



# List-Assisted RDD Sampling in Korea :Testing the Feasibility of National Surveys

SunWoong Kim, SangKyung Lee  
SungJoon Hong, SoHyung Park

Dongguk University & Hyundai Research Institute

**2009 World Association for Public Opinion Research  
September 11-13, Lausanne, Switzerland**

## Outline

- Previous Studies on Telephone Samples in Korea and U.S.
- Alternatives in Korea
- KOIS Database
- Incidence of Listed Phones in Residential Banks in Korea and U.S.
- Implementing GENESYS-type List Assisted RDD Sampling
- Computer-assisted Telephone Interviewing
- List Assisted RDD Frame Size
- Data Collection Process
- Survey Results
- Survey Cost
- Discussion

## Previous Studies on Telephone Samples in Korea

- Kim (2004)

“The Changes in Fixed Telephone Household Coverage Rates”

- Showed the changes in coverage rates in both fixed telephones and mobile phones
- The level of fixed telephone household noncoverage is much higher than 30% due to about 10% mobile only households.

## Previous Studies on Telephone Samples in Korea (cont.)

- Kim, Hong, and Park (2007)

“Household Noncoverage in Fixed Line Phone Surveys in Korea”

- The telephone directory frames may nationally cover under 65% of households.
- Due to nonresidential or nonworking numbers, the national or local undercoverage will be much higher than the levels presented.
- Directory or directory-based samples may produce considerably biased estimates because of the noncoverage problem.

## Previous Studies on Telephone Samples in Korea (cont.)

- Kang et al. (2008)

“Random Digit Dialing Telephone Survey and Major Findings”

- Used Early RDD sampling method after eliminating business numbers
- Attached four-digit numbers between 0000 and 9999 to area code-prefix combinations
- Used quota sampling rules to choose a respondent in a household

# Telephone Samples in the USA

## ◆ Household Level

### ● Cooper (1964)

Early RDD sampling: adding four-digit suffixes to known prefixes

- 75 to 80 percent are not assigned to a household

## Telephone Samples in the USA (cont.)

- Landon and Banks (1977)

Plus digit sampling (e.g. plus one sampling):

A list-assisted procedure in which a sample is selected from a directory and an integer is added to the last digit of the selected numbers

- It assumes that unlisted numbers are evenly mixed among listed numbers.

# Telephone Samples in the USA (cont.)

- Mitofsky(1970), Waksberg(1978)

Two stage RDD sampling (Mitofsky–Waksberg(M–W) technique):

Step1) One obtains each 100 bank by first drawing a valid combination of area code and prefix and then drawing the first two digits of the suffix.

Step2) 1 of the 100 numbers in each 100 bank is selected by drawing two digits for the last half of the suffix. The selected phone number in each 100 bank is dialed. If it is a residential number, the 100 bank is retained. If not, the 100 bank is discarded.



## Telephone Samples in the USA (cont.)

Step3) Additional phone numbers are drawn and dialed from each retained 100 bank until residential numbers are obtained in each 100 bank.

- The method is cumbersome to administer since it sometimes takes a fairly large number of callbacks to determine whether or not a telephone number is residential.

# Telephone Samples in the USA (cont.)

- Potthoff (1987)

Generalization of the M-W technique

: Choosing phone numbers per PSU in determining whether to retain the cluster

- Groves and Lepkowski (1986)

Dual-frame sampling

: Selecting a portion of a telephone sample from the frame of listed telephone numbers and the remainder from an RDD frame

## Telephone Samples in the USA (cont.)

- Casady and Lepkowski (1993)

Two-stratum list-assisted RDD design

: high-density stratum including 100 banks with one or more listed phone numbers and low-density stratum including all remaining phone numbers

– Superior to both original RDD and M-W technique

## Telephone Samples in the USA (cont.)

- Brick, Waksberg and Kulp (1995)

Described the coverage bias for a particular method of list-assisted RDD sampling developed by GENESYS, a commercial sampling vendor

Presented that only about 3 to 4 percent of all residential households are excluded

Became the standard for most list-assisted RDD samples

Handled by several commercial sampling vendors possessing the capacity to continuously update national list frames

# Telephone Samples in the USA (cont.)

## ◆ Person Level

How to select one eligible person (adult) in the household?

Simple quota sampling? Or random sampling?

Random sampling in most surveys is implemented by computer program-assisted method, which makes random selection by computer at the time of listing eligible persons in the household

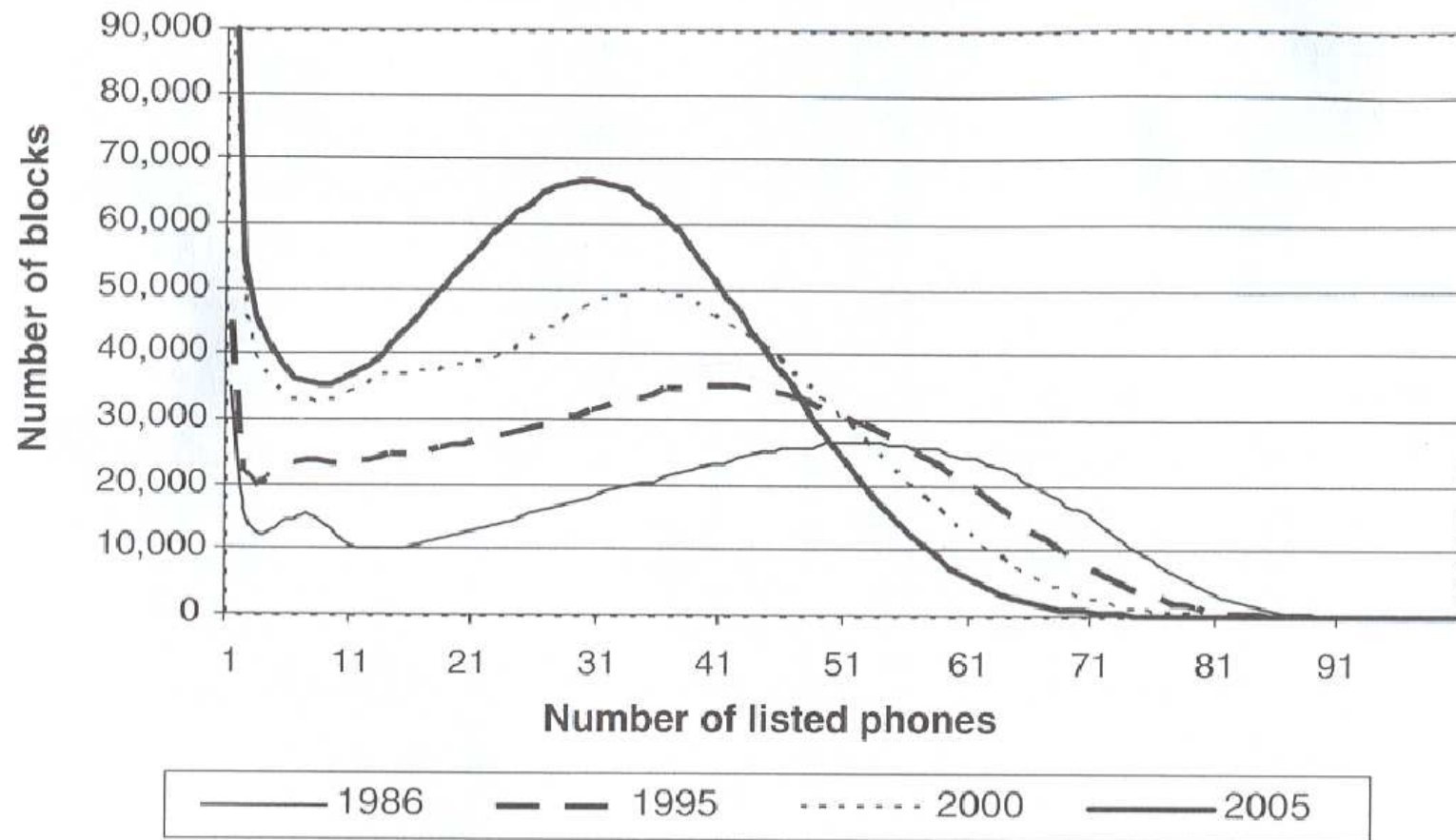
## Alternatives in Korea

- a. Early RDD sampling
- b. Plus digit sampling
- c. M-W technique
- d. Modified M-W technique
- e. Lepkowski-Groves technique
- f. Modified Lepkowski-Groves technique
- g. Casady-Lepkowski technique
- h. Modified Casady-Lepkowski technique
- i. GENESYS Sampling Systems
- j. GENESYS-type List-Assisted RDD Sampling
- k. New Techniques

# Korea Info Service (KOIS) Database, 2008

	Residential	Business
Listed Numbers	8,636,741 (53.6%)	4,438,243 (51.5%)
Unlisted Numbers	7,483,316 (46.4%)	4,172,069 (48.5%)
Total	16,120,057	8,610,312

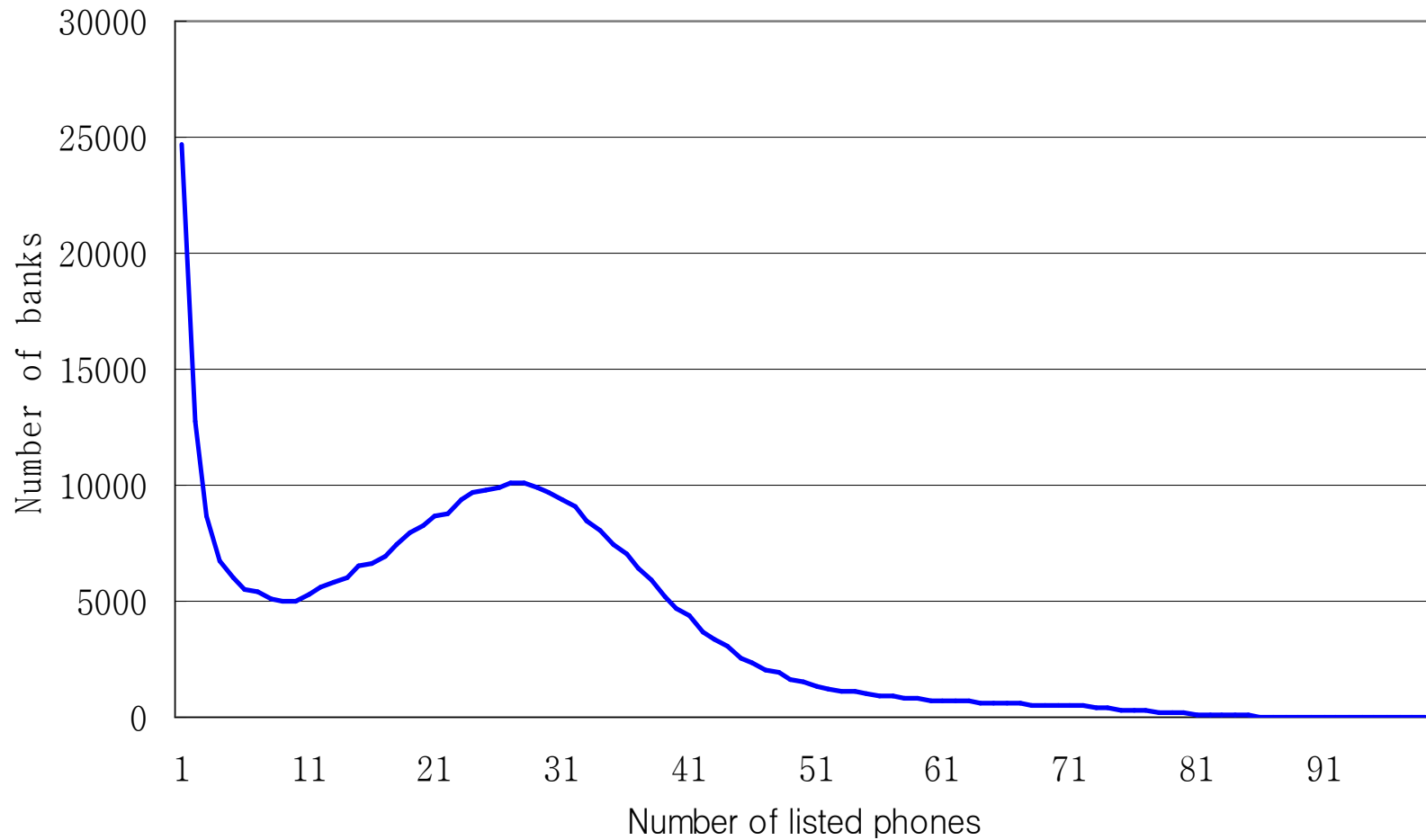
# Incidence of listed phones in working 100-banks (USA)



Source : Lepkowski et al. (2008) "Advances in Telephone Survey Methodology"

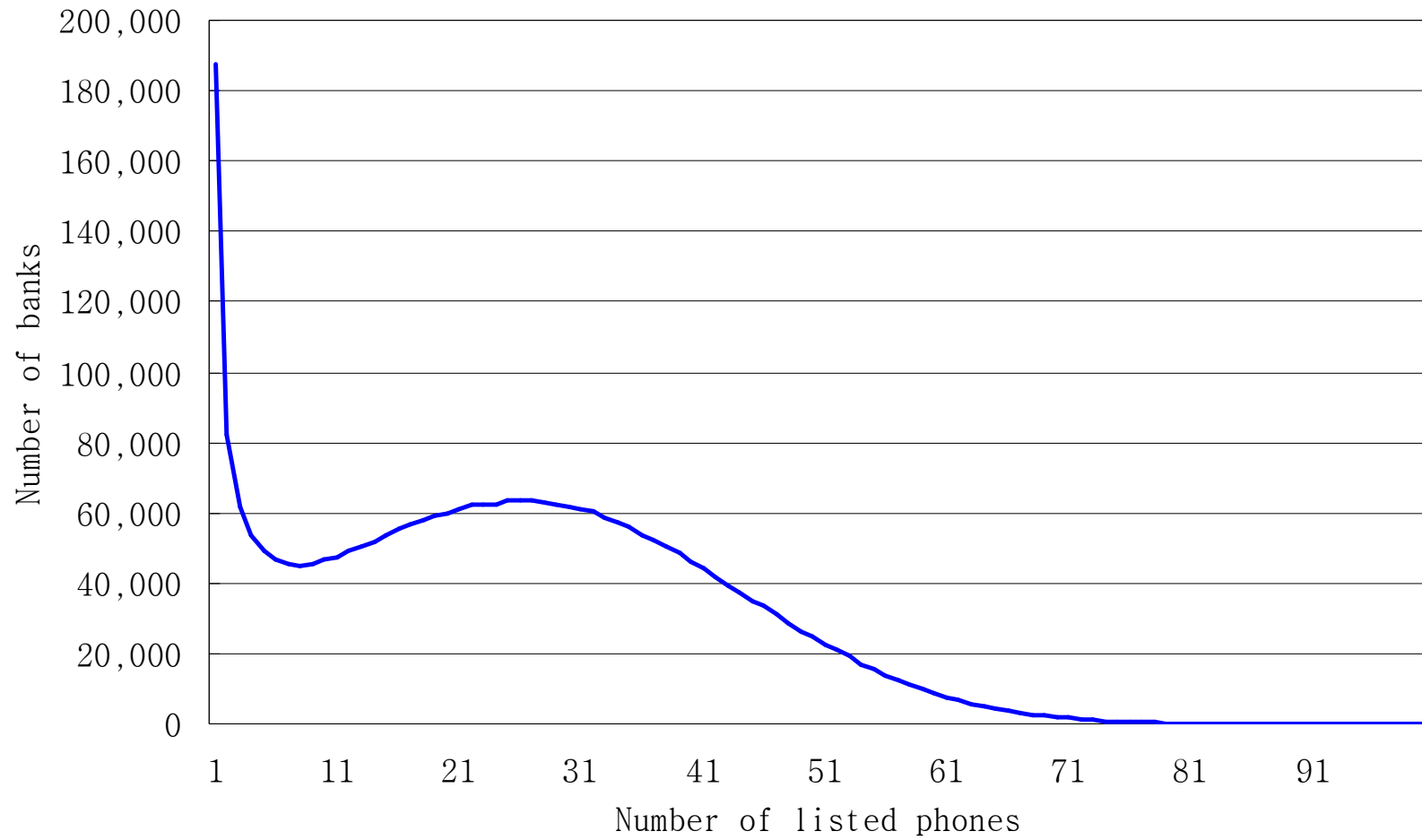


# Incidence of listed phones in working 100-banks (Korea, 2008)



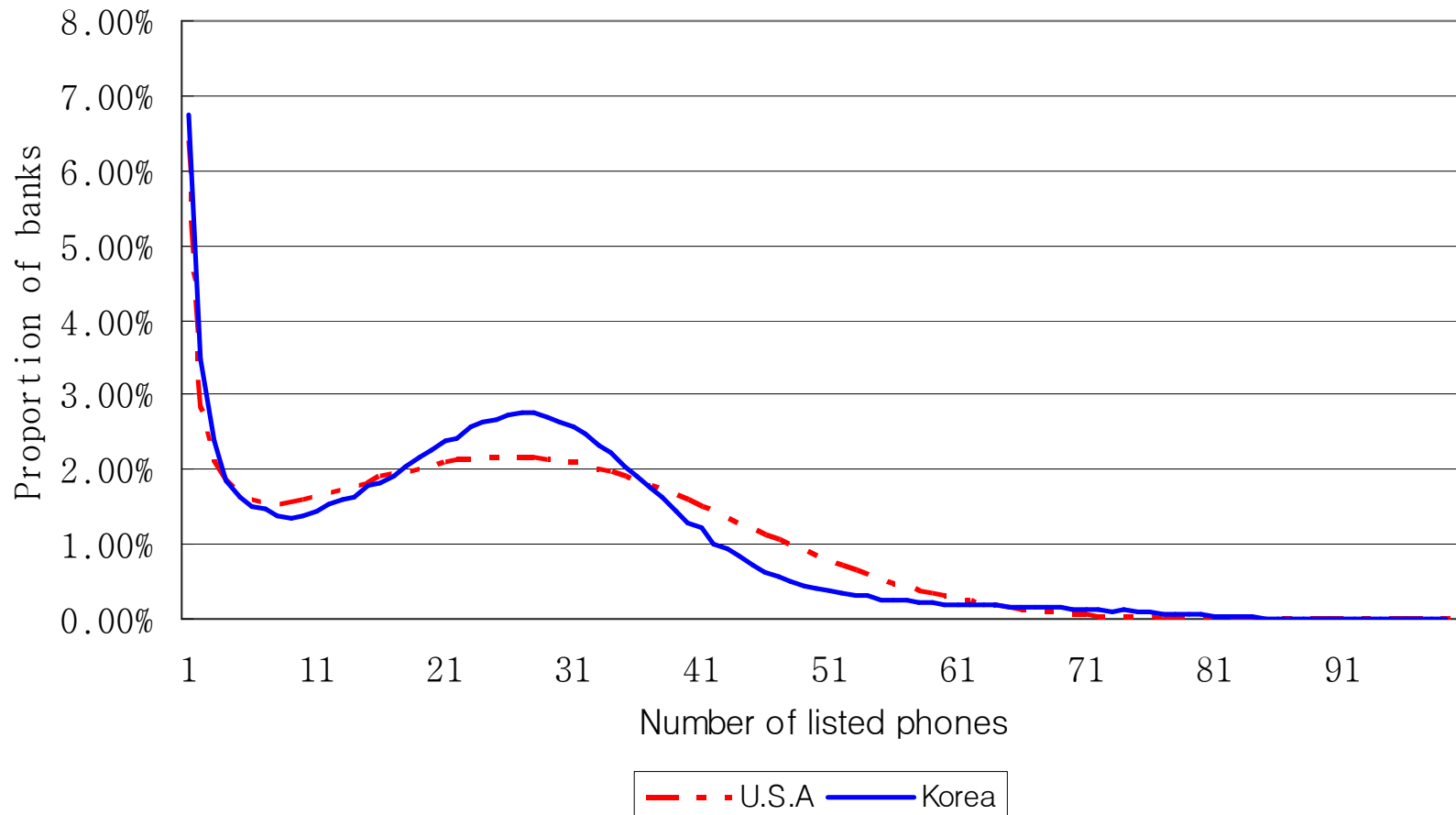
Source : KOIS (2008)

# Incidence of listed phones in working 100 banks (USA, 2008)



Source : Marketing Systems Group (2008)

# The rate of incidence of listed phones in working banks in Korea and USA (2008)



## Implementing List-Assisted RDD Sampling in Korea

1. All residential exchanges and working 100-banks are determined by using the KOIS telephone directories.
2. Establish a list-assisted RDD frame consisting of all listed numbers that are in 100-banks with one or more listed phone number.
3. Business numbers are eliminated from the RDD frame.
4. All exchanges and working 100-banks in the same area code are arrayed in ascending order.
5. A systematic sample with fractional intervals is selected from the phone numbers in the RDD frame.
6. Epsem (Equal Probability Selection Method) sample is generated in this design.

# The National Sustainability Study (NSS)

The NSS, which was conducted by the Survey Research Laboratory (SSL) in Dongguk University and Hyundai Research Institute, was designed to test the feasibility of list-assisted RDD.

(Sample Design)

- 1) The country was stratified into 7 metropolises and 9 provinces.
- 2) The households in each stratum were selected by using list-assisted RDD sampling. A respondent within the selected household was randomly selected.
- 3) The sample size of 504 was decided to estimate a proportion with a bound on the error of estimation of magnitude 3.7 percent point.
- 4) Neyman allocation was used to decide the sample size of each stratum.
- 5) Epsem sample is selected in each stratum.

## Computer-assisted Telephone Interviewing

- 1) The CATI system developed by Hyundai Research Institute was used.
- 2) The system involves a variety of tools for quality controls.
- 3) The system was redesigned for a new quality control.
- 4) The interviews by skillful interviewers were conducted for one week at the end of October, 2008.

## List-Assisted RDD Frame Size

No. of RDD 100-banks	Initial RDD Frame Size	Screened commercial numbers	Final RDD Frame Size
378,947	37,894,700	4,038,285	33,856,415

## Distribution of Dialed Numbers by Category

Category	Frequency	%	Cumulative %
Non-working (missing)	5,810	48.7	48.7
Unconnected (no ringing, silence)	83	0.7	49.4
Fax	272	2.3	51.7
Undetermined if residential	8	0.1	51.8
Business	788	6.6	58.4
Residential	4,967	41.6	100.0
Total	11,928	100.0	

Note. 'Residential' numbers may not be households.



## Distribution of Dialed Numbers by "Listed"

Category	Listed	Frequency		%	
Non-working	Y	907	5,810	7.6	48.7
	N	4,903		41.1	
Other	Y	16	83	0.1	0.7
	N	67		0.6	
Fax	Y	70	272	0.6	2.3
	N	202		1.7	
Undetermined if residential	Y	2	8	0.0	0.1
	N	6		0.1	
Business	Y	187	788	1.6	6.6
	N	601		5.0	
Residential	Y	1,677	4,967	14.1	41.6
	N	3,290		27.6	
Total	Y	2,859	11,928	24.0	100.0
	N	9,069		76.0	

## Distribution of Residential Numbers by Category

Category		Frequency	%
No answer*		1,571	31.6
Busy		140	2.8
Completed Interviews		504	10.1
Refusal - Receiver	Undetermined if it is a household **	1,826	36.8
	Household	329	6.6
	Both household and business	4	0.1
	Subtotal	2,159	43.5
Refusal - Respondent		164	3.3
No contact - Respondent		398	8.0
Ineligible - no adults		2	0.0
Broken Appointments		17	0.3
Determined if it is a household		12	0.2
Total		4,967	100.0

Note. \*, \*\*: Strong possibilities of being households

# Distribution of Residential Numbers by Listed

Category		Listed	Frequency	%	
No answer	Y	456	1,571	9.2	31.6
	N	1,115		22.4	
Busy	Y	28	140	0.6	2.8
	N	112		2.3	
Completed Interviews		Y	504	3.7	10.1
		N		319	
Refusal - receiver	Undetermined if it is a household	Y	1,826	13.3	36.8
		N		1,164	
	household	Y	329	2.4	6.6
		N		211	
	Both household and business	Y	4	0.1	0.1
		N		0	
	Subtotal	Y	2,159	15.8	43.5
		N		1,375	
Refusal – respondent		Y	164	1.3	3.3
		N		98	
No contact - Respondent		Y	398	3.0	8.0
		N		250	
Ineligible – no adults		Y	2	0.0	0.0
		N		2	
Broken Appointments		Y	17	0.1	0.3
		N		11	
Determined if it is a household		Y	12	0.1	0.2
		N		8	
Total		Y	4,967	33.8	100.0
		N		3,290	

# Household Hit-Rate

Category	Frequency	Hit-Rate	
		Including non-working, fax, and businesses (11,928)	Excluding non-working, fax, and businesses (4,967)
Household	1,417	11.9	28.5
Household and Business	13	0.1	0.3
Total	1,430	12.0	28.8

Note. 'Total' = 'Completed Interview's + 'Refusal-receiver (household, both household and businesses)' + 'Refusal-respondent' + 'No contact-respondent' + 'Ineligible-no adults' + 'Broken Appointments' + 'Determined if it is a household'

Note. Comparatively low hit rate due to 'No answer' and 'refusal-receiver'

## Comparison of Distribution of Household Sizes

Persons	List Assisted RDD		2005 Census
	Frequency	%	%
1	107	8.7	19.9
2	252	20.4	22.2
3	265	21.4	20.9
4	423	34.2	27.0
5	141	11.4	7.7
6	37	3.0	2.3
7	7	0.6	
8	3	0.2	
10	1	0.1	
Total	1,236	100.0	100.0

## Comparison of Sex and Age Distribution

	Male		Female		Total	
	Frequency	%	Frequency	%	Frequency	%
List Assisted RDD	186	36.9	318	63.1	504	100.0
2005 Census	17,148,184	49.0	17,816,855	51.0	34,965,039	100.0

	Age						
	20-29	30-39	40-49	50-59	60-69	70-79	80 and over
List Assisted RDD	9.9	21.8	28.6	15.9	11.9	8.3	3.6
2005 Census	21.0	23.5	22.9	14.7	10.2	5.8	1.9

## Number of callbacks for listing eligible persons

Number	Frequency	%	Cumulative %
0	496	41.9	41.9
1	256	21.6	63.5
2	176	14.9	78.3
3	109	9.2	87.5
4	69	5.8	93.3
5	39	3.3	96.6
6	23	1.9	98.6
7	9	0.8	99.3
8	4	0.3	99.7
9	3	0.3	99.9
14	1	0.1	100.0
Total	1,185	100.0	

## Number of callbacks for contact with sampled person

Number	Frequency	%	Cumulative %
0	273	40.8	40.8
1	152	22.7	63.5
2	87	13.0	76.5
3	65	9.7	86.2
4	41	6.1	92.4
5	24	3.6	96.0
6	14	2.1	98.1
7	8	1.2	99.3
8	1	0.1	99.4
9	3	0.4	99.9
14	1	0.1	100.0
Total	669	100.0	



## Number of callbacks from listing to contact of respondent

Number	Frequency	%	Cumulative %
0	638	95.4	95.4
1	12	1.8	97.2
2	5	0.7	97.9
3	3	0.4	98.4
4	2	0.3	98.7
5	4	0.6	99.3
6	2	0.3	99.6
7	2	0.3	99.9
8	1	0.1	100.0
Total	669	100.0	

## Complete Rate (listing to completing interview)

	listing	contact	complete
Frequency	1,185	669	504
%	100	56.4	42.5

## Comparison between listed and unlisted numbers : Sex

Sex	Listed Numbers	Unlisted Numbers	95% Interval Estimate of True Difference	
Male	39.8 %	36.9 %	-2.9 %	8.8 %
Female	60.2	63.1	-8.8	2.9

## Comparison between listed and unlisted numbers : Age

Age	Listed Numbers	Unlisted Numbers	95% Interval Estimate of True Difference	
20-29	11.7 %	12.0 %	-4.2 %	3.6 %
30-39	17.9	21.4	-8.3	1.2
40-49	27.5	28.1	-6.1	4.8
50-59	15.1	18.6	-7.9	1.0
60 and over	27.8	19.9	2.7	13.1

## Comparison between listed and unlisted numbers : Education

Education	Listed Numbers	Unlisted Numbers	95% Interval Estimate of True Difference	
None	5.0	2.9	-0.4	4.5
Elementary school graduate	14.3	8.7	1.8	9.5
Middle school graduate	12.1	8.5	-0.1	7.3
High school graduate	32.6	32.9	-5.8	5.3
Two-year-college graduate	9.3	12.1	-6.4	0.8
Four-year-college graduate or other	26.7	34.9	-13.6	-2.7

## Discussion

- Completed list assisted RDD design and randomized respondent selection
- Required stricter callbacks rules for the random selection of respondents
- Needed a study on bias due to the truncation of low-density stratum
- A study on differences in demographic characteristics between listed and unlisted numbers is useful
- Noncoverage problem due to at least 10% mobile only households should be solved
- The strategies for reducing sources of various errors must be developed
- Precision versus survey cost

## Acknowledgments

The authors wish to thank James M. Lepkowski, the chair of the Survey Methodology Program at the University of Michigan and Dale W. Kulp, the president of GENESYS Sampling Systems, for giving professional advice and encouragement to do this research project.

**Thank you!**