Cell Phone Reverse Directories – Promise or Peril?

Sarah Dipko[1] and Krista Freedman[1]
[1]Westat

Key Words: Cell phones, reverse directories, advance notification, sample purging

1. Introduction

Surveys of cell phone samples are becoming common in today's survey climate, in which coverage error due to cell phone-only households can be very large depending on the target population (Blumberg and Luke, 2007). Methods for conducting these surveys have seen great strides in development over the past five years; however two major problems have not been resolved. First, researchers have been unable to find a method of purging cell phone samples of non-working or non-residential numbers prior to data collection, resulting in inefficiency in working these samples relative to landline RDD samples. Second, no method of matching mailing addresses with cell phone numbers has been identified, which prevents mailing of advance letters and prenotification of the survey request. This second point is amplified by the fact that most cell phone studies offer financial incentives. Without a method of communicating news of incentives prior to making contact, cooperation rates are likely diminished as a result.

Several vendors now offer "reverse cell phone directory" services, in which name and address information connected to cell phone numbers is available for a fee. To test the accuracy of these services, an experiment was conducted in which cell phone numbers from two recently-conducted cell phone sample surveys were submitted for matching name and address information. The cell phone numbers selected for the experiment represented completed interviews with names and addresses obtained for incentive mailings, as well as non-working or non-residential results. One study was completed six months prior to the experiment solely in the state of California. The second study was nationwide in scope and in progress at the time of the experiment.

2. Challenges of Surveying Cell Phone Samples

Extensive research has been conducted in recent years on implementing surveys of cell phone samples. Recent guidelines for cell phone surveys issued by the American Association for Public Opinion Research (AAPOR) Cell Phone Task Force provide a thorough summary of the challenges inherent to this new form of RDD sample surveys (AAPOR Cell Phone Task Force, 2008). Key issues addressed in the guidelines include the inability to purge cell phone samples of non-working or non-residential numbers prior to data collection, and the absence of any viable database to match household mailing addresses to sampled cell phone numbers. These limitations reduce the efficiency of cell phone sample surveys relative to landline RDD sample surveys, and contribute to an increase in cost.

2.1 Inability to Purge Cell Phone Samples

Purging samples of non-working and non-residential phone numbers prior to the start of data collection is a standard feature of landline RDD telephone survey methodology, one

that has been refined and improved since 1995 (Battaglia, Ryan and Cynamon, 2005). This screening process can remove roughly half of the sample prior to data collection, with variation based on geographic coverage of the sample, resulting in significant cost savings due to reduced interviewer labor.

Absent a similar process for screening cell phone samples, significantly more labor must be expended in order to classify the status of all sampled numbers. Not only does this serve to increase cost, it also changes the content of the work for telephone interviewers, reducing the proportion of calls that result in contact or interviews with households. It has been estimated that the cost for cell phone surveys is roughly twice that of landline surveys (Keeter, Dimock, Kennedy, Best, and Horrigan, 2008). The lack of a screening technique to purge cell phone samples undoubtedly contributes significantly to this increase in cost.

2.2 Inability to Match Mailing Addresses to Sampled Numbers

Another common practice for landline RDD sample surveys is to match mailing addresses to sampled numbers, in order to mail advance letters to prenotify sampled households of the survey request. As noted in the AAPOR Cell Phone Task Force report, the databases and vendors that support this type of procedure for landline samples are not available for cell phone samples. Advance letters have been shown to increase survey participation for landline samples (De Leeuw, Hox, Korendijk, Lensvelt-Mulders, and Callegaro, 2006), but it has not yet been feasible to test this effect with cell phone samples. This inability to increase response via sending advance letters increases the cost of cell phone surveys, as larger samples are needed, increasing the labor required to dial the sample and convert a higher volume of refusals.

Some research has attempted to use text messaging as a form of prenotification for cell phone samples; however no increase in response due to text message prenotification has been demonstrated (Brick, Brick, Dipko, Presser, Tucker and Yuan, 2007). It is possible that mailed advance letters could serve to increase response for cell phone samples, but no such experiment has been feasible to date.

2.3 Research Questions

In the summer of 2008, the authors were asked to investigate the quality of "reverse cell phone directory" services currently being advertised. These services were typically offered in the form of paid searches on the Internet, conducted one at a time. Upon entry of a cell phone number, the services either returned matched information (address, name, etc.) or a "no match" result. Initial investigation of these services led to the formulation of our two research questions.

- Is it possible to purge cell phone samples of non-working numbers prior to data collection, identified by obtaining a "no match" result from a reverse cell phone directory service? Are these services reliable or is there too much error involved?

- Is it possible to prenotify cell phone sample households by matching mailing address information to cell phone numbers using a reverse cell phone directory service? Are the services able to identify a high proportion of known residential cell phone numbers? How much error can be expected in these matches?

3. Methods

Two reverse cell phone directory search vendors were selected for the experiment, one offering batch search jobs, the other offering one-at-a-time Internet searches. In July and August of 2008, samples were drawn from two cell phone sample surveys and were processed using the reverse directory vendors. Returned information from the searches was compared to survey information for each cell phone number, to evaluate the quality of the two reverse directory search vendors.

3.1 Selection of Reverse Cell Phone Directory Vendors

Our initial search for companies offering reverse cell phone directory services yielded numerous websites that appeared to reflect different company names. Upon further investigation, however, it became evident that many linked back to the same parent site or company. At the time of our experiment, 26 of the sites we found linked to Intelius, and 26 others either linked to ReversePhoneDetective.com or to its parent company (Public Data Solutions). In addition, six sites had CIS Worldwide as the parent company, and five linked back to USSearch.com.

After investigating costs and inquiring about the possibility of batch jobs with the four main vendors we had identified, we selected Intelius and ReversePhoneDetective.com for use in the experiment. These were the cheapest two options, and the only vendor that offered batch jobs with bulk discounts was ReversePhoneDetective.com.

Intelius charged $500 for 250 searches ($2.00 per search), and the searches were valid for one year. Any search not returning at least an owner name was free. Searches were performed one at a time on the Intelius website. An example return from an Intelius search is shown in Figure 1. When a match was returned often multiple names and addresses were yielded, complicating comparisons to validation data from the two case study survey samples. The nature of this search process makes it prohibitive in terms of implementing it with a large sample size of cell phone numbers.

ReversePhoneDetective.com was more expensive at $1,099.95 for 200 searches ($5.50 per search), and the searches were also valid for one year. Again, any search not returning at least an owner name was free. A list of phone numbers was sent to the company, and results were returned in a spreadsheet. The “batch” option was only available for 200 or more searches (i.e., less than 200 searches would have had to be done individually through the site). The returned information from this vendor was more straightforward to interpret, as a standard set of fields were provided, as listed in Table 1. The non-shaded rows reflect information about the cell phone number itself, the city/state/zipcode information here appeared to represent the origin of the cell phone or the billing center for the service provider. The shaded rows contained information about persons and/or addresses connected to cell phone numbers.

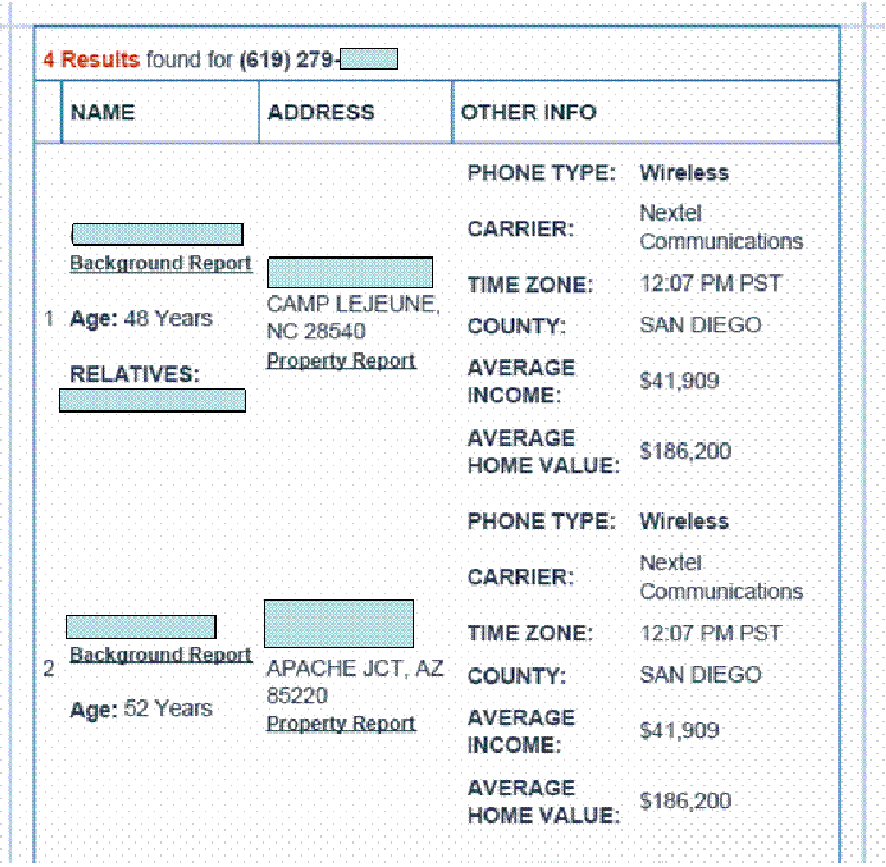

Figure 1: Example of return from Intelius cell phone reverse directory search

Table 1. List of Data Fields Returned from ReversePhoneDetective.com Cell Phone Reverse Directory Search Batch Job

Field	Description
Pk	ID number
Found	0/1 "found" or not
Phone	Phone number
Name_f	First name
Name_l	Last name
City	City
State	State
Zip	Zipcode
Carrier	Cell phone service carrier
Line_type	Landline or wireless
AddressA	Street address
AddressB	City, State and Zipcode

Note: Non-shaded rows refer to information about the actual cell phone number, and shaded rows reflect information about the person and/or address connected to a cell phone number.

3.2 Selection of Samples

The two samples of cell phone numbers used for this experiment are described below.

3.2.1 Case study number one

For our initial research conducted in July and early August of 2008, we selected 200 numbers from a cell phone sample used in a study conducted in the state of California. The numbers had reached final dispositions in January and February 2008, up to six months prior to the conduct of the experiment.

The phone numbers represented a mixture of outcomes at the screener level. We included 50 completes, 25 ineligibles due to phone status (respondents who also had landline phones), and 12 nonresidential outcomes, as all of these reflected situations in which name and mailing address had been obtained in order to mail incentives to cooperative screener respondents. In addition, 75 nonworking numbers, 13 noncontacts (numbers reaching repeated ring no answer results), and 25 cases with maximum call results were selected.

3.2.2 Case study number two

Due to concern over the limitations of the first sample, with phone numbers being searched several months after the coding of final dispositions, and the geography being limited to just one state, we selected a second sample of cell phone numbers for the experiment in August of 2008. It was possible to incorporate additional searches for the second sample under the original agreements with the two search vendors, as sufficient unused searches remained due to a lower than expected yield of matches from the first case study sample.

We drew the sample for our second case study from a nationwide cell phone study that was in operation at the time. We selected 100 completed screeners and 74 nonworking numbers from the cell phone numbers that had reached final dispositions during the month of August. Here also, the completes included name and mailing address for the purpose of providing respondent incentives, allowing validation of information returned from the search vendors.

3.3 Validation Against Survey Interview Data

Upon receiving the information from the two search vendors, results were compared to that on file for each cell phone number. Case-level coding of correspondence between the survey data and search results was conducted, and summarized in tabular form. This process was repeated for each search vendor, for both of the samples employed in the experiment. Results are described below.

4. Results

In this section, we present results of the reverse directory searches and validation coding separately for the two samples of cell phone numbers.

4.1 Case Study Number One

Results for case study number one are described in Table 2. The type of information returned from the searches is summarized first for the full sample, then separately for completes versus nonworking numbers, for the two search vendors.

Table 2. Validation Coding Results for Case Study Number One

	Intelius			ReversePhoneDetective		
Search returned:	All Cases (200)	CS/IP/NR (87)	NW (75)	All Cases (200)	CS/IP/NR (87)	NW (75)
Any data	79.0%	86.2%	72.0%	100.0%	100.0%	100.0%
Name and mailing address	33.0%	32.2%	36.0%	23.5%	35.6%	8.0%
Mailing address	33.0%	32.2%	36.0%	23.5%	35.6%	8.0%
Phone line type = Cellular	65.5%	74.7%	58.6%	74.0%	81.6%	68.0%
Match of returned data with survey data						
Name and mailing address		2.3%			1.1%	
First and last name		13.8%			12.6%	
Mailing address		5.7%			2.3%	

Most of the cell phone numbers yielded some information when searched, but the most frequently returned piece of information was the phone line type. For searches conducted through Intelius, roughly one-third of the phone numbers yielded a mailing address. The return was lower from ReversePhoneDetective, with one-quarter of the numbers yielding an address.

The return of an address does not guarantee that it is an accurate address. For the 87 cell phone numbers drawn from completed or ineligible/nonresidential interviews, we experienced a very low match rate between the address information returned from search vendors and that on file from the interview. The correct mailing address was returned for fewer than 6 percent of these cases from Intelius, and just over 2 percent from ReversePhoneDetective. Much of the mailing address information returned from the vendors appears to be in error.

Neither vendor was able to correctly identify nonworking cell phone numbers at an acceptable rate. Intelius returned name and mailing address information for roughly one-third of the 75 nonworking numbers. ReversePhoneDetective appeared better at performing this task; however this service still returned name and mailing address for 8 percent of these numbers.

4.2 Case Study Number Two

We were concerned that the age of the sample for case study number one may have affected our results. It was also possible that cell phones in the state of California might differ from the rest of the nation in a way that would affect our match rates for validation coding. For these reasons, we selected cell phone numbers that had reached final dispositions within the past month from a nationwide cell phone sample for case study number two. Results of this case study are described in Table 3.

Table 3. Validation Coding Results for Case Study Number Two

	Intelius			Reverse Phone Detective		
Search Returned	All Cases (174)	CS (100)	NW (74)	All Cases (174)	CS (100)	NW (74)
Any data	68.4%	89.0%	40.5%	100.0%	100.0%	100.0%
Name and mailing address	29.9%	38.0%	18.9%	23.6%	35.0%	8.1%
Mailing address	29.9%	38.0%	18.9%	23.6%	35.0%	8.1%
Phone line type = Cellular	53.4%	72.0%	28.4%	78.2%	76.0%	81.1%
Match of returned data with survey data						
Name & mailing address		0.0%			1.0%	
First & last name		17.0%			20.0%	
Mailing address		0.0%			1.0%	

The results of these searches were very similar to those obtained for case study number one. The rate of return of mailing address information was nearly unchanged. The percentage of cases for which mailing address information returned from the searches matched the information on file for completed interviews was lower when using fresher sample cases.

5. Conclusions

Results indicate that the information obtained from these services is not sufficiently accurate to be used for either sample purging or prenotification purposes, and in fact the potential exists for high levels of misclassification. If either of the search vendors from the experiment were used to purge samples of presumed nonworking numbers prior to data collection, a high percentage of valid residential numbers would never be contacted. Similarly, if the vendors were used to obtain addresses for mailing of prenotification letters, a high percentage of the advance letters would reach households unconnected to sampled cell phone numbers, and many households connected to sampled cell phone numbers would never receive the message.

Reverse cell phone directory search vendors do not appear to add to the telephone survey methodology "toolbox" at this point in time, from either a sample purging or prenotification perspective. Further investigation of these types of services is warranted, however, due to the cost implications of including cell phone number samples in RDD survey designs.

Acknowledgements

The authors would like to thank Mike Brick for requesting that we explore this avenue of research, and Trent Buskirk for inspiring him to ask us.

References

AAPOR Cell Phone Task Force (2008). Guidelines and Considerations for Survey Researchers When Planning and Conducting RDD and Other Telephone Surveys in

the U.S. With Respondents Reached via Cell Phone. Prepared for AAPOR Council by a Task Force operating under the auspices of the AAPOR Standards Committee. American Association for Public Opinion Research, April 2008.

Battaglia, Michael P., Meg Ryan, and Marcie Cynamon (2005). Purging Out-Of-Scope and Cellular Telephone Numbers from RDD Samples. In JSM Proceedings, AAPOR-ASA Section on Survey Research Methods. Alexandria, VA: American Statistical Association. 3794-3801.

Blumberg, Stephen J. and Julian V. Luke (2007). Coverage Bias in Traditional Telephone Surveys of Low-Income and Young Adults, Public Opinion Quarterly 71: 734-749

Brick, J. Michael, Pat D. Brick, Sarah Dipko, Stanley Presser, Clyde Tucker and Yangyang Yuan (2007). Cell Phone Survey Feasibility in the U.S.: Sampling and Calling Cell Numbers Versus Landline Numbers, Public Opinion Quarterly 71: 23-39.

De Leeuw, Edith, Joop Hox, Elly Korendijk, Gerty Lensvelt-Mulders, and Mario Callegaro (2006). The Influence of Advance Letters on Response in Telephone Surveys: A Meta-Analysis. Paper presented at the Second International Conference on Telephone Survey Methodology, Miami, FL.

Keeter, Scott, Michael Dimock, Courtney Kennedy, Jonathan Best, and John Horrigan (2008). Costs and Benefits of Full Dual-Frame Telephone Survey Designs. Paper presented at the Annual Conference of the American Association for Public Opinion Research, New Orleans, LA.