OSU PO No. RF01274446
RTI Project No. 0213324

May 16, 2012

# 2012 Ohio Medicaid Assessment Survey

## Sample Design and Methodology

**Submitted To**

**Ohio Colleges of Medicine Government Resource Center**
Attn: Timothy R. Sahr
157 Pressey Hall
Columbus, OH 43210
Telephone: (614) 366-3175
E-mail: timothy.sahr@osumc.edu

**Submitted By**

RTI International
P.O. Box 12194
Research Triangle Park, NC 27709-2194
http://www.rti.org/

**OSU PO No. RF01274446**
**RTI International Project No. 0213324**

This document is submitted in partial fulfillment of the requirements of Ohio State University Research Foundation Purchase Order RF01274446.

RTI International is a trade name of Research Triangle Institute.

# Table of Contents

## Section — Page



## Appendix



# List of Exhibits

## Number — Page

# 1. Objectives of the Sample Design

The 2012 Ohio Medicaid Assessment Survey (OMAS) will be a five-pronged design consisting of the following:

1. A list-assisted random digit-dialing (RDD) sample of landline numbers (base sample);
2. A high, medium, and low incidence African American RDD supplemental sample (African American oversample);
3. An Asian and Hispanic surname-based sample (Asian and Hispanic surname list samples);
4. A simple random sample of cell phone numbers (cell phone sample); and
5. An oversample of households with children (child oversample

## Sampling Plan

The OMAS sampling plan is a probability-based design with known probabilities of selection at each stage of selection. This design allows for inference to be made for the entire state of Ohio, as well as various subpopulations and regions of interest.

As we describe in this section, five separate samples will be allocated to meet the OMAS goals. The design will achieve the desired number of 22,355 completed interviews. For each of the five designs discussed previously, ***Exhibit 1*** summarizes the starting number of phone numbers that will be selected and the desired number of completed interviews for each sample type and with the child oversample split out separately to correspond with the sample sizes by county, discussed in ***Section 1.1.10***. The process of determining the starting number of telephone numbers selected is detailed in ***Section 1.1.11***.

### Exhibit 1. Proposed Sample Sizes by Type of Sample

| Type of Sample | Sample Size from Vendor | Completed Interviews (Eligible Respondents) |
|---|---|---|
| Base landline sample | 256,367 | 9,905 |
| Child oversample—landline | 82,720 | 3,760 |
| African American oversample | 73,012 | 2,400 |
| Hispanic surname sample | 11,538 | 641 |
| Asian surname sample | 12,820 | 641 |
| Cell phone sample | 123,920 | 4,068 |
| Child oversample—cell phone | 26,320 | 940 |
| Total | 586,697 | 22,355 |

## Population of Interest

The target population for the OFHS is the total, noninstitutionalized adult and child population residing in residential households in Ohio. Excluded from this population are adults and children

- in penal, mental, or other institutions;
- living on military bases covered by dedicated central office codes;

- living in other group quarters such as dormitories, barracks, convents, or boarding houses (with 10 or more unrelated residents);
- contacted at their second residence during a stay of less than 30 days;
- without access to a landline or cell phone;
- who do not speak English or Spanish well enough to be interviewed; and
- with physical or mental impairments that prevent them from completing an interview (as defined by the interviewer or by another member of the household), if a knowledgeable proxy is not available.

## Sampling Frames

The landline samples for the OMAS will consist of a random sample of telephone numbers from all current operating telephone exchanges in Ohio. MSG's Genesys system will be used to generate the full set of 100-blocks in Ohio. For the cell phone sample, the Telecorida Local Exchange Routing Guide will be used to identify the cell phone 1,000-banks in Ohio.

## General Sample Design

The 2012 OMAS will be a stratified simple random sample of telephone numbers in Ohio. There will be 105 unique strata in the 2012 OMAS. The sampling frame will first be stratified by type of phone (landline or cell). The landline frame will then be further split into 105 strata. Non-metropolitan counties will each be a stratum (81 strata). Each of the 7 metropolitan counties[1] will be further split into three strata based on the density of African Americans living in the Census tract (21 strata). Furthermore, all listed numbers with an Asian or Hispanic surname will be placed in their own stratum (2 strata). The cell phone frame will be a single statewide stratum (1 stratum).

Given the design of the OMAS, which is described in detail below, we anticipate design effects greater than 1 (i.e., the variance under the OMAS design divided by the variance under an SRS design will be greater than 1 due to clustering from oversampling areas with high concentrations of African Americans and disproportional allocation of sample across strata). In 2008, the total design effect (the design effect across all outcomes) for White non-Hispanics and African Americans was approximately 1.65, and the design effect for the proportion of uninsured White non-Hispanic and African American adults in Ohio was 2.7. For Asians and Hispanics, the total design effect was around 1.25, and the design effect for the proportion of uninsured adults was around 1.5. Based on the 2008 design effects and changes in the 2012 design, we will assume a total design effect of 2.0 and a design effect of 2.5 for the proportion of uninsured adults the entire state and African Americans, and we will assume a total design effect of 1.25 and a design effect of 1.5 for the proportion of the uninsured adults for Asians and Hispanics.

[1] The seven metropolitan counties include Montgomery, Summit, Cuyahoga, Franklin, Lucas, Stark, and Hamilton.

## Base Landline Sample

A random sample of 100-blocks will be selected. This sample will be selected through a list-assisted 1+block RDD method. Thus, we will work with MSG to remove any 100-blocks that do not contain any residential numbers. Based on the total desired number of completed interviews of 22,355, we anticipate obtaining 9,905 completed interviews in the base landline sample. To obtain these completed interviews, we will obtain an initial sample of 256,367phone numbers from MSG.

The initial sample of phone numbers will be stratified by the eight Medicaid Managed Care Regions in Ohio and the counties within the region. Any listed phone numbers associated with an Asian or Hispanic surname will be excluded. These phone numbers will be selected separately as discussed in ***Section 1.1.7***. Because the study's desire to create direct estimates for the Medicaid Managed Care Region, a balanced allocation of 1,275 completed interviews will be allocated to each region. The sample will then be proportionally allocated to counties within Medicaid Managed Care Regions to ensure representation from all 88 counties in Ohio. Within each stratum all phone numbers will have an equal probability of selection regardless of whether they are listed or unlisted. Although listed households have shown a higher propensity to respond, they are fundamentally different from unlisted households. Therefore, although there may be some advantages to oversampling listed households, we think the potential increase in bias is too large.

## African American Oversample

One key goal of the OMAS is to produce reliable probability-based estimates of the African American population. To achieve this, an oversample of telephone numbers in the seven high-density African American counties (Montgomery, Summit, Cuyahoga, Franklin, Lucas, Stark, and Hamilton) will be conducted. The proposed design will achieve an African American estimate with a margin of error (MOE) of +/− 5% by gender and age and a +/− 10% MOE for the seven largest metropolitan counties and family income level. The MOE is based solely on the total number of expected African American completed interviews, and includes 1,750 from the African American oversample, as well as an additional 1,516 completed interviews from the base landline and cell phone samples. Thus, the nominal sample size of African Americans is expected to be 3,316.

***Exhibit 2*** presents the achieved MOE for the estimate of the proportion of uninsured African Americans by key domain assuming a design effect of 2.5 for the expected nominal sample size. The design effect assumption is based on our experience with RDD surveys. The expected proportion of uninsured used for the MOE is based on the 2010 Ohio Family Health Survey (OFHS) estimates.

The African American oversample will not screen out non-African Americans. Based on prior experience, we expect that 25% of those contacted will be non-African American. Therefore, to obtain the targeted 1,750 African American interviews, we will need to complete an additional 600 interviews. To achieve this number, we will allocate 2,400 additional interviews to the seven high-density African American counties, which will require selecting an initial sample of 73,012 telephone numbers. This sample will be selected with the base landline sample. In other words, the base landline sample and

African American oversample will be drawn as a single sample rather than two separate samples to prevent the same phone number from being selected in each sample.

### Exhibit 2. Expected Maximum Margin of Error for African Americans by Domain of Interest

| Domain | | Expected Design Effect | Assumed Prevalence Estimate, % | Effective Sample Size | Nominal Sample Size | MOE (95%) |
|---|---|---|---|---|---|---|
| Gender | Male | 2.50 | 22.80 | 663 | 1658 | 3.20 |
| | Female | 2.50 | 22.80 | 663 | 1658 | 3.20 |
| Age | 0–18 | 2.50 | 6.00 | 597 | 1492 | 1.91 |
| | 19+ | 2.50 | 22.80 | 1326 | 3,316 | 2.26 |
| Family Income | ≤ 100% FPL | 2.50 | 25.80 | 265 | 663 | 5.28 |
| | 101 to ≤ 200% | 2.50 | 32.00 | 265 | 663 | 5.63 |
| | 201 to ≤ 300% | 2.50 | 24.00 | 265 | 663 | 5.15 |
| | 301 to ≤ 400% | 2.50 | 10.40 | 265 | 663 | 3.68 |
| | ≥ 400% FPL | 2.50 | 10.40 | 265 | 663 | 3.68 |
| Region | Cuyahoga County | 2.5 | 22.80 | 287 | 717 | 4.86 |
| | Franklin County | 2.5 | 22.80 | 221 | 553 | 5.54 |
| | Hamilton County | 2.5 | 22.80 | 225 | 562 | 5.49 |
| | Lucas County | 2.5 | 22.80 | 166 | 414 | 6.40 |
| | Montgomery County | 2.5 | 22.80 | 195 | 488 | 5.90 |
| | Stark County | 2.5 | 22.80 | 79 | 197 | 9.31 |
| | Summit County | 2.5 | 22.80 | 128 | 385 | 7.30 |

Because of the desire to produce an African American estimate for each of the seven largest urban counties, a balanced allocation of the African American oversample will be used. However, because the African American population in Start County is only 7.5% (according to the 2010 Census) and the largest concentration of African Americans in a Census tract is 60%, we will allocate less of the oversample to Stark County. Therefore, the design will allocate 300 completed interviews to Stark County and 350 completed interviews to the other six counties (from which we expect 50% of respondents to be African American in Stark County and 75% of respondents to be African American in the other six counties). Each county will then be further stratified into high-, medium-, and low-density African American areas. Current data from Claritas will be used to determine the percentage of African Americans in each phone exchange. Phone exchanges were stratified into three categories (high density, medium density, and low density). The categories were created in such a way to maximize the likelihood of obtaining the desired number of African American respondents while maintaining a reasonable unequal weighting effect. ***Exhibit 3*** presents the allocation of the African American oversample to the concentration strata in each county.

## Asian and Hispanic List Samples

Another goal of the OMAS is to obtain reliable probability-based estimates of Asians and Hispanics residing in Ohio. To ensure this, a random sample of telephone numbers associated with households linked to someone with either an Asian or Hispanic surname will be selected. A two-step

process will be used to create the list of Asians and Hispanics residing in Ohio. First, a database of all listed numbers in Ohio will be generated with associated name and telephone number. Second, a list of all possible Asian and Hispanic surnames will be generated. All persons in the first database with a surname listed in the second database will be included in the Asian and Hispanic lists from which a sample will be drawn.

### Exhibit 3. Allocation of African American Oversample Within Metropolitan Counties

| County | Minority Concentration | Population | AA Population | Sample | Total Completes | Expected AA Completes |
|---|---|---|---|---|---|---|
| Cuyahoga County, Ohio | | 1,280,122 | 385,204 | 10,641 | 350 | 263 |
| | Low | 947,219 | 109,224 | 1,972 | 76 | 9 |
| | Medium | 131,418 | 85,281 | 456 | 18 | 11 |
| | High | 201,485 | 190,700 | 8,212 | 257 | 243 |
| Franklin County, Ohio | | 1,163,414 | 251,766 | 10,701 | 350 | 263 |
| | Low | 1,009,214 | 143,127 | 1,205 | 50 | 7 |
| | Medium | 125,115 | 83,587 | 288 | 12 | 8 |
| | High | 29,085 | 25,052 | 9,207 | 288 | 248 |
| Hamilton County, Ohio | | 802,374 | 211,869 | 10,824 | 350 | 263 |
| | Low | 647,352 | 96,099 | 1,625 | 60 | 9 |
| | Medium | 94,830 | 62,284 | 404 | 15 | 10 |
| | High | 60,192 | 53,485 | 8,796 | 275 | 244 |
| Lucas County, Ohio | | 441,815 | 85,733 | 10,581 | 350 | 263 |
| | Low | 373,922 | 39,658 | 1,317 | 57 | 6 |
| | Medium | 47,029 | 27,512 | 266 | 12 | 7 |
| | High | 20,864 | 18,562 | 8,998 | 281 | 250 |
| Montgomery County, Ohio | | 535,153 | 112,328 | 10,569 | 350 | 263 |
| | Low | 436,467 | 34,904 | 1,329 | 58 | 5 |
| | Medium | 48,390 | 32,482 | 281 | 12 | 8 |
| | High | 50,296 | 44,941 | 8,959 | 280 | 250 |
| Stark County, Ohio | | 375,586 | 29,350 | 8,974 | 300 | 172 |
| | Low | 356,936 | 20,278 | 425 | 15 | 1 |
| | Medium | 9,857 | 3,774 | 73 | 2 | 1 |
| | High | 8,793 | 5,299 | 8,476 | 283 | 170 |
| Summit County, Ohio | | 541,781 | 80,490 | 10,722 | 350 | 263 |
| | Low | 499,763 | 50,181 | 1,557 | 62 | 6 |
| | Medium | 27,703 | 17,456 | 149 | 6 | 4 |
| | High | 14,315 | 12,853 | 9,017 | 282 | 253 |
| Total | | 5,140,245 | 1,156,740 | 73,012 | 2,400 | 1,750 |

Our design proposes achieving an MOE of +/− 5% by gender and age category (0 to 18 and 19 or older) for each ethnicity. ***Exhibits 4*** and ***5*** present the expected number of completed interviews per domain necessary to achieve the desired MOE for Asians and Hispanics, respectively. The MOEs are based solely on the total number of expected Asian and Hispanic completed interviews. This includes 641 completed interviews from each of the surname samples, as well as an additional 282 Asian completed

interviews and 553 completed Hispanic interviews from the base landline and cell phone samples. Thus, the nominal sample size for Asians will be 923 and the nominal sample size for Hispanics will be 1,194.

The MOE assumes a design effect of 1.5, an adult prevalence of the uninsured of 18.8%, and a child prevalence of the uninsured of 6.0%. The design effect and adult prevalence estimates are based on the 2010 OFHS. Furthermore, the MOE is based on the total number of expected interviews among Asians and Hispanics, including interviews obtained from the surname sample, the base landline sample, and the cell phone sample. The design meets the desired precision goals for all subpopulations.

### Exhibit 4. Estimated Maximum Margin of Error for Asians by Domain of Interest

| Domain | | Expected Design Effect | Assumed Prevalence Estimate, % | Effective Sample Size | Nominal Sample Size | MOE (95%) |
|---|---|---|---|---|---|---|
| Gender | Male | 1.50 | 18.80 | 308 | 461 | 4.38 |
| | Female | 1.50 | 18.80 | 308 | 461 | 4.38 |
| Age | 0-18 | 1.50 | 6.00 | 277 | 415 | 2.80 |
| | 19 or older | 1.50 | 18.80 | 615 | 922 | 3.09 |

### Exhibit 5. Estimated Maximum Margin of Error for Hispanics by Domain of Interest

| Domain | | Expected Design Effect | Assumed Prevalence Estimate, % | Effective Sample Size | Nominal Sample Size | MOE (95%) |
|---|---|---|---|---|---|---|
| Gender | Male | 1.50 | 18.80 | 398 | 597 | 3.84 |
| | Female | 1.50 | 18.80 | 398 | 597 | 3.84 |
| Age | 0-18 | 1.50 | 6.00 | 358 | 537 | 2.46 |
| | 19 or older | 1.50 | 18.80 | 796 | 1,194 | 2.72 |

Based on the desired level of precision, 641 completed interviews from Asians and 641 completed interviews from Hispanics in their respective surname samples will be obtained. For each of the ethnic surname samples, screening will be conducted so that only members of the appropriate ethnic group are interviewed. Based on prior experience, it will be assumed that 15% of numbers listed on the Hispanic surname list and 30% of numbers listed on the Asian surname list will be screened out because the contacted number is not associated with a Hispanic or Asian person, respectively. Based on these assumptions a random sample of 11,538 telephone numbers from the Hispanic surname list and 12,820 telephone numbers from the Asian surname list will be selected.

Because a list of all persons with a listed telephone number in Ohio with an Asian or Hispanic surname is being used as a frame, the sample of telephone numbers will be selected by simple random sample. The sample will not be stratified, but rather randomly selected at the statewide level. Therefore, we expect counties with a higher Asian or Hispanic population to have an increased sample in proportion to their Asian and Hispanic populations. Furthermore, because screening will be conducted, persons selected in a surname strata that are contacted, but do not belong to the desired ethnic group will not be

asked to participate in the survey. Therefore, these individuals have a zero probability of selection. Although potential for bias may be introduced, prior OMAS surveys determined that this bias is minimal.

## Cell Phone Sample

The cell phone sample will be a random sample of phone numbers from cellular-dedicated 1,000-banks. The cell phone sample is an important component to the 2012 OMAS design. Based on the latest available data, as of June 2011, 31.6% of all households use only cell phones (Bloomberg and Luke, 2011). Furthermore, an even greater percentage are "mostly" cell phone users, which means that even though they have a landline in their household, our interviewers are likely to only reach them through their cell phone. Studies have shown that cell phone only and mostly cell phone individuals skew toward younger adults. Therefore, it is critical to include a reasonably sized cell phone sample to generate accurate estimates for the state of Ohio. To minimize any potential bias by excluding cell phone respondents, 25.6% of the sample will be allocated to the cell phone sample, which translates into 5,008 completed interviews. The cell phone sample will be an overlapping sample with the landline sample in that we will include those residents that have both a landline and a cell phone. To achieve the desired number of completed interviews, we will select an initial sample of 150,240 cell phone numbers.

## Households with Children Oversample

The OMAS will oversample households with children. The oversample will consist of 4,700 additional completed numbers. The oversample will be allocated such that 3,760 of the interviews will be conducted by landline and 940 of the interviews will be completed by cell phone. The landline and cell phone samples will be selected simultaneously with their respective samples to ensure there is no overlap between the samples. Accordingly, the sample will be allocated to strata in the same manner as the base landline sample and cell phone sample.

Based on census information, 30% of households have at least one child residing there. However, after accounting for the oversample, 45% of responding households are expected to have a child in residence. To achieve this constraint, our design will subsample from households with only adults (i.e., some households with only adults will not be asked to participate in the study). As shown in ***Appendix A***, the subsampling rate for the landline samples will be 77.2% and the subsampling rate for the cell phone sample will be 80.3%. To achieve the child oversample, an additional 82,720 landline numbers and 20,860 cell phone numbers will be selected.

## Expected Number of Completed Interviews per County and Minimum Number of Interviews per County

Under the design, the landline base sample, African American oversample and households with children oversample will be the only portions of the design allocated to each specific county. However, based on the distribution of the population, we can estimate the expected sample yield from the statewide samples (i.e., the cell phone, cell phone with child oversample, Asian, and Hispanic samples). Based on a total sample size of 22,355, our main goal is to be able to produce direct estimates for each of the eight Medicaid Managed Care Regions in Ohio. However, we anticipate that we will also be able to produce

direct estimates for several counties, especially the more metropolitan counties that will have additional samples through the African American oversample. Based on our design, ***Exhibit 5*** presents our expected sample yield by county and sample type. Our design produces similar results to what is estimated in the solicitation. Furthermore, while the design does not allow for county-level estimates for all 88 counties in Ohio, a minimum number of completed interviews is associated with each county to ensure representation from the entire state. Our design sets the minimum number of completed interviews per county at 30 interviews. As seen in ***Exhibit 6***, once all sample types are taken into account, all counties meet the minimum target sample size. To increase the likelihood of achieving the minimum sample sizes, the sample of phone numbers for the landline will allocated such that counties with historically low response rates in the 2008 and 2010 OFHSs will have more phone numbers allocated to them while counties with a relatively higher response rate will have fewer phone numbers allocated to them.

## Exhibit 6. Expected Sample Distribution by County

| County | Base Sample | Landline Child Over-Sample | African American Over-Sample | Asian Surname Sample | Hispanic Surname Sample | Cell Phone Sample | Cell Child Over-Sample | Total Sample |
|---|---|---|---|---|---|---|---|---|
| Adams County, Ohio | 20 | 8 | 0 | 2 | 1 | 10 | 2 | 43 |
| Allen County, Ohio | 103 | 39 | 0 | 5 | 3 | 38 | 9 | 197 |
| Ashland County, Ohio | 43 | 16 | 0 | 3 | 1 | 19 | 4 | 86 |
| Ashtabula County, Ohio | 54 | 21 | 0 | 5 | 8 | 36 | 8 | 132 |
| Athens County, Ohio | 122 | 46 | 0 | 3 | 1 | 23 | 5 | 200 |
| Auglaize County, Ohio | 45 | 17 | 0 | 2 | 1 | 16 | 4 | 85 |
| Belmont County, Ohio | 133 | 50 | 0 | 3 | 2 | 25 | 6 | 219 |
| Brown County, Ohio | 32 | 12 | 0 | 2 | 1 | 16 | 4 | 67 |
| Butler County, Ohio | 262 | 99 | 0 | 22 | 17 | 130 | 30 | 560 |
| Carroll County, Ohio | 23 | 9 | 0 | 1 | 0 | 10 | 2 | 45 |
| Champaign County, Ohio | 45 | 17 | 0 | 1 | 1 | 14 | 3 | 81 |
| Clark County, Ohio | 152 | 58 | 0 | 6 | 5 | 49 | 11 | 281 |
| Clermont County, Ohio | 141 | 53 | 0 | 9 | 6 | 70 | 16 | 295 |
| Clinton County, Ohio | 30 | 11 | 0 | 2 | 1 | 15 | 3 | 62 |
| Columbiana County, Ohio | 240 | 91 | 0 | 5 | 3 | 38 | 9 | 386 |
| Coshocton County, Ohio | 70 | 27 | 0 | 1 | 0 | 13 | 3 | 114 |
| Crawford County, Ohio | 23 | 9 | 0 | 2 | 1 | 15 | 3 | 53 |
| Cuyahoga County, Ohio | 684 | 260 | 350 | 92 | 124 | 452 | 104 | 2,066 |
| Darke County, Ohio | 58 | 22 | 0 | 2 | 1 | 19 | 4 | 106 |
| Defiance County, Ohio | 38 | 14 | 0 | 2 | 7 | 14 | 3 | 78 |
| Delaware County, Ohio | 92 | 35 | 0 | 7 | 4 | 61 | 14 | 213 |
| Erie County, Ohio | 41 | 15 | 0 | 4 | 5 | 27 | 6 | 98 |
| Fairfield County, Ohio | 78 | 29 | 0 | 6 | 3 | 52 | 12 | 180 |
| Fayette County, Ohio | 16 | 6 | 0 | 1 | 0 | 10 | 2 | 35 |
| Franklin County, Ohio | 617 | 234 | 350 | 97 | 85 | 410 | 95 | 1,888 |
| Fulton County, Ohio | 42 | 16 | 0 | 2 | 7 | 15 | 3 | 85 |
| Gallia County, Ohio | 58 | 22 | 0 | 2 | 1 | 11 | 3 | 97 |

(continued)

## Exhibit 6. Expected Sample Distribution by County (continued)

| County | Base Sample | Landline Child Over-Sample | African American Over-Sample | Asian Surname Sample | Hispanic Surname Sample | Cell Phone Sample | Cell Child Over-Sample | Total Sample |
|---|---|---|---|---|---|---|---|---|
| Geauga County, Ohio | 50 | 19 | 0 | 5 | 4 | 33 | 8 | 119 |
| Greene County, Ohio | 179 | 68 | 0 | 11 | 6 | 57 | 13 | 334 |
| Guernsey County, Ohio | 76 | 29 | 0 | 2 | 1 | 14 | 3 | 125 |
| Hamilton County, Ohio | 571 | 217 | 350 | 44 | 30 | 283 | 65 | 1,560 |
| Hancock County, Ohio | 73 | 28 | 0 | 5 | 5 | 26 | 6 | 143 |
| Hardin County, Ohio | 31 | 12 | 0 | 1 | 1 | 11 | 3 | 59 |
| Harrison County, Ohio | 30 | 11 | 0 | 1 | 0 | 6 | 1 | 49 |
| Henry County, Ohio | 27 | 10 | 0 | 1 | 4 | 10 | 2 | 54 |
| Highland County, Ohio | 31 | 12 | 0 | 2 | 1 | 15 | 3 | 64 |
| Hocking County, Ohio | 16 | 6 | 0 | 1 | 0 | 10 | 2 | 35 |
| Holmes County, Ohio | 34 | 13 | 0 | 1 | 0 | 15 | 3 | 66 |
| Huron County, Ohio | 32 | 12 | 0 | 2 | 4 | 21 | 5 | 76 |
| Jackson County, Ohio | 63 | 24 | 0 | 1 | 0 | 12 | 3 | 103 |
| Jefferson County, Ohio | 132 | 50 | 0 | 3 | 4 | 25 | 6 | 220 |
| Knox County, Ohio | 32 | 12 | 0 | 2 | 1 | 21 | 5 | 73 |
| Lake County, Ohio | 123 | 47 | 0 | 12 | 18 | 81 | 19 | 300 |
| Lawrence County, Ohio | 118 | 45 | 0 | 3 | 1 | 22 | 5 | 194 |
| Licking County, Ohio | 88 | 34 | 0 | 8 | 4 | 59 | 14 | 207 |
| Logan County, Ohio | 24 | 9 | 0 | 2 | 1 | 16 | 4 | 56 |
| Lorain County, Ohio | 161 | 61 | 0 | 11 | 39 | 106 | 24 | 402 |
| Lucas County, Ohio | 427 | 162 | 350 | 23 | 47 | 156 | 36 | 1,201 |
| Madison County, Ohio | 23 | 9 | 0 | 2 | 1 | 15 | 3 | 53 |
| Mahoning County, Ohio | 531 | 202 | 0 | 8 | 23 | 84 | 19 | 867 |
| Marion County, Ohio | 35 | 13 | 0 | 2 | 2 | 23 | 5 | 80 |
| Medina County, Ohio | 92 | 35 | 0 | 7 | 7 | 61 | 14 | 216 |
| Meigs County, Ohio | 45 | 17 | 0 | 1 | 0 | 8 | 2 | 73 |
| Mercer County, Ohio | 40 | 15 | 0 | 2 | 1 | 14 | 3 | 75 |
| Miami County, Ohio | 113 | 43 | 0 | 6 | 3 | 36 | 8 | 209 |
| Monroe County, Ohio | 28 | 11 | 0 | 1 | 0 | 5 | 1 | 46 |
| Montgomery County, Ohio | 590 | 224 | 350 | 33 | 21 | 189 | 44 | 1,451 |
| Morgan County, Ohio | 28 | 11 | 0 | 1 | 0 | 5 | 1 | 46 |
| Morrow County, Ohio | 18 | 7 | 0 | 1 | 0 | 12 | 3 | 41 |
| Muskingum County, Ohio | 163 | 62 | 0 | 5 | 1 | 30 | 7 | 268 |
| Noble County, Ohio | 28 | 11 | 0 | 1 | 0 | 5 | 1 | 46 |
| Ottawa County, Ohio | 40 | 15 | 0 | 2 | 4 | 15 | 3 | 79 |
| Paulding County, Ohio | 19 | 7 | 0 | 1 | 2 | 7 | 2 | 38 |
| Perry County, Ohio | 19 | 7 | 0 | 1 | 0 | 13 | 3 | 43 |
| Pickaway County, Ohio | 29 | 11 | 0 | 2 | 1 | 20 | 5 | 68 |
| Pike County, Ohio | 16 | 6 | 0 | 2 | 0 | 10 | 2 | 36 |
| Portage County, Ohio | 130 | 49 | 0 | 6 | 5 | 57 | 13 | 260 |
| Preble County, Ohio | 47 | 18 | 0 | 2 | 0 | 15 | 3 | 85 |

(continued)

## Exhibit 6. Expected Sample Distribution by County (continued)

| County | Base Sample | Landline Child Over-Sample | African American Over-Sample | Asian Surname Sample | Hispanic Surname Sample | Cell Phone Sample | Cell Child Over-Sample | Total Sample |
|---|---|---|---|---|---|---|---|---|
| Putnam County, Ohio | 33 | 13 | 0 | 1 | 6 | 12 | 3 | 68 |
| Richland County, Ohio | 100 | 38 | 0 | 5 | 3 | 44 | 10 | 200 |
| Ross County, Ohio | 42 | 16 | 0 | 5 | 1 | 28 | 6 | 98 |
| Sandusky County, Ohio | 59 | 22 | 0 | 3 | 11 | 22 | 5 | 122 |
| Scioto County, Ohio | 42 | 16 | 0 | 3 | 1 | 28 | 6 | 96 |
| Seneca County, Ohio | 55 | 21 | 0 | 3 | 6 | 20 | 5 | 110 |
| Shelby County, Ohio | 54 | 21 | 0 | 2 | 1 | 17 | 4 | 99 |
| Stark County, Ohio | 303 | 115 | 300 | 16 | 16 | 133 | 31 | 914 |
| Summit County, Ohio | 437 | 166 | 350 | 34 | 21 | 191 | 44 | 1,243 |
| Trumbull County, Ohio | 468 | 178 | 0 | 8 | 9 | 74 | 17 | 754 |
| Tuscarawas County, Ohio | 75 | 28 | 0 | 4 | 3 | 33 | 8 | 151 |
| Union County, Ohio | 28 | 11 | 0 | 4 | 2 | 18 | 4 | 67 |
| Van Wert County, Ohio | 28 | 11 | 0 | 1 | 1 | 10 | 2 | 53 |
| Vinton County, Ohio | 25 | 10 | 0 | 0 | 0 | 5 | 1 | 41 |
| Warren County, Ohio | 151 | 57 | 0 | 17 | 10 | 75 | 17 | 327 |
| Washington County, Ohio | 117 | 45 | 0 | 4 | 1 | 22 | 5 | 194 |
| Wayne County, Ohio | 92 | 35 | 0 | 4 | 2 | 40 | 9 | 182 |
| Williams County, Ohio | 37 | 14 | 0 | 2 | 2 | 13 | 3 | 71 |
| Wood County, Ohio | 121 | 46 | 0 | 6 | 9 | 44 | 10 | 236 |
| Wyandot County, Ohio | 22 | 8 | 0 | 1 | 1 | 8 | 2 | 42 |
| Total | 9,905 | 3,760 | 2,400 | 641 | 641 | 4,068 | 940 | 22,355 |

## Starting Sample Size of Telephone Numbers

In order to achieve the desired number of completed interviews detailed in Exhibit 6, a response ratio factor is applied to the desired number of completed interviews to obtain the starting number of telephone numbers that will be purchased from MSG. The ratios vary by stratum type (i.e., landline, cell phone, surname sample). This average ratio is based on previous OMAS experience. However, based on the 2008 OFHS, we recognize that persons across strata do not respond at the same rate. Therefore, based on the response rates from 2008, the ratio used to determine the starting number of selected phone numbers is adjusted to account for the varying response propensities across strata. The adjustment applied to the average rate is the ratio of the average 2008 response rate and the response rate within the stratum in 2008. For the landline RDD samples (i.e., base landline, African American oversample, landline child oversample) an average response rate of 22:1 is used. For cell phone samples (base cell phone, child oversample), a ratio of 32:1 is used due to lower response rates in cell phones. For the Asian surname sample a ratio of 20:1 is used. For the Hispanic surname sample a ratio of 18:1 is used. The Asian and Hispanic surname samples use different ratios because the accuracy rate in identifying a person in the correct minority group in the Asian surname list is lower than in the Hispanic surname list. ***Exhibit 7*** presents the adjusted ratio and starting sample sizes for each of the 105 stratum.

## Exhibit 7. Adjusted Response Ratios and Starting Sample Size by Stratum

| Stratum | Stratum Description | Desired Completed Interviews | Adjusted Response Ratio | Starting Sample Size |
|---|---|---|---|---|
| 1 | Adams County, Ohio | 28 | 24 | 672 |
| 2 | Allen County, Ohio | 142 | 22 | 3,124 |
| 3 | Ashland County, Ohio | 59 | 21 | 1,239 |
| 4 | Ashtabula County, Ohio | 75 | 24 | 1,800 |
| 5 | Athens County, Ohio | 169 | 21 | 3,549 |
| 6 | Auglaize County, Ohio | 62 | 34 | 2,108 |
| 7 | Belmont County, Ohio | 183 | 23 | 4,209 |
| 8 | Brown County, Ohio | 44 | 20 | 880 |
| 9 | Butler County, Ohio | 362 | 27 | 9,774 |
| 10 | Carroll County, Ohio | 32 | 20 | 640 |
| 11 | Champaign County, Ohio | 62 | 21 | 1,302 |
| 12 | Clark County, Ohio | 210 | 24 | 5,040 |
| 13 | Clermont County, Ohio | 194 | 28 | 5,432 |
| 14 | Clinton County, Ohio | 42 | 20 | 840 |
| 15 | Columbiana County, Ohio | 331 | 24 | 7,944 |
| 16 | Coshocton County, Ohio | 96 | 22 | 2,112 |
| 17 | Crawford County, Ohio | 32 | 21 | 672 |
| 18 | Cuyahoga County, Ohio - Low Density | 774 | 26 | 20,124 |
| 19 | Cuyahoga County, Ohio - Medium Density | 115 | 26 | 2,990 |
| 20 | Cuyahoga County, Ohio - High Density | 406 | 32 | 12,992 |
| 21 | Darke County, Ohio | 80 | 21 | 1,680 |
| 22 | Defiance County, Ohio | 52 | 22 | 1,144 |
| 23 | Delaware County, Ohio | 127 | 25 | 3,175 |
| 24 | Erie County, Ohio | 56 | 27 | 1,512 |
| 25 | Fairfield County, Ohio | 107 | 27 | 2,889 |
| 26 | Fayette County, Ohio | 21 | 24 | 504 |
| 27 | Franklin County, Ohio - Low Density | 789 | 24 | 18,936 |
| 28 | Franklin County, Ohio - Medium Density | 104 | 24 | 2,496 |
| 29 | Franklin County, Ohio - High Density | 309 | 32 | 9,888 |
| 30 | Fulton County, Ohio | 58 | 20 | 1,160 |
| 31 | Gallia County, Ohio | 80 | 23 | 1,840 |
| 32 | Geauga County, Ohio | 68 | 34 | 2,312 |
| 33 | Greene County, Ohio | 246 | 27 | 6,642 |
| 34 | Guernsey County, Ohio | 104 | 22 | 2,288 |
| 35 | Hamilton County, Ohio - Low Density | 695 | 27 | 18,765 |
| 36 | Hamilton County, Ohio - Medium Density | 108 | 27 | 2,916 |
| 37 | Hamilton County, Ohio - High Density | 334 | 32 | 10,688 |
| 38 | Hancock County, Ohio | 100 | 21 | 2,100 |
| 39 | Hardin County, Ohio | 43 | 22 | 946 |
| 40 | Harrison County, Ohio | 42 | 19 | 798 |
| 41 | Henry County, Ohio | 37 | 21 | 777 |
| 42 | Highland County, Ohio | 43 | 21 | 903 |
| 43 | Hocking County, Ohio | 21 | 24 | 504 |
| 44 | Holmes County, Ohio | 47 | 23 | 1,081 |
| 45 | Huron County, Ohio | 44 | 22 | 968 |

(continued)

## Exhibit 7. Adjusted Response Ratios and Starting Sample Size by Stratum (continued)

| Stratum | Stratum Description | Desired Completed Interviews | Adjusted Response Ratio | Starting Sample Size |
|---|---|---|---|---|
| 46 | Jackson County, Ohio | 87 | 20 | 1,740 |
| 47 | Jefferson County, Ohio | 182 | 21 | 3,822 |
| 48 | Knox County, Ohio | 44 | 21 | 924 |
| 49 | Lake County, Ohio | 170 | 33 | 5,610 |
| 50 | Lawrence County, Ohio | 163 | 25 | 4,075 |
| 51 | Licking County, Ohio | 122 | 25 | 3,050 |
| 52 | Logan County, Ohio | 33 | 21 | 693 |
| 53 | Lorain County, Ohio | 222 | 27 | 5,994 |
| 54 | Lucas County, Ohio - Low Density | 555 | 23 | 12,765 |
| 55 | Lucas County, Ohio - Medium Density | 75 | 23 | 1,725 |
| 56 | Lucas County, Ohio - High Density | 309 | 32 | 9,888 |
| 57 | Madison County, Ohio | 32 | 34 | 1,088 |
| 58 | Mahoning County, Ohio | 732 | 26 | 19,032 |
| 59 | Marion County, Ohio | 48 | 21 | 1,008 |
| 60 | Medina County, Ohio | 127 | 29 | 3,683 |
| 61 | Meigs County, Ohio | 62 | 18 | 1,116 |
| 62 | Mercer County, Ohio | 55 | 21 | 1,155 |
| 63 | Miami County, Ohio | 155 | 23 | 3,565 |
| 64 | Monroe County, Ohio | 39 | 34 | 1,326 |
| 65 | Montgomery County, Ohio – Low Density | 721 | 23 | 16,583 |
| 66 | Montgomery County, Ohio – Medium Density | 86 | 23 | 1,978 |
| 67 | Montgomery County, Ohio – High Density | 357 | 32 | 11,424 |
| 68 | Morgan County, Ohio | 39 | 18 | 702 |
| 69 | Morrow County, Ohio | 25 | 24 | 600 |
| 70 | Muskingum County, Ohio | 225 | 23 | 5,175 |
| 71 | Noble County, Ohio | 39 | 20 | 780 |
| 72 | Ottawa County, Ohio | 55 | 20 | 1,100 |
| 73 | Paulding County, Ohio | 27 | 24 | 648 |
| 74 | Perry County, Ohio | 27 | 24 | 648 |
| 75 | Pickaway County, Ohio | 40 | 34 | 1,360 |
| 76 | Pike County, Ohio | 21 | 24 | 504 |
| 77 | Portage County, Ohio | 179 | 29 | 5,191 |
| 78 | Preble County, Ohio | 64 | 22 | 1,408 |
| 79 | Putnam County, Ohio | 46 | 25 | 1,150 |
| 80 | Richland County, Ohio | 138 | 27 | 3,726 |
| 81 | Ross County, Ohio | 58 | 21 | 1,218 |
| 82 | Sandusky County, Ohio | 82 | 22 | 1,804 |
| 83 | Scioto County, Ohio | 58 | 22 | 1,276 |
| 84 | Seneca County, Ohio | 76 | 21 | 1,596 |
| 85 | Shelby County, Ohio | 75 | 23 | 1,725 |
| 86 | Stark County, Ohio - Low Density | 412 | 28 | 11,536 |
| 87 | Stark County, Ohio - Medium Density | 13 | 32 | 416 |
| 88 | Stark County, Ohio - High Density | 293 | 30 | 8,790 |
| 89 | Summit County, Ohio - Low Density | 618 | 25 | 15,450 |

(continued)

## Exhibit 7. Adjusted Response Ratios and Starting Sample Size by Stratum (continued)

| Stratum | Stratum Description | Desired Completed Interviews | Adjusted Response Ratio | Starting Sample Size |
|---|---|---|---|---|
| 90 | Summit County, Ohio - Medium Density | 37 | 25 | 925 |
| 91 | Summit County, Ohio - High Density | 298 | 32 | 9,536 |
| 92 | Trumbull County, Ohio | 645 | 24 | 15,480 |
| 93 | Tuscarawas County, Ohio | 103 | 24 | 2,472 |
| 94 | Union County, Ohio | 39 | 24 | 936 |
| 95 | Van Wert County, Ohio | 39 | 19 | 741 |
| 96 | Vinton County, Ohio | 35 | 34 | 1,190 |
| 97 | Warren County, Ohio | 209 | 30 | 6,270 |
| 98 | Washington County, Ohio | 162 | 20 | 3,240 |
| 99 | Wayne County, Ohio | 127 | 22 | 2,794 |
| 100 | Williams County, Ohio | 51 | 22 | 1,122 |
| 101 | Wood County, Ohio | 167 | 20 | 3,340 |
| 102 | Wyandot County, Ohio | 31 | 21 | 651 |
| 103 | Cell phone | 5,008 | 30 | 150,240 |
| 104 | Asian Surname | 641 | 20 | 12,820 |
| 105 | Hispanic Surname | 641 | 18 | 11,538 |
| | Total | 22,355 | | 586,697 |

### Creation of Sample Replicates

Once each of the samples is selected, the selected telephone numbers will be formed into replicates containing 50 telephone numbers. Sample will be released such that the expected sample yield will be representative of the entire state.

### Selection of Respondents Within a Household

Among the households contacted through a landline, one adult (i.e., person 19 years old or older) will be randomly selected using the modified birthday method. Among those contacted through a cell phone, the owner of the phone (if 19 years old or older) will be selected. Persons contacted on an unexpected phone type (i.e., a landline sample number that is a cell phone or vice versa) will be considered ineligible for the study.

Furthermore, in households with children, one child will be randomly selected. However, rather than having the child complete a survey, a proxy respondent that is knowledgeable about the child will be identified to complete the survey for the child. Ideally, this adult will be the same as the one selected to complete the adult survey, but it can be someone different if the randomly selected adult indicates he/she cannot accurately respond for the child.

### Statewide Precision

Because of the total sample size, it will not be possible to create county-level estimates of children's health insurance status within all 88 Ohio counties. With the inclusion of the child oversample,

estimates for both children and adults will be produced in each of the eight Medicaid Managed Care Regions.

In addition to precision targets for African Americans, Asians, and Hispanics, the OMAS would like to achieve an MOE of +/− 3% at the state level by gender, age category, family income category, and region. Our target MOE assumes a design effect of 2.5, an average adult prevalence of the uninsured of 18.8% (this rate is varied by family income based on the 2010 OFHS), and a child prevalence of the uninsured of 6.0%.Taking into account our proposed sample design, ***Exhibit 8*** presents the expected nominal sample sizes,[2] effective sample sizes,[3] and MOE[4] rates by each of these categories. ***Exhibit 8*** shows that all 18 estimates will achieve the desired MOE.

## Exhibit 8. Estimated Margin of Error for State-Level Estimates by Domain of Interest

| Domain | | Population Count | Population Distribution, % | Expected Design Effect | Assumed Prevalence Estimate, % | Effective Sample Size | Nominal Sample Size | MOE (95%) |
|---|---|---|---|---|---|---|---|---|
| Gender[a] | Male | 5,632,156 | 48.8 | 2.50 | 18.80 | 4,366 | 10,914 | 1.16 |
| | Female | 5,904,348 | 51.2 | 2.50 | 18.80 | 4,576 | 11,441 | 1.13 |
| Age[b] | 0–18 | 3,067,126 | 26.6 | 2.50 | 6.00 | 1,070 | 2,675 | 1.42 |
| | 19–34 | 2,173,075 | 18.8 | 2.50 | 18.80 | 1,684 | 4,211 | 1.87 |
| | 35–54 | 3,222,022 | 27.9 | 2.50 | 18.80 | 2,498 | 6,244 | 1.53 |
| | 55-64 | 1,452,266 | 12.6 | 2.50 | 18.80 | 1,126 | 2,814 | 2.28 |
| | 65 and up | 1,622,015 | 14.1 | 2.50 | 18.80 | 1,257 | 3,143 | 2.16 |
| Family Income | < 100% FPL | 1,828,407 | 15.8 | 2.50 | 26.70 | 1,417 | 3,543 | 2.30 |
| | 100 to <=149% | 1,056,539 | 9.2 | 2.50 | 28.00 | 819 | 2,047 | 3.08 |
| | 150 to <=199% | 1,109,668 | 9.6 | 2.50 | 20.00 | 860 | 2,150 | 2.67 |
| | 200 to <=250% | 1,048,326 | 9.1 | 2.50 | 13.40 | 812 | 2,031 | 2.34 |
| | 251 to <=299% | 1,048,326 | 9.1 | 2.50 | 13.40 | 812 | 2,031 | 2.34 |
| | 300 to <=399% | 1,732,768 | 15.0 | 2.50 | 5.33 | 1,343 | 3,358 | 1.20 |
| | >=400% FPL | 3,712,470 | 32.2 | 2.50 | 5.00 | 2,878 | 7,194 | 0.80 |

(continued)

[2] The nominal sample size is the expected number of completed interviews based on a simple random sample design. It is defined as the product total sample size (22,355) and the expected proportion of the population, based on census figures, for the subpopulation of interest. For example, the nominal sample size for males is 22,355*0.488=10,914.

[3] The effective sample size is the expected number of completed interviews after accounting for the complex survey design. It is defined as the nominal sample size divided by the expected design effect. For example, for males the effective sample size is 10,914/2.0=5,457.

[4] The margin of error (MOE) is the product of the standard error and the critical value (for at 95% MOE the critical value is 1.96). The standard error is defined as sqrt[(p*(1-p))/(n-1)] where p is the assumed prevalence estimate and n is the effective sample size taking into account the survey design. For males, the 95% MOE is 1.04%.

**Exhibit 8. Estimated Margin of Error for State-Level Estimates by Domain of Interest (continued)**

| Domain | | Population Count | Population Distribution, % | Expected Design Effect | Assumed Prevalence Estimate, % | Effective Sample Size | Nominal Sample Size | MOE (95%) |
|---|---|---|---|---|---|---|---|---|
| Region[c] | Metropolitan[c] | 6,279,360 | 54.4 | 2.50 | 18.80 | 4,867 | 12,168 | 1.10 |
| | Appalachian[d] | 1,803,217 | 15.6 | 2.50 | 18.80 | 1,398 | 3,494 | 2.05 |
| | Rural non-Appalachian[e] | 1,541,904 | 13.4 | 2.50 | 18.80 | 1,195 | 2,988 | 2.22 |
| | Suburban[f] | 1,912,023 | 16.6 | 2.50 | 18.80 | 1,482 | 3,705 | 1.99 |

[a] 2010 census data

[b] 2010 American Community Survey 1-year estimates

[c] Metropolitan counties include Allen, Butler, Cuyahoga, Franklin, Hamilton, Lorain, Lucas, Mahoning, Montgomery, Richland, Summit, and Stark.

[d] Appalachian counties include Adams, Ashtabula, Athens, Brown, Belmont, Carroll, Clermont, Columbiana, Coshocton, Gallia, Guernsey, Harrison, Highland, Hocking, Holmes, Jackson, Jefferson, Lawrence, Meigs, Monroe, Morgan, Muskingum, Noble, Perry, Pike, Ross, Scioto, Trumbull, Tuscarawas, Vinton, and Washington.

[e] Rural non-Appalachian counties include Ashland, Champaign, Clinton, Crawford, Darke, Defiance, Erie, Fayette, Hancock, Hardin, Henry, Huron, Knox, Logan, Marion, Mercer, Morrow, Ottaway, Paulding, Preble, Putnam, Sandusky, Seneca, Shelby, Van Wert, Warren, Wane, Williams, and Wyandot

[f] Suburban Counties include Auglaize, Clark, Delaware, Fairfield, Fulton, Geauga, Greene, Madison, Medina, Miami, Lake, Licking, Pickaway, Portage, Union, and Wood.

# 2. Design-Based Weights and Post-Survey Adjustments

The design-based weights for an individual selected for the OMAS is the inverse probability of selection of that individual. An individual's probability of selection is based on the OMAS design, which is a three-stage design.

1. First stage: stratified SRS of phone numbers[5]
2. Second stage: subselection of adult only households; all households with children selected
3. Third stage: subselection of adult from landline household; all cell phone frame respondents selected

## Notation

The following notation will be used in document:

$\pi_{h1}$ = first stage probability of selection in stratum h

$\pi_{h2}$ = second stage probability of selection in stratum h

$\pi_{hi3}$ = third stage probability of selection for individual i in stratum h

[5] The OMAS is stratified by landline and cell phone frames. The landline frame is stratified by county, listed numbers with an Asian surname, and listed numbers with a Hispanic surname. The seven urban counties are further stratified by high, medium, and low minority populations.

$\pi_{hj4}$ = fourth stage probability of selection for child in household j in stratum h

$n_h$ = number of phone numbers sampled in stratum h

$N_h$ = number of eligible phone numbers in population in stratum h

$s_h$ = second stage subsampling rate in stratum h for households or individuals without children

$c_j$ = the number of families in household j.

## First Stage Probability of Selection

In the first stage of selection a random sample of phone numbers will be selected within each stratum. Within each stratum each phone number will have an equal probability of selection. Ineligible sampled phone numbers (e.g., non-working numbers, business phone numbers) are identified and removed from the population count. This leads to the resulting probability

$$\pi_{h1} = \frac{n_h}{N_h}$$

## Second Stage Probability of Selection

In the second stage of selection a subsample of households or individuals (in the case of the cell phone sample) without children will be selected. Households or individuals with children will be selected with certainty. This leads to the resulting probability

$$\pi_{h2} = \begin{cases} s_h & \text{if household or individual does not have a child} \\ 1 & \text{otherwise} \end{cases}$$

Where $s_h$ is defined as follows

$s_h = \frac{a_h^e}{a_h}$ where $a_h^e$ is the expected number of households or individuals without a child selected under a SRS in stratum h and $a_h$ is the number of households or individuals without a child selected after accounting for the oversample of households or individuals with children in stratum h[6].

## Third Stage Probability of Selection

The third stage of selection will select an adult respondent for the OMAS. Respondents identified on the landline frame will have one person 19 years or older living in the household selected at random using the nearest birthday technique. Adult (19 years old or older) respondents from the cell phone frame will be selected with certainty. This leads to the resulting probability

[6] See appendix for details on subsampling rate.

$$\pi_{hi3} = \begin{cases} 1/k_j & \text{if respondent i from landline frame} \\ 1 & \text{if person i from cell phone frame} \end{cases}$$

where $k_j$ is the number of adults living in household j.

## Fourth Stage Probability of Selection (Child Sample Only)

The fourth stage of selection will account for the fact that a household may have multiple families with children and the child will be selected only from the family from which the adult respondent is a member. This weight will only be applied to estimate the number of children in households. The resulting probability will be denoted as follows:

$$\pi_{hj4} = 1/c_j$$

## Design-Based Weights

Based on our design there will be separate design-based weights for adults and children.

The design-based weight for adult person *i* is the inverse of the product of the first three stages' probability of selection. In other words,

$$w_{hi} = \left(\frac{1}{\pi_{h1}}\right)\left(\frac{1}{\pi_{h2}}\right)\left(\frac{1}{\pi_{hi3}}\right)$$

The design-based weight for a child is the inverse of the product of all four stages of selection. In other words,

$$w_{hi} = \left(\frac{1}{\pi_{h1}}\right)\left(\frac{1}{\pi_{h2}}\right)\left(\frac{1}{\pi_{hi3}}\right)\left(\frac{1}{\pi_{hj4}}\right)$$

## Post-survey Adjustments

Upon the completion of data collection, several post-survey adjustments will be applied to the design based weights to minimize any potential bias and ensure that estimate represent the target population. These adjustments include

Adjustment for eligibility status
Adjustment for nonresponse
Adjustment for multiple phone numbers
Adjustment for the number of persons within a household (landline only)
Adjustment for dual-frame design (landline and cell phone)
Poststratification to population control totals

The OMAS design is an overlapping dual-frame design. Respondents will be selected independently from the landline frame and the cell phone frame, regardless of their phone use status (i.e., regardless of whether they receive calls only on a landline, only on a cell phone, or on both a landline and a cell phone). This design is often referred to as a "cell-any" design. Because of the overlap between the frames, care must be taken to properly account for dual use respondents (those who receive calls on both a cell phone and a landline). Dual use respondents have multiple ways of coming into the sample, and this multiplicity must be accounted for in the weighting process.

The calculation of base weights, eligibility adjustments, nonresponse adjustments, adjustments for multiple phone numbers, and adjustments for the number of persons within the household (as appropriate) will be performed independently within each sampling stratum within each frame (cell and landline). Although not all surveys implement nonresponse adjustments prior to merging data from multiple frames, it is essential for at least two reasons: there is different auxiliary information available in each frame, and causes of nonresponse and resulting bias are likely different based on sampling frame.

After the creation of the base household weight, each sampled telephone number will be assigned to one of four categories: respondent, nonrespondent, unknown, and ineligible. Because only a portion of the telephone numbers in the unknown category correspond to eligible housing units, we will adjust the weights of unknown telephone numbers on the basis of the screening eligibility rate of telephone numbers with known eligibility status. We will then remove ineligible telephone numbers from the frame and perform a separate nonresponse adjustment for each frame so that the weight of respondents will account for the weights of nonrespondents and the prorated weights of unknowns. The nonresponse adjustment model will include variables that are known about both respondents and nonrespondents (e.g., census data). After making the nonresponse adjustment, we will retain only respondents on the file.

Once the file has been reduced to responding households, a second adjustment will be performed to account for households with multiple telephone numbers (and thus multiple probabilities of selection). To make this adjustment, the nonresponse-adjusted household-level base weight will be divided by the number of residential landlines available at the household for the landline RDD sample (including households with children oversample) and minority oversamples. Similarly, an adjustment will be made to the cell phone sample (including households with children oversample) to account for individuals with multiple cell phones. To prevent excessive unequal weighting effects and the subsequent variance inflation, it is recommended that the number of telephone lines associated with an individual be truncated at a maximum level to be agreed upon between RTI and OSU. In 2008, the maximum level was set at three (Duffy et al., 2008). We will work with OSU to determine if this is still an appropriate number. This adjustment produces the final household level weight and sets the stage for the creation of person-level weights.

After performing the initial weight adjustments, we will classify respondents on the landline and cell frames based on their phone usage (cell phone only, dual users, and landline only). Dual use respondents can be further classified into three categories: cell mostly, landline mostly, and true dual

users (those who receive about half of their calls on landlines and about half of their calls on cell phones). An analysis based on survey responses will be used to determine the appropriate number of phone use categories to incorporate in the weighting process, carefully weighing the benefits and drawbacks from the use of multiple dual service domains. Respondents in the cell phone only and landline only groups could only have come into the sample on a single frame. Respondents with both cell phones and landlines (dual users) could have come into the sample from either frame.

Two different approaches can be used to adjust for this multiplicity. One approach is to use a single-frame estimation technique, where the dual use respondents' probabilities of selection are calculated for both frames. For many studies, this approach is very straightforward and accurate. Alternatively, the weights of dual users can be adjusted with a composite weighting adjustment, as introduced by Hartley (1962). With this approach, the weights of respondents selected on the landline frame would be multiplied by a compositing factor $\lambda$ ($0< \lambda<1$), and the weights of respondents selected on the cell phone frame would be multiplied by $(1- \lambda)$. There are multiple methods that can be used to determine the appropriate compositing factor $\lambda$ (Kennedy, 2011; Xia et al., 2010; AAPOR Cell Phone Task Force). RTI has experience choosing the appropriate value of $\lambda$, and has implemented both sample-driven approaches and approaches that are independent of the sample and stable over time (which minimizes bias and controlling for weight variation). For OMAS, the choice of compositing factor may depend on whether variance or bias is the bigger concern (Brick, 2006; Kennedy, 2011). An analysis of the survey data will be conducted to determine the most appropriate method for the OMAS.

After adjusting the weights of dual users to account for the overlapping design, we will combine the landline and cell phone respondents into a single file and will poststratify the weights to known population totals. We will include phone usage totals for the state of Ohio from the National Health Interview Survey (http://www.cdc.gov/nchs/data/nhsr/nhsr039.pdf) in the poststratification to ensure that persons of all phone usage types are appropriately represented in the sample. In addition to poststratifying by phone usage, we will poststratify to other known population totals such as county, age, gender, education, Medicaid status, and race. The generalized exponential modeling (GEM; Folsom & Singh, 2000), available in RTI's SUDAAN software package, will be used for the poststratification adjustments. GEM will allow us to perform poststratification adjustments for telephone usage and demographic characteristics simultaneously. In addition, it allows fit criteria to be adjusted and weight trimming to occur within a single step so that the impact of trimming and the poststratification adjustment on the unequal weighting effect can be determined. This model can be tailored by collapsing poststratification cells, adjusting model convergence criteria, and adjusting the amount of trimming to minimize the unequal weighting effect while maintaining the statistical validity of the weights.

## Response Rates

The 2012 OMAS will calculate response rates using two approaches. First, the traditional AAPOR RR3 will be calculated. Because of the subsampling of households with no children a screening response rate (SRR) and interview response rate (IRR) will be calculated. Second, a new approach which calculates a separate response rate for adults and children will be calculated. These two rates will be

compared for comparability. Both of these rates will be calculated for landline respondents, cell phone respondents, and all respondents (landline and cell combined).

This section defines how each of these rates will be calculated

The following notation will be used throughout this section

- C: complete interview
- P: partial complete of interview
- R: refused
- U: eligibility unknown
- $e_U$: proportion of eligible

AAPOR RR3

Because screening will occur during the OMAS interview, the AAPOR response rate 3 will be calculated in two steps. First, a screening rate ($SRR_3$) in which the proportion of unknown ineligible is accounted. Second, an interview response rate ($IRR_3$) conditional on getting through the screening. The final response rate ($RR_3$) is the product of the $SRR_3$ and $IRR_3$.

The SRR3 is defined as:

$$SRR_3 = \frac{C + P}{C + P + R + e_u U}$$

The IRR3 is defined as:

$$IRR_3 = \frac{C}{C + P}$$

Thus, the product of these two response rates equals the traditional AAPOR RR3:

$$RR_3 = \frac{C}{C + P + R + e_u U}$$

Separate response rate for households with adults and households with children

For the following, the subscript k is used to denote numbers from households with children, $n_k$ to denote numbers from households without children.

Since the questions are asked in a certain order, there are two levels of eligibility. First, phone line eligible, if this is a residential phone line:

$$e_l = \frac{\text{\# of residential phone lines}}{\text{\# of phone numbers called}}$$

Second, child eligible, if this household has children:

$$e_k = \frac{\#\text{ of households with children}}{\#\text{ of residential phone lines}}$$

So, there are two levels of unknowns:

$U_k$=Unknown child eligible (this is residential phone line, but we don't know if they have children),

$U_l$=Unknown phone line eligible (we don't know if this is a residential phone line)

Now, for the households with children,

$$RR_{3,k} = \frac{C_k}{C_k + P_k + R_k + e_k U_k + e_k e_l U_l}$$

For the households without children

$$RR_{3,nk} = \frac{C_{nk}}{C_{nk} + P_{nk} + R_{nk} + (1 - e_k) U_k f + (1 - e_k) e_l U_l f}$$

where f is the fraction of subsampling of households without children

To get a single RR3, we can pool the above two rates:

$$\frac{C_k + C_{nk}}{C_k + P_k + R_k + e_k U_k + e_k e_l U_l + C_{nk} + P_{nk} + R_{nk} + (1 - e_k) U_k f + (1 - e_k) e_l U_l f}$$

And simple algebra shows that this rate should be in between the above two.

# 3. References

AAPOR Cell Phone Task Force (2010). New considerations for survey researchers when planning and conducting RDD telephone surveys in the U.S. with respondents reached via cell phone numbers. http://aapor.org/AM/Template.cfm?Section=Cell_Phone_Task_Force&Template=/CM/ContentDisplay.cfm&ContentID=2818

Brick, M., Dipko, S., Presser, S., Tucker, C., & Yuan, Y. (2006). Nonresponse bias in a dual frame sample of cell and landline numbers. *Public Opinion Quarterly, 70*(5), 780-793.

Duffy, T., Muzzy, S., Bausch, S., Ferketich, A., Iachan, R., Hannah, K.M., et al. (2008). *2008 Ohio Family Health Survey: Methodology Report.*

Folsom, R. E., & Singh, A. C. (2000). *The Generalized Exponential Model for sampling weight calibration for extreme values, nonresponse, and poststratification*. Paper presented at the Joint Statistical Meetings of the American Statistical Association, Indianapolis, IN.

Hartley H.O. (1962), Multiple- frame surveys. *Proceedings of the Social Statistics Section, American Statistical Association*, pp. 203-206.

Kennedy, C. (2011). *An evaluation of popular weighting approaches in dual frame RDD surveys*. Presented at American Association for Public Opinion Research, Phoenix, AZ.

Xia, K., Pedlow, S., & Davern, M. (2010). Dual-frame weights (landline and cell) for the 2009 Minnesota Health Access Survey. In *JSM Proceedings, Survey Research Methods Section,* Alexandria, VA: American Statistical Association. http://www.amstat.org/sections/srms/proceedings/y2010/Files/308018_60874.pdf

# Appendix A: Subsampling Rates of Households with No Children

Let $m_{h-L}$ and $m_{h-C}$ be the total number of completed interviews in stratum h from the landline frame (L) and cell phone frame (C), respectively. Similarly, let $m_{h-L,C}$, $m_{h-L,A}$, $m_{h-C,C}$ and $m_{h-C,A}$ be the total number of completed interviews in stratum h from the landline frame and cell phone frame among households (or individuals in the cell phone sample) with children and without children, respectively, under a design that oversamples households with children. Also, let $m^e_{h-L,C}$, $m^e_{h-L,A}$, $m^e_{h-C,C}$, $m^e_{h-C,A}$ be the expected number of completed interviews in stratum h from the landline frame and cell phone frame among households (or individuals in the cell phone sample) with children and without children, respectively, under a design that does not oversample households with children (i.e., a proportional design to the number of households with children in the population).

Furthermore, let $p_{C-L}$ and $p_{C-C}$ be the expected proportion of households (or individuals) with a child on the landline and cell phone frames, respectively. Based on experience during the 2008 and 2010 OFHS, we will assume that $p_{C-L} = 28\%$ of households reached through a landline have a child in the house and $p_{C-C} = 35\%$ of individuals reached through a cell phone have a child.

Based on these, the number of completed interviews from the landline frame, when oversampling households with children, can be expressed in terms of the number of completed interviews from households without children, $m_{h-L,A}$, and households with children, $m_{h-L,C}$ as follows

$$m_{h-L} = m_{h-L,A} + m_{h-L,C} = \left[ m_{h-L,B} * \left(1 - p_{C-L}\right) \right] + \left[ \left( m_{h-L,B} * p_{C-L} \right) + m_{h-L,O} \right]$$

Where $m_{h-L,B}$ is the number of completed interviews from the base sample and $m_{h-L,O}$ is the number of completed interviews from the DOH oversample.

Whereas, the expected number of completed interviews from the landline frame, based on a proportional design can be expressed as

$$m^e_{h-L} = m^e_{h-L,A} + m^e_{h-L,C} = \left[ m^e_{h-L,B} * \left(1 - p_{C-L}\right) + m_{h-L,O} * \left(1 - p_{C-L}\right) \right] + \left[ \left( m_{h-L,B} * p_{C-L} \right) + \left( m_{h-L,O} * p_{C-L} \right) \right]$$

Similar expressions can be derived for the cell phone sample.

Based on these expressions, the proportion of households without any children (i.e., $s_h$) that should be kept in order to achieve the desired oversample is

$s_{h-L} = \frac{m_{h-L,A}}{m^{e}_{h-L,A}}$ for the landline sample and $s_{h-C} = \frac{m_{h-C,A}}{m^{e}_{h-C,A}}$ for the cell phone sample.

***Exhibit A-1*** displays the current sample sizes by frame and the resulting subsampling rate.

## Exhibit A-1. OMAS Subsampling Rate of Adult Households with No Children

| Frame Type | Base Sample Size | Oversample Size | Subsampling Rate |
|---|---|---|---|
| Landline | 13,587 | 3,760 | 0.772 |
| Cell | 4,068 | 940 | 0.803 |