GfK Custom Research
North America

Academic Paper

Cell phone and Landline – Considerations for Sample Design, Estimates, Weighting, and Costs

By Robert Benford, Trevor Thompson, Chris Fleury, Geoff Feinberg, Barry Feinberg, Nicole Speulda, and Annie Weber

Paper originally presented at the 64th annual conference of the
American Association for Public Opinion Research
Hollywood FL; May 14, 2009

Background and Motivation

Reducing potential coverage and non-response bias strongly influences our sample design approach for the AP-GfK polls. These polls are intended to gather opinions about issues, political and otherwise, that the adult population in our country faces. These polls require sound design within pre-determined cost parameters. A relatively current summary of the issues researchers face when considering the inclusion of cell phone samples with landline samples is available to guide best practices (Lavrakas, Shuttles, Steeh, Fienberg 2007). However, no clear cut guidance exists for some important design components such as how many elements of each frame to include in the total, whether or not the cell phone frame should include only CPO elements through screening, and what is the appropriate method for weighting. Kennedy (2007) ties together past work in weighting (Brick, Dipko, Presser, Tucker, and Yangyang 2006) along with her own approach. Brick et al's work provides the basis for dealing with the frame overlap in a reasonable way through a composite estimator (Hartly 1962) that is used in Kennedy's dual frame weighting as well as ours in this paper. As in Kennedy's case, we are dealing with population based inference.

It is reasonable to expect that some design decisions be based on hypothesized potential error from coverage or non-response. For example, if one were to undertake research related to risky behaviors, it would be prudent to include cell phone sample given the differences in characteristics and behaviors noted by (Blumberg and Luke 2008)[1]. However, at this point in time, it is simply not known when estimates might be biased from the exclusion of cell phone samples in our surveys. For example, Keeter, Kennedy, Clark, Tompson, Mokrzycki (2007) itemize a substantial list of estimates that do not suffer had they not included a cell sample in the design. The 2004 horse race estimates were a non-issue as far as cell samples is concerned (Keeter 2004). It appears, however, that portions of the population reached by cell phone were an important design consideration in 2008 horse race estimates (PEW 2008). In short, the dynamics of the adult population and their selection of phone service may be akin to hitting a moving target.

It has been well documented that cell phone and landline samples yield distributions of people and households that differ demographically and perhaps in other important ways as well (Blumberg 2008). It is important to gain an understanding of how these differences could impact political estimates, in particular presidential candidate preference, as this could provide guidance with the element selection process as well as the approach to weighting the data in general. The goal is to develop sampling and weighting methods that produce approximately unbiased estimates and at reasonable costs that could be tolerated by pollsters and researchers constructing sample designs that include cell phone samples. Absent a "truth" or reference on political measures, we are left with striving for achieved and balanced samples that match well-established population estimates such as those found within the Current Population Survey (CPS), National Health Interview Survey, American Community Survey, etc. The CPS 2008 March Supplement is used mainly due to its malleability in constructing profiles for weighting purposes. Additionally, MRI data is used as the reference estimate for patterns of phone service by Census region.

Although we note trends regarding support for Obama when cell phone samples are included, the main focus of this paper is not on specific estimates, as the gamut of estimates one might

[1]This is the most current of the string from the NHIS Early Release Program

measure is inexhaustible. Rather, we address the implications of design decisions in achieved samples, weighting, and costs.

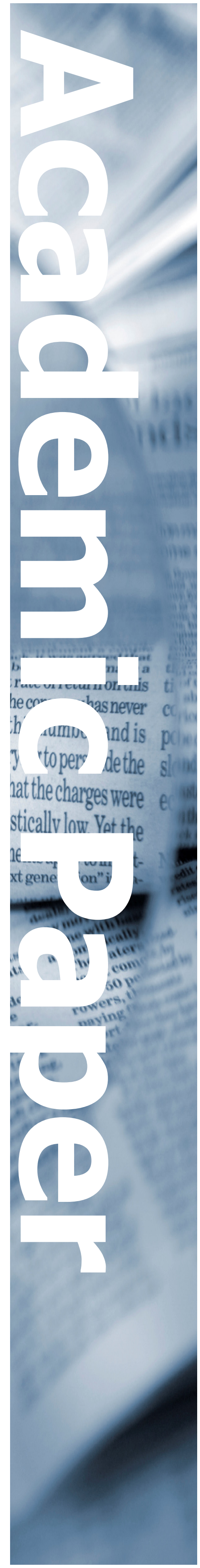
Data Source

Four AP-GfK Roper public opinion polls, two pre-election and two post-election[2] provide that data for analysis. All four polls employ dual frame sampling methodologies with 80% of the achieved sample coming from landline RDD using Survey Sampling International's EPSEM A without a business purge and 20% of the achieved sample from RDD Cell phone sample. It should be noted that in the pre-election polls, the 80/20 design is achieved at the likely voter subset, which means that the total sample for these two polls is not perfectly 80/20. Both landline and cell phone samples were purchased from SSI.

AP-GfK polls average between 15 and 20 minutes among landline sample and between 17 and 23 minutes for the Cell phone sample. Potential respondents are informed that interviewers are calling on behalf of the Associated Press[3]. The population of inference is the United States adult population (18 or older) in all 50 states and the District of Columbia. A full methodology as well as poll questions can be found at: http://www.ap-gfkpoll.com/index.html.

Minor differences exist between the results reported on the website and the results reported in this paper. This is due to differences in weighting decisions regarding the handling of extreme weights. One objective in this paper was to control weight extremes somewhat through limiting the number of adults and number of lines to a maximum of two.

[2] Pre-election polls ran September 27-30 and October 16-20, 2008. Post-election polls ran November 6-10 and December 3-8, 2008.
[3] All four polls were conducted at Eastern Research Services.



Results

Design Comparisons

Our sample design allows for the partitioning of the sample into three mutually exclusive components. That is, the traditional landline sample can be isolated from those contacted by cell phone sample and the cell phone sample can be further disaggregated into cell phone only and cell phone with a landline as others have done (Brick et al 2006, Kennedy 2007). The first analysis is to compare the four polls combined to evaluate departures from a reference source. Table 2 shows the estimates for a set of demographics common to the four polls. Ideally, if any sample matches this profile the achieved sample would be deemed self-weighting. That is, we routinely see departures in achieved samples from our reference source for weighting, compute weights, and declare the weighted sample approximately unbiased.

Table 2 can be used to assess three sample designs:

1. Use of only a traditional landline sample (column 3)
2. Stratified: Traditional landline with CPO (column 4)
3. Dual frame approach (column 5)

When the achieved sample over represents the reference source it appears as a positive number while under represented demographic categories are negative. To examine departures from the reference source, the absolute values of the differences are summed. This is done for common demographics across the four polls and for the demographic variables employed in weighting.

With only a few exceptions (as might be expected by chance), the sample aligns better with the reference source as we incorporate the cell phone only and then the cell phone with a landline into the design. From this perspective, the dual frame design is preferred as net differences from CPS improve across all variables including those used for weighting. Of course, this is limited to constraints imposed by the original design of 20% cell phone frame and 80% landline frame where the portion of CPO is an uncontrolled subset of the cell frame (31% of cell sample is CPO in Table 2). For the CPO stratum to be proportional to size, the sample size for that stratum would need to be slightly more than doubled.

Weights are computed in two stages. First, a pre-weight[4] is computed to account for different selection probabilities. This is followed by demographic balancing using a rim weighting procedure. A need for pre-weights arises for several reasons. First, the selection probabilities of each frame differ. In addition, the landline sample takes into account the number of adults 18 or older in the household as well as the number of lines that can connect to any one adult in each sample frame. Further, in the dual frame design, a multiplicity adjustment is needed to account for the overlap in the two frames. For the dual frame approach the pre-weights are computed:

Landline Pre-weight Dual Frame:

$$\frac{\text{Frame Size}}{\text{Sample Size}} \times \frac{1}{\text{Frame RR}} \times \frac{\#\ \text{Adults}^{5}}{\#\ \text{Landlines}^{5}} \times \lambda^{dual}$$

Cell phone Pre-weight Dual Frame:

$$\frac{\text{Frame Size}}{\text{Sample Size}} \times \frac{1}{\text{Frame RR}} \times \frac{1}{\#\ \text{Cells}^{5}} \times (1-\lambda)^{dual}$$

This is similar to the person-based approach used by Kennedy (2007) with some exceptions. First, the post stratification adjustment employed in the AP-GfK polls is with respect to the

[4] This is referred to as a composite weight in Brick et al 2006 and Kennedy 2007.
[5] Capped at 2 for all pre-weights.

total frame, not by strata such as region. In addition, the adjustment for the number of adults is not applied to the cell phone frame. Our rationale for these amendments is twofold: First, the

percentage of those sharing a cell phone is rather low (14% by our estimates). Second, to precisely capture phone usage would require substantial questionnaire time that was not feasible.

Table 1: Comparison of the Distribution of Type of Phones Service

	Cell phone no landline	Cell phone with a landline	Landline with a Cell phone	Landline no Cell phone
Unweighted	6.8	14.9	59.4	18.9
Pre-weighted λ = 0.8	22.9	9.5	49.2	18.4
Pre-weighted λ = 0.5	23.9	24.7	32.1	19.2
MRI	17.8	66.8		15.4

Pre-weights as computed above for the dual frame approach use the mixing parameter λ, where λ is equal to the proportion of frame overlap[6]. Frame overlap is about 20% cell phone and 80% landline; therefore, λ is set at 0.8. The exponent dual is set to 0 when the frames do not overlap and 1 when they do. Table 1 shows the impact of the pre-weight and its mixing parameter on the phone service distribution compared to the estimated distribution from MRI data. Setting the mixing parameter λ = 0.8 reflects the achieved frame overlap and results in the correct contribution to estimates from the overlap portion of the sample. This does overstate the distribution for those who have only one type of phone service. We also test λ = 0.5 to see the impact on the phone service distribution. There is little change in the distribution if there is only one type of phone service; however, the overlap distribution is very different. While we find no impact on pre-election estimates in the test, there is a difference

[6] We adopt nomenclature per Brick et al 2006 and Kennedy 2007 for the mixing parameter.

in the weighting variability of the pre-weight with $\lambda = 0.5$ resulting in a decrease in terms of efficiency of 7 to 8 percentage points.

The mixing parameter is dropped for pre-weighting computations for traditional landline and for concatenating the CPO stratum to the landline stratum. Further, traditional landline is only adjusted by Census region since phone service becomes irrelevant.

Columns 6, 7, and 8 in Table 2 show the pre-weights applied to each respective sample approach. As expected the pre-weights improve upon the achieved demographic distributions. The dual frame approach is the more appealing design choice given the improvement in departures from the CPS estimates; however, that does not negate the approach that simply adds a cell phone stratum to the landline stratum. That addition of CPO to landline is a substantial improvement in the achieved distribution over landline alone.

Pre-weights become a multiplier entering the demographic balancing through rim weighting. That is, the sample now needs to be adjusted to the reference profile with respect to selected demographics. Our polls adjust the following:

1. Age by Sex by Race where age categories are 18-29, 30-49, 50-64, 65+ and Race is Black and all other races
2. Census Region X Phone Service[7]
3. Hispanic and non-Hispanic[8]
4. Educational Attainment as Less than High School, High School Graduate, Some College or Technical School, and Four Year College Graduate or higher

The data are simultaneously adjusted to these marginal distributions through an iterative proportional fitting or raking procedure. Final estimates for the polls are based on these

[7] Phone service by Census region is from MRI data, Fall 2008.
[8] Surveys are conducted in English or Spanish based on the respondent's preference.

weights and process. The distribution of final weights for polls is analyzed and decisions about the range of the weights and trimming or capping are sometimes taken. For comparison trimming of the weights has not been invoked so that one can more easily examine how the weights perform.

Table 2: Departures from CPS for Frame Combinations Achieved Sample and Pre-Weights – Variables used in final weighting polls in bold

1	2	3	4	5	6	7	8
- Under Represented	**CPS**[9]	**Landline (3,336)**	**Landline + CPO (3,626)**	**Dual Frame (4,262)**	**Pwt Landline**	**Pwt Landline+ CPO**	**Pwt Dual Frame**
Northeast	18.6	0.2	-0.2	-0.8	-0.2	-1.4	-1.6
Midwest	22.1	-0.4	0.1	0.3	-0.6	1.1	1.4
South	36.7	-1.3	-1.2	-0.4	-1.1	-0.7	-1.0
West	22.6	1.5	1.3	0.9	1.8	1.1	1.2
Less than HS	14.3	-8.8	-8.7	-9.0	-9.0	-8.8	-8.6
HS Grad	30.9	-8.5	-8.5	-8.8	-8.5	-8.5	-8.2
Some College	27.9	-1.2	-1.2	-0.7	-0.4	-0.5	-0.3
College Grad +	26.9	18.5	18.3	18.5	18.0	17.8	17.0
Hispanic	13.5	-7.2	-7.0	-6.9	-6.6	-6.2	-5.7
Non-Hispanic	86.5	7.2	7.0	6.9	6.6	6.2	5.7
Own	71.6	7.0	3.7	3.4	8.0	-2.9	-3.3
Rent/Other	28.4	-7.0	-3.7	-3.4	-8.0	2.9	3.3
Large Central metro	29.6	-7.2	-6.2	-4.1	-7.1	-4.0	-2.7
Large Fringe metro	24.1	0.6	0.0	-0.6	0.6	-1.3	-2.0
Medium metro	19.8	2.0	1.9	1.6	1.9	1.4	1.4
Small metro	9.5	1.2	1.3	1.0	1.2	1.6	1.3
Micropolitan	10.3	1.4	1.5	1.2	1.4	1.5	1.5
Noncore	6.5	2.2	1.8	1.1	2.3	0.9	0.7
White/Other	88.2	4.4	4.3	3.9	4.8	4.3	3.9
Black	11.8	-4.4	-4.3	-3.9	-4.8	-4.3	-3.9
Male	48.4	-4.8	-3.5	-2.0	-4.4	-0.4	0.3
Female	51.6	4.8	3.5	2.0	4.4	0.4	-0.3
18-29	22.0	-14.5	-11.5	-10.4	-13.7	-4.3	-4.1
30-49	37.4	-8.8	-8.2	-6.5	-7.0	-5.4	-5.4
50-64	24.3	13.1	11.4	10.8	12.8	7.5	7.5
65+	16.4	10.2	8.3	6.1	7.9	2.2	2.0

[9] Distributions for urbanization come from: Table 6. Number of counties and percentage of population in each of the urbanization levels of the NCHS Urban-Rural Classification

1 Adult	33.4	6.1	5.5	3.5	-7.9	-6.3	-4.8
2 Adults	51.9	-9.8	-9.4	-7.7	0.1	-0.9	-2.3
3 Adults	14.7	3.7	3.9	4.2	7.8	7.1	7.1
CPO[10]	17.8		-9.8	-11.0		8.0	5.1
Both	66.8	-5.4	3.0	7.5	-4.4	-9.7	-8.1
LLO	15.4	5.4	6.8	3.5	4.4	1.7	3.0
All Variables		168	147	131	159	112	108
Weighting Variables		120	108	99	113	81	78

Table 3: Departures from CPS for Frame Combinations Final Weights and Simulated Dual Frame 50-50 Design – Variables used in final weighting polls shown in bold

1 - Under Represented	**2** **CPS**[11]	**3** **Final Landline**	**4** **Final Landline+ CPO**	**5** **Final Dual (20/80) Frame**	**6** **Dual 50/50 Frame**	**7** **Pwt Dual 50/50 Frame**	**8** **FNLWT 50/50**
Northeast	18.6	1.5	0.1	0.1	-2.1	-2.2	0.1
Midwest	22.1	0.2	0.5	0.5	1.1	1.9	0.4
South	36.7	-1.7	-0.3	-0.3	0.7	-0.7	-0.2
West	22.6	0.0	-0.3	-0.3	0.3	0.9	-0.3
Less than HS	14.3	0.0	0.0	0.0	-9.2	-8.3	0.0
High School Grad	30.9	0.0	0.0	0.0	-9.2	-7.7	0.0
Some College	27.9	0.0	0.0	0.0	0.0	0.2	0.0
College Grad +	26.9	0.0	0.0	0.0	18.5	15.8	0.0
Hispanic	13.5	0.0	0.0	0.0	-6.6	-5.3	0.0
Non-Hispanic	86.5	0.0	0.0	0.0	6.6	5.3	0.0
Own	71.6	-3.6	-7.1	-8.6	-1.3	-5.8	-9.8
Rent/Other	28.4	3.6	7.1	8.6	1.3	5.8	9.8
Large Central metro	29.6	-6.0	-4.3	-2.1	-0.1	-0.3	0.8
Large Fringe metro	24.1	-1.0	-1.5	-2.5	-2.1	-3.2	-4.3
Medium metro	19.8	1.3	1.2	1.1	1.0	1.1	0.6
Small metro	9.5	1.0	1.1	1.2	0.6	1.0	1.2
Micropolitan	10.3	2.0	1.7	1.5	1.0	1.6	1.8
Noncore	6.5	2.9	2.0	1.2	-0.2	-0.1	0.0
White/Other	88.2	0.0	0.0	0.0	3.3	3.3	0.0
Black	11.8	0.0	0.0	0.0	-3.3	-3.3	0.0
Male	48.4	0.0	0.0	0.0	1.7	1.9	0.0
Female	51.6	0.0	0.0	0.0	-1.7	-1.9	0.0
18-29	22.0	0.0	0.0	0.0	-5.1	-1.9	0.0
30-49	37.4	0.0	0.0	0.0	-3.7	-4.7	0.0
50-64	24.3	0.0	0.0	0.0	8.0	6.2	0.0

[10] Cell phone only, both, and landline only estimates from MRI Fall 2008.
[11] Distributions for urbanization come from: Table 6. Number of counties and percentage of population in each of the urbanization levels of the NCHS Urban-Rural Classification

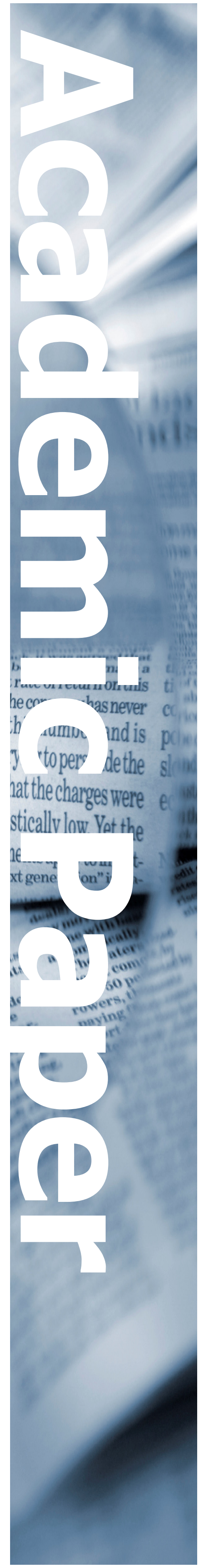
65+	16.4	0.0	0.0	0.0	0.8	0.4	0.0
1 Adult	33.4	-9.6	-7.0	-5.3	0.3	-3.3	-3.6
2 Adults	51.9	-4.6	-4.3	-5.3	-5.1	-3.4	-5.5
3 Adults	14.7	14.2	11.2	10.6	4.8	6.7	9.1
CPO[12]	17.8		0.0	0.0	-2.3	6.3	-0.1
Both	66.8	-1.7	0.1	0.1	5.3	-10.1	0.1
LLO	15.4	1.7	0.0	0.0	-3.1	3.8	0.0
All Variables		54	50	49	100	104	48
Weighting Variables		4	1	1	82	72	1

Table 3 (previous page) shows the final demographic balancing. As expected the absolute differences for variables used in weighting approaches zero. Since MRI data are used to adjust phone service by region, some differences exists due to differences between MRI and the CPS on Census region. However, potential bias still exists in that household owners are under represented, a finding consistent with Kennedy (2007) and single adult households are under represented although we see consistent improvement with CPO and dual frame approaches. Improvement in the distribution of single adult households is most likely due to connecting with these hard to find at home respondents. We also note that the dual frame design improves under coverage of Large Central metro areas.

Cell samples improve the demographic distribution in comparison to the CPS reference source. To understand whether or not an increase in the elements from the cell frame in equal proportion to the landline frame could improve the demographic distribution further, we simulate a 50/50 dual frame design. Columns 6 through 8 of Table 3 model a 50/50 dual frame design. Additional cell phone data is provided by simulation through a replication process. This particular scenario yields a mixed sample that is closer to CPS estimates and is comparable in accounting for non-response when the pre-weights are applied.

Pre-election Estimates

Although our primary purpose is to look at design considerations, it is useful to know the impact of these decisions on pre-election estimates. Candidate preferences based on design



variants from two pre-election polls are shown in Table 4. Although not significant, weighted estimates of preference for Obama increase as the design moves from landline through to full dual frame dual frame estimates Obama's advantage at 2.9 percentage points greater over McCain than the landline only sample. For reference, the final Obama margin is +7.6, with Obama getting 52.0% of the vote compared to McCain's 44.4%.

Included in Table 4 are the unweighted leaned candidate preferences for each sample design. In every instance cell phone sample increases in support for Obama and weighting adjusts Obama support up and McCain support down.

Table 4: Average Candidate Leaned Preference for two Pre-Election Polls (unweighted estimates in parentheses)

	Dual Frame			**CPO + Landline**			**Landline Only**		
	Obama	McCain	Obama Margin	Obama	McCain	Obama Margin	Obama	McCain	Obama Margin
Population	2,261			1,885			1,735		
	49.7%	37.8%	+11.9	48.7%	37.9%	+10.8	47.0%	38.9%	+8.1
	(46.5)	(42.1)	+4.4	(46.1)	(41.5)	+4.6	(45.6)	(41.6)	+4.0
Registered Voters	2,041			1,706			1,597		
	48.6	40.0	+8.6	47.8	40.3	+7.5	46.7	41.0	+5.7
	(46.0)	(42.9)	+3.1	(45.5)	(42.6)	+2.9	(45.0)	(42.6)	+2.4

Design Effects or Weighting Efficiencies

Sample design choices have implications for variability due to weighting. Table 5 shows the weight efficiencies and deffs due to weighting. Final weights are shown along with a decomposition of the two component parts of the weights described earlier. Not surprising, pre-weights for the landline only design are the most efficient and that the dual frame design, by its complex nature is the least efficient. If we isolate demographics we see improvements in efficiency for either design that includes cell phone sample. Given the unweighted comparison against CPS earlier shows dual frame as the more favorable, it is expected that demographic weighting would perform better in the dual frame design than in other options.

[12] Cell phone only, both, and landline only estimates from MRI Fall 2008.

Deffs shown in Table 5 are larger, that is, variability due to weighting is greater, than results shown by Kennedy (2007) in her Service method. We find similar comparisons in two of four polls where the deff is greater for dual frame (no screener) than for CPO (screener). There are several possible explanations for the greater variability shown here. Data that Kennedy analyzes has a larger cell phone sample, thus weights do not have to work as hard to bring service in line (less variability on this factor). Another possibility, even though most variables used in weighting are the same, is that there are differences in how those variables are categorized in the raking. One last differentiating factor may be in the pre-weight differences noted earlier. Regardless, in Kennedy's analysis as well as the analysis here, the core demographics employed in weighting are a constant across the three sample designs.

Table 5: Weighting Efficiencies and Design Effect due to Weighting across several polls

		Efficiencies due to Weighting			**Deffs due to Weighting**		
Poll	Weight Type	Landline Only	CPO + Landline	Dual Frame	Landline Only	CPO + Landline	Dual Frame
P1	Pre weight	90%	71%	69%	1.11	1.41	1.45
	Balancing Weight	44%	54%	55%	2.26	1.87	1.82
	Final Weight	42%	50%	48%	2.38	1.95	2.08
P2	Pre weight	90%	69%	57%	1.11	1.45	1.75
	Balancing Weight	47%	53%	56%	2.13	1.90	1.80
	Final Weight	44%	49%	49%	2.30	2.03	2.03
P3	Pre weight	88%	63%	61%	1.14	1.58	1.64
	Balancing Weight	50%	52%	55%	2.00	1.92	1.82
	Final Weight	46%	49%	48%	2.19	2.06	2.08
P4	Pre weight	87%	62%	58%	1.15	1.61	1.72
	Balancing Weight	59%	65%	65%	1.70	1.53	1.53
	Final Weight	54%	59%	57%	1.86	1.65	1.75

NOTE: Efficiencies are computed as the sum of the weights divided by the sum of the squared weights. No trimming or treatment of the weights is present to show the true results.

Table 6 compares the dual frame 20/80 design against two modeled designs. First, the CPO portion is replicated so that CPO portion of the sample is nearly proportionate to the population estimate and then weighted. This design is more efficient than the dual 20/80 design. A dual 50/50 design is also modeled that results in the CPO portion of the cell frame is nearly proportional to the population estimate and here we get final weights that exceed in efficiency either design with a cell phone sample.

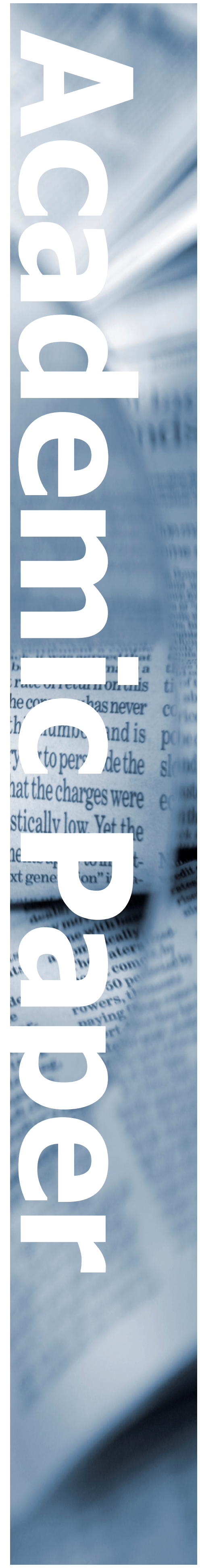
Table 6: Comparison of Dual 20/80 and Modeled Dual 50/50

	CPO	Dual 20/80	Dual 50-/50
Pre Weight	63%	64%	81%
Balancing weight	60	57	70%
Final Weight	55	50	63%

Cost Model

Conventional thinking depicts the cost of completing interviews via a cell phone frame at some ratio to the cost of a landline interview. One example might be that the cost per interview (CPI) by cell phone is two times that of the landline CPI. It needs to be recognized that this ratio of costs varies. A ten minute interview might be twice the CPI in the cell frame versus the landline CPI but this ratio is smaller for say a 15 minute interview. This is because the costs of cell phone interviews are incremental in nature.

The core questionnaire content is common regardless of sample frame. Incremental costs are found in additional screening questions necessary for those contacted on a cell phone; additional questions are needed to gather information for distribution a reimbursement, the reimbursement, and the cost differential in cell versus landline samples. The choice to screen for cell phone only persons has that as an incremental cost as well given a certain percentage of the cell frame contains these persons of interest. Contact rates, cooperation and other sample disposition rates differ by cell frame and landline as well, which may be thought of in terms of a ratio. Thus, a cost model that covers variants in interview length or population members is complex with core fixed costs, incremental costs assigned to the cell frame



interviews, and variable costs between the two sample types. However, ratio in one design to another make understanding sample frame design decisions easier to comprehend.

AP-GfK polls average about 15 minutes for the landline interview. Additional costs then result in a ratio of 1.8:1, cell phone CPI to landline CPI. Further, we estimate the incremental costs of screening for CPO persons and get a ratio of 2.9:1. These ratios lend themselves to an allocation model across the range of designs described in this paper.

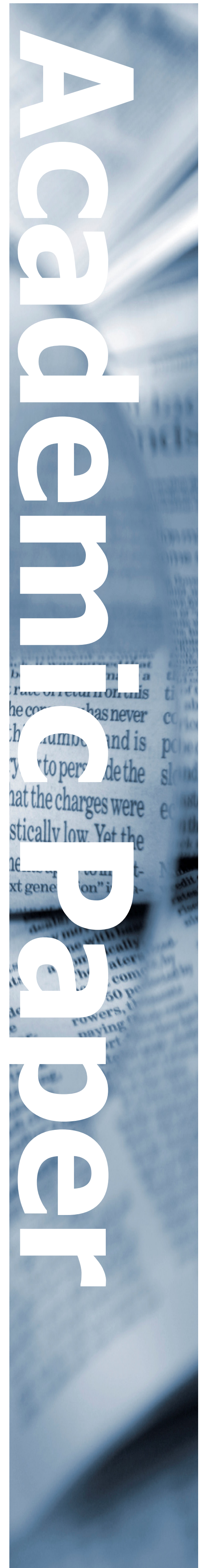
If we start with a landline interview the cost can be expressed as the sample size times the cost of obtaining each interview with n as the sample size and C as the cost per interview:

$$^{n}\text{landline} * {}^{C}\text{landline}$$

Costs for obtaining interviews from a cell frame are higher due to higher sample cost, manual dialing, asking additional questions to ascertain safety and age, and offering a re-imbursement along with collecting the relevant information and processing re-imbursements.[13] Our experience tells us that the cost of a cell phone interview is a little less than double a landline interview. Conceptually then the cost is:

$$({}^{n}\text{landline} * {}^{C}\text{landline}) + ({}^{n}\text{Cell phone} * 1.8\,{}^{C}\text{landline})$$

While this might be useful to compute the total cost, another approach is to understand the difference in cost from a base of landline cost. Total cost then is dependent on the allocation of sample[14]. To understand the relative cost of our design decision we substitute n for each



[13] Re-imbursement is $5.
[14] Costs do not include project management and statistical support in developing weights and estimation.

sample frame with the portion of sample allocation. This is A for landline and 1-A for cell frame:

$$(^{AC}\text{landline}) + ((1\text{-}A)1.8^{C}\text{landline})$$

If A = 1 then the cost is entirely that of a landline sample. Our 80/20 dual frame design then is 16% greater than a landline only sample cost. This is:

$$(0.80^{C}\text{landline}) + (0.20)*\ 1.8^{C}\text{landline}) = 1.16^{C}\text{landline}$$

Similarly, if we screen for CPO we estimate that the CPO is about three times the cost of a landline interview[15]. If we set CPO at 10% of the sample the cost is 20% greater than a landline only sample:

$$(0.90^{C}\text{landline}) + ((0.10)\ 2.9^{C}\text{landline}) = 1.19^{C}\text{landline}$$

A 50% landline, 50% cell frame design is 40% more than a landline only design:

$$(0.50^{C}\text{landline}) + ((0.50)1.8^{C}\text{landline}) = 1.4^{C}\text{landline}$$

A comparable coverage solution to the 50/50 dual frame is to sample CPO proportionate to population estimates at 17.8% which is:

$$(0.822^{C}\text{landline}) + ((0.178)\ 2.9^{C}\text{landline}) = 1.34^{C}\text{landline}$$

[15] A large part of the cost of screening for CPO may be related to how simple or how complex that approach to screening of this population is, i.e., number of questions to define a person as CPO.

This may be a useful way to understand the relative cost of sample design decisions in contrast to traditional landline designs. However, these should be put in the context of efficiency (Table 7).

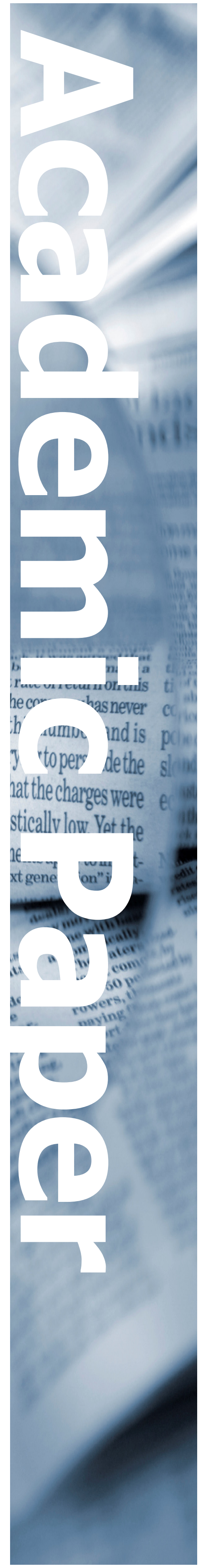
Table 7: Comparison of Efficiencies, Relative Costs, Effective Sample (n=1,000)

	Efficiency	Relative Cost to landline	Effective Sample Size	Ratio of Effective Sample to Landline	Weighted Cost for effective n=1,000
Landline only	0.465	1.00	465	1.00	2.15
CPO w/landline	0.523	1.19	523	1.12	2.28
Dual 80/20	0.500	1.16	500	1.08	2.32
Dual 50/50	0.630	1.40	630	1.35	2.22
CPO proportional	0.550	1.34	550	1.18	2.44

Table 7 shows the importance of accounting for precision needs or tolerances around estimates based on design decisions and when weights are used to approximate unbiased estimates. Although landline samples are the least expensive, they result in the least amount of effective sample size. If, for example, precision is needed at +/-3.1% at the 95% confidence level, then effectively, a sample of 1,000 is needed. Aside from dampening the variability through trimming weights to lift the effective sample size, which would affect each design comparably, collecting a larger unweighted sample to achieve the desired effective sample size can be considered. The last column in Table 7 shows the adjusted costs to achieve this goal. While landline, with all its coverage issues, is still the least expensive, the best buy is a dual 50/50 design.

Discussion

Landline samples are the least expensive but depart significantly from demographic estimates. Keeter et al (2007)[16] document well the diminishing younger adult population in surveys important to some survey estimates. Options for improving upon landline samples are to add



[16] See Figure 1, p 774.

CPO to improve coverage or go with full dual frame approaches to improve coverage as well as potential non-response when we connect through cell phones with those that we might not in the landline frame. Dual frame approaches get us closest to certain demographic population distributions, however, due to the complex nature of dealing with frame overlap the overall weighting for these designs is the least efficient unless one employs designs where the cell frame proportion of the total sample is more in line with population estimates.

Although there is a thirst for the "right" answer to weighting adjustments in cell/landline dual frame surveys, we assert that there is not one right answer. There are an inexhaustible number of variables, combinations of those variables, and the categories that variables can be sorted into. As an example, we chose to leave home ownership and number of adults out of the weighting algorithm even though they remain potential sources of bias. In sum, the use of weights often represents a trade-off: weights can reduce bias, but they can also lead to inflated sampling variances in our estimates.

In this paper we use efficiency as a more straightforward form of understanding the implications of inflation of sampling variances due to weighting. Statisticians can quantify the impact of weighting on sample estimates in terms of "effective sample size" or efficiency rather than inflated variances as design effect. By computing an efficiency based on the variability due to weighting we both quantify the variance inflation effects of weighting in readily understood terms, and facilitate significance testing that takes into account possibly inflated variances.

Conclusions

Dual frame designs are an improvement in achieved samples that match more closely robust population estimates. This is a desirable outcome. Weights can be computed that correct for selection probabilities, deal with frame overlap, and adjust achieved samples to robust population estimates in a more than adequate fashion albeit with inflated variance due to weighting based on sample design decisions and the achieved sample distribution along, in particular, those variables used in weighting data. Greater weighting variability is found in the dual frame designs, but these designs can guard against coverage error and the potential for non-response among those elements we may or may not get through a landline sample. Dual frame designs appear to be critical to precision in estimates when the weighting variables are covariates of the survey estimates. Design decisions not only have implications for financial costs but costs in effective sample size because of weighting variability.

It is difficult to know *a priori* precisely what types of estimates are biased in the absence of cell phone samples, but it is clear that some estimates are affected. Our advice is to error on the side of better precision through achieving demographic coverage when it is possible that estimates will be biased and lead to incorrect inferences.

References

Blumberg SJ, Luke JV, *Cell phone Substitution: Early Release of estimates from the National Health Interview Survey, January-June 2008*, National Center for Health Statistics. Available from: http://www.cdc.gov/nchs/nhis.htm. December 17, 2008

Brick, J. Michael, Sarah Dipko, Stanley Presser, Clyde Tucker, and Yangyang Yuan, *Nonresponse Bias in Dual Frame Sample of Cell and Landline Numbers*, Public Opinion Quarterly, Special Issue 2006, Volume 70 / Number 5:780-793.

Hartley, Herman O. 1962. "Multiple Frame Surveys." *Proceedings of the Social Statistics Section, American Statistical Association*, 203-206

Keeter, Scott, *The Impact of Cell Phone Noncoverage Bias on Polling in the 2004 Presidential Election*, Public Opinion Quarterly, Spring 2006, Volume 70 / Number 1: 88-98.

Keeter, Scott, Courtney Kennedy, April Clark, Trevor Tompson, and Mike Mokrzycki, *What's Missing from National Landline RDD Surveys? The Impact of the Growing Cell-Only Population*, Public Opinion Quarterly, Special Issue 2007, Volume 71 / Number 5:772-792.

Kennedy, Courtney, *Evaluating the Effects of Screening for Telephone Service in Dual Frame RDD Surveys*, Public Opinion Quarterly, Special Issue 2007, Volume 71 / Number 5: 750-771.

Lavrakas, Paul J, Charles D. Shuttles, Charlotte Steeth, and Howard Fienberg, *The State of Surveying Cell Phone Numbers in the United States: 2007 and Beyond*, Public Opinion Quarterly, Special Issue 2007, Volume 71 / Number 5: 840-854.