Appendix C1

November 14, 2011

MEMORANDUM FOR   Robert Cage
Chief, Revision Planning and Special Projects Branch
Division of Consumer Prices and Price Indexes
Bureau of Labor Statistics

Through: Cheryl R. Landman
Chief, Demographic Surveys Division
U. S. Census Bureau

From: Ruth Ann Killion
Chief, Demographic Statistical Methods Division
U. S. Census Bureau

Final DSMD Signed

Prepared by: Stephen Ash
Chief, Victimization and Expenditures Branch
Demographic Statistical Methods Division

1

Kathlene Garland
Victimization and Expenditures Branch
Demographic Statistical Methods Division

1

Subject: 2011 Telephone Point of Purchase Survey: High-Level Discussion
of Weighting for a Landline and Cell Phone Sample Design

This memorandum discusses how the Demographic Statistical Methods Division (DSMD) calculates sample weights for a Telephone Point of Purchase Survey (TPOPS) sample design that uses samples selected from both a landline and cell phone frame. This discussion of the weighting is intended as a high-level review and is not intended to be a complete specification or requirement. We discuss the *what* and the *how*, but also make sure we explain the *why*.

The discussion begins with a background section that sets up the remainder of the paper. The rest of the paper explains each of the major components of the final sample weights:

- Calculation of Base Weights
- Adjustment for Multiple Phones
- Adjustment for Non-Interviews
- Multiple Frame Adjustment or Combining the Landline and Cell Phone Samples
- Ratio Adjustments to Known PSU Totals
- Ratio Adjustments to Known First-Stage Strata Totals

The final section shows how we put together the base weights and all of the weighting adjustment factors. The memorandum also includes two attachments to help the reader. The first attachment is a glossary of terms used throughout the memorandum. The second attachment is an example that illustrates the two ratio adjustments.

We need to revise the weighting methodology because TPOPS is augmenting the current landline frame with a cell phone frame beginning with the Q122 sample. This is necessary because there is an accumulation of evidence that random digit dialing (RDD) surveys that solely use landline frames are having increasingly poor coverage and differential representativeness. Much of the recent under coverage can be associated with people using cell phones in place of landlines, which is referred to as the "cell phone only" universe (Tucker *et al*. 2004), (Keeter *et al*. 2007) or equivalently the "wireless only" universe (Blumberg *et al*. 2010). We need to include the cell phone only universe in TPOPS because the spending habits of households in the cell phone only universe are most likely different from the spending habits of households in the current landline frame and the general population of interest, the civilian non-institutional population of the U.S.

1. Background to the Sample Design

The background section discusses all of the relevant features of the sample design of TPOPS. The features include the present design and aspects of the new design that uses the cell phone frame. We begin with a discussion of notation that is used throughout the memorandum. We then define the universe of interest and the frames used to enumerate the universe. Next, we discuss relevant aspects of the sample design of TPOPS and some general assumptions about the interviewing.

1.1 Notation

We now define some general notation used in our more technical discussion of the weighting.

U universe of interest

F_{LL} the set of households associated with (or equivalently have telephone numbers on) the landline frame

F_{Cell} the set of households associated with the cell phone frame

U_A the subset of the universe of households that is covered by (or equivalently have telephone numbers on) the landline frame, but not by the cell phone frame – "landline only" households

U_B the subset of the universe of households that is covered by the cell phone frame – "cell phone only" households

U_{AB} the subset of the universe of households that have telephone numbers that are covered by both the cell phone and land line frames -- "dual use" households

W universe of telephone numbers (landline and cell) associated with households

G_{LL} the set of telephone numbers on the landline frame

G_{Cell} the set of telephone numbers on the cell phone frame

W_A the subset of the landline frame that are considered landline only telephone numbers

W_B the subset of the cell phone frame that are considered cell phone only telephone numbers

W_{AB} the intersection of the landline and cell phone frames

1.2 The Universe of Interest for TPOPS

The universe of interest for TPOPS is the civilian non-institutional population of the U.S. Figure 1 provides a representation of the universe of interest and the frames used to enumerate the TPOPS universe.

*Figure 1: Representation of the Universe of Households
in terms of Landline and Cell Phone Frames*
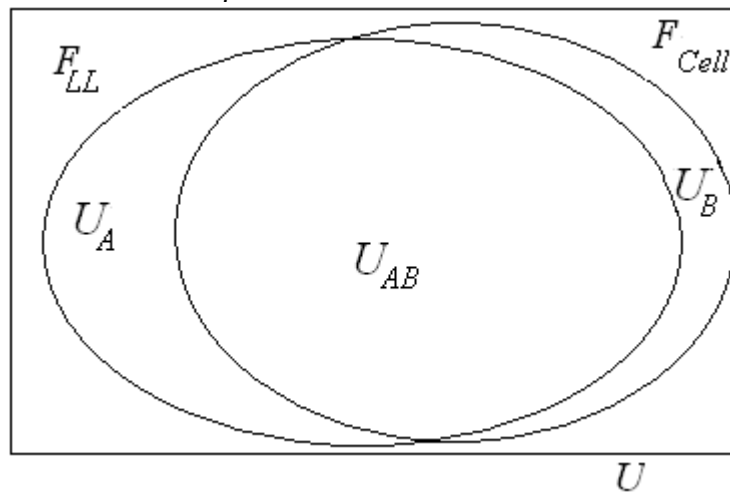
In Figure 1, the outside square labeled with U represents the universe of interest. Within the universe U, there are two ellipses labeled F_{LL} and F_{Cell}, which define the set of households associated with the landline frame and cell phone frame, respectively.

The two frames, F_{LL} and F_{Cell}, have units that are common and not in common. The intersection of the two frames is labeled U_{AB} and represents the subset of households which have both a landline and a cell phone. The subset of the universe labeled U_A represents the landline only households or the households which are on the landline frame, have a landline phone, but do not have a cell phone. Similarly, the subset of the universe labeled U_B represents the cell phone only households or the households which are on the cell phone frame, have a cell phone, but do not have a landline telephone. Under coverage is represented by the units outside the two ellipses but within the rectangle. Not included in Figure 1 is the set of households that have no telephone number.

We note that Figure 1 is similar to the figure provided by Link *et al*. (2007) and Lohr (2009; p. 71).

In Figure 1, a given household is one unit so it is in only one set, for example, a given household cannot be in both U_A or U_B.

Both the landline and cell phone frames are not lists of households, but lists of telephone numbers that are associated with the households. As Wolter *et al*. (2010) explains, we select our sample from the set of telephone numbers associated with the households of the U.S. which is represented in Figure 2.

Figure 2: Representation of the Universe of Telephone Numbers associated with Households in terms of Landline and Cell Phone Frames
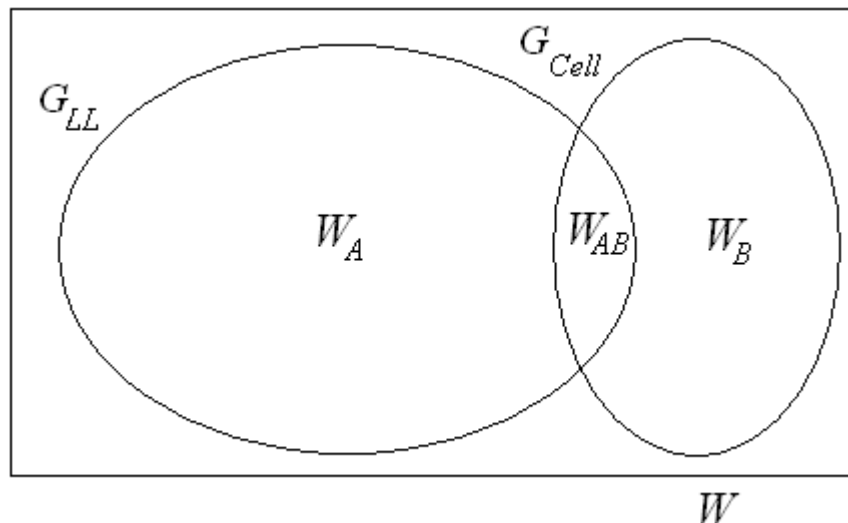
In Figure 2, the outside square labeled with W represents the set of all telephone numbers associated with the universe of interest. Within W, there are two ellipses labeled G_{LL} and G_{Cell}, which define the set of telephone numbers associated with the landline frame and cell phone frame, respectively.

Although Figure 2 is similar to Figure 1, there is one important difference. Since a given household can have several telephone numbers associated with it, multiple telephone numbers in Figure 2 can be associated with the same household. A given household can have a telephone number in W_A, W_B, and W_{AB} or not in either G_{LL} and G_{cell}. This difference will be important in our discussion of how we account for multiple telephones in the weighting.

1.3 Review of TPOPS First-Stage Sample Design

Unlike most RDD surveys, TPOPS is a two-stage sample design where most RDD surveys have a one-stage sample design. RDD surveys are a more natural fit with national non-clustered surveys since interviewing can be completed from anywhere in the U.S. including a centralized facility. However, the primary sampling units (PSUs) of TPOPS will later be visited by Bureau of Labor Statistics (BLS) staff to conduct the pricing for the selected items in outlets identified by TPOPS, so it is better cost-wise for TPOPS to be a two-stage survey.

In the first stage of TPOPS, a sample of PSUs is selected to be representative of the U.S. The PSUs are a single county or a group of counties. In the 2000 or the current sample design, the first-stage sample includes 87 sample PSUs (Greenlees 2004).

1.4 Review of the Current TPOPS Second-Stage Sample Design

Since 1996, TPOPS has been a list-assisted Random Digit Dialing (RDD) survey. Here, list-assisted refers to the method for generating the frame, which is a list of telephone numbers for each PSU. The frame for a list-assisted survey includes all telephone numbers which are in working banks that include at least one telephone number listed in the white pages. We say a working bank (or 100-bank) is defined as the set of 100 telephone numbers that have the same area code (digits 1-3), pre-fix (digits 4-6), and first two digits of the line number (digits 7-10) or simply the same first eight digits of a telephone number. A working bank is also referred to as a 100-bank because it is a bank of 100 telephone numbers.

We select a sample of cell phone numbers from a cell phone frame. The cell phone frame consists of 1,000-banks that have been assigned as cellular banks to a certain switch or wire center. The numbers inside the banks are exclusively cell phone numbers that a cellular telephone provider owns. For example, a cellular telephone provider could own all phone numbers in the bank 123-456-7XXX. In the 1,000-banks, the set of cell phone numbers have the same first seven digits of a telephone number.

The switch or wire center is a cell phone tower and is the basic unit of geography for the cell frame. When a new cell phone number is activated, it is assigned a number that is associated with the nearest wire center. Using this frame, we have a higher likelihood of identifying a respondent near their residence.

Metro areas can have several wire centers surrounding it whereas rural areas could have wire centers covering several counties. For certain PSUs that have few or no wire centers, TPOPS will expand the area to include wire centers from neighboring counties to compensate.

1.5 Interviewing Assumptions

We next discuss four assumptions that are important to how interviews will be conducted.

A1: Implicit within the sample design is the understanding that we select a sample of telephone numbers, but the unit of interest is the household. The link between telephone numbers and households is not simply one-to-one because a telephone number can be associated with several households due to call forwarding. Similarly, a single household can be reached with several telephone numbers. A weighting adjustment is applied to the base weights to account for the differential probabilities of selection for households.

A2: With respondents from both the landline and the cell phone universe, TPOPS will ask about all spending habits of all persons within the contacted household. With the landline universe, the household was generally expected to be the unit of interest since landlines are implicitly associated with the entire household. The assumption is more important to the addition of cell phones, since cell phones are implicitly associated with a person and not a household. Within the questions we ask respondents, it needs to be clear that we are asking about the spending of the entire household associated with the cell phone that we have contacted, and not only the spending of the owner of the cell phone.

In the future, this assumption could be examined by comparing the number of outlets or total expenditures per household of the landline and cell phone samples. If the number of outlets collected from the cell phone frame is smaller than the number of outlets collected from the landline frame, the assumption may need to be revisited.

A3: We are not collecting the telephone numbers of all the telephones associated with a given respondent household. We will ask respondents to report the number of landlines and cell phones associated with the household. Asking for every telephone number would be burdensome and probably viewed as intrusive by respondents.

A4: We will not ask enough questions during the TPOPS interview to identify the "wireless mostly" and "landline mostly" subgroups as is done by Blumberg *et al*. (2010). The National Health Interview Survey (NHIS) asks the respondent to consider all of the telephone calls his or her family receives and to report whether "all or almost all calls are received on cell phones, some are received on cell phones and some on regular phones, or very few or none are received on cell phones." (Blumberg *et al*. 2010). TPOPS has chosen not to ask these types of additional questions in order to manage respondent burden.

2. Base Weights

Base weights for TPOPS reflect the probability of selection for both the first- and second-stage sample designs. The "overall base weight" for unit k, or simply BW_k, is the product of the inverse of the first- and second-stage probabilities of selection, i.e.,

$$BW_k = \begin{cases} \left[\pi_{hi}^{(1)} \times \pi_i^{(2,LL)} \right]^{-1} & k \in F_{LL} \\ \left[\pi_{hi}^{(1)} \times \pi_i^{(2,cell)} \right]^{-1} & k \in F_{cell} \end{cases}$$

For the rest of this section, we separately discuss how the first- and second-stage probabilities of selection are defined.

First-Stage Probabilities of Selection

Since the first-stage units were selected with probability proportional to size where the size was the number of people in the PSU, the first-stage probability of selection for PSU i of strata h is

$$\pi_{hi}^{(1)} = \frac{M_{hi}}{M_h}$$

where

h index on the first-stage strata

M_h number of people in stratum h

M_{hi} number of people in PSU i of stratum h

The values of M_h and M_{hi} are obtained from prior decennial census counts.

Second-Stage Probabilities of Selection

For the landline sample, the second-stage probability of selection for PSU i is

$$\pi_i^{(2,LL)} = \frac{n_i^{(LL)}}{u_i^{(LL)} \times 100}$$

and similarly for the cell phone sample, the second-stage probability of selection for PSU i is

$$\pi_i^{(2,cell)} = \frac{n_i^{(cell)}}{u_i^{(cell)} \times 1,000}$$

where $u_i^{(LL)}$ is the number of landline 100-banks and $u_i^{(cell)}$ is the number of cell phone working 1,000-banks. Note that $\pi_i^{(2,LL)}$ and $\pi_i^{(2,cell)}$ for both the landline and the cell phone frame are sampling fractions. The numerator is the sample size and the denominator is the size of the universe. For the landline frame, the size of the universe is the number of 100-banks or u_i^{LL} times 100 or the number of telephone numbers within each 100- bank. For the cell phone frame, the size of the universe is the number of 1,000-banks or u_i^{LL} times 1,000 or the number of telephone numbers within each 1,000-bank. We do not include the subscript h in order to simplify the notation and since only one PSU is selected within each stratum.

In section 4, we discuss how we combine the units from the two samples.

Is there a better way to resolve geographic screen-outs?

In this next section, we discuss a new paradigm for geographic screen-outs. This includes how we consider screen-outs in weighting and estimation. With this new way of thinking, we can justify interviewing all units that are in a sample PSU, even if they were originally selected in a different sample PSU.

We begin by reviewing how we have thought about screen-outs. Generally, we have selected units in a given PSU and expected that all the units should be from the same PSU. If we found units from counties outside the given PSU, we "screened them out," which makes them ineligible or out-of-scope. We knew we would have telephone numbers from other counties, but we considered geographic screen-outs as "errors." This way of thinking is represented in Figure 3.

Figure 3: Old Paradigm for Two-Stage RDD Sample Design

A	C	E
B	D	F

The assumption that Figure 3 represents is that the frame for each PSU was expected to be "pure" or only include those telephone numbers of the given PSU.

A different way of thinking about the landline and cell phone frames admits to the messiness of telephone numbers in both landlines and cell phones. This is important because portability of telephone numbers will only increase. With our different way of thinking, we say that the majority of phone numbers associated with each frame for a given PSU are telephone numbers that are in the same PSU. However, there are residual telephone numbers associated with possibly every other PSU in the universe. Using the simple example presented in Figure 4, the majority of numbers in the frame for PSU *A* are in PSU *A*, but there are also telephone numbers from all the other PSUs also: *B, C, D, E and F*.

A consequence of this paradigm is that if we wanted to get a complete accounting of the telephone numbers from PSU *A*, we would need to go to the frames of all of the PSUs, since at least a residual amount of a PSU's telephone numbers is in every PSU frame.

Since this is a sample survey and not a complete census, this is not a problem. If our sample included PSUs *B* and *E*, we could interview all the units from PSUs *B* and *E* ignoring which frame they came from. We would use the first-stage probability of selection of the location that they were originally selected in and not their new location.

The first-stage sample design of TPOPS allows us to make an estimate of all the units associated with PSUs *B* and *E*.

Figure 4: New Paradigm for Two-Stage RDD Sample Design

A benefit of this new paradigm is that we do not waste sample. Continuing with the previous example, in the old paradigm; we would have not interviewed a respondent from PSU *B* that was selected in PSU *E*. Similarly, we also would have not interviewed a respondent from PSU *E* that was selected in PSU *B*. In the new paradigm, we interview both cases because it is necessary for estimation.

Movers Between Interviews

Between the first and the last interview of TPOPS, some respondents will move. If we apply the same rationale of the previous section, then we use the original probability of selection and enumerate the household with respect to its new location. For example, if a unit was originally selected in the Washington DC metro area sample (A312), but its occupants moved to a unit that was also originally selected in Philadelphia (A102) between the second and third interview, we would collect their outlets for the 1st and 2nd interview with respect to Washington DC and collect their outlets for the 3rd and 4th interview with respect to Philadelphia. With all four interviews, we use the unit's original overall probability of selection that includes $\pi_i^{(1)}$ from the Washington DC PSU. All weighting adjustments would be applied according to their location at the time of the interview.

However, if a unit moved to a county that was not in a sample PSU, it would be ineligible and not interviewed. We will not continue to screen them for subsequent interviews.

Call Forwarding

A common theme throughout the weighting is that the probabilities of selection and the sample weights are determined by the type of frame – either landline or cell – and not the type of telephone reached. This is also true with call forwarding. If we reach a respondent with a landline telephone

number *X* that was forwarded to cell phone number *Y*, the frame of *X* still serves as the frame for the sample unit.

3. Adjustment for Multiple Phones

TPOPS already accounts for households with multiple telephone numbers associated with a given household with the Multiple Phone Number Factor (*MPNF*). The *MPNF* would be expanded to account for multiple cell phones in a household.

Why is the MPNF necessary? If a unit has multiple telephone numbers (landline, cell, or both), we could have contacted the unit in multiple different ways. This means that they have an increased chance of being selected into the survey, i.e., their probability of selection is increased. The *MPNF* adjusts the probability of selection so that it reflects the multiple ways we could select the unit.

The *MPNF* is also the factor that allows us to sum over the sample of telephone numbers and produce estimates of households. The factor can do this because it accounts for the multiple telephone numbers that we could have used to contact the single household.

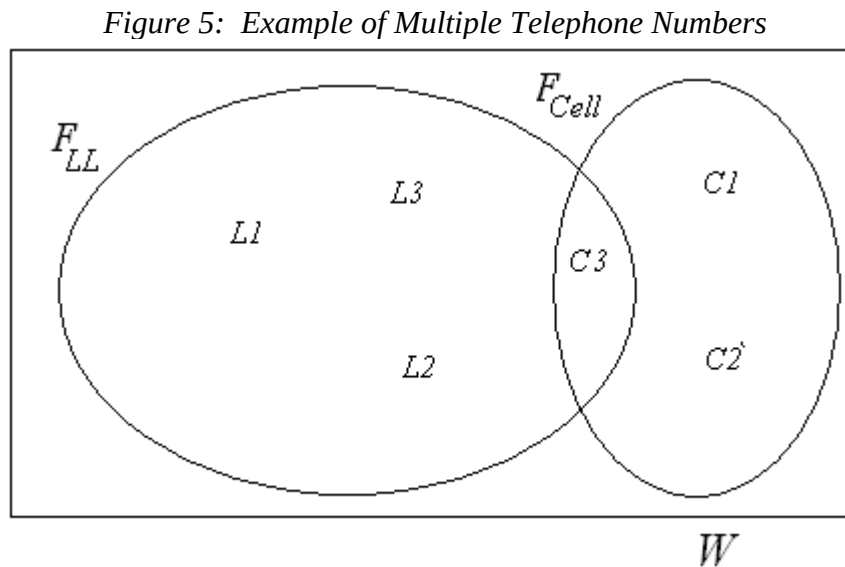To facilitate the discussion on this topic, we provide Figure 5.

Figure 5: Example of Multiple Telephone Numbers

Figure 5 shows the six telephone numbers for a hypothetical household. *L*1, *L*2, and *L*3 are landline telephone numbers and *C*1, *C*2, and *C*3 are cell phone telephone numbers. All of the telephone numbers in Figure 5 come from the same household. C3 is a cell phone number that is listed on both frames (landline and cell).

In a perfect world, if we selected one of the units *L*1, *L*2, *L*3, or *C3* from F_{LL}, we would multiply their probability of selection by 4 to adjust for the fact that there are four different telephone numbers that we could have selected and reached the same household. Similarly, if we selected one of the units *C*1, *C*2, or *C*3 from F_{Cell}, we would multiply their probability of selection by 3 to adjust for the fact that there are three different telephone numbers that we could have selected and reached the same household.

However, we do not have all of the information needed to adjust in this way because we do not know if a telephone number from a given frame is also in the other frame – see interviewing assumption A3. For our example, we would know that the household has three landlines and three cell phones, but we would not know which telephone numbers were in the intersection of the two frames. We recommend adjusting cell phones with the count of cell phones associated with the household and similarly adjusting landline telephones with the count of landlines associated with the household as:

$$MPNF = \begin{cases} 1/K_{LL}, & k \in F_{LL} \\ 1/K_{Cell}, & k \in F_{Cell} \end{cases}$$

where

K_{LL} the number of landlines associated with a given household
K_{Cell} the number of cell phones associated with a given household

This adjustment is not completely accurate. It implicitly assumes that F_{LL} and F_{cell} are disjoint. However, it should be reasonable for the following reasons: (1) the intersection of F_{LL} and F_{Cell} is currently very small and (2) most cell phones are in F_{Cell} and most landlines are in F_{LL}. This may change in the future as the portability of cell phones will increase the W_{AB}.

Note that the *MPNF* is applied to both frames but is defined with respect to the frame from which the sample unit was selected. The described revision of the *MPNF* would apply to both solutions 1 and 2 described later in section 4.

Call Forwarding and the Number of Telephones by Type

If someone uses call forwarding on a regular basis, then we want all of the telephone numbers that can be used to contact the household. For example, if someone has a landline number forwarded to a cell phone, we want both telephone numbers to be counted in the totals K_{LL} and K_{Cell} because we could have reached the household by selecting either telephone number. In the reserve case, a cell phone number forwarded to a landline phone number, we similarly want both numbers counted in the totals K_{LL} and K_{Cell}.

When call forwarding involves both a business telephone number and a home or telephone number, things become complicated. For example, if we call a household and the household has a telephone number that is forwarded to a business, we want the respondent to include the number in the count of their landline telephones. Instead of trying to unravel the use of every telephone number used by the household, we will instead rely on the respondent to not count business numbers – forwarded or not. The alternative is to ask respondents many more questions in order to clarify the use of every telephone number.

3. Adjustment for Non-Interviews

We considered two ways to complete the non-interview adjustment: either right before or right after combining the sample from the two frames. We recommend right before based on two basic considerations of nonresponse adjustments.

C1. What information is available for both completed interviews and non-interviews? This is a consideration because nonresponse adjustments can only use variables that are known for both completed interviews and non-interviews.

C2. What variables best stratify the sample in terms of creating nonresponse strata that are homogeneous with respect to their probability of responding to the survey and heterogeneous between nonresponse strata?

For the first consideration, there is not much information known about all of the non-interviews. For immediate hang-up non-interviews, we do not know anything except the frame from which we selected them. So, we are limited by our lack of auxiliary variables.

Based on this limitation of the available auxiliary variables that we can get, we will use the non-interview adjustment cells defined by Table 1.

Table 1: Cells for Non-Interview Adjustment

	Actual Interview Number			
Type of frame?	1	2	3	4
Landline				
Cell Phone				

Within each cell g, we will calculate the non-interview adjustment factor as

$$NAF_g = \frac{C_g + D_g}{C_g}$$

C_g the weighted sum of the completed interviews for cell g

D_g the weighted sum of the non-interviews for cell g

We would prefer to use the actual type of phone contacted, which we think would be more predictive of response/nonresponse than frame. However, we would have to speak with the respondent to determine the actual type of phone. This of course is an issue for non-interviewed households, many of which we do not directly speak to.

4. Multiple Frame Estimation or Combining the Landline and Cell Phone Samples

We begin this discussion by considering the estimator of a total. The technical discussion reviews how we currently estimate the total and then discusses how we can improve it by combining the sample from the landline with a sample from the cell phone frame. Combining the sample from the two frames increases the overall coverage of the survey.

More Notation

For the discussion of combining the cell phone and landline frames, we define some additional notation.

s_A the sample of telephone numbers selected from W_A

s_B the sample of telephone numbers selected from W_B

$s_{AB,LL}$ the sample of telephone numbers selected from W_{AB} using G_{LL}

$s_{AB,cell}$ the sample of telephone numbers selected from W_{AB} using G_{cell}

s_{AB} the sample of telephone numbers selected from W_{AB}

$n_{AB,LL}$ sample size for $s_{AB,LL}$

$n_{AB,cell}$ sample size for $s_{AB,cell}$

k index of the units of the universe

y_k the variable of interest for unit k

w_k sample weight for unit k

w_k^{LL} sample weight for unit k in the landline sample

w_k^{cell} sample weight for unit k in the cell phone sample

The weights w_k^{LL} and w_k^{cell} represent the survey weights using all the previously discussed base weights and adjustment factors including the multiple phone number and nonresponse factors.

Combining Samples from Two Frames

The general statistic of interest is a total of some variable of interest, which can be represented as

$$Y = \sum_{k \in U} y_k$$

The current estimator that uses the landline frame only can be stated in terms of the partitioning of U represented in Figure 1 as

$$Y' = Y^{(A)} + Y^{(B)}$$

where the totals from the landline frame are defined as

$$Y^{(A)} = \sum_{k \in U_A} y_k$$

and

$$Y^{(B)} = \sum_{k \in U_B} y_k \; .$$

A simple expression for an estimator of $\hat{Y}^{'}$ is

$$\hat{Y}^{'} = \hat{Y}^{(A)} + \hat{Y}^{(AB,LL)}$$

where the estimators of the total from the landline frame are defined as

$$\hat{Y}^{(A)} = \sum_{k \in s_A} w_k^{LL} y_k$$

and

$$\hat{Y}^{(AB,LL)} = \sum_{k \in s_{AB,LL}} w_k^{LL} y_k \; .$$

Note that $Y^{(A)}$ and $Y^{(B)}$ are both sums over households and their estimators $\hat{Y}^{(A)} = \sum_{k \in s_A} w_k^{LL} y_k$ and $\hat{Y}^{(AB,LL)}$ are sums over telephone numbers. We can do this because the *MPNF* accounts for the multiple telephone numbers associated with a given household.

The estimator $\hat{Y}^{'}$ only uses a sample selected from the landline frame. The weights that TPOPS currently uses (Killion 2010) are designed for this estimator.

The cell phone frame will allow us to improve the coverage of TPOPS by including the cell phone only universe or U_B. A general statistic for the total that uses both frames, which can be represented as the set $\left[F_{LL} \cup F_{cell} \right]$, can be stated as

$$Y^{''} = \sum_{k \in \left(F_{LL} \cup F_{cell} \right)} y_k$$

or alternatively in terms as the partitions of $\left[F_{LL} \cup F_{cell} \right]$ as

$$Y^{''} = Y^{(A)} + Y^{(B)} + Y^{(AB)} \qquad .$$

The total for the cell phone only universe is defined as

$$Y^{(B)} = \sum_{k \in U_B} y_k$$

and an estimator of the total using a sample from the cell phone frame is

$$\hat{Y}^{(B)} = \sum_{k \in s_B} w_k^{cell} y_k \ .$$

Although we do not have complete coverage of U, the addition of the cell phone frame will have

better coverage than the landline frame alone.

What should we do with the intersection of the two frames? How should we estimate $Y^{(AB)}$? There are two options, which we will now discuss.

Solution 1: Estimate the intersection with the landline sample only

The first solution is to only use the sample from one of the two frames. Here, we would only use the sample from the landline frame in the estimator

$$\hat{Y}^{(AB1)} = \sum_{k \in s_{AB,LL}} w_k^{LL} y_k \ .$$

Since we do not know whether a unit is in U_{AB}, for either frame, until we talk to the respondent; we need to call all of the selected sample units from both frames. We can stop the interview with estimator $\hat{Y}^{(AB, LL)}$ once we identify that the respondent has both a landline telephone and cell phone and they were selected from F_{cell}.

We could have also chosen to use the data from the cell phone sample alone. We chose the data from the landline because we plan to allocate more sample to the landline sample than the cell phone sample. This means that the landline sample will provide a better estimate in terms of sampling variability.

Solution 2: Estimate the intersection from both the landline and cell samples

The second solution uses the sample units that are in U_{AB} from both frames. Then the estimator becomes

$$\hat{Y}^{(AB2)} = \alpha_{LL} \hat{Y}^{(AB,LL)} + \alpha_{Cell} \hat{Y}^{(AB,cell)}$$

where

$$\hat{Y}^{(AB,cell)} = \sum_{k \in s_{AB,Cell}} w_k^{cell} y_k \ ,$$

and α_{LL} and α_{Cell} are suitably chosen. For $\hat{Y}^{(AB2)}$ to be an unbiased estimator of $Y^{(AB)}$, we need

$0 \le \alpha_{LL} \le 1, 0 \le \alpha_{Cell} \le 1,$ and $\alpha_{LL} + \alpha_{cell} = 1.$

The optimal choice of α_{LL} and α_{cell} with respect to reducing the variance of $\hat{Y}^{(AB2)}$ is

$$\alpha_{LL}=\frac{v\left(\hat{Y}^{(AB,LL)}\right)}{v\left(\hat{Y}^{(AB,cell)}\right)+v\left(\hat{Y}^{(AB,LL)}\right)} \tag{1}$$

and

$$\alpha_{Cell}=\frac{v\left(\hat{Y}^{(AB,cell)}\right)}{v\left(\hat{Y}^{(AB,cell)}\right)+v\left(\hat{Y}^{(AB,LL)}\right)}. \tag{2}$$

Given that we initially do not know the variance of the $\hat{Y}^{(AB,LL)}$ and $\hat{Y}^{(AB,cell)}$, we can use the following expressions.

$$\alpha_{LL}=\frac{\left(1/n_{AB,cell}\right)}{\left(1/n_{AB,LL}\right)+\left(1/n_{AB,cell}\right)}=\frac{n_{AB,cell}}{n_{AB,LL}+n_{AB,cell}} \tag{3}$$

and

$$\alpha_{Cell}=\frac{\left(1/n_{AB,cell}\right)}{\left(1/n_{AB,LL}\right)+\left(1/n_{AB,cell}\right)}=\frac{n_{AB,cell}}{n_{AB,LL}+n_{AB,cell}} . \tag{4}$$

The assumption behind (3) and (4) is that the variances of $\hat{Y}^{(AB,LL)}$ and $\hat{Y}^{(AB,cell)}$ can be expressed as generally as σ^2/n for both the landline and cell phone sample. Further, we assume that σ^2 is a constant for both the landline and cell phone sample; consequently, the only part of the variance that is variable is the sample size.

Combining Multiple Frames Solution for TPOPS

TPOPS uses solution 2. In both solutions, the sample selected from W_{AB} from G_{cell} cannot be identified prior to the interview. Considerable time, cost, and effort is also spent to contact a respondent due to the high screen-out rate of cell phone numbers. We think it is the best use of our resources to use all of the contacted respondents.

With solution 2, we would initially combine the estimators $\hat{Y}^{(AB,LL)}$ and $\hat{Y}^{(AB,cell)}$ using (3) and (4),
and then later when enough information is available to estimate the variance, we would use (1) and (2).

6. Ratio Adjustment to Known PSU Totals

TPOPS recently (1st quarter of 2011) started using two separate ratio adjustments to known totals in the weighting. The first adjusts the estimated count of persons within the PSU to the known total for the PSU. The known totals are derived from the 5-year American Community Survey (ACS) data files.

To reduce the variability due to the cell phone sample, we divide the ACS population totals with estimates derived from NHIS [Blumberg *et al*. 2011]. We will divide the population by telephone status: wireless only, landline only, or dual use. See Attachment B for an example of this ratio adjustment and the ratio adjustment described in the next section.

The weighting adjustment factor takes the form of

$$POPAF_{hic} = \frac{\widetilde{N}_{hic}}{\hat{N}_{hic}}$$

where

i index on the first-stage PSU of stratum h

c index of the ratio adjustment cell c (wireless only, landline only, dual use)

$\widetilde{N}_{hic}$ total number of people in cell c for PSU i in stratum h

$\hat{N}_{hic}$ estimated total number of people from TPOPS in cell c for PSU i in strata h

The known total is derived from two sets of survey estimates as

$$\widetilde{N}_{hic} = \breve{N}_{hi}\,\hat{p}_{hic}$$

where

$\breve{N}_{hi}$ estimated total number of people from ACS for PSU i in stratum h

$\hat{p}_{hic}$ estimated proportion of telephone status c derived from NHIS for PSU i in stratum h

The estimates of $\breve{N}_{hi}$ are derived from the ACS five-year data file and the estimates of $\hat{p}_{hic}$ are derived from the landline and cell phone usage estimates in Blumberg *et al*. (2011). Since the NHIS estimates are at the state or selected county level, weighted averages of the estimates will be used to derive the PSU-level estimates of $\hat{p}_{hic}$ where the weights will be proportional to the eligible population of the county or group of counties.

The estimates of $\hat{N}_{hic}$ from TPOPS will be derived as

$$\hat{N}_{hic} = \sum_{k \in s_{hic}} w_k^{\iota} z_k$$

where

$$w_k^{\dot{\iota}}$$
the weight that incorporates the base weight and all of the weighting adjustments calculated up to this point in the weighting

z_k number of reported people in the household

Collapsing of the cells will be completed when there are less than 30 completed interviews in a cell or the ratio adjustment is greater than two. We collapse with fewer than 30 completed interviews to ensure that $\hat{N}_{hic}$ is a reasonable estimate – reasonable with respect to the variance. We collapse when the weighting adjustment factor $POPAF_{hic}$ is greater than 2.0 to avoid large weights that can lead to large variances.

We do note that NHIS provides the telephone status estimates by state, selected counties or groups of selected counties. This means that the telephone status estimates are not exactly aligned with the counties or groups of counties that comprise the PSUs of TPOPS. PSU-level estimates will be derived by taking weighted averages of the telephone use estimates where the weight will be proportional to the estimated size of the counties that comprise a given PSU.

7. Ratio Adjustments to Known First-Stage Strata Totals

The second weighting adjustment that TPOPS has recently started using is an adjustment to known first-stage stratum totals, which we refer to as *STRATAF*. Like *POPAF*, the factor *STRATAF* is a ratio adjustment defined with respect to the count of persons in the first-stage stratum as

$$STRATAF_h = \frac{N_h}{\hat{N}_h}$$

where

N_h total number of people from TPOPS for PSU *i* in stratum *h*

$\hat{N}_h$ estimated total number in people from ACS for PSU *i* in stratum *h*

The estimates of $\hat{N}_h$ will be derived as

$$\widetilde{N}_h = \sum_{k \in s_h} w_k^{\dot{\iota}} z_k$$

where

$$w_k^{\dot{\iota}}$$
the weight that incorporates the base weight and all of the weighting adjustments calculated up to this point in the weighting

z_k the number of reported people in the household

Collapsing of the cells is completed when there are less than 30 completed interviews in a cell or the ratio adjustment is greater than two. The reasoning for the collapsing is the same as discussed in the previous section for *POPAF*.

We considered forming cells with the *STRATAF* and not forming cells with *POPAF*. This would be more natural, but we thought that the weighted averages of cell phone usage for the PSU would be better than the stratum estimate.

8. Putting It All Together

The final weight for TPOPS can be summarized as

$$FW_k = \begin{cases} BW_k \times MPNF_k \times NAF_g \times POPAF_{hic} \times STRATAF_{hd}, & k \in U_A \ or \ k \in U_B \\ BW_k \times MPNF_k \times NAF_g \times \alpha_{LL} \times POPAF_{hic} \times STRATAF_{hd}, & k \in F_{LL} \ and \ k \in U_{AB} \\ BW_k \times MPNF_k \times NAF_g \times \alpha_{cell} \times POPAF_{hic} \times STRATAF_{hd}, & k \in F_{cell} \ and \ k \in U_{AB} \end{cases}$$

Miscellaneous

This memorandum is stored in the directory "M:\ADC-LEDSP\VEB\TPOPS_Final Memos" with the name "2011-01 TPOPS Discussion of Weighting for Cell Phones, v0.2.docx".

References

Blumberg, S.J., Luke, L.V., Genesh, N., Davern, M.E., Boudreaux, M.H., Soderberg, K. (2010). National Health Statistics Report, No. 39, "Wireless Substitution: State-level Estimates From the National Health Interview Survey, January 2007-June 2010," dated April 20, 2011.

Greenlees, J. (2004). Bureau of Labor Statistics Memorandum, "PSU Rotation Requirement for TPOPS, 2004 thru 2014," from John S. Greenlees to Chester E. Bowie.

Keeter, S., Kennedy, C., Clark, A., Tompson, T., and Mokrzycki, M. (2007). "What's Missing from National Landline RDD Surveys? The Impact of the Growing Cell-Only Population," Public Opinion Quarterly, 71, 5, 814-839.

Killion, R.A. (2010). "TPOPS: Q104 Weighting Specification" Working paper, dated April 1, 2010.

Link, M.W., Battaglia, M.P., Frankel, M.R., Oborn, L., and Mokdad, A.H. (2007). "Reaching the U.S. Cell Phone Survey Results with an Ongoing Landline Telephone Survey," Public Opinion Quarterly, 71, 5, 814-839.

Lohr, S.L. (2009). "Multiple Frame Surveys," *from Handbook of Statistics, Sample Surveys: Design, Methods and Applications, 29A,* Elsevier, 71-88.

Tucker, C., Brick, J.M., and Meekins, B. (2007). "Household Telephone Service and Usage Patterns in the United States in 2004: Implications for Telephone Samples," Public Opinion Quarterly, 71, 1, 3-22.

Wolter, K.M., Smith, P., and Blumberg, S.J. (2010). "Statistical foundations of cell-phone surveys," Survey Methodology, 36, 203-215.

cc: R. Cage   (BJS)
        M. Saxton
        S. Stanley
        E. Bergmann
        C. Laskey   (DSD)
        D. Pepe
        A. Okon
        J. Arthur
        P. Flanagan   (DSMD)
        T. Lee

Glossary

Base Weight – is how many units that unit represents before all the adjustments.

Call Forwarding – is a telephone service that allows a subscriber to have incoming calls forwarded to a different number.

Eligible/Ineligible – A unit is eligible if it is in the universe of interest. For TPOPS, phone numbers that are associated with civilian non-institutional households in the U.S. are eligible. Government, business, and military numbers telephone numbers are ineligible.

Frame – are lists of telephone numbers that are associated with the households. The landline and cell phone frames are subsets of the universe.

Geographic Screen-out – is a phone number that is not residential and not inside the geographic boundaries of the PSU.

List-assisted – is a list of telephone numbers for each PSU which is used to generate the frame. The frame includes all one hundred telephone numbers of any working bank where there is at least one listed residential telephone number in the white pages.

Non-interview – is when an attempt to interview a phone number was made but was not completed due to a refusal, ring no answer, answering machine, etc.

Outlet – is the location where consumers purchase goods and services.

Portability – is the household's ability to move and keep his/her cell or landline telephone number with them to their new location.

Primary Sample Unit (PSU) – is a geographic area that includes a county or a group of counties.

Random Digit Dialing (RDD) – is a method for selecting a random sample of telephone numbers.

Telephone Number – is a 10-digit number (aaa) – bbb – cccc where aaa is the area code, bbb is the prefix, and cccc is the line number.

Wire Center – is a cell phone tower and is the basic unit of geography for the cell phone frame.

Working Bank – is the set of of 100 landline telephone numbers.

Universe or Universe of Interest – In finite population sampling, the universe of interest, or simply the universe, is the well-defined set of units for which we would like to produce an estimate for. For TPOPS, the universe of interest is the set of non-institutional civilian households in the U.S.

Example of the Ratio Adjustment Factors

We consider a single stratum from TPOPS that has two PSUs: Columbus, OH and Green Bay, WI. **Please note that the totals within the example are made-up and do not match any totals of any publications.** Table B1 shows the counties associated with each PSU of the example.

Table B1: Example Strata and PSUs

PSU Name	FIPS County Code	County Name	1990 Population Count
Columbus, OH	39041	Delaware	66,929
	39045	Fairfield	103,461
	39049	Franklin	961,437
	39089	Licking	128,300
	39097	Madison	37,068
	39129	Pickaway	48,255
Green Bay, WI	55009	Brown	194,594

In the first-stage sample design, TPOPS selects one PSU per stratum. For our example, we have selected the Green Bay PSU. Using the totals from Table B2, we calculate the first-stage probability of selection $\pi_{hi}^{(1)}$ for Green Bay as $0.126356 = 194,594 / 1,540,044$.

Table B2: Example PSU Totals

PSU	M_{hi}	$\pi_{hi}^{(1)}$
Columbus, OH	1,345,450	0.873644
Green Bay, WI	194,594	0.126356
Stratum Total	1,540,044	

Next, we start with the NHIS estimates of phone usage. The first column of Table B3 lists the types of phone usage from NHIS and the second column provides each type's associated estimated proportion. For TPOPS to use the NHIS proportions, we first drop the "no service" category, since we want the proportions to apply to all phones in service. We also combined three categories into "dual use." Dropping "no service" and combining dual use is represented in column 2 of Table B3. Since we dropped "no service", we adjusted or "rescaled" the final TPOPS proportions so that they would sum to 1.0. The fourth column presents the rescaled estimates of our example.

Table B3: Phone Usage Estimates

Original Categories from NHIS	Proportions from NHIS	Before Rescaling	Proportions for TPOPS	Categories for TPOPS
Wireless only	0.188	0.188	0.193	Wireless only
Landline only	0.052	0.052	0.053	Landline only
Dual use	0.733	0.733	0.753	Dual Use
Wireless mostly	0.193			
Landline mostly	0.140			
No service	0.027	not used		
Total	1.000	0.973	1.000	

The proportions from Table B3 were then applied to the ACS population estimate for the PSU, which is represented in the second column of Table B4. In our example, ACS provided the estimate of 201,456 for the number of persons in Green Bay WI. The third column of Table B5 is the TPOPS pre-estimate of the number of persons in Green Bay using the base weight and all of the other weighting adjustments calculated up to the *POPAF*. The *POPAF* is calculated as the ratio of the ACS estimated with the NHIS proportions applied with the TPOPS pre-estimate, which is represented in the fourth column of Table B5.

Table B4: Calculation of POPAF

	ACS Estimate $\left(\check{N}_{hic}\right)$	TPOPS pre-estimate $\left(\hat{N}_{hic}\right)$	POPAF
Wireless Only	38,925	27,918	1.394228
Dual Use	10,766	14,890	0.723071
Landline Only	171,765	143,315	1.058962
Total	201,456	186,123	

The final weighting adjustment factor *STRATAF* is calculated as the ratio for the known total for the stratum with the TPOPS pre-estimate – the estimate using the base weight and all the other weighting adjustments including *POPAF*.

Table B5 concludes our example with the Columbus OH and Green Bay WI Stratum. A single factor is calculated for *STRATAF* that adjusts the TPOPS estimate of the stratum total to the ACS estimate of the stratum total.

Table B5: Calculation of POPAF

	ACS Estimate $\left(\breve{N}_h\right)$	TPOPS pre-estimate $\left(\hat{N}_h\right)$	STRATAF
Columbus, OH and Green Bay, WI Stratum	1,654,321	1,594,352	1.037613098

Because *POPAF* was previously applied, we know that the TPOPS estimate of N_h is the ACS estimate divided by the probability of selection, i.e., 1,594,352 = 201,456 / 0.126356.