Attachment 9. Address-based Sampling Frame

RTI has conducted a significant body of research on the use of address-based frames for household surveys, particularly regarding issues of coverage compared with traditional "counting and listing" approaches. (Iannacchione, Staab, and Redden, 2003; Iannacchione et al, 2007; McMichael, Ridenhour, and Shook-Sa, 2008; Iannacchione et al, 2010; Shook-Sa et al, 2010; Staab and Iannacchione, 2003). Our research has shown that the coverage of the household population in North Carolina using mailing addresses is comparable to the coverage of counting and listing in urban areas, but is somewhat lower in rural areas (Iannacchione et al, 2007).

Currently, the sampling frame for the NCVS is maintained by the U.S. Census Bureau. As such, it is subject to Title XIII restrictions which do not allow it to be shared with research contractors. In contrast, mailing addresses are offered to the public by the U.S. Postal Service (USPS) through a nonexclusive license agreement with qualified private companies. One such company is Valassis Direct Mail, Inc. In July 2010, the Valassis Lists product accounted for all but 35,000 of the more than 137 million residential mailing addresses on the U.S. Postal Service Computerized Delivery Sequence (CDS) File. The CDS File contains all postal delivery points serviced by the USPS. In addition to the CDS file, the USPS makes available the No-Stat file, a file of mailing addresses that supplement the CDS file with both active and vacant addresses that are excluded from the CDS file.

Selecting the Sample. The sampling frame will be comprised of addresses from the CDS and No-Stat files. We will select a two-stage sample of addresses. In the first stage, we will select a probability proportional to size (PPS) sample of 64 primary sampling units (PSUs). Primary sampling units (PSUs) are five-digit ZIP codes and the size measure is the number of addresses in the PSU. After selecting PSUs, we will randomly assign them to one of the four mode and incentive conditions. In the second stage of sample selection we will select a sample of addresses from each sampled PSU and append phone numbers to as many sampled addresses as possible. All English-speaking adults residing at the selected addresses will be eligible to complete the interview.

Although it is not unreasonable to assume that virtually every household in the United States has a mailing address, not all mailing addresses are suitable for in-person household surveys because interviewers must be able to locate a mailing address "on the ground." Households with city-style[1] mailing addresses are considered locatable for in-person household surveys and constitute the vast majority of elements on the CDS file. Households with mailing addresses that are not locatable include those with simplified rural addresses[2] and households that only receive mail through residential Post Office (P.O.) Boxes. In addition, the CDS file contains some addresses that are incomplete. Drop points are addresses where mail is delivered to a single location for multiple units. The CDS file contains the drop point address and the number of drop units but does not include drop unit descriptors. The No-Stat file contains drop-unit descriptors for a

[1] A city-style mailing address contains a street name and number as well as city, state, and ZIP Code.
[2] A simplified rural address does not have a street address. Mail delivery is based on the resident's name, city, state, and ZIP Code. Typically, simplified rural addresses are assigned to all households on a rural carrier route.

portion of drop units on the CDS. Thus inclusion of drop points that are not contained on the No-Stat file would require in-field sample selection procedures.

The sampling frame for this study consists of 16,567,614 active, complete, locatable residential mailing addresses within NC, OH, PA, and VA. These addresses were derived from Valassis' May, 2011 CDS and No-Stat Files[3]. We did not include vacant and seasonal addresses on the frame because the target population is limited to occupied households. We excluded the 170,703[4] active drop units for which we could not identify complete drop unit addresses from the frame because including them would require additional field selection procedures.

In order to maintain an EPSEM (equal-probability-of-selection-method) sample, the same number of addresses will be selected from each PSU. To achieve this, each PSU must have at least n / nc addresses, where n = the total sample size, and nc = the number of clusters. For the SCV, n=3,840 and nc=64, so each PSU must contain at least 60 addresses. However, in addition to selecting 60 primary addresses from each PSU, we will select 30 additional addresses as a hold sample. This hold, or supplemental sample, will be released if response rates are lower than expected and additional completed interviews are required. To allow for the selection of a hold sample, and to prevent every address within a PSU from being selected, we set the minimum size as 120 addresses. The 256 zip codes (containing 15,657 addresses) with fewer than 120 addresses were excluded from the frame.

The total number of active locatable addresses excluded from the frame (i.e. drop points without drop unit designators and addresses in zip codes below the minimum size criterion) represents 1.1% of active locatable addresses in the four states.

The number of mailing addresses on the SCV frame compares favorably with the estimated number of occupied housing units in these four states which was 16,423,552[5] in 2010. Table 9-1 compares the number of addresses on the SCV frame to the 2010 Census estimate for the number of occupied housing units for each of the four states in the SCV. It also provides the estimated English-speaking adult household population and the number of zip codes on the frame for each state.

Matching Telephone Numbers to Addresses. Selected addresses will be matched to land line and cell telephone number databases maintained by commercial vendors. Matching addresses to cell number databases will enable us to increase telephone coverage of cell-phone-only households[6]. A two-tiered approach will be used to attain the highest rate of correct telephone number matches. First, the sample will be batched through a telephone number and name appending process. By appending name and telephone number, an additional identifier will be developed to

[3] The two types of active, locatable supplemental addresses contained on the No-Stat file are locatable city-style addresses for P.O. Box throwbacks on rural and highway contract carrier routes and locatable city-style addresses including unit type and number for approximately 20 percent of the units within drop points.
[4] This assumes that all drop units on the No-Stat file correspond to drop units on the CDS file.
[5] Source: 2010 Census.
[6] Blumberg and Luke (2011) report that persons living in over 45 percent of U.S. households have little or no chance of being contacted for a landline telephone survey: http://www.cdc.gov/nchs/nhis.htm.

ensure that the end results are active telephone lines. This batch process will match the sampled address with the occupants currently believed to be associated with that address.

Table 9-1. SCV State Summary

State	English-Speaking Household Population 18+[1]	Occupied Housing Units[2]	Total Addresses on Frame[3]	Number of 5-Digit ZIP codes on Frame
North Carolina	6,372,180	3,745,155	3,801,620	711
Ohio	8,180,115	4,603,435	4,694,717	988
Pennsylvania	8,886,405	5,018,904	4,976,512	1,277
Virginia	5,362,377	3,056,058	3,094,765	761
Total	**28,801,077**	**16,423,552**	**16,567,614**	**3,737**

[1] 2009 American Community Survey 1-Year Estimates. (B16004, B15001, B07013, B09001)
[2] 2010 Census (H1)
[3] Addresses consist of active, complete, locatable mailing addresses in zip codes above the minimum size criterion

The second tier of the matching process will use the names, telephone numbers, and addresses to conduct a comprehensive telephone search. We will obtain up to three verified telephone numbers for the provided address (including cell phone numbers) as well as the date of the most recent association of that number and person with the sampled address. Using these data, we can determine the most likely current occupants.

Based on previous research, we expect to associate a name and telephone number with between 50 and 60 percent of the sampled addresses. As part of the experiment, match rates will be monitored as well as the proportion of incorrect or nonworking numbers.