Assuring Quality in Dual Frame RDD National or Sub-national Surveys using Cell Phone Numbers without Area Codes in South Korea

Sun-Woong Kim, Young-Je Woo, Nam-Hoon Kim

Survey & Health Policy Research Center
Dongguk University

2017 Joint Statistical Meetings, Baltimore, Maryland July 29 – August 3

dongguk UNIVERSITY SHPRC

OUTLINE

1. Purpose of Study
2. Dual Frame RDD Sample Design
3. National Tobacco Survey using Dual Frame RDD
4. Multi-stage Weighting Process
5. Study Results
6. Concluding Remarks

1. Purpose of Study

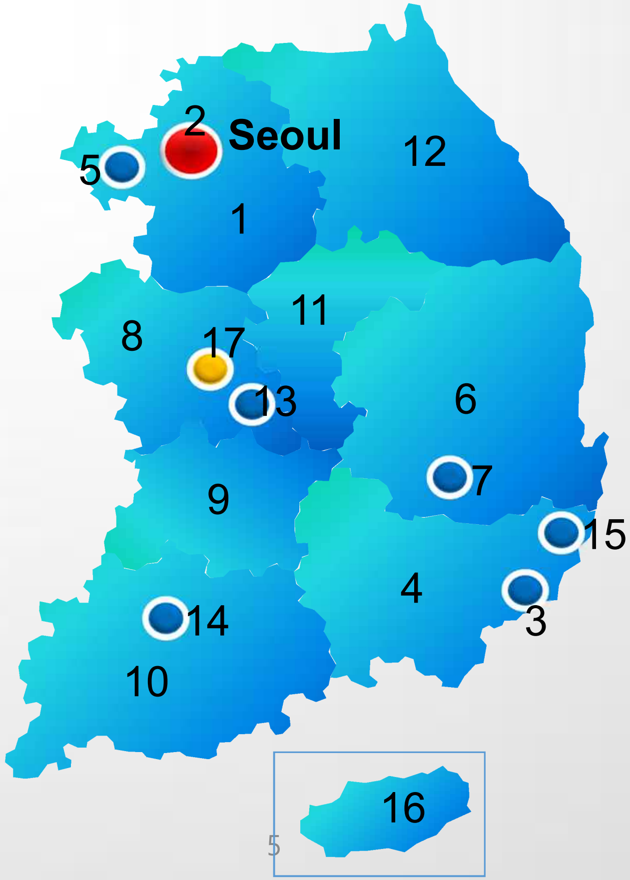
Country-specific characteristics of the cell phone numbering system have led to a variety of drawbacks in conducting dual-frame RDD telephone surveys (Engel *et al*., 2015).

- In the U.S. the area code of cell phone numbers has become increasingly unreliable for surveys at the state or local level due to the regional portability of the numbers (Christian, Dimock, & Keeter, 2009).

- In South Korea cell phone numbers raise a serious concern about RDD sampling and estimation at both national level and sub-national levels.

- It is because different from landline numbers, cell phone numbers do not involve 17 area codes useful to select random or stratified samples based on geographies. Instead of area codes, they have only one mobile prefix “010”.

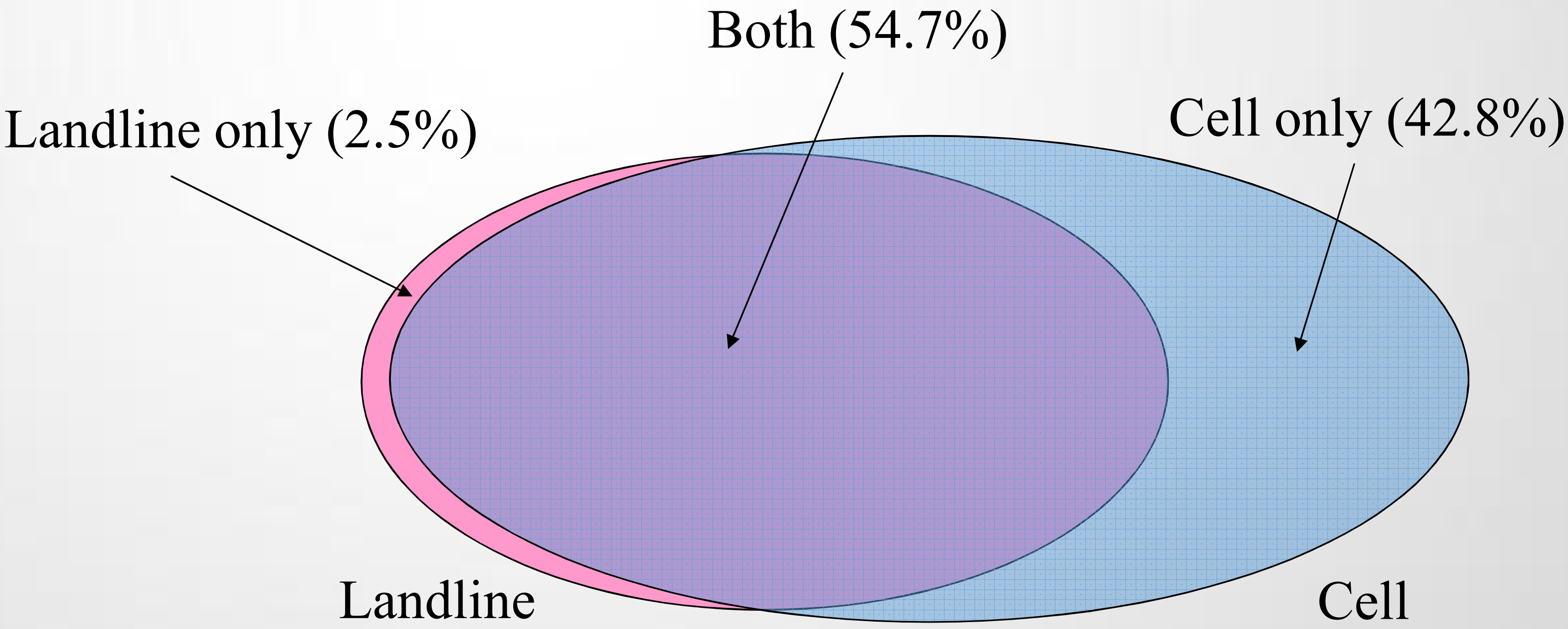
010 XXXX YYYY



- We present dual frame RDD design to overcome this problem due to cell phone numbers without area codes.

2. Dual Frame RDD Sample Design

2016 Overlapping Dual Frame Coverage (individuals)



dongguk UNIVERSITY SHPRC

2. Dual Frame RDD Sample Design (Cont.)

Using to avoid coverage bias due to cell phone-only populations

Landline RDD Frame (size 33,900,000)
: List-Assisted RDD Sampling

Based on 1+ listed 100-banks
(Kim *et al.* , 2012, IJPOR)

Cell RDD Frame (size 74,240,000)
: Single Stage Epsem RDD Sampling*

Based on 10,000-banks

*Each number in cell phone RDD frame
is strictly selected with equal probability.

3. National Tobacco Survey using Dual Frame RDD

Purpose	Provide estimates of rates of tobacco use at the provincial and national level
Sampling & sample size	Cell phones: Random sampling & 1,801 (60%) Landline phones: Stratified random sampling & 1,202 (40%) Dual: Mixed sampling &: 3,003 (100%)
Sampling individuals	Randomly selecting one among individuals aged 19 or over using each RDD sample number regardless of landline or cell phone numbers
Data Collection	May – July, 2016 Up-to-date CATI system At least 10-12 call attempts to noncontact numbers (including weekends)

4. Mutli-stage Weighting Process

RDD Initial Weights

Landline (with strata)

$$W_{initial,h} = \frac{\text{total number of landline numbers in List-Assisted RDD frame}}{\text{total number of landline numbers selected randomly}}$$

Cell (without strata)

$$W_{initial} = \frac{\text{total number of cellular numbers in RDD frame}}{\text{total number of cellular numbers selected randomly}}$$

Removal of out-of-scope numbers (including those not in service)

Landline (with strata)

$$A_{1,h} = \begin{cases} 0 & \text{if out of scope} \\ P_{in\text{-}scope,h} & \text{if unresovoed} \\ 1 & \text{otherwise} \end{cases}$$

Cell (without strata)

$$A_{1} = \begin{cases} 0 & \text{if out of scope} \\ P_{in\text{-}scope} & \text{if unresovoed} \\ 1 & \text{otherwise} \end{cases}$$

Informant non-response adjustment

Landline (with strata)

$$A_{2,h} = \frac{\text{sum of weights for all sampled landline numbers}}{\text{sum of weights for informant landline numbers}}$$

Cell (without strata)

$$A_{2} = \frac{\text{sum of weights for all sampled cell numbers}}{\text{sum of weights for informant cell numbers}}$$

Lepkowski & Kim (2005) originally illustrated and Park & Kim (2011) developed the theory.

Person-level weight

Landline only or cell only person

$$A_3 = \frac{1}{\sum_{j=1}^{\alpha_{ik}} 1/\beta_{ijk}}$$ landline phone: $i=1$, cell phone: $i=2$

α_{ik} : Number of phone i's to be reached to respondent k

β_{ijk} : Number of adults who use jth phone i *with* respondent k

Landline and cell person

$$A_3 = \frac{1}{\sum_{j=1}^{\alpha_{1k}} 1/\beta_{1jk} + \sum_{j=1}^{\alpha_{2k}} 1/\beta_{2jk} - \sum_{j=1}^{\alpha_{1k}} 1/\beta_{1jk} \sum_{j=1}^{\alpha_{2k}} 1/\beta_{2jk}}$$

The values of α_{ik} and β_{ijk} are obtained from informants and respondents using *clearly formulated specific questions* about eligible persons and devices.

Person-level non-response adjustment

$$A_4 = \frac{\text{sum of weights for all selected persons}}{\text{sum of weights for respondents}}$$

Post-stratification

Landline

$$A_{5,h} = \frac{\text{population estimate in a post-stratum (age and gender)}}{\text{sum of weights of respondents in a post-stratum (age and gender)}}$$

Cell

$$A_{5} = \frac{\text{population estimate in a post-stratum (self-report location, age and gender)}}{\text{sum of weights of respondents in a post-stratum (self-report location,age and gender)}}$$

Final weight

Landline (with strata)

$$W_{final} = W_{initial,h} \cdot A_{1,h} \cdot A_{2,h} \cdot A_{3} \cdot A_{4} \cdot A_{5,h}$$

Cell (without strata)

$$W_{final} = W_{initial} \cdot A_{1} \cdot A_{2} \cdot A_{3} \cdot A_{4} \cdot A_{5}$$

5. Study Results

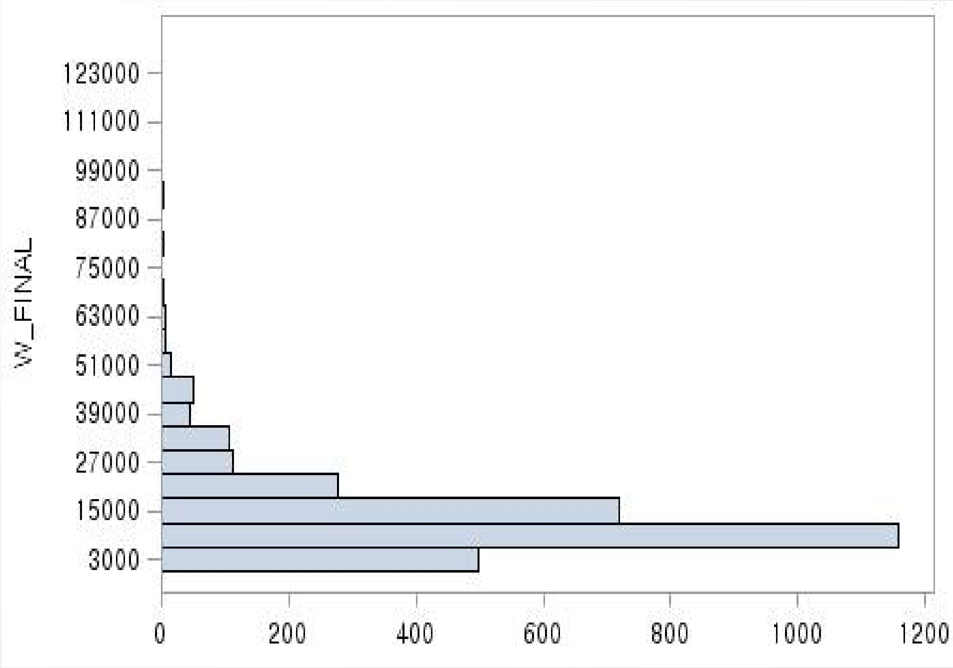
- **Response Rates**

Frame	RR1	RR5
Landline	10.2%	44.2%
Cell	18.4%	36.9%
Dual	13.9%	39.5%

▪ Unequal Weighting Effect

Kish (1965) and Biemer & Christ (2008)

$$1 + CV^2 = 1.6$$



- **Area Distribution of Respondents by Frame (unweighted %)**

Area	Cell %	Landline %	Dual %	Population %
1	21.9	24.1	22.8	23.7
2	28.4	17.0	23.8	19.8
3	5.4	6.6	5.9	7.1
4	4.7	7.4	5.8	6.4
5	4.9	3.9	4.5	5.7
6	4.1	6.7	5.2	5.4
7	4.5	3.8	4.2	4.9
8	3.4	5.2	4.1	4.1
9	3.3	5.5	4.2	3.6
10	3.3	3.9	3.6	3.6
11	2.6	3.4	2.9	3.1
12	2.7	3.8	3.2	3.1
13	3.9	2.9	3.5	3.0
14	2.6	2.3	2.5	2.9
15	2.1	1.8	2.0	2.2
16	1.6	1.3	1.5	1.1
17	0.6	0.2	0.4	0.4
Total	100.0	100.0	100.0	100.0

▪ Gender Distribution of Respondents by Frame

Frame	Male/Female %	Population %
Cell	60.6 / 39.4	
Landline	37.0 / 63.0	49.5 / 50.5
Dual	51.2 / 48.8	

- **Age Group Distribution of Respondents by Frame**

Age group	Cell %	Landline %	Dual %	Population
19 - 29	30.9	12.9	23.7	17.6
30 - 39	21.1	7.1	15.6	18.3
40 - 49	17.5	19.5	18.2	21.1
50 - 59	16.4	20.4	18.0	19.9
60 - 69	8.7	18.7	12.7	12.1
70 or over	5.4	21.4	11.8	11.0

- **Gender Distribution of Respondents by Frame in Seoul**

Frame	Male/Female %	Population %
Cell	54.3 / 45.7	
Landline	31.4 / 68.6	48.6 / 51.4
Dual	47.8 / 52.2	

- Age Group Distribution of Respondents by **Frame** in **Seoul**

Age group	Cell %	Landline %	Dual %	Population
19 - 29	32.8	7.8	25.7	19.1
30 - 39	21.9	6.4	17.5	19.9
40 - 49	17.6	25.0	19.7	20.0
50 - 59	15.8	20.1	17.0	19.0
60 - 69	6.6	16.7	9.5	12.4
70 or over	5.3	24.0	10.6	9.6

- **Smoking Rates and Standard Errors by Area (weighted %)**

Area	Num. of Respondents	Smoking Rates %	Standard Error %
1	684	18.1	1.7
2	716	19.2	1.8
3	177	18.9	3.7
4	173	19.0	3.7
5	135	19.4	5.5
6	155	26.8	4.4
7	127	19.9	4.7
8	124	19.0	4.5
9	125	23.2	4.5
10	107	17.8	4.7
11	87	17.6	4.6
12	95	13.9	4.6
13	106	17.2	4.1
14	75	11.4	4.5
15	60	19.7	6.4
16	44	12.4	4.6
17	13	20.9	13.7
Nation	**3,003**	**18.9**	**0.9**

6. Concluding Remarks

- Although cell phone numbers do not involve area codes, single stage epsem sampling in cell RDD frame is very useful to select representative samples in national or sub-national surveys.

- The quality of information on eligible persons and devices used in calculating “person-level weight” especially depends on the responses from informants and respondents. Thus, it is important that the interviewers should be trained to conduct excellent interviews.

- More efficient weighting procedures for reducing the weighting effect will be examined in the next study.

THANK YOU

Contact at: sunwk@dongguk.edu