WORKING PAPER 1

MATHEMATICA
Policy Research

Working PAPER

BY JOHN HALL, BARBARA LEPIDUS CARLSON, AND KAREN CYBULSKI

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

December 2011

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 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. In this paper, we discuss contact and cooperation rates, along with number of calls per complete, by sample frame. In addition, we provide information about landline and cell telephone usage by sample type and compare characteristics among the following telephone usage categories:

- Cell only
- Cell mostly
- Some of each
- Landline mostly
- Landline only

In this comparison we will also include demographics, health status, and insurance coverage.

Background and Introduction

Between 1990 and 2010, the number of cell phone subscribers in the United States rose from some 5 million to about 300 million (Infoplease.com and NationMaster.com) (Table 1).

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

Year	Households (millions)[a]	Cellular Subscribers[b] (millions)	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	[c]	24.5	14.9
2010	117.5	302.9	29.7	12.9

[a]Rawlings (n.d.), Simmons and O'Neil (2001).
[b]Note that there can be multiple subscribers per household.
[c]We did not find this estimate published but expect it exists.

As of 2000, the prevalence of cell-only households—those with no 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 dialing (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.[1] As a result, RDD for cell phones is now possible.

We discuss a national RDD telephone survey, the Health Tracking Household Survey (HTHS) (previously known as the Household Component of the Community Tracking Study), which added a cell phone RDD sample to the most recent round of data collection. The 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, which is funded by the Robert Wood Johnson Foundation, is sponsored and designed by the Center for Studying Health System Change (CSHSC) (www.hschange.org). Mathematica Policy Research assisted in the design and is responsible for sampling and data collection.

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.

Results

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

[1]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 telephone survey, we would need to match telephone numbers to addresses or 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.

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
Within-household handoff success rate	87.9	79.1	85.6

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. Under study procedures, cell phone numbers were ineligible if they belonged to a child, but no such exclusion existed 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 difference in rates of "any refusal" for the two samples was substantial (38 percent for landline and 46 percent for cell) but not nearly as large as the difference in the final cooperation rate. However, 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[2] was almost two-thirds higher for the cell phone sample: these were hours spent dialing and actively interviewing.

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 attempted to get a separate "core" interview from a family

[2]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.

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. We did find a higher handoff success rate for the landline sample than for the cell phone sample: among households in the landline sample, 88 percent of the attempted handoffs were successful, while among households in the cell phone sample, the rate was 79 percent (Table 2).

In Table 3, we show the final unweighted[3] 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. The much lower cooperation rates in the cell phone sample (shown above in Table 2) contributed to the markedly lower response rates at the household level and in the cumulative household-family response rates.

Table 3. HTHS Round 6 Response Rates[4]

	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[5] 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. (We do not cover the details of the weighting process 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 of the sample group (cell phone or landline), we asked:

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

Of the 5,659 completed household interviews from the landline sample, 229 respondents (4 percent) said we had reached them on a cell phone. And of the 1,937 completed household interviews from the cell phone sample, 51 respondents (3 percent) said we had reached them on a landline. The questions that followed were based on which type of phone we

[3]The unweighted and weighted response rates are virtually identical within sample type.

[4]The response rate was computed using the American Association for Public Opinion Research (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.

[5]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. We excluded active duty military from the study.

reached them on, 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 8 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 relocate to a place with a 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 respondents reached on cell phones the following question:

"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 10 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?"**

For those reached on cell phones, the questions were slightly different:

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 classify each household into one of the following five categories:

- Cell only
- Cell mostly
- Some of each
- Landline mostly
- Landline only

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

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.

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

The percentages in Table 4 are unweighted. In the remainder of this section, we present weighted distributions and means. The weighting process for the HTHS survey involved a number of 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 λ (where $0 < \lambda < 1$) and those from the cell phone sample by $1-\lambda$.[6] This adjustment addresses the overlap in 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

[6]For this study, we decided on a value of $\lambda = .55$. Derivation of this value is beyond the scope of this paper.

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, ethnicity [whether Hispanic], 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.

Also in Table 5, we show 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 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[a]	1.25	1.37	1.31	1.22	1.25
Mean persons per household[a]	2.18	2.73	2.76	2.14	2.09
Mean children per household[a]	0.61	0.71	0.71	0.39	0.45

[a] $p < .05$ (design-adjusted Analysis of Variance [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 (Percentages)

	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 dif-

ferential 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 CSHSC.

There are statistically significant differences in these sociodemographic characteristics by phone usage category. The way in which demographic characteristics are summarized here is a bit different than published findings from some other surveys. However, 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, July-December 2010; Blumberg and Luke 2011), especially in terms of race, ethnicity (whether Hispanic), 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[a]	40.3	44.4	48.3	57.7	53.0
Percentage Hispanic[a]	22.3	10.9	7.2	8.1	17.7
Percentage white[a]	63.5	77.7	81.2	84.4	72.3
Percentage male[a]	49.7	47.4	45.7	44.2	42.9
Percentage working last week[a]	47.2	61.3	56.9	40.9	30.9
Percentage in poverty[a]	25.1	10.5	7.5	9.5	25.4
Percentage in poor health[a]	5.5	3.0	3.4	5.1	7.8
Percentage uninsured[a]	31.2	13.6	11.2	12.3	20.4

[a] $p < .05$ (design-adjusted chi-square or ANOVA).

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

Future Work

We plan to write a follow-up paper describing the procedures 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. Available at http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201106.pdf. Accessed July 2011.

Infoplease.com. "Cell Phone Subscribers in the U.S., 1985–2008." Available at http://www.infoplease. com/ipa/A0933563.html. Accessed July 2011.

NationMaster.com. "Media Statistics. Mobile Phone Subscribers by Country." Available at http://www.nationmaster.com/graph/med_mob_pho_sub-media-mobile-phone-subscribers&date=1990. Accessed July 2011.

Rawlings, Steve W. "Population Profile of the United States: Households and Families." Washington, DC: U.S. Census Bureau. Available at http://www.census.gov/population/www/pop-profile/hhfam.html. Accessed July 2011.

Simmons, Tavia, and Grace O'Neill. "Households and Families: 2000 (Census 2000 Brief)." Washington, D.C.: U.S. Census Bureau, 2001. Available at http://www.census.gov/prod/2001pubs/c2kbr01-8.pdf. Accessed July 2011.

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. Available at http://www.hschange.org. Accessed July 2011.

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.

About the Series

Policymakers require timely, accurate, evidence-based research as soon as it's available. Further, statistical agencies need information about statistical techniques and survey practices that yield valid and reliable data. To meet these needs, Mathematica's working paper series offers policymakers and researchers access to our most current work.

For more information, contact John Hall, senior statistician, at jhall@mathematica-mpr.com.

www.mathematica-mpr.com

Improving public well-being by conducting high-quality, objective research and surveys

PRINCETON, NJ - ANN ARBOR, MI - CAMBRIDGE, MA - CHICAGO, IL - OAKLAND, CA - WASHINGTON, DC

MATHEMATICA
Policy Research