity (ref: White)				
Black	-0.223***	-0.082*	-0.123	0.109**
Hispanic	-0.401***	0.005	0.043	0.135***
Other	-0.598***	0.129**	-0.047	0.004

*** p <.001 ** p <.01 * p <.05

Stock/Flow: The ADAM sample design categorized the arrestee population into two categories -- stock and flow. Stock referred to those arrestees who had already been booked before the interviewers arrived each day. Flow comprised those arrestees who arrived while the interviewers were in the booking facility.

The modeling results show that the "flow" arrestees were less likely not to be contacted but were more likely to refuse to participate. Among those who participated in the survey interviews, "flow" arrestees were more likely to refuse to provide urine sample for testing for their actual drug use behavior. Once interviewed, "flow" arrestees were more likely than "stock" arrestees to underreport drug use.

Interviewer Characteristics: Interviewer's age, gender, and race/ethnicity appear to influence the survey process at many stages. Arrestees to be interviewed by the female interviewers were more likely not to be contacted. They were, however, less likely to refuse to participate or provide urine sample. After giving consent for interviews, arrestees interviewed by females were more likely to underreport drug use than arrestees interviewed by male arrestees. Arrestees interviewed by interviewers more than 30 years old were less likely to refuse to participate or refuse to provide urine sample than arrestees interviewed by younger interviewers. Non-white minority interviewers were more likely to get the eligible arrestees contacted. Compared with arrestees interviewed by whites, arrestees interviewed by blacks were less like to refuse to be interviewed but more likely to underreport, arrestees interviewed by Hispanics were also more likely to underreport and arrestees interviewed by other interviewers were more likely to refuse to be interviewed. In separate analysis that included interactions, it is found that female interviewers had lower refusal rate than male interviewers when the interviewees were male arrestees, and had higher refusal rate when the interviewees were female arrestees.

Felony Status: Arrestee noncontact status is negatively related to felony status; refusal is positively related to felony status. In analytic results not shown, felony status is positively to key ADAM outcome variables – cocaine, heroin, and methamphetamine uses.

7. Propensity Score (PS) and Nonresponse Bias

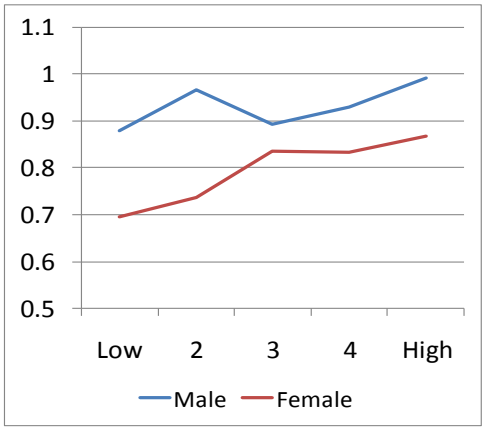
The nonresponse bias of the population mean for a variable *y* is a function of the covariance between survey variables of interest and the response propensity (e.g., Little and Rubin 2002)

$$Bias(\bar{y}_r) \approx \frac{\sigma_{yp}}{\bar{P}}$$

The basic underlying premise of deterministic models is that nonresponders (*m*) may be different from responders (*r*) on characteristics associated with key variables measured in the survey. If response propensity is uncorrelated with the variable in question, then nonresponse will not lead to bias but just an increase in sampling error and the inference upon which the results will be based is smaller, due to the nonresponse (e.g., Bethlehem, 2002; Lessler & Kalsbeek, 1992).

Figure 1: Selected Survey Variables by Noncontact Propensity Strata

a. Number of arrests last year



b. Average number of days per month used crack cocaine past year

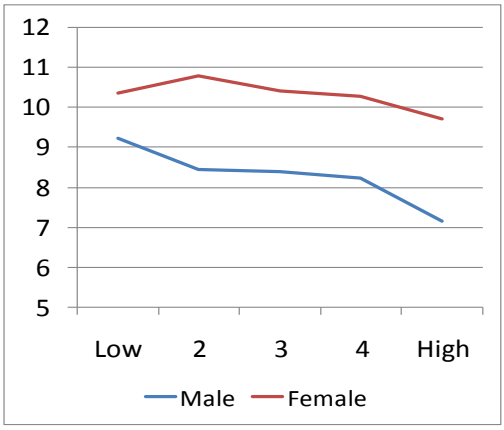
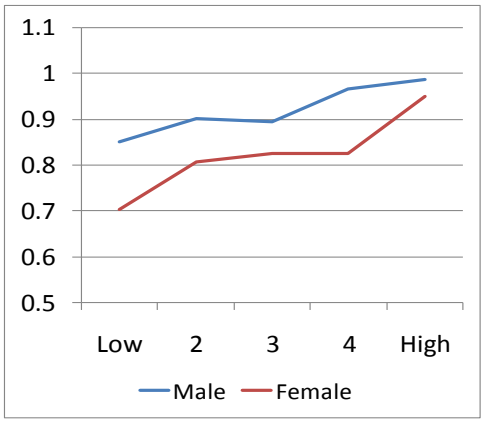


Figure 2: Selected Survey Variables by Survey Refusal Propensity Strata

a. Number of arrests last year



b. Average number of days per month used crack cocaine past year

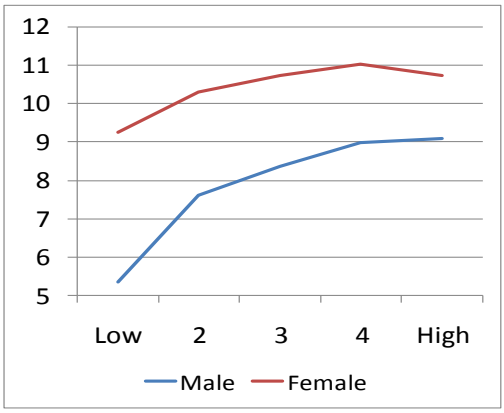
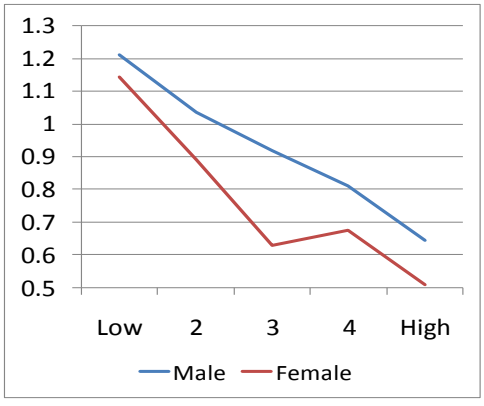
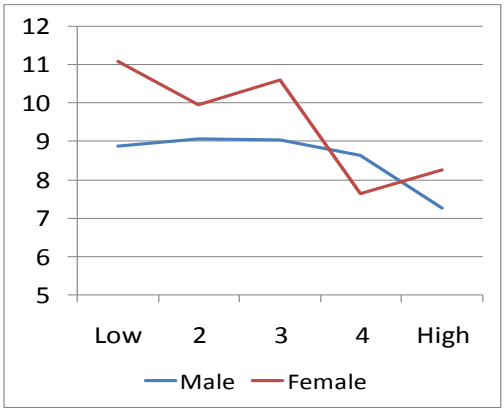


Figure 3: Selected Survey Variables by Urine Sample Refusal Propensity Strata

a. Number of arrests last year



b. Average number of days per month used crack cocaine past year



The size of nonresponse bias is known to depend on the amount of statistical association between response probabilities (p_i) and the measurement of interest (Y_i) in the population (Lessler and Kalsbeek, 1992). To reduce the dimensionality of the survey stage-specific outcome to single scalar quantity, separate propensity scores (PS) are derived. To estimate the propensity score, the logistic regressions were used to obtain the propensity scores through the predicted probability of noncontact, refusal, biosample decline, and underreporting. For example, the propensity score for survey refusal is the probability that an arrestee with characteristics X does not respond to the survey. The estimated propensity scores were divided into five strata, with the quintiles of the estimates propensity scores as the cutoffs for the different strata.

To assess the potential bias caused by the differential likelihoods of noncontactability, survey response, and urine cooperation, a series of bivariate relationships were investigated. Figures 1-3 show how different strata of three major propensity scores were associated with different magnitudes of two exemplified survey variables of interest – the number of times the arrestees were arrested in the year prior to the interview, and the average days per month the arrestees used crack cocaine in the past year. These figures reveal that the propensities of noncontactability, survey response, and urine cooperation were associated with these two important and typical survey variables in the ADAM surveys. Because of these associations, it is inferred that the estimations on survey variables such as these two typical measures without adjustments of the survey nonresponses would be biased. On the other hand, since the nonresponses and the survey variables were shown as being related, the biases can be adjusted and controlled under the premises that these nonresponses can be treated as being conditionally missing at random.

8. Conditionality of Survey Stages

Each survey response stage imposes a condition upon the subsequent stage. To investigate the relationship among the stages, a correlation matrix is set up as in the following Table.

Table 2: Relationships among the Survey Nonresponse and Under-reporting Propensities

	Non-Contact	Survey Refusal	Urine Sample Refusal	Any Under-reporting
Non-Contact	1			
Survey Refusal	-0.312***	1		
Urine Sample Refusal	-0.244***	0.462***	1	
Any Under-Reporting	-0.121***	0.449***	-0.062***	1

Note: *** *p <.001*

Overall for the ADAM arrestees, survey refusal propensity is positively associated with the urine sample refusal propensity and the propensity of under-reporting of illicit drugs.

Non-contact propensity is negatively associated with both the survey and urine sample refusal propensities and also negatively associated with the under-reporting propensity. There exists a small negative yet significant correlation between the urine sample refusal propensity and the propensity to underreport drug use.

9. Sensitivity Analysis

To test whether the overall picture captured through the pooled multiple years' national data can be applied to local urban areas, sensitivity analysis was performed on one local area. The results listed below are based on the localized analysis using the ADAM 2000-2003 data collected from Denver, Colorado.

Table 3 shows that the major correlation pattern for Denver remains the same as that for the national data, with two exceptions: (1) there was a negative relation between survey refusal and urine sample refusal ($r=-0.255$, $p<.001$); and (2) there was a positive relationship ($r=0.142$, $p<.001$) between urine sample refusal propensity and the underreporting propensity.

Table 3: Relationships among the Survey Nonresponse and Under-reporting Propensities among Arrestees at Denver, Colorado

	Non-Contact	Survey Refusal	Urine Sample Refusal	Any Under-reporting
Non-Contact	1.000			
Survey Refusal	-0.578***	1.000		
Urine Sample Refusal	-0.465***	-0.255***	1.000	
Any Under-Reporting	-0.113***	0.314***	0.142***	1.000

Note: *** *p <.001*

Table 4: Coefficient for Respondent Characteristics in Logistic Model Predicting Likelihood of Under-Reporting – *The Denver Site Study*

	Under-reporting
Age 20 or younger	-0.115
Age 26 -34	0.676***
Age 35 - 44	1.037***
Age 45 or older	1.243***
Female	0.754***
Having high school diploma	-0.335*
Employed (full or part)	-0.179
Married	0.285
Drug dependence	0.281
Having stable residence	0.017
Ever jailed past year	0.240

*** p <.001 ** p <.01 * p <.05

Table 4 further shows that age, gender, and education were among the significant factors predicting the propensity of arrestees' underreporting propensity.

10. Conclusions and Discussions

Estimates on selected survey variables differ across noncontact, survey refusal, and bio-sample refusal propensity strata. Recruiting high propensity nonresponding sampled arrestees is expected to improve estimates. The strong and positive correlation between the survey refusal propensity and under-reporting of illicit behaviors is alarming, suggesting that the truthfulness of self-reports must be considered in conjunction with the recruiting efforts, in order to reduce the total error. Adjustments need to be made not only for the biases caused by nonresponses but also for the measurement error such as underreporting. These adjustments should be made with a consideration of the full process involving multiple stages at which conditionality is embedded and interrelated.

The nonresponse pattern analysis helps to develop the logistic regression models to predict contactability, survey cooperation, and bio-sample cooperation. These models produce propensity scores (Rosenbaum and Rubin, 1983) that can be used as multiplier adjustments to the initial survey weights (based on the probability of selection) (David, Little, Samuhel, and Triest, 1983; Little, 1986) which may be implemented in future studies which may also use the categories of the propensity score to test the relationship between an independent variable of concern and a designated new outcome variable. Bias due to various stages of nonresponses can be considered and analyzed next. For example, the measurement bias due to under-reporting may be obtained through the standardized difference in the means of logit(PS) between true true-reports and under-reports.

Further investigations can be made to test alternative models in nationally-representative sample with further environment-arrestee and interviewer-arrestee interaction terms, and making multivariate adjustments such as (1) weighting through the use of the inverse of the PS; (2) using PS directly in the models to control for selection bias; and (3) regress within PS strata, etc. For further multivariate modeling, the measurement error can be tested as a function of response propensity. . One limitation of the current study is that the group of frame and survey process variables constitutes only a small part of the variables that may explain and predict the contactability and the likelihood of survey cooperation. Additional (or unrecorded) paradata, such as the data capturing the process of contacting the arrestees for interviews, could be very useful, and need to be considered in future data collection design for the arrestee population.

References.

American Association for Public Opinion Research (AAPOR) (2001) *Standard Definitions: Final Disposition of Case Codes and Outcome Rates for RDD Telephone Surveys and In-person Household Surveys*

Bailey, L., Moore, T.F., Bailar, B. A. (1979) An interviewer variance study for the eight impact cities of the National Crime Survey Cities Sample. *Journal of the American Statistical Association*, 73, 16-23.

Bethlehem, J. (2002) Weighting nonresponse adjustments based on auxiliary information. Pp. 275-288 in R. M. Groves, D.A. Dillman, J. L. Eltinge, and R. J.A. Little. ed. *Survey Nonresponse*. New York: Wiley.

David, M., Little, R., Samuhel, M., and Triest, R. (1983) Nonrandom nonresponse models based on the propensity to respond. Pp. 168-173 in *Proceedings of the Business Economic Statistical Section*. Alexandria, VA: American Statistical Association.

Groves, R.M. (2006) Nonresponse rates and nonresponse bias in household survey. *Public Opinion Quarterly*, 70, 646-675.

Groves, R. and Couper, M. (1998) *Nonresponse in Household Interview Surveys.* New York: Wiley.

Handon, R. H. and Marks, E. S. (1958) Influence of the interviewer on the accuracy of survey results. *Journal of the American Statistical Association*, 53, 635-655.

Lessler, J. T. and W. D. Kalsbeek (1992) *Nonsampling Error in Surveys*. New York: Wiley.

Little, R.J.A. (1986) Survey nonresponse adjustments for estimates of means. *International Statistical Review*, 54, 2, 139-157.

Office of Management and Budget (OMB) (1999) *Implementing Guidances for OMB Review of Agency Information Collection*. Draft, June 2, Washington, D.C.: Government Printing Office.

Office of Management and Budget (OMB) (2006) *Questions and Answers when Designing Surveys for Information Collections.* Washington, D.C.: Office of Information and Regulatory Affairs.

Rosenbaum, P.R. and Rubin, D.B. (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 1, 41-55.