Assessing Representativeness in RDD Surveys: Coverage and Nonresponse in the Behavioral Risk Factor Surveillance System

R. Sowmya Rao[1], Michael W. Link[2], Michael P. Battaglia[3], Martin R. Frankel[4], Pamela Giambo[5], Ali H. Mokdad[2]
[1]MGH Biostatistics Center and Institute for Health Policy
[2]Centers for Disease Control and Prevention
[3]Abt Associates Inc.
[4]Abt Associates Inc. and Baruch College, CUNY
[5]Westat

Keywords: Random-digit dialing, Non-response bias, Non-coverage bias, Current Population Survey

Introduction

Survey researchers are increasingly concerned about potential bias in random-digit dialed (RDD) surveys resulting from frame non-coverage and unit non-response. Households with no landline telephones, as well as those with only cellular telephones are excluded from the RDD sample frame (approximately 5 percent of the population). The ability of the population to move their telephone numbers almost anywhere in the country or to convert them into cellular telephones makes assessment of frame non-coverage at the sub-national level (e.g., state level) difficult because the RDD sample is drawn based on the area codes/central office codes. Unit non-response is an issue in any of the various survey modes (mail, telephone, in-person) but response rates to RDD surveys have been declining in the last decade (Curtin et al. 2005) due to growth in screening technologies, privacy concerns, telemarketing, and refusals.

In order to evaluate the degree to which non-coverage and unit non-response contribute to under-representation of important subgroups in RDD surveys, the Behavioral Risk Factor Surveillance System (BRFSS) -- a monthly RDD survey administered by all the states with assistance from the Centers for Disease Control and Prevention (CDC) to collect health-related information – is used as an example.

Methods

BRFSS is an important survey, which generates state-specific prevalence estimates among adults of the major health conditions and behavioral risks associated with pre-mature morbidity and mortality. Details of the survey can be found in Mokdad et al. (2003) or at www.cdc.gov/brfss.

We were interested in evaluating non-coverage and non-response in six states (California, Illinois, North Carolina, New Jersey, Texas and Washington), which were participating in a separate BRFSS pilot study designed to test techniques for improving coverage and reduce non-response (Link et al. 2005a, 2005b). Five of these states have experienced state-level response rates at or below 40% over the past several years (with North Carolina being the exception). From the 2003 BRFSS and the March 2003 Current Population Survey (CPS), we identified the following socio-demographic variables of interest that are common to both surveys: age, sex, education, marital status, race/ethnicity, employment status, household income, number of children in household, type of household, and MSA versus non-MSA. Person weights were used to obtain the weighted frequencies. For the BRFSS, the person weight used does not include the final post-stratification adjustments. The total sample sizes of adults across the six states were: BRFSS – 55,181, CPS – 33,386.

We compared the distributions of the socio-demographic variables for six states from the BRFSS with the distribution of the same variables from the CPS but are only presenting results for IL in the interest of space. Table 1 displays the breakdown of the sample size by age and sex for BRFSS and CPS for the six states. We compared the unweighted and design-weighted frequencies from BRFSS to the weighted frequencies from the CPS and computed the difference in the percentages.

Since we noticed that males were under-represented in the BRFSS, we computed the probability of selecting a male versus a female depending on the household composition (type of household). We also looked at the cross-tabulation of household composition by sex of the respondent.

We further attempted to assess how much of the under (over)-representation was due to non-response versus non-coverage due to non-telephone households. In order to do this, first we compared the unweighted and design-weighted frequencies of the socio-demographic variables from BRFSS with weighted frequencies from

only telephone households in the CPS and computed the differences, DIFFBRF_CPS_TEL. This does not reflect differences due to the exclusion of non-telephone households from the BRFSS sample. Then we computed the proportion of the difference using the entire CPS sample (DIFFBRF_CPS), due to the difference using only telephone households (DIFFBRF_CPS_TEL). These proportions have been denoted as PROP_NONRESPONSE_CPS.

$$PROP\_NONRESPONSE\_CPS = abs\left(\frac{DIFFBRF\_CPS\_TEL}{DIFFBRF\_CPS}\right)x100.$$

Where,

PERCENT_CPS_TEL = Proportion of only telephone household sample from 2003 CPS in a specific category of the variable.

DIFFBRF_CPS = (PERCENT_BRF – PERCENT_CPS).

DIFFBRF_CPS_TEL=(PERCENT_BRF_TEL – PERCENT_CPS_TEL).

We plan to assess the degree to which current BRFSS post-stratification methods could be improved to reduce coverage and non-response bias by computing new weights that would take into account other variables in the post-stratification stage of the weights, and incorporate an adjustment to compensate for the exclusion of non-telephone households (Frankel et al. 2003).

Results

Table 2 displays the unweighted and design-weighted frequencies of the socio-demographic variables in the BRFSS and the difference in the distribution with the CPS for IL. Some differences that were noticeable in the comparison of unweighted frequencies disappeared when the weights were taken into account (e.g. the under-representation of the youngest age group (18-24) goes from 4.0% to 1.2%).

Although the data is not shown for all six states, we summarize the results for all the states as follows: The youngest age group (*18-24*) is highly under-represented in NC, NJ, TX and WA. In CA and IL, they are under-represented but not by a substantial amount. Males are substantially under-represented in all six states. The least educated (*Did not graduate from high school*) are under-represented while the highly educated (*Graduated from college or technical school*) are over-represented. As would be expected, the magnitude of representation differs by state. Compared to the CPS, non-Hispanic whites are over-represented in all the states. Hispanics are under-represented in CA and TX, African-Americans are under-represented in IL, NC, NJ, and TX, and Asians are under-represented in all six states. Those who have never been married are under-represented in each of the six states while individuals who are married are over-represented in all states except CA. Those who are unemployed are over-represented in CA, NJ, TX and WA. The highest income category (*$50,000+*) is under-represented in all the states. In CA and TX the category *<$15,000* is over-represented while this is under-represented in all the other states. Compared to the CPS, there is an over-representation of households with no children. Households with only one woman were over-represented in all states except IL. Households with only 1 man and 1 woman were over-represented in CA and WA. MSA was under-represented in CA and NJ, while it was over-represented in WA.

Weighting had no appreciable effect on the estimates in the cross-tabulation of type of household with sex of the respondent. In each of the six states, women seemed more likely to be interviewed than men. In households with equal numbers of men and women, the proportion of female respondents ranged from 55%-70%. In the other categories where the number of men to women is unequal, the proportion of male and female respondents is unlike our expectations. For example, among households with two men and one woman we would expect the proportion of male respondents to be 66% which is not what is observed.

Table 3 displays the results in Illinois of the proportion of non-response that can be attributed to non-coverage due to non-telephone households. The results indicate that:

(i) Having a proportion less than 50% (e.g., age groups 18-24 and 25-29) suggests that most of the overall difference is due to the non-coverage of non-telephone households.

(ii) When the proportion is between 50% and 100%, (e.g. age group 30-34), then most of the overall difference is due to non-response.

(iii) Note that some of the proportions (PROP_NONRESPONSE_CPS) are greater than 100 (e.g., age group 35-39). This implies the difference due to non-response (-1.41%) is larger than the overall difference (-1.23%), where the overall difference reflects non-response and non-coverage of non-telephone households.

Discussion

The use of the Current Population Survey as a benchmark proved very useful because the CPS is a high response rate survey that includes non-telephone households. We infer from the results of our analysis that the under- or over-representation of different socio-demographic characteristics varies by states, but that in general the younger age groups, men, and those with a lower education are under-represented. We also found evidence of differential interview non-response related to gender – in households with equal numbers of men and women there were fewer male respondents than expected. A method to decompose the overall difference into components due to non-response and non-coverage was developed. For many socio-demographic characteristics most of the overall difference was due to non-response. Notable exceptions included adults age 18-24 and 25-29 years, married adults, adults with household incomes less than $15,000, and MSA status.

References

Curtin, Richard, Presser, Stanley, and Singer, Eleanor. 2005. Changes in Telephone Survey Nonresponse Over the Past Quarter Century. *Public Opinion Quarterly*. Vol. 69: 87-98.

Frankel, Martin R., Srinath, K.P., Hoaglin, David C., Battaglia, Michael P., Smith, Philip J., Wright, Robert A., and Khare, Meena. 2003b. Adjustments for non-telephone bias in random-digit-dialling surveys. *Statistics in Medicine*, Volume 22, pp. 1611-1626.

Link, M., M. Battaglia, M. Frankel, L. Osborn, and A. Mokdad. 2005a. Effectiveness of Address-based Sampling Frame Alternative to RDD: BRFSS Mail Survey Experiment Results. *2005 Proceedings of the American Statistical Association*, Survey Research Methods Section [CD-ROM], Alexandria, VA: American Statistical Association.

Link, M., M. Battaglia, P. Giambo, M. Frankel, A. Mokdad, and S. Rao. 2005b. Assessment of Address Frame Replacements for RDD Sampling Frames. *2005 Proceedings of the American Statistical Association*, Survey Research Methods Section [CD-ROM], Alexandria, VA: American Statistical Association.

Mokdad, Ali H., Donna Stroup, and H. Wayne Giles. 2003. "Public Health Surveillance for Behavioral Risk Factors in a Changing Environment: Recommendations from the Behavioral Risk Factor Surveillance Team." *MMWR* 2003;52 (No. RR-9):1-12.

Table 1: Sample size of adults in the six states for BRFSS and CPS by age and sex.

Variable	California		Illinois		New Jersey		North Carolina		Texas		Washington	
Age	**BRFSS**	**CPS**	**BRFSS**	**CPS**	**BRFSS**	**CPS**	**BRFSS**	**CPS**	**BRFSS**	**CPS**	**BRFSS**	**CPS**
18 – 24	377	1619	431	637	560	437	583	359	514	1016	1411	327
25 – 29	392	1077	392	445	656	329	624	307	492	720	1256	227
30 – 34	401	1262	479	592	907	400	791	367	659	802	1466	249
35 – 39	455	1405	524	607	1145	475	864	360	588	798	1600	282
40 – 44	475	1405	613	628	1301	519	924	374	631	796	1956	365
45 – 49	434	1228	545	578	1251	445	933	293	590	691	1967	297
50 – 54	403	981	494	452	1107	362	914	267	551	589	1974	252
55 – 59	352	745	437	323	916	256	845	200	467	438	1637	176
60 – 64	291	515	348	269	787	190	692	156	407	321	1419	136
65 – 69	252	370	266	191	678	145	700	133	365	233	1103	84
70 – 74	230	319	281	195	679	131	576	123	283	229	1012	64
75 – 79	192	262	226	120	560	126	457	86	212	173	839	65
80+	206	295	227	186	564	163	482	91	226	177	980	79
Missing	15		4		194		70		50		24	
Sex												
Male	1843	5488	2046	2438	4506	1871	3478	1467	2297	3331	7316	1262
Female	2632	5995	3221	2785	6799	2107	5977	1649	3738	3652	11328	1341
Total	**4475**	**11483**	**5267**	**5223**	**11305**	**3978**	**9455**	**3116**	**6035**	**6983**	**18644**	**2603**

Table 2: Percent distribution of the socio-demographic variables in the 2003 BRFSS and its difference from the percent distribution in the March 2003 CPS – unweighted, and using only design weights in Illinois.

Variable	BRFSS				
	COUNT	PERCENT		DIFFBRF_CPS[1]	
Age		**Unweighted**	**Design Weight**	**Unweighted**	**Design Weight**
18 – 24	431	8.19	11.47	-4.01	-1.18
25 – 29	392	7.45	7.99	-1.07	-0.58
30 – 34	479	9.10	9.41	-2.23	-1.83
35 – 39	524	9.96	10.20	-1.67	-1.24
40 – 44	613	11.65	12.06	-0.38	0.72
45 – 49	545	10.36	10.62	-0.71	0.37
50 – 54	494	9.39	9.50	0.73	0.63
55 – 59	437	8.30	8.00	2.12	1.68
60 – 64	348	6.61	5.98	1.46	0.89
65 – 69	266	5.05	4.36	1.40	0.43
70 – 74	281	5.34	4.14	1.61	0.21
75 – 79	226	4.29	3.41	2.00	0.92
80 or older	227	4.31	2.87	0.75	-1.03
Sex					
Male	2046	38.85	40.42	-7.83	-7.27
Female	3221	61.15	59.58	7.83	7.27
Education					
Did not graduate HS	501	9.53	10.34	-5.1	-3.06
HS Grad	1508	28.69	29.62	-4.62	-4.03
Attended College/Tech School	1464	27.85	28.10	2.18	1.31
College/Tech School Grad	1783	33.92	31.94	7.56	5.78

[1]DIFFBRF_CPS = (PERCENT_BRF – PERCENT_CPS).

Table 2: Percent distribution of the socio-demographic variables in the 2003 BRFSS and its difference from the percent distribution in the March 2003 CPS – unweighted, and using only design weights in Illinois (…contd).

Variable	BRFSS				
	COUNT	PERCENT		DIFFBRF_CPS[1]	
Race/Ethnicity		**Unweighted**	**Design Weight**	**Unweighted**	**Design Weight**
White only*	4061	77.23	73.56	12.88	4.60
Black/African-American only*	541	10.29	10.41	-6.10	-5.57
Asian only*	106	2.02	2.42	-1.43	-1.24
Native Hawaiian/PI only*	9	0.17	0.18	0.13	0.12
AI/Alaskan Native only*	41	0.78	0.73	0.66	0.64
Other race only*	7	0.13	0.08	-	-
Multiracial*	40	0.76	0.77	0.26	0.27
Hispanic	453	8.62	11.86	-6.55	1.10
Marital Status					
Married	2848	54.23	60.16	-3.40	2.42
Divorced	629	11.98	8.85	2.92	-0.10
Widowed	578	11.01	7.11	4.90	0.90
Separated	109	2.08	1.89	-0.34	-0.25
Never Married	961	18.30	18.87	-6.50	-6.10
A member of an unmarried couple	127	2.42	3.13	-	-
Employment Status					
Unemployed	240	4.56	5.00	-0.15	0.18
Not unemployed	5027	95.44	95.00	0.15	-0.18
Income					
Less than $15,000	420	9.21	8.51	-1.36	-2.80
$15,000 - < $25,000	764	16.75	16.20	7.10	6.30
$25,000 - < $35,000	663	14.54	13.84	4.31	3.01
$35,000 - < $50,000	792	17.36	17.94	2.70	3.39
$50,000 or more	1922	42.14	43.51	-12.75	-9.89
Number of children in Household					
None	3212	61.06	56.46	10.65	2.07
One child	792	15.06	17.03	-5.05	-1.21
>=2 children	1256	23.88	26.51	-5.61	-0.86
Household Type					
Only one man	493	9.36	5.03	3.65	-1.22
Only one woman	1129	21.44	11.66	10.18	-0.18
One man and One woman	2554	48.49	50.23	-0.60	0.63
> One man and No women	90	1.71	2.43	-0.47	0.13
Two men and One woman	273	5.18	7.71	-3.43	-0.15
Three men and One woman	62	1.18	2.50	-0.81	0.63
>=3 men and fewer women	30	0.57	1.69	-0.58	0.49
Equal number of men and women	103	1.96	4.00	-2.68	-0.42
>One woman and No men	186	3.53	4.14	0.07	0.94
Two women and One man	281	5.34	7.72	-3.57	-1.25
Three women and One man	40	0.76	1.49	-0.62	0.40
>=3 women and fewer men	26	0.49	1.38	-1.13	-0.01
MSA/non-MSA					
MSA	4289	81.43	82.19	-8.61	-0.35
Non-MSA	978	18.57	17.81	8.61	0.35

*Non-Hispanic.

[1]DIFFBRF_CPS = (PERCENT_BRF – PERCENT_CPS).

Table 3: Proportion of the difference between the BRFSS and CPS due to non-response in Illinois.

Variable	DIFFBRF_CPS[1]	DIFFBRF_CPS_TEL[2]	PROP_NONRESPONSE_CPS[3]
Age			
18 – 24	-1.18	-0.50	42.53
25 – 29	-0.58	-0.09	16.27
30 – 34	-1.83	-1.69	92.47
35 – 39	-1.23	-1.41	114.52
40 – 44	0.72	0.50	70.04
45 – 49	0.37	0.25	67.64
50 – 54	0.63	0.43	68.01
55 – 59	1.68	1.58	93.93
60 – 64	0.89	0.79	88.85
65 – 69	0.43	0.31	73.53
70 – 74	0.21	0.20	95.41
75 – 79	0.92	0.80	86.93
80 or older	-1.03	-1.17	113.03
Sex			
Male	-7.27	-7.04	96.89
Female	7.27	7.04	96.89
Education			
Did not graduate HS	-3.06	-2.40	78.56
HS Grad	-4.03	-3.83	94.97
Attended College/Tech School	1.31	0.96	73.43
College/Tech School Grad	5.78	5.27	91.17
Race/Ethnicity			
White only*	4.60	3.04	66.10
Black/African-American only*	-5.57	-4.51	80.88
Asian only*	-1.24	-1.25	100.47
Native Hawaiian/PI only*	0.12	0.12	97.19
AI/Alaskan Native only*	0.64	0.64	99.12
Multiracial*	0.27	0.27	100.92
Hispanic	1.10	1.61	146.20
Marital Status			
Married	2.42	0.66	27.31
Divorced	-0.10	0.08	78.48
Widowed	0.90	0.91	100.07
Separated	-0.25	-0.09	34.89
Never Married	-6.10	-4.69	76.85
Employment Status			
Unemployed	0.18	0.35	192.23
Not unemployed	-0.18	-0.35	192.23

*Non- Hispanic

[1]DIFFBRF_CPS = (PERCENT_BRF – PERCENT_CPS).

[2]DIFFBRF_CPS_TEL = (PERCENT_BRF_TEL – PERCENT_CPS_TEL).

[3] $PROP\_NONRESPONSE\_CPS = abs\left(\frac{DIFFBRF\_CPS\_TEL}{DIFFBRF\_CPS}\right) x100.$

Table 3: Proportion of the difference between the BRFSS and CPS due to non-response in Illinois (contd…).

Variable	DIFFBRF_CPS[1]	DIFFBRF_CPS_TEL[2]	PROP_NONRESPONSE_CPS[3]
Income			
Less than $15,000	-2.80	-1.16	41.39
$15,000 - < $25,000	6.30	6.68	106.16
$25,000 - < $35,000	3.01	2.74	91.15
$35,000 - < $50,000	3.39	3.64	107.37
$50,000 or more	-9.89	-11.91	120.34
Number of children in Household			
None	2.07	2.46	118.76
One child	-1.21	-1.32	109.20
>=2 children	-0.86	-1.14	132.19
Household Type			
Only one man	-1.22	-0.64	52.43
Only one woman	-0.18	0.63	354.84
One man and One woman	0.63	-0.52	82.49
> One man and No women	0.13	0.58	451.04
Two men and One woman	-0.15	-0.08	56.33
Three men and One woman	0.63	0.60	94.84
>=3 men and fewer women	0.49	0.54	111.89
Equal number of men and women	-0.42	-0.70	166.39
> One woman and No men	0.94	0.90	94.86
Two women and One man	-1.25	-1.65	132.18
Three women and One man	0.40	0.44	108.17
>=3 women and fewer men	-0.01	-0.09	1591.10
MSA/non-MSA			
MSA	-0.35	0.05	14.00
Non-MSA	0.35	-0.05	14.00

[1]DIFFBRF_CPS = (PERCENT_BRF – PERCENT_CPS).

[2]DIFFBRF_CPS_TEL = (PERCENT_BRF_TEL – PERCENT_CPS_TEL).

[3] $PROP\_NONRESPONSE\_CPS = abs\left(\frac{DIFFBRF\_CPS\_TEL}{DIFFBRF\_CPS}\right)x100.$