

## **RDD Unplugged: Findings from a Household Survey Using a Cell Overlap Design**

John Hall,<sup>1</sup> Barbara Lepidus Carlson,<sup>2</sup> Karen CyBulski<sup>1</sup>

<sup>1</sup>Mathematica Policy Research, Inc., P.O. Box 2393, Princeton, NJ 08543-2393

<sup>2</sup>Mathematica Policy Research, Inc., 955 Massachusetts Ave., Suite 801, Cambridge, MA 02139

### **Abstract**

Sampling the U.S. residential population using list-assisted random digit dialing (RDD) of landline telephone numbers has become problematic because of the increasing proportion of the population that is reachable only through cell phones. To address this coverage problem, round 6 of the Health Tracking Household Survey (HTHS6) employed an RDD dual-frame “cell overlap design”: samples were selected from landline and cell frames, and interviews were attempted with all contacted households. Other approaches sometimes used to address the coverage issue include address-based sampling and dual-frame RDD designs in which the cell frame is screened for cell-only households. HTHS6 asked a series of questions about telephone usage from respondents in both the landline and cell sample frames. This paper will discuss contact and cooperation rates, along with number of calls per complete, by sample frame. In addition, this paper will provide information about landline and cell telephone usage by sample type and compare characteristics among the various telephone usage categories (cell only, cell mostly, some of each, landline mostly, landline only), including demographics, health status, and insurance coverage.

**Key Words:** random digit dialing, cell phone, dual-frame, sampling

### **1. Background and Introduction**

Between 1990 and 2010, the number of cell phone subscribers in the United States rose from about 5 million to about 300 million (Infoplease.com and NationMaster.com) (Table 1). As of 2000, the prevalence of cell phone-only households—those without any landline—was less than one percent (Tucker et al. 2007) but had risen to nearly 30 percent by 2010 (Blumberg and Luke 2011).

This drastic change in cell phone usage has significantly affected the coverage of surveys that use random digit dial (RDD) sampling, a common sampling technique for surveys targeting the general population. Because of this trend in cell phone usage over the last decade, using only a landline-based RDD sample results in reducing the coverage of the population. RDD surveys attempt to cover most or all of the population, but until recently one could only obtain RDD samples for landline telephone numbers from sampling vendors. In 2003, Survey Sampling, Inc. made RDD cell phone samples available to its clients. Another major vendor of RDD samples, Marketing Systems Group, Inc., followed suit two years later.<sup>1</sup> As a result, RDD for cell phones is now possible.

---

<sup>1</sup> Another way some surveys have addressed the coverage issue is to use address-based sampling (ABS), which makes use of the U.S. Postal Service’s Delivery Sequence File. For a

**Table 1: Trends in Cell Phone Usage in the United States, 1990–2010**

Year	Households (million)*	Cellular Subscribers** (million)	Cell-Only Households (percentage)	Landline-Only Households (percentage)
1990	91.9	5.3	---	---
1994	97.1	24.1	<1	---
2000	105.5	109.5	<1	---
2004	112.0	158.7	5.0	42.2
2006	114.4	233.0	12.8	29.6
2008	116.8	262.7	20.2	17.4
2009	117.2	***	24.5	14.9
2010	117.5	302.9	29.7	12.9

\*Rawlings (n.d.), Simmons and O'Neil (2001).

\*\*Note that there can be multiple subscribers per household.

\*\*\*We did not find this estimate published but expect it exists.

This paper will discuss a national RDD telephone survey, the Health Tracking Household Survey (previously known as the Household Component of the Community Tracking Study), that added a cell phone RDD sample to the most recent round of data collection. The Health Tracking Household Survey (HTHS) is a periodic telephone survey of U.S. households and their interaction with the health care system, including insurance coverage, access to care, and use of health care. The first HTHS survey was completed in 1996. The survey is funded by the Robert Wood Johnson Foundation, and it is sponsored and designed by the Center for Studying Health System Change ([www.hschange.org](http://www.hschange.org)). Mathematica Policy Research, Inc. assisted in the design and is responsible for sampling and for data collection.

## 2. Methods

The current study uses data from the sixth round of the survey, conducted between April 2010 and March 2011. Because we could no longer ignore the decreasing coverage of samples based on landline RDD, we decided to introduce the cell phone RDD sample for round 6. Accounting for the relative cost and variance associated with each, we determined the optimal allocation between the landline and cell phone samples: 75 percent of the completed interviews should come from the landline sample and 25 percent from the cell sample. We released a total of 53,738 telephone numbers, about 40 percent of which were from the cell phone sampling frame, and obtained 7,596 completed household interviews, 25.5 percent of which were from the cell phone frame.

---

*(continued)*

telephone survey, we would need to match telephone numbers to addresses or to use multiple modes. However, the address match rates would be insufficient for our purposes here. In addition, this interview cannot be self-administered by mail or web because it is lengthy, has complex skip patterns, and requires handoffs to other household members. The HTHS can only be completed by phone.

### 3. Results

The cooperation rate and level of effort was different between the two sample types, as we expected (Table 2):

**Table 2: Completion Statistics for Round 6 of the Health Tracking Household Survey**

	Landline Sample	Cell Sample	All Households
Sampled Phone Numbers	32,069	21,669	53,738
Household Completes	5,659	1,937	7,596
Completed on Other Phone Type	229	51	280
Ineligibility Rate	45.7	51.8	48.2
Contact Rate Among Possible Eligibles	64.0	64.4	64.2
Cooperation Rate Among Contacted	50.8	28.8	42.5
Interviewer Hours Per Family Complete	1.9	3.1	2.2

Most ineligible phone numbers in both samples were unassigned, nonworking, or nonresidential numbers, but the cell sample had a slightly higher ineligibility rate than did the landline sample. The main reason for this is that the landline RDD sample is “list-assisted,” meaning that phone numbers are only sampled from series of numbers in which at least one phone number is a published residential number. However, there is no list-assisted sampling for cell phone numbers, so more numbers are unassigned. In addition, our sampling vendor, Marketing Systems Group, has a procedure for prescreening phone numbers to detect nonworking and nonresidential numbers, but the option is only available for landline numbers, not cell numbers. Finally, we were able to match many phone numbers to addresses for the landline sample, allowing us to send an advance letter along with a \$5 bill and a promise of a larger incentive check upon completion of the survey. But we found that address matching for the cell sample had a low success rate, so we stopped trying to match addresses after the first cell phone sample release. Cell phone numbers are also ineligible if they belong to a child, but no such exclusion exists for household landlines.

The contact rates were remarkably similar between the landline and cell phone samples. Among those with whom we made contact, we saw a much higher cooperation rate in the landline sample compared to the cooperation rate in the cell sample. The two samples had fairly comparable rates of “any refusal” (38 percent for landline and 46 percent for cell), but during the first round of refusal conversion, we found that the cell phone cases were almost twice as unlikely to answer our call, which meant that the refusing cell phone cases were less likely to be converted from a refusal to a complete. As expected, the number of interviewer hours per complete at the family level<sup>2</sup> was almost two-thirds higher for the cell phone sample: these were hours spent dialing and actively interviewing.

---

<sup>2</sup> We did a separate core interview with each family unit within a household, defining families as those persons who would typically be covered under a single health insurance policy.

The HTHS survey first requires that a household respondent enumerate the household members and their ages and relationships. Next, the computer-assisted telephone interviewing program organizes the household into family units. The program then randomly selects one child per family unit. We attempt to get a separate “core” interview from a family informant in each family unit and to get a separate “self-response module” interview from each adult in the household. We were concerned that these handoffs to multiple respondents within the household would be more problematic for the cell phone sample. Although we did see a higher success rate for the landline sample, the difference was not as substantial as expected: 88 percent versus 79 percent for the cell phone sample.

Table 3 shows the final unweighted<sup>3</sup> response rates for each sample type at the household level, family level, and cumulative; we have estimated the number of eligible phone numbers for those with undetermined status.

**Table 3: HTHS Round 6 Response Rates<sup>4</sup>**

	Household	Family (marginal)	Family (cumulative)
Landline Sample	47.3	96.0	45.4
Cell Sample	31.3	92.9	29.1

Although many other surveys introducing cell phone RDD samples screen out any people who could have also been reached on a landline (choosing to capture cell-only households with the additional sample), we chose to implement a cell overlap design in which all cell phone responses were retained if respondents were otherwise eligible for the survey<sup>5</sup> and the cell phone number did not belong to a child. In so doing, we maximized the number of respondents and minimized screening costs, but we also had to contend with the chances of dual selection into both the landline and cell RDD samples in the weighting process. (Details of the weighting process are not covered in this paper.) To obtain the information necessary for weighting, we added a number of questions about telephone usage to the survey instrument.

First, regardless from which sample the telephone number came, we asked,

**Is XXX-YYY-ZZZZ a cellular telephone?**

Of the 5,659 completed household interviews from the landline sample, 229 (four percent) said we had reached them on a cell phone. And of the 1,937 completed household interviews from the cell phone sample, 51 (three percent) said we had reached them on a landline. The phone questions that followed were based on which type of

---

<sup>3</sup> The unweighted and weighted response rates are virtually identical within sample type.

<sup>4</sup> The response rate was computed using AAPOR standard definition Response Rate 3 (AAPOR 2011). The calculation of the eligibility rate was done as described in Strouse et al. 2007 for Round 5.

<sup>5</sup> To be eligible, a phone number had to be associated with a working residential phone on which people received nonbusiness calls. Further, to be eligible for the survey, the household had to contain at least one civilian adult.

phone we reached them, not the phone type from which they were sampled. If we reached a cell phone, we then asked,

**For classification purposes, can you tell me what state you are living in now?**

We found that almost eight percent of respondents we reached on cell phones had an area code serving a state different from the one in which they lived. This points to another issue with sampling cell phones, within-state migration to locations served by different area codes. With landlines, we know something about the location of households served by sampled telephone numbers because of the area code and exchange. However, people can continue using their cell phone number even if they move to someplace with a completely different area code; further, a cell phone number's exchange (the three numbers following the area code) does not have any geographical significance.

Another complication of including the cell phone sample is logistics. We asked all respondents reached on a cell phone,

**Your safety is important to me. Are you driving in a car, walking down the street, in a public place or other location where talking on the phone might distract you or jeopardize your safety and/or confidentiality?**

**IF YES: I would like to call you at a more convenient time. IF NEEDED: If you would prefer that I call you at another telephone number, I can do that too.**

Finally, toward the end of the questionnaire, we asked a series of questions about phone usage. The following ten questions were asked of those reached on a landline:

- 1. Are there any telephone numbers INSIDE your home, other than XXX-YYY-ZZZZ, that people receive calls on but that are NOT cell phones?**
- 2. How many of these additional phone numbers are ONLY used for business purposes or to connect a computer or fax machine?**
- 3. Do you or any other adults in your household have a working cell phone?  
IF YES: How many cell phones do you and other adults in your household have?**
- 4. How many of these cell phones are ONLY for business purposes?**
- 5. How many adults in the household have a cell phone they receive personal calls on?**
- 6. Do you share a cell phone for receiving personal calls with other adults in the household?**
- 7. Thinking about all the calls your household receives, how many of these calls are received on cell phones? Would you say that**
  - all or almost all 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?**
- 8. Not counting cell phones, has your household been without telephone service for two weeks or more during the past 12 months?**

**9. For how long was your household without telephone service in the past 12 months?**

**10. When your household was without telephone service, did someone in your household have a working cell phone?**

The versions asked of those reached on cell phones were similar but not worded exactly the same. For example, instead of asking, “**Are there any telephone numbers INSIDE your home, other than XXX-YYY-ZZZZ, that people receive calls on but that are NOT cell phones?**” we asked, “**Are there any telephone numbers INSIDE your home that people receive calls on but that are NOT cell phones?**” Instead of asking, “**Do you or any other adults in your household have a working cell phone? IF YES: How many cell phones do you and other adults in your household have?**” we asked, “**Not counting XXX-YYY-ZZZZ, how many working cell phones do you and other adults in your household have?**”

Using the responses to some of these questions, we were able to quantify the number of landlines and cell phones in each household and to classify each household into one of the following five categories:

1. Cell only
2. Cell mostly
3. Some of each
4. Landline mostly
5. Landline only

Table 4 shows the distribution of phone usage by type of phone reached.

**Table 4: Phone Usage by Type of Phone Reached (Unweighted Percentages)**

	Landline Phone	Cell Phone	All Households
	n = 5,481	n = 2,115	n = 7,596
Cell Phone Only	0	61	17
Cell Phone Mostly	12	7	11
Some of Each	33	21	29
Landline Mostly	25	11	21
Landline Only	30	0	22
Total	100	100	100
Percentage of All Households With Two or More Landlines	2.2	0.7	1.7
Percentage of All Households With Two or More Cell Phones	39.2	58.8	44.7

Of those we reached on a landline, about a third were landline only, a quarter landline mostly, and a third some of each. But for those reached on a cell phone, nearly two-thirds were in cell phone-only households. Remarkably, there was a higher percentage of cell

phone mostly households among the landline respondents than among the cell phone respondents.

The percentages expressed in Table 4 are unweighted. In this section, we present weighted distributions and means. The weighting process for the HTHS survey has a number of steps and stages. We constructed weights for households, family units and individuals. To construct household-level weights, we performed the following steps:

- Calculated the probability of selection of each phone number within each type of sample (landline or cell).
- Adjusted for whether the eligibility of the phone number was determined.
- Adjusted for household eligibility and response among known households.
- Adjusted for multiple chances of selection within phone type.
- Poststratified the household count so that it added up to the total number of households in the contiguous United States and so that the cell-only and landline-only portions of the sample accounted for 27 and 13 percent of the households, respectively.
- For the rest of the households, multiplied the dual-user households from the landline sample by  $\lambda$  (where  $0 < \lambda < 1$ ) and those from the cell phone sample by  $1-\lambda$ .<sup>6</sup> This addresses the overlap population between the two sample types.

To construct family-level weights, we started with the household weight through the previous step, then adjusted for nonresponse among secondary families in responding households. No sampling was done at the family level—all families within a selected household were included. To construct person-level weights, we adjusted for the random selection of one child per family. Other than that, no sampling was done at the person level; because there was one informant per family, there was essentially no person-level nonresponse within responding families. The final steps were to poststratify the person-level weights by demographic characteristics (such as gender, age group, Hispanic ethnicity, race, and education) and then to trim any outlier weights.

Table 5 shows some characteristics of household composition by phone usage category; these are weighted by the final household-level weight as described above.

Table 5 shows statistically significant differences across the phone usage categories in terms of the mean number of families and persons per household, although these differences are not meaningfully large. The mean number of children per household is much lower in the landline-mostly and landline-only households, but this is likely

---

<sup>6</sup> For this study, we decided on a value of  $\lambda = .55$ . Derivation of this value is beyond the scope of this paper.

because of the higher ages of those in such households, which we address in Table 7 below.

**Table 5: Household Composition by Phone Usage Category**

	Cell Only	Cell Mostly	Some of Each	Landline Mostly	Landline Only
Sample Size	1,274	829	2,278	1,585	1,630
Weighted Percentage	26.9	11.8	28.7	19.4	13.2
Mean Families Per Household*	1.25	1.37	1.31	1.22	1.25
Mean Persons Per Household*	2.18	2.73	2.76	2.14	2.09
Mean Children Per Household*	0.61	0.71	0.71	0.39	0.45

\* $p < .05$  (design-adjusted ANOVA).

Table 6 shows phone usage distribution by region of the country. This table is weighted by the final household-level weight.

**Table 6: Phone Usage Category by Census Region**

	Northeast	Midwest	South	West
Cell Only	22.6	27.4	29.7	25.9
Cell Mostly	10.5	10.5	12.3	13.7
Some of Each	31.9	28.1	27.8	27.6
Landline Mostly	22.1	20.9	17.7	18.2
Landline Only	12.9	13.1	12.5	14.6
	100.0	100.0	100.0	100.0

$p = .006$  (design-adjusted chi-square).

Phone usage proportions vary slightly by region. The South has the highest proportion of cell-only households, and the Northeast has the lowest. The West has the highest proportion of landline-only households, and the South has the lowest.

Table 7 shows sociodemographic characteristics at the person level and is limited to adults in the sample. This particular table is weighted by the penultimate person-level weight as described earlier. It has all adjustments (including the cell overlap adjustment) except for the final poststratification adjustments and trimming, which could confound the weighted estimates presented here. Keep in mind that these estimates are not fully adjusted for differential nonresponse patterns among various demographic groups not explicitly accounted for in the weighting process. Because of this, the estimates presented here may not match the official estimates published by the Center for Studying Health System Change.

There are statistically significant differences in these sociodemographic characteristics by phone usage category. Although the way in which demographic characteristics are summarized here is a bit different than published findings from some other surveys, these findings, particularly those for the cell only population, are similar to those found in other national surveys such as the National Health Interview Study (NHIS; Blumberg and Luke 2011), especially in terms of race, ethnicity, gender, and uninsured rate. We see that among adults, the mean age is lowest for those in cell-only households and highest for those in landline-mostly and landline-only households. Those in the cell-only and landline-only “tails” of the phone usage continuum are more likely to be Hispanic,



nonwhite, in poverty, in poor health, and uninsured. In particular, the cell-only population, which would not have been reachable without the addition of the cell RDD sample, has the highest uninsured rate (31 percent), and those in dual-usage households have the lowest uninsured rate (11 to 14 percent). Given the key outcomes from this particular survey, excluding the cell-only population (and much of the cell-mostly population) would have underestimated the percentage of uninsured persons in the United States.

**Table 7: Weighted Person-Level Sociodemographic Estimates (Adults)**

	Cell Only	Cell Mostly	Some of Each	Landline Mostly	Landline Only
Sample Size	2,206	1,747	4,858	2,906	2,686
Weighted Sample (Thousands)	58,504	29,850	71,449	40,221	27,640
Mean Age*	40.3	44.4	48.3	57.7	53.0
Percentage Hispanic*	22.3	10.9	7.2	8.1	17.7
Percent White*	63.5	77.7	81.2	84.4	72.3
Percentage Male*	49.7	47.4	45.7	44.2	42.9
Percentage Working Last Week*	47.2	61.3	56.9	40.9	30.9
Percentage in Poverty*	25.1	10.5	7.5	9.5	25.4
Percentage in Poor Health*	5.5	3.0	3.4	5.1	7.8
Percentage Uninsured*	31.2	13.6	11.2	12.3	20.4

\* $p < .05$  (design-adjusted chi-square or ANOVA).

#### 4. Conclusion

With this paper, we hope to add to the existing published findings on this important emerging issue as it pertains to survey data collection. These findings, particularly those for the cell-only population, are similar to those found in other national surveys such as the NHIS. However, we have expanded the findings by creating and presenting findings for more refined phone usage categories, such as cell only, cell mostly, some of each, landline mostly, and landline only.

We found that people in landline-only and cell-only households are more likely than those in mixed-use households to be Hispanic, nonwhite, lower income, in poor health, and uninsured. Those in cell-only households are younger, with more children in the household. In fact, people in landline-mostly households today look similar to what those in landline-only households used to look like: older, white, and female, with fewer children. We found that a very high percentage of the cell phone sample was cell phone only. We also found that handoffs to other respondents in the household were relatively successful, though not as easy as in the landline sample.

As expected, the cell phone sample was much more expensive to collect data from and its response rate was lower, but we allocated the sample in such a way as to minimize the cost and the variance while minimizing the bias due to undercoverage. We believe that our choice of a cell overlap (rather than cell only) design was more efficient. Had we excluded mixed-used households from the cell sample, we would have lost 825 completed household interviews.

## 5. Future Work

We plan to write a follow-up paper that describes the procedures we used in creating the composite weight that accounts for the overlap between the RDD landline and RDD cell phone samples. We will also compare re-weighted estimates excluding the cell-only households, and those excluding the cell phone sample, to see what kind of bias we would have experienced had we not included the cell sample.

## References

- The American Association for Public Opinion Research. 2011. “Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 7th edition. AAPOR.
- Blumberg, Stephen J., and Julian V. Luke. “Wireless Substitution. Early Release of Estimates from the National Health Interview Survey, July to December 2010.” Washington, DC: National Center for Health Statistics, June 2011. Retrieved from <http://www.cdc.gov/nchs/nhis.htm>
- Infoplease.com. “Cell Phone Subscribers in the U.S., 1985–2008.” Retrieved from <http://www.infoplease.com/ipa/A0933563.html>
- NationMaster.com. “Media Statistics. Mobile Phone Subscribers by Country.” Retrieved from [http://www.nationmaster.com/graph/med\\_mob\\_pho\\_sub-media-mobile-phone-subscribers&date=1990](http://www.nationmaster.com/graph/med_mob_pho_sub-media-mobile-phone-subscribers&date=1990)
- Rawlings, Steve W. “Population Profile of the United States: Households and Families.” Washington, DC: U.S. Census Bureau, n.d. Retrieved from <http://www.census.gov/population/www/pop-profile/hhfam.html>
- Simmons, Tavia, and Grace O’Neill. “Households and Families: 2000 (Census 2000 Brief).” Washington, D.C.: US Census Bureau, 2001. Retrieved from <http://www.census.gov/prod/2001pubs/c2kbr01-8.pdf>
- Strouse, Richard, Barbara Carlson, John Hall, and Karen CyBulski. “Community Tracking Study Household Survey Methodology Report 2007 (Round 5). Washington, DC. Technical Publication No. 72, June 2009. Retrieved from <http://www.hschange.org>.
- Tucker, Clyde, J. Michael Brick, and Brian Meekins. “Household Telephone Service and Usage Patterns in the United States in 2004: Implications for Telephone Samples.” *Public Opinion Quarterly*, vol. 71, spring 2007, pp. 3–22.