Washington State Population Survey

Recommendations for the 2008 Survey

Final Report

Authors:

Jill Montaquila
David Ferraro
Tom Krenzke

October 15, 2007

Prepared for:

Washington State Office
Of Financial Management
Olympia,
Washington 98504

Prepared by:

WESTAT
1650 Research Boulevard
Rockville,
Maryland 20850

ACKNOWLEDGMENTS

The authors are grateful to the following contributors to this report: Jing Kang for processing WesVar; David Morganstein and Daniel Levine for thorough review and insights; and Mike Brick for valuable advice throughout the task.

TABLE OF CONTENTS

Chapter			Page
1	INTRODUCTION		1-1
2	METHODOLOGY		2-1
3	RECOMMENDATIONS FOR INCREASING OPERATIONAL EFFICIENCY AND/OR IMPROVING RESPONSE RATES		3-1
	3.1	Prior Notification Letter	3-1
	3.2	Incentives	3-2
	3.3	Pre-Screening of Telephone Numbers	3-2
	3.4	Survey Introduction	3-3
	3.5	Number of Attempts Guideline	3-4
	3.6	Call Scheduling	3-9
	3.7	Refusal Conversion Letter	3-10
	3.8	Subsampling Nonmailables	3-11
	3.9	Ever Refused Analysis	3-12
	3.10	Subsample Refusal Prior to 2nd Attempt	3-14
	3.11	Providing a Toll-Free Number for Respondents to Call to Complete Survey	3-14
	3.12	Classification of Cases for Refusal Conversion	3-15
4	RECOMMENDATIONS FOR REDUCING NONRESPONSE OR COVERAGE BIAS		4-1
	4.1	Eligibility Questions	4-1
	4.2	Quotes	4-2
	4.3	Tabulations	4-2
	4.4	Interviewer Awards/Incentives	4-3
	4.5	Cell-Only Households	4-3
	4.6	Disposition Codes	4-4
	4.7	Multiple Telephone Number Adjustment	4-5
	4.8	Imputation	4-6
5	RECOMMENDATIONS FOR SURVEY STRUCTURE		5-1
	5.1	Increase the Sample Size in One Region	5-1
	5.2	Call Scheduling and Treatment of Cases	5-1
	5.3	Reserve Sample	5-2
	5.4	Item Nonresponse Reports	5-3
	5.5	Pretest Survey Instrument	5-3
	5.6	Enhance Quality Control Procedure	5-4
	5.7	Variance Estimation	5-6
	5.8	Capture the Sort Order (for Variance Estimation)	5-6
	5.9	Confidentiality Procedures	5-7
	5.10	American Community Survey	5-7

TABLE OF CONTENTS (Continued)

Chapter		Page
6	INFORMATION REVIEWED, WITHOUT RECOMMENDATIONS	6-1
	6.1 Advanced Letter	6-1
	6.2 Editing	6-1
	6.3 Timing of the Selection of Telephone Numbers	6-1
	6.4 Computation of Response Rates	6-1
	6.5 Hard and Soft Refusals	6-1
	REFERENCES	R-1

Table		
1	Number of attempts by total, eligible, and completed households: 2006	3-5
2	Comparison of the Number of Attempts: 2006	3-6
3	Number of attempts by sample disposition: 2006	3-10
4	Ever refused by total, eligible, and completed households: 2006	3-12
5	Comparison of the Ever Refused: 2006	3-13
6	Comparison of the Do Not Call List refusals: 2006	3-16
7	Ever refused for specific disposition codes: 2006	4-5

Figure		
1	Estimates by number of attempts	3-7

1. INTRODUCTION

The Washington State Population Survey (WSPS) is designed to examine and monitor the health and welfare of all civilian and armed forces household members within the geographic boundaries of Washington state, and is a valuable source of information that serves as a basis for important decisions by state policymakers, in planning, evaluating, and implementing programs to improve the health and welfare of its citizens. The continuing, biennial survey is managed by the Office of Financial Management (OFM), and administered by the Gilmore Research Group. Over 7,500 household respondents were interviewed by telephone in the 2006 WSPS, which provided data on over 17,000 persons. The telephone numbers for households were selected using a list assisted random-digit dialing (RDD) method, using differential sampling rates across strata defined by the eight regions of the state.

As is typical with other RDD surveys, the WSPS estimates are subject to potential bias due to nonparticipation, which could affect the quality of the results. Further, in recent years, the WSPS has seen its response rate decline to about 27 percent for the 2006 survey, from 42.5 percent in 2000. As a result, the OFM initiated a nonresponse bias analysis study and an overall review of the survey procedures, in order to investigate possible problem areas, and to gather recommendations that could lead to improved response rates and more reliable estimates.

The nonresponse bias analysis on the 2006 WSPS data, Ferraro and Krenzke (2007), is hereafter referred to as the 2006 Nonresponse Bias Analysis (NRBA) report. It indicates the potential bias for estimates relating to metropolitan statistical area status, education attainment, race/ethnicity, marital status, household income, and home ownership, some of which are characteristics of cell phone only households. As discussed in the aforementioned report, these characteristics could be integrated into the weighting process to reduce the potential bias due to nonresponse and undercoverage.

This second report is the result of a more broadly focused review of the 2006 WSPS structure and how the survey is conducted. Section 2 begins with a discussion of the methodology used in our review, and the subsequent sections cover a series of recommendations. Our review identified approaches, listed in Section 3 that can be implemented to improve response rates and identified opportunities for increased operational efficiency. In addition to the weighting process proposal outlined in the 2006 NRBA report, Section 4 documents other potential sources of bias and provides additional recommendations for reducing such bias. Further, our review of the WSPS survey procedures leads to a number of suggested improvements in the survey structure, as shown in Section 5. Section 6 presents other aspects that were reviewed, but for which we currently have no recommendations.

Within Sections 3 through 5, the recommendations are sorted in priority order based on the perceived impact on reducing non-sampling error with other issues considered such as cost, timing, and other practical benefits. Lastly, along with these recommendations, Westat is prepared to provide helpful advice and counsel to OFM and to Gilmore Research, through OFM, with the hope of improving the WSPS.

2. METHODOLOGY

One of the goals of this task is to offer suggestions on how to improve response rates and to reduce the potential for bias. To explain further, we introduce the following expression for nonresponse bias for a sample mean ($\bar{y}_R$):

$$Bias(\bar{y}_R) = (1 - W_R)(\bar{Y}_R - \bar{Y}_N),$$

where W_R is the weighted unit response rate, $\bar{Y}_R$ is the population mean of the respondents, and $\bar{Y}_N$ is the population mean for the nonrespondents. The formula shows that there are two components of the bias expression. While the response rate (first component) is universally recognized as a measure of survey quality, it is not by itself a good indicator of nonresponse bias. The difference between participants and nonparticipants (second component) is just as important. Theoretically, even if the response rate is 27 percent, when there is no difference in the mean of the characteristic *y* between participants and nonparticipants, bias does not exist. In practice, the second component is unknown; however, proxies (auxiliary data) are used to estimate the difference. Weighting adjustments are used to reduce nonresponse bias. Therefore, sources of bias are identified and ways to reduce bias are implemented.

We used two general approaches to identify and support recommendations: theory-driven and data-driven. In the theory-driven approach, we reviewed several aspects of the total survey process, including the population to be covered, the timing of operations, sample design and selection, training, telephone protocols and follow-up procedures, editing, quality control procedures, estimation applications, as well as any available information related to response and nonresponse, and coverage. In this analysis, Westat weighed its understanding of the WSPS structure and how it is conducted against its current best methods.

In the data-driven approach, data (e.g., number of attempts, never refused/ever refused) were used to help inform the recommendations. For example, a chi-square test on attempts versus WSPS survey items, determines if the distribution of the survey item is independent of the number of attempts. This helps to determine if the estimate changes due to the extra attempts. WSPS final population weights were used in the analysis, and replicate weights were used to adequately reflect the effect of the sample design (stratification of households by regions and clustering of persons within households) on variance estimates.

3. RECOMMENDATIONS FOR INCREASING OPERATIONAL EFFICIENCY AND/OR IMPROVING RESPONSE RATES

Our review has suggested a number of recommendations for increasing both response rates and operational efficiency. Implementing more efficient approaches may allow for resources to be shifted to increasing response rates. As seen in the 2006 NRBA report, the response rate among cases with mailable addresses is significantly higher than for other cases. Therefore, a good approach is to improve upon the number of mailable addresses. Each of the recommendations in this section could potentially increase response rates – either directly or indirectly.

3.1 Prior Notification Letters

The 2006 Data Collection Report states that:

For the 58,705 RDD sample observations purchased overall, about 51,202 remained after purging known businesses and non-working telephone numbers. Of these 51,202 numbers, approximately 15,285 could be matched with a name and address. Prior notification letters were sent to the 12,977 (25% of 51,202) respondents with valid mailing addresses.

In our experience this is a very low match rate, e.g. Westat averages over 60% for national studies, after excluding the business and nonworking numbers. Also, if the difference between 12,977 and the 15,285 is due to unusable/incomplete addresses, then we would suggest considering other vendors for the reverse append.

<u>Recommend</u>

Consider other vendors for the reverse append, since we have found large variations in match rates and address quality among vendors. It would well behoove OFM to extend an invitation to a number of such vendors, asking for statements of their success in matching.

<u>Benefit</u>

A higher match rate would enable prior notification letters to be mailed to a significantly larger proportion of households, which would be expected to increase the initial cooperation rate.

3.2 Incentives

At the present time, incentives to sampled households are not currently offered in the WSPS; however, extensive research has shown that incentives contribute to improved response rates. In an experiment conducted by Brick, et al (2005), initial cooperation rates increased by roughly three percent from the incentive provided with each advanced letter without adding to the overall cost of the survey.

Recommend

Consider an incentive provided to each household to complete. Also, consider offering an incentive to boost success in refusal conversion.

Benefit

As indicated, incentives have been shown to boost response rates without adding to the overall cost of the survey.

3.3 Pre-Screening of Telephone Numbers

In the 2006 Data Collection Report, it is noted that:

"For the statewide sample, a total of 58,705 telephone numbers were ordered … We received 51,202 telephone numbers from SSI after they purged for known non-working and business telephone numbers."

This purge rate (of about 13%) appears much to low for list-assisted RDD samples. Our review of recent national RDD surveys show purge rates around 40 percent or greater.

Recommend

Unless the use of predictive dialing is planned for the 2008 survey, again we would propose considering other prescreening services.

Benefit

A higher purge rate will reduce the number of unproductive telephone numbers to be attempted by interviewers. In addition to reducing cost, this may positively impact interviewer morale and save interviewing time.

3.4 Survey Introduction

The overarching suggestion is to keep the introduction as short and to the point as possible. Emphasize you are calling for the governor up front, and briefly indicate what the purpose is, i.e., it is particularly important that the respondent hears and understands that the state of Washington is requesting the help of its residents.

> Hello, this is [INTERVIEWER NAME] from Gilmore Research [Does "from Gilmore Research" need to be in there?] calling on behalf of the governor's budget office. PAUSE: [Why the pause? This practically invites the respondent to hang up or interject that they aren't interested.] We are conducting a survey [we usually refer to it as a "research study" rather than a "survey."] funded by the Washington State Legislature [Does the "funded by...Legislature" need to be in there? In the interest of keeping the intro as short as possible, consider removing this clause.] to <<learn from local residents about their employment situation, education, family income, and housing. This information>> [Suggest replacing the part between <<and>> with "gather information that" will be used by the state to help direct programs that will expand the economy and provide new jobs, to plan for schools and train teachers, and to plan for adequate housing based on family needs. [Make this description of the programs more concise.]

Recommend

Hello, this is [INTERVIEWER NAME], calling on behalf of the governor's budget office (or whomever OFM sees best). We are conducting a research study to gather information that will be used by the state to help direct programs that will provide new jobs, plan for education, and plan for adequate housing.

Benefit

This introduction is shorter and to the point. Westat usually does not mention ‘Westat’. We have not had issues or feedback regarding that approach. We suspect that most of the time caller ID

doesn’t say ‘Westat’, but sometimes it does. Other times it may say “Unavailable” or something else. As an example, a national household education study introduction begins with, "Hello, this is [INTERVIEWER NAME] and I'm calling for the U.S. Department of Education..." The interviewer does not claim to be calling *from* the U.S. Dept. of Education--they simply say they're calling *for* the DoE.

3.5 Number of Attempts Guideline

The number of attempts for the total, eligible, and completed cases (at the household level) is shown in Table 1. The low completion rate (2% of eligibles) for households over 21 attempts comes at a high cost per completion. The estimates of some key statistics for cases completed in 1 to 21 attempts were compared to the overall estimates (Table 2). All key statistics are not statistically significant, with the exception of labor force status, however, the absolute bias (-0.17) and relative bias (-0.3%) are very small and is no concern. Weighting adjustments would also likely reduce the differences since, in theory, the cases completed in over 21 attempts would be treated as nonrespondents during the weighting process.

Another way to look at this is shown in Figure 1. The four graphs have three lines to illustrate differences in estimates by the number of calls. One line shows the final estimate for all cases. A second line shows the estimate by call attempts (attempts are in groups of three, 1-3, 4-6, etc.) which tends to bounce above and below the final estimate. The last line shows the estimate for the cumulative number of calls and when it converges to the final estimate. Medicare, which has a highly significant difference at 14 attempts, converges to the final estimate at 21 attempts. The other variables shown converge at fewer attempts.

<u>Recommend</u>

Consider making at most 21 attempts, for all non-purged cases, to contact a household even though a small number of completed cases will be lost.

<u>Benefit</u>

The completion rate for over 21 attempts is very small which comes at a high cost of dialing numbers many times. The savings could be allocated to more cost effective areas of the survey that could potentially more than make up the difference in completes.

Table 1. Number of attempts by total, eligible, and completed households: 2006

Attempts	Total	Eligible	Complete
1	17525	2718	1721
2	5490	2712	1281
3	4037	2630	925
4	2800	2168	624
5	2249	1794	460
6	1949	1643	381
7	1596	1338	297
8	1274	1037	216
9	1038	875	169
10	932	773	163
11	836	654	126
12	743	589	113
13	604	481	94
14	541	430	61
15	461	371	71
16	402	314	53
17	427	327	59
18	410	317	45
19	354	263	35
20	392	293	28
21	333	265	35
22	323	254	18
23	277	217	19
24	213	153	10
25	166	121	17
26	158	103	12
27	136	100	6
28	156	110	6
29	178	126	13
30	2205	1957	11
31	910	868	7
32	94	78	3
33	145	127	1
34	132	110	0
35	317	282	2
36	129	106	0
37	197	160	0
38	255	224	0
39	269	242	0
40	46	41	0
41	2	2	0
42	1	1	0

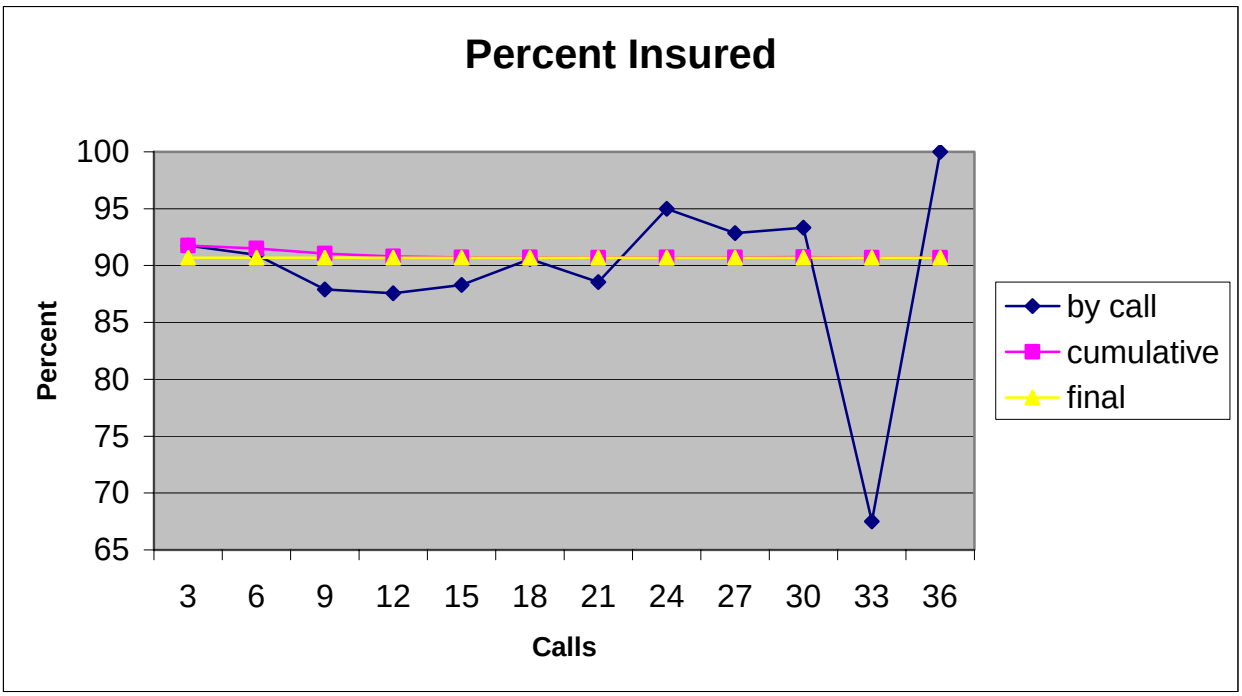
SOURCE: Office of Financial Management, Washington State Population Survey (WSPS) 2006.

Table 2. Comparison of the Number of Attempts: 2006

Characteristic	Overall estimate	Number of Attempts 1-21	Number of Attempts Over 21	Chi-square	
				Statistic	p-value
Insured	90.69	90.69	90.87	0.00	0.9604
Received Medicare	12.98	13.02	10.49	0.98	0.3226
Received Medicaid	13.94	13.98	11.38	0.41	0.5232
Education (25 and over)					
Less than high school	6.42	6.47	3.08	6.11	0.0695
High school	24.04	24.00	26.74		
Some college	29.74	29.87	21.59		
College graduate	39.81	39.66	48.60		
Race					
Hispanic	8.87	8.92	5.50	3.81	0.3873
White	76.77	76.71	80.03		
Black	3.36	3.39	1.12		
American Indian/Alaskan Native	1.45	1.47	0.09		
Native Hawaiian/Other PI	1.53	1.53	1.70		
Asian	5.24	5.20	7.80		
Others	2.79	2.77	3.76		
Labor force status (15 and over)					
Employed	61.52	61.34	72.12	12.22	0.0012
Unemployed	4.45	4.50	1.46		
Not in labor force	34.04	34.16	26.42		
Household income as percent of Federal Poverty Level					
0-99%	8.91	8.88	11.08	1.15	0.8520
100-199%	14.65	14.69	12.34		
200-299%	15.18	15.23	12.06		
300-399%	15.72	15.75	14.22		
400% and over	45.54	45.46	50.29		
Own/rent					
Rent	23.67	23.69	22.49	0.04	0.8494
Own	76.33	76.31	77.51		
Household type					
Husband-wife family household	70.15	70.30	60.33	1.86	0.4450
Other family household	18.58	18.49	24.11		
Non-family household (2 or more persons)	1.84	1.83	2.50		
Single-person household	9.43	9.37	13.06		

NOTE: Detail may not sum to totals because of rounding.

SOURCE: Office of Financial Management, Washington State Population Survey (WSPS) 2006.



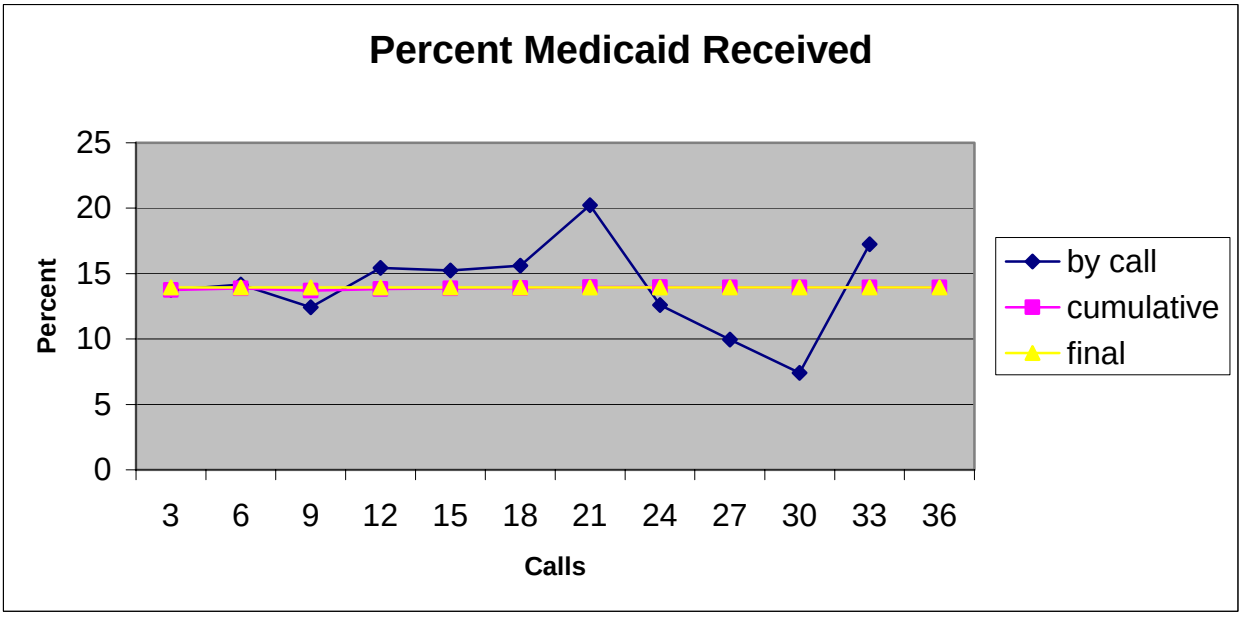
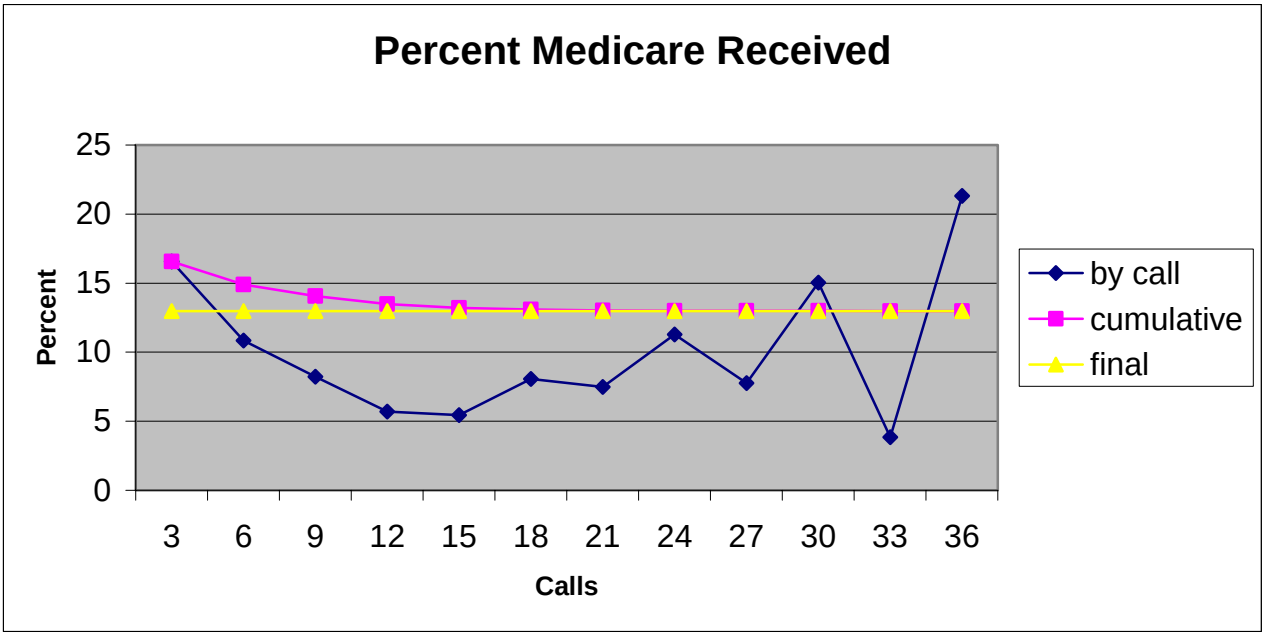


Figure 1. Estimates by number of attempts



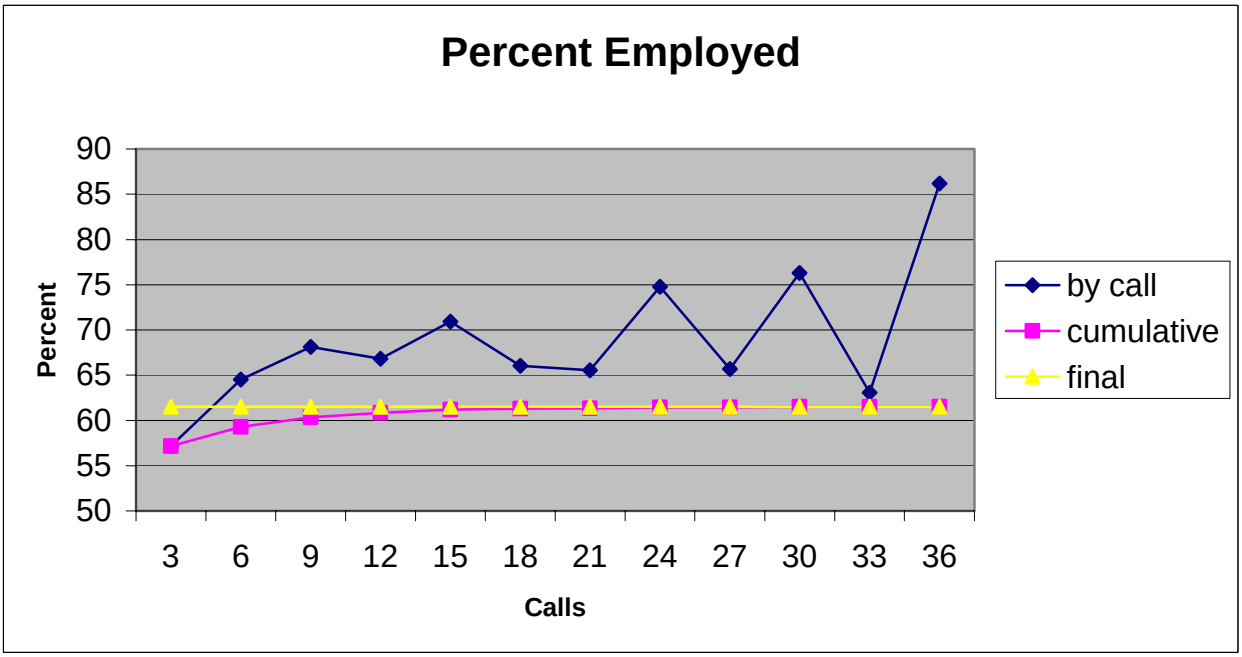


Figure 1. Estimates by number of attempts (Continued)

3.6 Call Scheduling

Table 3 shows the distribution of the number of call attempts for no answer, answering machine, and busy signal cases. Households were not worked equally in these codes and many were not attempted enough times.

Recommend

We recommend considering increased automation of the number of attempts per case in the call scheduler and to require a certain number of attempts based on the prior call history.

Benefit

Building such rules into the scheduler might increase the number of households completed or at least contact made, and also help to ensure that cases are worked equally.

Table 3. Number of attempts by sample disposition: 2006

Sample disposition	DISPOS	Attempts	Frequency
Answering Machine	05	1-5	40
		6-10	129
		11-15	137
		16-20	206
		21-25	211
		26-30	1019
		31-35	783
		36+	453
No answer	04	1-5	109
		6-10	266
		11-15	262
		16-20	279
		21-25	262
		26-30	805
		31-35	428
		36+	211
Busy Signal	06	1-5	21
		6-10	50
		11-15	31
		16-20	40
		21-25	62
		26-30	267
		31-35	203
		36+	100

SOURCE: Office of Financial Management, Washington State Population Survey (WSPS) 2006.

3.7 Refusal Conversion Letter

In the 2006 WSPS, there was no special mailing to refusal cases in advance of refusal conversion attempts.

Recommend

We recommend a special mailing (to mailable cases) prior to attempting refusal conversion. This letter would be aimed at addressing the specific concerns often raised during refusal (e.g., confidentiality, the “Do Not Call” list, etc.). More about “Do Not Call” later.

Benefit

Studies conducted by Westat have demonstrated the effectiveness of sending a letter prior to attempting refusal conversion, even when a pre-notification letter had previously been sent. This, in addition to the measures raised under 'Pre-Screening of Telephone Numbers,' would be expected to positively affect response rates. In a national survey conducted by Westat earlier this year, a refusal conversion rate (with up to two refusal conversion attempts) of over 30 percent was attained.

3.8 Subsampling Nonmailables

To achieve gains in operational efficiency in RDD studies, Westat developed a technique of oversampling RDD telephone numbers that have mailable addresses (i.e., RDD-generated telephone numbers for which a commercial supplier can match addresses). Telephone numbers with mailable addresses have higher contact and cooperation rates, so oversampling these numbers translates into the ability to direct resources to harder-to-complete cases (Brick, Judkins, Montaquila, and Morganstein, 2002).

Recommend

To implement the two-phased sampling approach, first select the sample as usual. Next, conduct the purging process and then send the non-purged records to a commercial vendor to obtain addresses. Next subsample the telephone numbers from the non-purged, non-mailable telephone numbers.

Retain the subsampling factor so that proper household base weights can be made to account for the nonmailable cases that were deselected. For example, suppose three out of four nonmailable cases are retained in the subsampling approach, and then adjust the household base weights of the retained nonmailable cases by a factor of 4/3.

Benefit

Resources can be redirected into other ways to improve the sample, improve response rates, and work on hard-to-complete cases.

3.9 Ever Refused Analysis

Table 4 shows the ever refused variable (EVERREF) for the total, eligible, and completed cases at the household level. The table also shows the number of times that cases refused and whether a case ever refused. Only two households were completed at the third refusal conversion attempt.

The estimates of some key statistics for cases that never refused were compared to cases that ever refused (Table 5). Though most of the differences are small, this reinforces that refusal conversion is needed and is the proper protocol. Westat generally attempts up to two refusal conversions.

Recommend

Consider dropping the third refusal conversion, and consider the next recommendation under 'Subsample Refusal Prior to 2nd Attempt'.

Benefit

The significant cost savings can be allocated to more cost effective areas of the survey (e.g., advanced letters, refusal letters, better data on mailables, more sample in one region).

Table 4. Ever refused by total, eligible, and completed households: 2006

	Total	Eligible	Complete
EVERREF			
Missing	35,574	13,455	5,669
1	4,963	4,017	1,160
2	5,748	5,487	251
3	4,417	4,415	2
Times refused			
Zero	4,417	13,455	5,669
One	15,128	13,919	1,160
Two	10,165	9,902	251
Three	4,417	4,415	2
Ever refused			
No	35,574	13,455	5,669
Yes	15,128	13,919	1,413

SOURCE: Office of Financial Management, Washington State Population Survey (WSPS) 2006.

Table 5. Comparison of the Ever Refused: 2006

Characteristic	Overall estimate	Never refused	Ever refused	Chi-square	
				Statistic	p-value
Insured	90.7	90.8	90.5	0.14	0.7097
Received Medicare	13.0	13.1	12.5	0.30	0.5823
Received Medicaid	13.9	13.3	16.3	4.79	0.0287
Education (25 and over)					
Less than high school	6.4	6.3	6.8	9.47	0.0234
High school	24.0	23.8	24.8		
Some college	29.7	29.0	32.3		
College graduate	39.8	40.8	36.0		
Race					
Hispanic	8.9	8.6	9.9	4.70	0.4064
White	76.8	77.2	75.2		
Black	3.4	3.0	4.5		
American Indian/Alaskan Native	1.4	1.4	1.7		
Native Hawaiian/Other PI	1.5	1.5	1.8		
Asian	5.2	5.6	4.0		
Others	2.8	2.8	2.7		
Labor force status (15 and over)					
Employed	61.5	62.0	59.7	2.18	0.3221
Unemployed	4.4	4.2	5.2		
Not in labor force	34.0	33.7	35.1		
Household income as percent of Federal Poverty Level					
0-99%	8.9	8.1	11.9	7.77	0.0670
100-199%	14.7	14.4	15.4		
200-299%	15.2	15.4	14.3		
300-399%	15.7	15.9	15.1		
400% and over	45.5	46.2	43.3		
Own/rent					
Rent	23.7	23.2	25.4	1.06	0.3025
Own	76.3	76.8	74.6		
Household type					
Husband-wife family household	70.1	70.0	70.8	3.68	0.1829
Other family household	18.6	18.2	19.8		
Non-family household (2 or more persons)	1.8	1.8	2.0		
Single-person household	9.4	10.0	7.3		

NOTE: Detail may not sum to totals because of rounding.

SOURCE: Office of Financial Management, Washington State Population Survey (WSPS) 2006.

3.10 Subsample Refusal Prior to 2nd Attempt

The WSPS had over 10,000 cases submitted for refusal conversion. In many cases, there were 3 attempts. The industry standard is one attempt, although two attempts are common at Westat. Also, the refusal conversion rate of 8 percent for WSPS is relatively low.

Recommend

We suggest that one refusal conversion attempt be applied to all initial refusals, except hostile and profane refusals. Then select a subsample of cases (rate = 0.60) prior to the second refusal conversion attempt. The random subsample can be done right after the initial selection of telephone numbers, thus creating a flag that identifies cases to work the 2nd attempt. The cases selected for the 2nd refusal attempt are released into the field first. Then the other 40 percent, ones for which a second attempt is not done, are released a couple of weeks later. It will be important to take into account the subsampling rate when computing base weights for households. For example, with a subsampling rate of 0.60, multiply the household base weight by 1.00/0.60 for only the cases that are retained for the 2nd refusal conversion attempt.

Benefit

Subsampling refusals can be used to reduce the total calendar time required to conduct the WSPS survey. It has the potential to reduce data collection costs and improve response rates. These potential benefits need to be evaluated more carefully before we can be more precise about them.

3.11 Providing a Toll-Free Number for Respondents to Call to Complete Survey

For the 2006 survey, a toll-free number was provided for respondents to call if they had questions about the survey, but this could not be used to actually complete the survey.

Recommend

Provide (in the prior notification letter) a toll-free number that respondents may use to dial in to complete the survey.

Benefit

In addition to the marginal cost savings associated with these inbound calls (i.e., due to interviews being completed as a result of the respondent calling in to complete the interview *before* any further outbound calls are attempted), it is likely that providing a toll-free number the respondent can call to complete the survey also serves to add legitimacy to the survey.

3.12 Classification of Cases for Refusal Conversion

For the 2006 survey, the Data Collection Report states that:

The purpose of refusal conversion is to improve the response rates for the study and decrease the non-response error. In addition, individuals who first refuse may have different attitudes and behaviors from more cooperative respondents. Conversions were attempted on soft refusals – that is, cases in which a mild or polite refusal occurred. Stronger refusals and those requesting to be placed on a "Do Not Call" list were not re-contacted, as IRB requires that participation in the interview be voluntary.

Further, the data in Table 6 shows that the households on the "Do Not Call" list are different from the eligible; therefore, motivating the need to re-contact. The auxiliary data shown here are characteristics shown in the 2006 NRBA report to have potential for bias.

Table 6. Comparison of the Do Not Call List refusals: 2006

Characteristic	Eligible	Do not call	Chi-square	
			Statistic	p-value
Metro status flag				
Metropolitan statistical area	81.6	84.5	16.31	0.0001
Not in metropolitan statistical area	18.4	15.5		
Percentage in exchange that are college graduates				
Low: less than 22.9	30.9	25.7	16.96	0.0002
Medium: 22.9 to 31.9	33.3	34.3		
High: greater than 31.9	35.8	39.9		
Percentage renters				
Low: less than 26.9	32.6	37.7	14.19	0.0008
Medium: 26.9 to 36.2	30.8	28.6		
High: greater than or equal to 36.2	36.6	33.7		
Percentage Income 100K and up				
Low: less than 11.8	25.5	20.3	28.24	0.0000
Medium: 11.8 to 18.2	35.9	35.6		
High: greater than 18.2	38.5	44.1		
Percentage income between 1K and 10K (inclusive)				
Low: less than 5	37.6	43.8	23.36	0.0000
Medium: 5 to 7.9	35.1	34.2		
High: greater than 7.9	27.3	22.0		

NOTE: Detail may not sum to totals because of rounding.

SOURCE: Office of Financial Management, Washington State Population Survey (WSPS) 2006.

Recommend

We recognize that there may be IRB constraints, but would be interested in learning more about how refusals are classified. Our standard approach attempts refusal conversion on firm refusals (as well as mild refusals), just not on hostile, abusive or threatening refusals. With regard to the "Do Not Call" Registry, the Federal Trade Commission exempted survey research from the restrictions imposed by the Do Not Call Implementation Act. Thus, Westat will provide appropriate responses to a respondent's comment that he/she is on the "Do Not Call list", which may convert that respondent.

Benefit

In addition to the refusal conversion letter recommended under ‘Refusal Conversion Letter’, this would be expected to positively affect response rates. In a national survey conducted by Westat earlier this year, a refusal conversion rate (with up to two refusal conversion attempts) of over 30 percent was attained. Reaching these cases will reduce the bias due to nonresponse beyond what the weighting adjustments could do.

4. RECOMMENDATIONS FOR REDUCING NONRESPONSE OR COVERAGE BIAS

4.1 Eligibility Questions

The introduction asks to speak to a person *currently living at this residence*, then asks if they own more than one home… if the response is a single home, the state residence questions are skipped. We are concerned that it is confirmed that the person is a current resident of the *home*, but not whether the person is a current resident of the *state*.

INT04

> The person I need to speak with is the person currently living at this residence who is <<18 years of age or older>> [Suggest replacing words between <<and>> with "at least 18 years old"] and who knows the most about the household finances. Would that be you or someone else?

INT07

> Thank you for being willing to speak with me today, but we are only interviewing Washington residents.
>
> [So people with more than 1 residence whose primary residence is in another state get screened out. But people with a single residence that may be in another state do not.]

Recommend

Ask county of residence even for those with just a single home, unless WSPS derives eligibility (whether a WA resident or not) based on zip code.

Benefit

Accurate coverage of the target population will result.

4.2 Quotas

It is our understanding that quotas were established and interviewing stopped when the quotas were reached. Further, release groups of 200 cases were formed. However, as seen through data sent to us, those coded as Over Quota were attempted (261 cases). The Over Quota cases were treated as ineligible.

Recommendation

Work all cases released thoroughly and completely. A quota sampling approach deviates from a probability sample. The impact of quota sampling is that the cases that were released, but not worked (or treated as ineligible as in WSPS), were not given a chance of selection, thus creating a sampling bias.

Benefit

Adherence to probability based sampling reduces or eliminates sample bias, undercoverage or other nonsampling error. It also increases credibility with regards to statistical inferences concluded from the WSPS data.

4.3 Tabulations

It is our understanding that for household level estimates, the tabulation uses the person weight related to the reference person from each household. This is not a widely accepted approach since the person weight has been adjusted for the person characteristics (such as age, gender, and race) of all individuals of each reporting household.

Recommend

WSPS may choose to use a nonresponse adjusted household weight for tabulations. A step further is to poststratify the nonresponse adjusted household weight using the reference person characteristics (such as race).

Benefit

Use of the household level weight more accurately estimates household items.

4.4 Interviewer Awards/Incentives

For the 2006 survey, interviewers were given awards/incentives for completed cases.

Recommend

We would caution against giving awards/incentives for completed cases. Instead, we suggest rewarding interviewers based on other measures of productivity and quality.

Benefit

The overall benefit could be positive effects on both response rates and interviewer morale. For example, refusal converters have a very important, challenging task and would be expected to get relatively fewer completes due to the nature of their cases. Additionally, if there is an incentive for completed cases, an interviewer may push too hard on initial contact making refusal conversion on that case more difficult.

4.5 Cell-Only Households

For the 2006 survey, the Data Collection Report states that:

"There is a growing segment of the population that is moving away from using "landline" telephones and is completely dependent upon cell phones for personal telephone communication. This segment is not included in the sampling frame for general RDD surveys because they do not use a traditional home telephone. Gilmore Research is working with the Washington State DOH and the Centers for Disease Control in an experiment to understand the benefits and drawbacks to conducting random surveys among cell phone only households. Without more information we are not recommending a modification in the sampling of the SPS to include cell phones, but we would like to share with OFM what we learn from this experiment as plans are made for future studies."

Recommend

We agree that it is worth considering supplementing the sample with a sample of cell phone numbers. Westat has been engaged in research on this topic (see, for example, Brick et al. 2007) and has followed the research conducted by other organizations, such as Arbitron, on this topic. We recommend that in considering whether to include cell phone numbers in the WSPS sample, OFM also consider the results of these other studies (in addition to the study Gilmore Research is undertaking).

Benefit

A broader review of studies on this topic may shed additional light on both statistical and operational aspects of sampling cell phone numbers.

4.6 Disposition Codes

Table 7 shows "ever refused" for specific disposition codes. The standard practice is that cases coded as final no answer or answering machines are assigned the same result on every attempt. Attempting refusal conversion on any attempt implies that human contact was made at some point in the calling sequence and residential status established.

Recommend

If human contact is made on any attempt, do not code the household as no answer or answering machine no matter the result of the final attempts. Once human contact is made, classify the household as residential unknown eligibility.

Benefit

When coded appropriately, this will reduce the amount of coverage bias in the estimates, since a portion of the answering machines and no answers are estimated to be residential.

Table 7. Ever refused for specific disposition codes: 2006

Sample disposition	DISPOS	EVERREF	Frequency
Answering Machine	5	missing	1541
		1	819
		2	257
		3	5
Spanish Answering Machine	34	missing	28
		1	12
		2	7
No answer	4	missing	2548
		1	341
		2	89
Spanish No answer	35	missing	25
		1	13
		2	6
Busy Signal	6	missing	628
		1	110
		2	34
		3	2
Over quota	81	missing	77
		1	150
		2	33
		3	1
Blocked call	15	missing	390
		1	38
		2	12

SOURCE: Office of Financial Management, Washington State Population Survey (WSPS) 2006.

4.7 Multiple Telephone Number Adjustment

Households having multiple telephone numbers for technologies such as personal computers and fax machines have become common. If a household has more than one telephone number used for residential purposes, the household could have been sampled on any of these numbers which has sampling and weighting implications. In a recent survey, about 5.5 percent of completed screeners received weight adjustments for having multiple phone numbers.

Recommend

Ask additional questions to determine the number and usage of telephone numbers in each household (Roth, Montaquila, and Brick, 2001). As recommended in the paper, add a weighting step to

adjust the weight by the number of eligible telephone numbers in a household, as determined by the survey responses to the additional questions. Based on an evaluation reported in Massey and Botman (1988), the adjustment was capped at 2 (i.e., divide by 2 if the household has two or more eligible phone numbers). But in recent years, the adjustment has been capped at 3 (i.e., if exactly two eligible phone numbers, divide by 2; if three or more eligible phone numbers, divide by 3).

Benefit

A multiple telephone number adjustment accounts for the multiple chances of selecting the household and reduces the associated bias.

4.8 Imputation

It is our understanding that missing values are imputed for the following variables: age, race, sex, insurance related items, labor force, and income. Many of these items are constrained through relationships with others in the households or with reported values from other highly correlated items. Some resulting imputations are logically deduced or drawn from a uniform distribution within tightly constrained boundaries. The hotdeck approach is used to fill in remaining missing values for several variables from donors with the same values of other data items.

Recommend

Typically, some modeling is used to identify covariates for use in imputation. Classification algorithms or regression models are two approaches to identify key covariates. A classification algorithm can identify pockets of the population that have unique characteristics in terms of what is being imputed (e.g., high incomes in an area with a high concentration of young adult renters in a particular region). The main objective is to use correlated variables to help define the hotdeck imputation cells. For one approach, after logical imputation steps, a classification algorithm (such as Search) can be used to create hotdeck cells. The algorithm will create cells that best explain the variation in the variable to be imputed. The algorithm will determine the cells from a long list of predictor variables (some of which may not be used), and it can be based on a minimum cell size. Variables such as Region and Metropolitan Statistical Area (MSA) status may be useful in imputing income, labor force, race, age, and so forth.

To reduce the amount of attenuation in resulting correlations between variables, a sequential hotdeck approach can be used. For example, income may be useful to impute age or race. Typically an

imputation attempt will start with the variable (call it item A) with the least amount of missing data for which key covariates (i.e., complete items such as frame variables) are identified. As stated above, classification tree algorithms or regression modeling can be used to search for the best set of auxiliary variables for each survey item to be imputed. These key predictor variables are used to group the records into n-partitions and hot deck imputation is processed[1]. Then the process is repeated for the next item (B), which can now use item A as a predictor. The exhaustive process continues for all items to be imputed. While widely accepted, this sequential approach falls somewhat short of maintaining the covariance structure among survey items, unless a lot of time and effort is used in modeling relationships and finding key variables to include in defining hotdeck cells for each variable to be imputed. However, it is better than a non-sequential approach.

Benefits

The benefit of using the classification algorithm is increased precision of the imputed values and a sequential approach reduces the amount of attenuation in correlations. The use of the classification algorithm may allow for more variables to be considered and to contribute to the cell formation, while maintain stopping rules on the number of cases in each hot deck cell.

[1] Alternative approaches to the hot deck have included nearest neighbor, predictive mean matching, regression imputation

5. RECOMMENDATIONS FOR SURVEY STRUCTURE

5.1 Increase the Sample Size in One Region

The variability of an estimate from a survey can be described by using the *design effect*. The term "design effect" is used to describe the variance of a sample estimate under a particular sample design relative to the variance of the estimate under a simple random sample with the same sample size. Design effects are used to evaluate the efficiency of the sampling design and estimation procedure utilized to develop the estimates. The estimated design effect due to the differential sampling rates is 1.35, which is due mainly to the low sampling rate in the Other Puget Metro region.

Recommend

By adding sample to this region or reallocating sample from other regions (which may not be feasible to hit targets) the design effect can be reduced substantially. For example, if the sample size in the Other Puget Metro region was increased from 5,559 to 8,000 cases, the estimated design effect drops to 1.18 (13% decrease in variance). In general when designing an RDD survey, initial sample sizes are determined based on precision requirements and expected residential and response rates by region (in this case). While this may produce an efficient design by region, it must be balanced against the precision requirements of the state estimates, which means having similar sampling rates for each region. Balancing these needs and their cost implications is rarely easy, and we are available to assist you. We also understand you may split Other Puget Metro in three parts. We recommend setting a minimum sample size based on precision and then allocate proportionally to the population.

Benefit

Reducing the design effect also reduces the standard errors on estimates. This in turn improves the ability to detect differences by increasing the effective sample size.

5.2 Call Scheduling and Treatment of Cases

Based on information provided in the 2006 Data Collection Report and the OFM Callback Priorities Document, it appears case calling priorities are based only on the number of prior attempts

(with the exception of Spanish language cases), and it also appears there may be a fair amount of manual managing of cases.

Recommend

We recommend considering increased automation of priorities in the call scheduler. (With the caveat that there may be more automation than we have been able to garner from the documentation provided to us.) For example, rules could be built into the scheduler to ensure some balance in the days of the week and times of day that calls are attempted on a case. Additionally, priorities could be adjusted so that, for example, cases with prior human contact would have higher priority than cases with no prior human contact.

Benefit

Building such rules into the scheduler might reduce the amount of manual review and handling of cases required, and also helps to ensure that attempts are distributed across days of the week and times of day. This might also allow for cases to be worked more efficiently.

5.3 Reserve Sample

The 2006 Data Collection Report states that:

Of the 51,202 telephone numbers we received, 42,649 were released into the field during the first 10 weeks of calling. In order to reach the goal for all the regions except Region 7, an additional 8,553 telephone numbers were ordered from SSI. Of the 8,553 numbers received, 494 of these numbers were not released during the survey because they were not needed to reach the regional goals.

Recommend

Westat understands that attempts were made to order enough initial sample to cover the target sample sizes, and that an unusual shortfall occurred that led to a supplemental sample. Westat typically orders a slightly larger sample of telephone numbers than anticipated (i.e., building in about a 10 to 20% reserve), and subsample in advance to designate a particular reserve wave. We release the basic sample (i.e., the part we expect to need) in waves (where a wave is a random subsample), to control the

yield. The cost is marginal for the extra sample, and in some cases, the reserve sample is not released. If WSPS approaches it differently, we recommend following guidelines stated above.

Benefit

For a very low cost, a reserve sample will be on hand for release if needed. The sample can be released much more quickly than if a new sample of telephone numbers had to be ordered. Additionally, the sample is "cleaner" in that it is a single sample from the survey population, rather than two independent samples, selected at different rates and possibly containing duplicates.

5.4 Item Nonresponse Reports

It is our understanding that item nonresponse rates are not computed and reported on.

Recommend

That item nonresponse rates be produced and reviewed. Typically, if the item response rate is less than 70 percent, the item reported data should be scrutinize by running crosstabs on the item response indicator with other variables to determine if there is potential for bias, then caution users. This should be done even for items that are imputed.

Benefit

Data users can understand the quality of the data, and make decisions about its usefulness. Problematic items may lead to discussions on how to improve the item response rate, perhaps through rewording the question or substituting it.

5.5 Pretest Survey Instrument

For the 2006 survey, the Data Collection Report states that:

The telephone questionnaire pretest was conducted between January 12 and January 21, 2006, completing 100 interviews with actual respondents.

Recommend

If substantial revisions to the questionnaire are made for the 2008 survey (from the version that was used in the 2006 survey); an independent pretest sample should be selected and not folded in with the main study sample. Otherwise, treat this as a "predictor sample" that is part of the main study sample, as was done in 2006.

Benefit

The American Association for Public Opinion Research (AAPOR) includes as one of its "Best Practices" for survey research the pretesting of survey instruments and procedures. In order for a pretest to be effective, there must be sufficient time to modify the instruments and procedures between the pretest and the main data collection.

5.6 Enhance Quality Control Procedures

The following are some recommended QC procedures relating to data collection, sampling, and weighting. There are most likely recommendations listed that are already in place for WSPS. It is our understanding with regard to data collection that WSPS supervisors and project managers listen to interviews, review frequency reports at end of field period and early on in surveys, and review weekly data collection updates. For sampling and weighting, it is our understanding that more checks should be done.

Recommended for Data Collection

- Ensure that report numbers regarding frequencies of final disposition codes are derivable from the survey control data file. This provides a quality control and audit trail system that facilitates a clean and efficient process.
- Compare the percent distribution of disposition codes from prior year on a weekly basis.
- Compare various interim rates to past years and use to predict final sample yield and final rates.
- Monitor other aspects of each interviewer's work to detect and correct weaknesses as quickly as possible, such as the inefficient use of time or resources.

- Maintain a communication system to assure that all problems are addressed quickly and uniformly. This system assures complete documentation of each problem and its resolution.

Benefits

Early signs provide an opportunity to identify and correct. Also, prediction of final sample yields allow for more timely release of release groups, and perhaps a shorter field period, reducing survey costs.

Recommended for Sampling

- Compare the percent distribution of counts of selected phone numbers among auxiliary variables across survey years.

- Compute a preliminary base weight using the vendor reports on number of working banks and auxiliary data, and make estimates to compare against prior year estimates for subgroups.

Benefit

Obvious errors can be detected prior to data collection.

Recommended for Weighting

- Review the computer code used to calculate and assign weights, checking input file and variable names, and output file and variable names.

- Compare the weighted sample counts and/or totals with the corresponding frame counts, and other control totals whenever available.

- Calculate descriptive statistics on the weights, including extreme values, the sum of the weights, the number of cases with missing weights, average weights within selected and relevant subgroups after each weighting step, and compare to the previous weighting step.

- Compute preliminary estimates to compare to prior year estimates early on to identify and check into any large deviations.

- Check variance estimates. Preliminary estimates of characteristics can be produced for the entire population as well as subdomains.

Benefits

When several stages of adjustment are used to produce final weights, it is essential to check the weighting procedures carefully to correct any errors in the preparation of weighting specifications or computer programming.

5.7 Variance Estimation

It is our understanding that the SAS Proc SurveyFreq is used for tabulations and for estimates of standard errors. SurveyFreq uses a widely accepted approach, called Taylor Series, for variance estimation.

Recommend

Although SurveyFreq accounts for aspects in the sample design when estimating variances, it does not account for the impact that the weighting process has on variances. Replication approaches, such as the paired jackknife, can be used to account for variation due to nonresponse and benchmarking adjustments.

Benefit

For lower response rates, the variance contribution from the weighting process becomes larger; therefore, the resulting estimates reflect the true variance more accurately.

5.8 Capture the Sort Order (for Variance Estimation)

Recommend

The sort order is useful in capturing the systematic sampling approach used in selecting phone numbers.

Benefit

Implicitly, there are benefits to a systematic sampling approach similar to stratification. The sort order allows one to form key variance estimation variables in order to accurately account for this variance reduction.

5.9 Confidentiality Procedures

It is our understanding that there is a de-identification approach used prior to the release of public use data. That is, direct identifiers, such as names, addresses, any other IDs, are removed prior to data release.

Recommend

In addition to de-identification, we recommend the preparation of cross tabulations of reporting variables with other factual data items, and identify any records with a high risk of disclosure due to small cell counts. Also, if other public use data are available, search any external data available and identify whether or not anyone can use such data to match to WSPS in order to identify individuals. Then coarsen (recode, collapse categories, suppress variables) items to lower disclosure risk, then use data perturbation techniques to create more uncertainty to create reasonable doubt as to the derived identity of individuals and households.

Benefits

Such steps in a data confidentiality analysis will help to create a disclosure safe public use file.

5.10 American Community Survey

In 2010 the ACS will be replacing the Census long-form, including health items. It may be considered that the WSPS could reduce sample sizes in 2010 due to the large amount of data available. Some evaluation will be needed to better understand the relationship between WSPS and ACS items, estimates, and timelines.

6. INFORMATION REVIEWED WITHOUT RECOMMENDATIONS

6.1 Advanced Letter

The advanced letter was reviewed and there are no recommendations.

6.2 Editing

We reviewed the edits that were sent to us and find them acceptable. It is our understanding that basically for open-ended questions, changes were made if needed. It does not look as if edits are flagged, which is sometimes requested by our clients. More information about our editing system is provided in the section on 'Other Shared Information'.

6.3 Timing of the Selection of Telephone Numbers

It is our understanding that phone numbers are selected 3 weeks prior to data collection. It is important to minimize this time interval in order to have the most recent phone number data possible. Three weeks is an appropriate time interval to us.

6.4 Computation of Response Rates

Westat reviewed the computation of response rates. Although some recommendations were provided about the mapping of disposition codes in this report, we have no recommendations at this time about the computation of response rates. A more thorough review will occur if time and resources permit.

6.5 Hard and Soft Refusals

We intended to look closely at soft refusals; however, we recall that the coding in the survey control file is inconsistent with published tabulations. We looked at the hostile refusals and found that the degree of consistency in distribution with the general population has mixed results. There is no good way to reach this population, and weighting adjustments are used to help reduce bias to the extent possible.

REFERENCES

Brick, J.M., Judkins, D., Montaquila, J., and Morganstein, D. (2002). Two-phase list-assisted RDD sampling. *Journal of Official Statistics*, 18(2), 203-215.

Brick, J.M., Brick, P.D., Dipko, S., Presser, S., Tucker, C., and Yuan, Y. (2007). Cell phone survey feasibility in the U.S.: Sampling and calling cell numbers versus landline numbers. *Public Opinion Quarterly*, 71(1), 23-39.

Brick, J.M., Montaquila, J., Hagedorn, M., Roth, S., and Chapman, C. (2005). Implications for RDD Design from an Incentive Experiment. *Journal of Official Statistics*, 21(4), 571-589.

Ferraro, D. and Krenzke, T. (2007). Washington State Population Survey: Nonresponse Bias Analysis on the 2006 Survey. Prepared for the Washington State Office of Financial Management.

Massey, J.T. and Botman, S.L. (1988). Weighting adjustments for random digit dialed surveys. In: Groves, R.M.; Biemer, P.P.; Lyberg, L.E.; Massey, J.T.; Nicholls II, W.L.; Waksberg, J. (eds). 1988. *Telephone Survey Methodology*. New York: John Wiley & Sons. Pp. 143-160.

Roth, S.B., Montaquila, J., and Brick, J.M. (2001). Effects of telephone technologies and call screening devices on sampling, weighting and cooperation in a random digit dialing (RDD) Survey. *ASA Proceedings of the Joint Statistical Meetings.*