Address-Based Sampling and the National Survey on Drug Use and Health: Evaluating the Effects of Coverage Bias

Katherine Morton[1], Joseph McMichael[1], Jamie Ridenhour[1]
Jonaki Bose[2]
[1]RTI International, 3040 Cornwallis Road, Research Triangle Park, NC 27709
[2]SAMHSA, Office of Applied Studies, 1 Choke Cherry Road, Rockville, MD 20850

Abstract

A common concern of survey researchers is whether the coverage properties of an address-based sampling (ABS) frame create outcome bias in the estimates from in-person surveys. This paper evaluates basic demographics and several drug use and mental health measures obtained from 1,725 respondents in a probability sample of 200 area segments from the National Survey on Drug Use and Health (NSDUH). The evaluation compares outcomes from respondents covered by the NSDUH's field enumerated (FE) frame to those covered by an ABS frame derived from the United States Postal Service Computerized Delivery Sequence (USPS CDS) file. After post-stratifying the weights to populations with known ABS undercoverage, we test for significant differences in outcomes between the two frames.

Key Words: bias, coverage, in-person surveys, ABS

1. Introduction

The National Survey on Drug Use and Health (NSDUH) is a large-scale survey which provides the federal government with national, state and substate data on substance use and mental health in the civilian, non-institutionalized population aged 12 or older. Data are collected annually from approximately 140,000 households and 67,500 persons. Currently, the NSDUH uses field enumeration to obtain lists of dwelling units (DUs) within sampled areas or segments.[1]

The NSDUH field enumeration process begins more than one year in advance of the survey year with the production of maps for sampled areas. Approximately 370 field "listers" then enumerate sampled areas from April through November of the year preceding each annual survey.

Prior research has shown that address-based sampling (ABS) frames can be constructed for a fraction of the cost of field enumerated frames and can significantly reduce the time required for frame construction. In addition, in order to control field enumeration costs,

---

[1] Segments for the NSDUH comprise one or more adjacent census blocks that in combination meet or exceed the minimum requirement of 100 DUs in rural areas or 150 DUs in urban areas. For a segment to be classified as rural, all of the census blocks in the segment have to be rural. If one or more of the segment's blocks are urban, the segment is also urban.

sampled areas are typically formed to be geographically small which may introduce intracluster correlation.

While ABS is attractive from a cost and time perspective, there are some drawbacks. First, ABS is known to undercover certain populations such as persons residing in group quarters and rural areas. In areas with some ABS coverage, a frame supplementation procedure such as the Check for Housing Units Missed (CHUM) procedure[2] can be used to pick up missing dwelling units and increase coverage (McMichael et al, 2008). In areas with little to no ABS coverage, the CHUM procedure becomes very similar to a field enumeration operation and there is no cost benefit to using the ABS frame. These areas can be pre-identified and field enumerated during the sampling process by using a coverage prediction algorithm. Segments where adequate ABS coverage is predicted are assigned to the ABS frame supplemented with the CHUM procedure, while field enumeration supplemented with the HOI procedure is retained in segments where ABS coverage is expected to be poor. This hybrid field enumeration and ABS approach is currently being investigated for the NSDUH.

This paper describes results from a field study which examined the coverage of a hybrid frame in a sample of area segments from the NSDUH. In particular, we describe differences in prevalence estimates between the ABS frame and NSDUH's field enumerated frame.

2. Mailing List Field Study

The Mailing List Field Study (MLFS) was conducted in a subsample of 200 NSDUH segments[3] which were originally fielded in the first quarter of 2009. The MLFS sample comprises 3,878 sampled dwelling units (SDUs) which were screened and deemed eligible[4] for the NSDUH. Among these eligible SDUs, 1,725 completed interviews were available for use in the analysis.

For the MLFS, we attempted to match the addresses of the 3,878 eligible SDUs to a list of mailing addresses purchased from a commercial vendor. SDUs whose mailing address did not initially match to the ABS list were followed up with a telephone call or in-person field check to verify or correct the mailing address of the SDU in order to determine whether they were on the ABS list. In addition, during the field check, the field interviewer determined if the SDU could be picked up by the CHUM procedure (Iannacchione et al, 2010). Upon completion of the field work, we identified the subset of persons who were covered by the ABS frame and the CHUM procedure. Table 1 displays the number of covered DUs and persons by frame type.

---

[2] The CHUM procedure was developed by RTI International for ABS frames and is analogous to the half-open interval (HOI) procedure used to supplement field enumerated frames.

[3] Segments in Alaska and Hawaii were excluded from the MLFS sampling frame.

[4] An eligible SDU is either a housing unit (HU) for a single household or a noninstitutional group quarters unit (GQU) where at least one civilian aged 12 years or older will be residing for the majority of the calendar quarter.

Table 1. Dwelling Unit and Person Sample Sizes by Frame Type

	FE + HOI	ABS + CHUM	ABS Only
Eligible DUs	3,878	3,728	3,229
Interviews	1,725	1,650	1,402

*FE=Field Enumeration; HOI=Half-Open Interval; ABS=Address-Based Sampling; CHUM=Check for Housing Units Missed

3. Coverage Bias Analysis

Prior to analyzing the results, the weights for persons covered by the ABS frame and CHUM procedure were poststratified using the general exponential model (GEM) for sampling weight calibration (Folsom and Singh, 2000). Namely, we poststratified the weights to external population counts[5] by age group (12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older), race (Hispanic or Latino, not Hispanic or Latino White, not Hispanic or Latino Black, and not Hispanic or Latino Other), gender, and rural or urban.[6] Two-way interactions of age, race, and gender were also included in the model.

Using the poststratified weights, we then compared basic demographics and several drug use and mental health measures for respondents covered by NSDUH's field enumerated frame supplemented with the HOI (FE/HOI) to those covered by the ABS frame and CHUM procedure (ABS/CHUM). Table 2 presents the results of these comparisons. Of the 31 FE/HOI and ABS/CHUM comparisons made, only three were significant at the 0.05 level (college graduates, persons with federal poverty threshold greater than or equal to 200%, and persons residing in group quarters). Three additional comparisons were significant at the 0.10 level (past year treatment for illicit drugs, total family income less than $20,000, and persons between 100 and 199 percent of the federal poverty threshold).

We also compared estimates among persons covered by the FE/HOI frame to those covered by the ABS frame only (i.e. assuming no frame supplementation). Slightly more comparisons were significant at the 0.05 level (past year treatment for illicit drugs, past year treatment for alcohol use, college graduates, total family income less than $20,000, percent of federal poverty threshold greater than or equal to 200%, group quarters, and core-based statistical areas [CBSAs]) and at the 0.10 level (past month use of cigarettes). In addition, when comparing persons covered to the ABS/CHUM frame to those covered by the ABS frame only, past year treatment for illicit drug use, past year treatment for alcohol use, college graduates, total family income less than $20,000, group quarters, and CBSAs were statistically significant. These findings emphasize the importance of a frame supplementation procedure such as the CHUM.

---

[5] The control totals used in the poststratification adjustment were 2009 population estimates obtained from the U.S. Census Bureau.

[6] The rural or urban variable was included in the model because rural areas are known to have poor ABS coverage.

Table 2. Comparison of Results from FE and ABS Frames

Variable	Mean (FE +HOI)	Mean (ABS + CHUM)	Mean (ABS Only)	FE - ABS/ CHUM	FE - ABS Only	ABS/ CHUM - ABS Only
Past Month Use:						
Cigarettes	0.237	0.231	0.226	0.006	0.011[b]	0.005
Alcohol	0.513	0.513	0.514	0.000	-0.001	-0.001
Illicit Drugs	0.072	0.071	0.071	0.002	0.001	0.000
Illicit Drugs except MJ	0.035	0.032	0.032	0.002	0.003	0.000
Cocaine	0.006	0.004	0.003	0.002	0.002	0.000
Past Year Dependence:						
Illicit Drugs	0.020	0.018	0.019	0.002	0.001	-0.001
Alcohol	0.031	0.028	0.028	0.002	0.003	0.000
Past Year Treatment:						
Illicit Drugs	0.005	0.005	0.006	0.000[b]	-0.001[a]	-0.001[a]
Alcohol	0.007	0.007	0.008	0.000	-0.001[a]	-0.001[a]
Past Year Mental Health (Age 18+):						
Serious Psych. Distress	0.130	0.128	0.128	0.002	0.002	0.000
Major Depressive Episode	0.058	0.059	0.057	-0.001	0.001	0.002
Education (Age 18+):						
Less than high school	0.125	0.122	0.120	0.003	0.005	0.002
High school graduate	0.361	0.359	0.358	0.002	0.003	0.001
Some college	0.270	0.271	0.264	-0.001	0.006	0.007
College graduate	0.245	0.249	0.258	-0.004[a]	-0.013[a]	-0.009[a]
Total Family Income:						
Less than $20,000	0.160	0.153	0.145	0.007[b]	0.015[a]	0.008[a]
$20,000 - $49,999	0.334	0.338	0.342	-0.004	-0.008	-0.004
$50,000 - $74,999	0.223	0.222	0.226	0.000	-0.003	-0.003
$75,000 or more	0.284	0.287	0.288	-0.003	-0.004	-0.001
% of Federal Pov. Threshold:						
<100%	0.111	0.111	0.106	0.001	0.006	0.005
100-199%	0.167	0.160	0.164	0.007[b]	0.002	-0.004
>=200%	0.711	0.719	0.726	-0.008[a]	-0.015[a]	-0.007
Group Quarter	0.012	0.011	0.004	0.001[a]	0.007[a]	0.006[a]
% Owner-occupied:						
>=50%	*	*	*	*	*	*
10 - <50%	*	*	*	*	*	*
<10%	0.036	0.036	0.035	0.001	0.002	0.001
CBSA	0.927	0.928	0.943	-0.001	-0.017[a]	-0.016[a]
Region:						
Northeast	*	*	*	*	*	*
North Central	0.196	0.197	0.201	-0.001	-0.005	-0.004
South	0.342	0.340	*	0.002	*	*
West	*	*	*	*	*	*

*=low precision, estimate not reported; a=p-value<0.05; b=pvalue<0.10

FE=Field Enumeration; HOI=Half-Open Interval; ABS=Address-Based Sampling; CHUM=Check for Housing Units Missed; MJ=Marijuana; Psych=Psychological; Pov=Poverty

4. Conclusions

Estimates based on the ABS-only frame were limited to small but statistically significant differences when compared to the FE frame. Using the CHUM to supplement the ABS frame helped to mitigate some of these differences.

Note that because the estimates based on the FE frame and the estimates based on the ABS frame share a large portion of their cases, these comparisons have the statistical power to declare very small differences in the overall prevalence estimates statistically significant (Table 3). For example, for a prevalence estimate of 0.01, a difference of 0.002 can be detected with 80 percent power and a significance level of 0.10 assuming an ABS coverage rate of 95 percent.

Table 3. Detectable Difference in Estimates With At Least 80 Percent Power and a
Significance Level of 10 Percent, Assuming 200 Segments and 1,725 Persons

Assumed Coverage Rate	True Value over All Dwelling Units	Detectable Change in Estimate for Omitting Noncovered Dwelling Units
0.95	0.01	0.0019
0.95	0.02	0.0024
0.95	0.08	0.0042
0.95	0.25	0.0062
0.95	0.50	0.0068

Hybrid frames such as the one being investigated for the NSDUH would share an even larger proportion of cases with the field enumerated frame since segments below the designated threshold would be field enumerated. Therefore, the hybrid field enumerated and ABS frame would have even less coverage bias. For the hybrid frame being investigated for the NSDUH, we examined coverage bias among several subgroups, including rural areas and group quarters. We found no substantive differences in the estimates.

Acknowledgements

References

Folsom, R.E., & Singh, A.C. 2000. The generalized exponential model for sampling weight calibration for extreme values, non-response, and post-stratification. In *JSM Proceedings*, Survey Research Methods Section, Alexandria, VA: American Statistical Association. 598-603.

Iannacchione, V., K. Morton, J. McMichael, B. Shook-Sa, J. Ridenhour, S. Stolzenberg, D. Bergeron, J. Chromy, and A. Hughes. 2010. The best of both worlds: a sampling frame based on address-based sampling and field enumeration. In *JSM Proceedings,* Survey Research Methods Section, Alexandria, VA: American Statistical Association.

McMichael, J.P., J.L. Ridenhour, and B.E. Shook-Sa. 2008. A robust procedure to supplement the coverage of address-based sampling frames for household surveys. In *JSM Proceedings,* Survey Research Methods Section, Alexandria, VA: American Statistical Association. 4329-4335.