

# Young Respondents Wanted: Report on Parallel Surveys Using CATI and Cell Phone Application

---

2018 WAPOR-ASIA Conference

Pei-shan Liao  
RCHSS, Academia Sinica

Ya-hui Yang and Chin-yen Hsiao  
School of Business Administration, Fu Jen Catholic  
University

# Outline

- Recent issues in telephone surveys (CATI)
  - Low response rate and coverage issues
  - Disproportional share of young respondents
- Purpose of this study
  - Increasing young respondents in CATI
  - Comparability of survey data from a cell-only sample
- Research design
  - 2016 National Hakka Population Survey
  - Adjustment of selection probability in CATI
  - A parallel survey on a cell-only sample
- Concluding remarks

# RECENT ISSUES IN TELEPHONE SURVEYS (CATI)

---

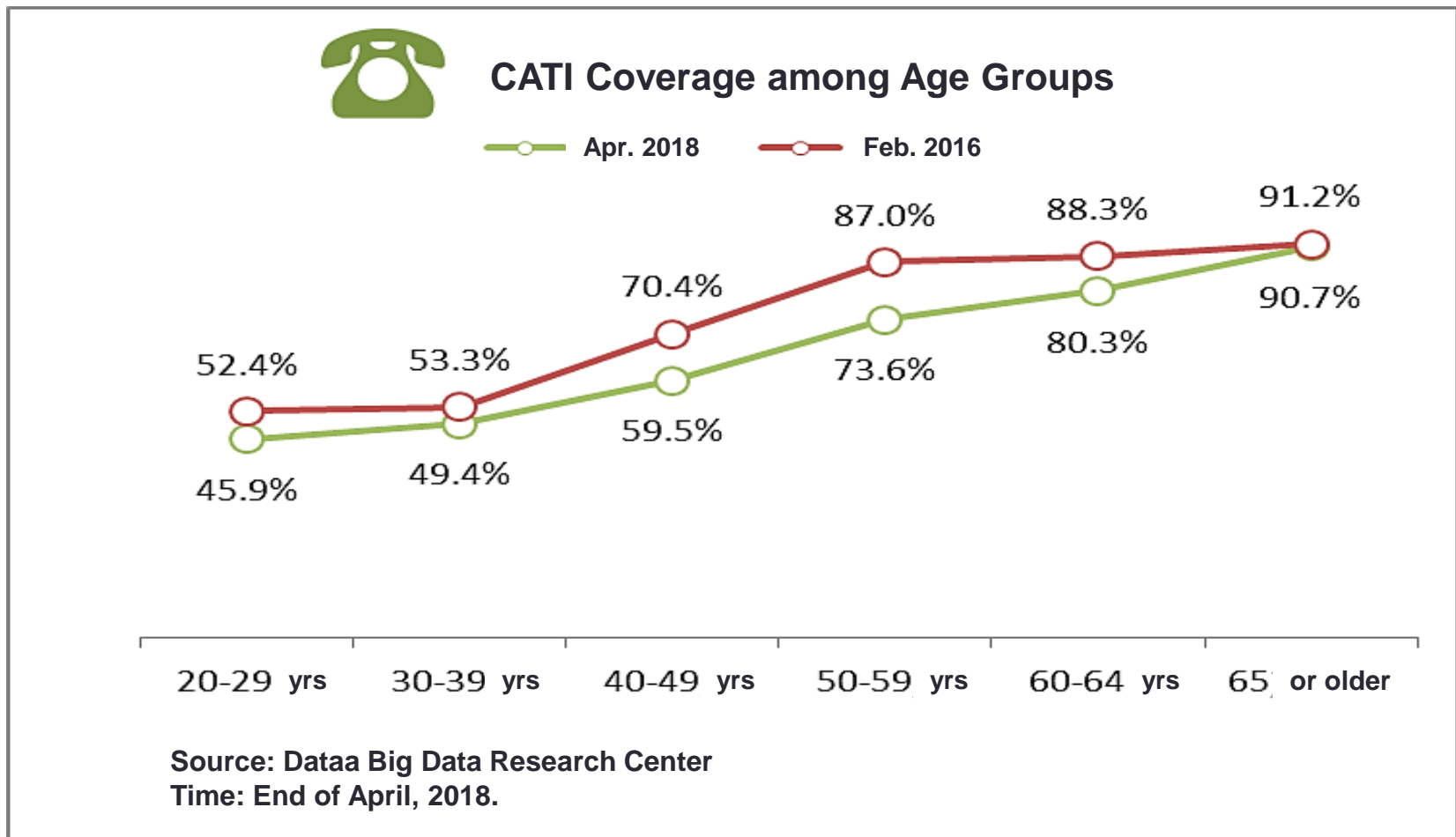
# Low response rate and coverage issues

- Increasing refusals
  - Fraud calls, being reluctant to take surveys, and awareness of personal information are the main reasons.
- Change of life style
  - People with shift work and late night activities are difficult to be reached during the regular survey hours.
- Less use of household landlines
  - Younger generation and those who are heavy users of cell-phone do not answer household landlines.
- Increase of cell-only population
  - Cell-only population are mostly younger generation and those who simply cannot be reached by calling landlines.

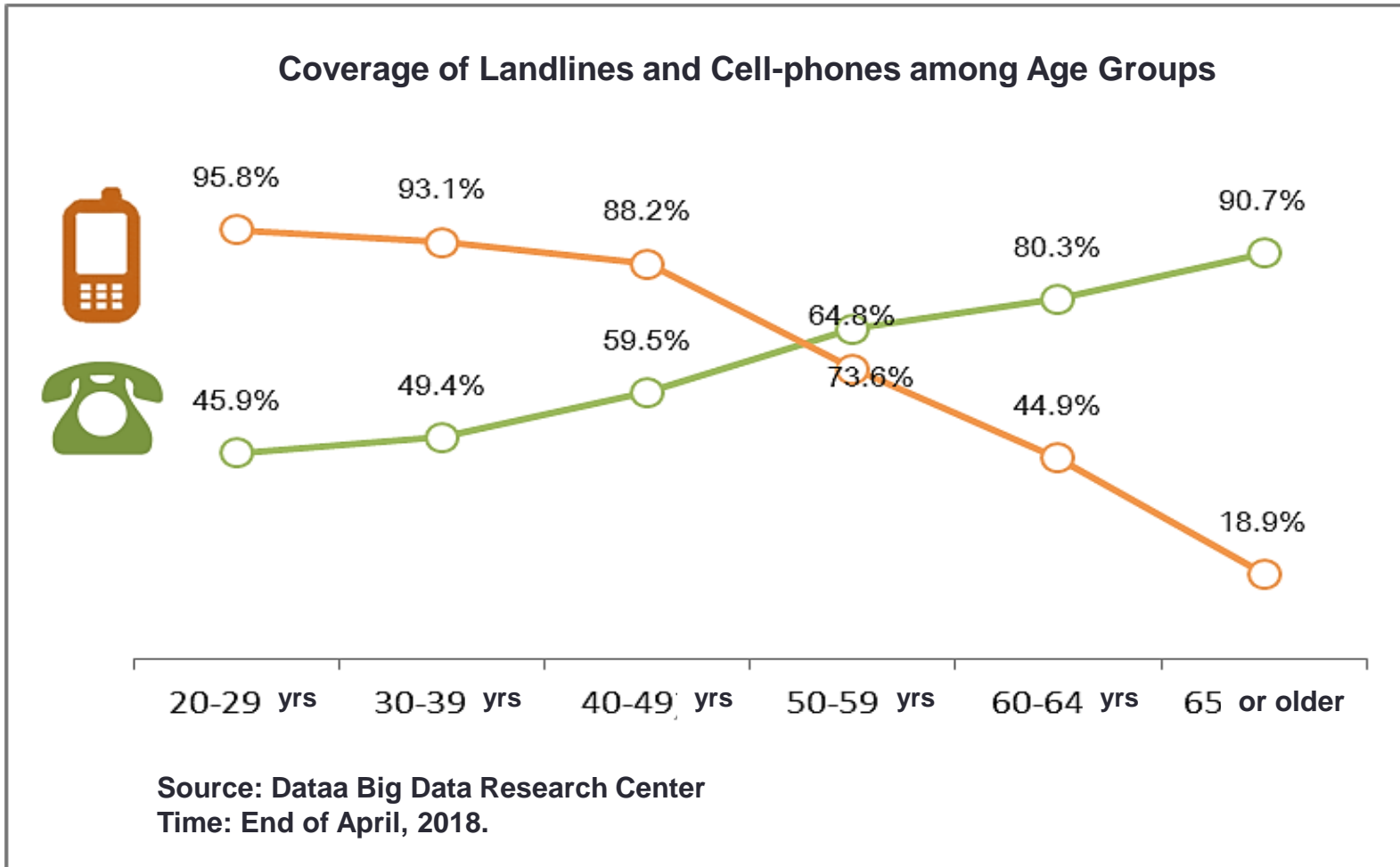
# Cell-only Population in Taiwan

- Development of technology and change of life style contribute to the increase of cell-only population.
- According to Chu and Huang (2015), traditional landline can only reach 73% of the population. In other words, 27% of them are not covered in CATI surveys using household landlines.
- Based on the 2016 data released by Dataa, 29% of the people aged 20 or older cannot be reached by household landlines. CATI coverage rate for those aged 50 or older is more than 80% but less than 55% for those who are 20 to 39 years old.

# Low Coverage for People 20-39 Years Old



# Coverage: Landlines vs. Cell Phones



# Disproportional share of young respondents

Results of the 2014 National Hakka Survey as An Example.

Age	Population		Sample distribution before weighting		Sample distribution after weighting	
	No.	%	n	%	n	%
Total	23,373,517	100.00	78,174	100.00	78,174	100.00
9 years or under	2,026,924	8.7	7,956	10.2	6,779	8.7
10-19 years old	2,876,629	12.3	9,438	12.1	9,621	12.3
20-29 years old	3,252,033	13.9	5,889	7.5	10,971	13.9
30-39 years old	3,939,272	16.8	6,936	8.9	13,178	16.8
40-49 years old	3,671,985	15.7	10,421	13.3	12,281	15.7
50-59 years old	3,539,321	15.1	16,050	20.5	11,837	15.1
60 years or older	4,068,353	17.4	21,484	27.5	13,607	17.4

Compared to the population in 2014:

**-6.4 %**

**-7.9%**





# Purpose

- Increasing young respondents in CATI
  - In particular those who are 20 to 39 years old.
  - By means of adjusting the selection probability of the young respondents during the within-household sampling procedure.
- Comparability of survey data from a cell-only sample
  - By conducting a parallel survey on a cell-only sample using cell-phone application to better understand the dis/similarity of sample composition and survey results.

# RESEARCH DESIGN

---

# The 2016 National Hakka Population Survey

- The main purpose of this regularly conducted survey is to obtain estimates and distributions of Hakka population for policy applications.
- The 2016 survey used computer-assisted telephone interviews (CATI) to collect data from the general public, and then to identify Hakka among the respondents.
- A four-stage stratified systematic sampling, with equal allocation within each PSU (borough/township), is adopted to obtain a nationally representative sample.

# Adjustment of Selection Probability

- A within-household sampling approach was employed to select one individual in each household without age limit. The procedure is adjusted to increase the probability of being selected among the younger respondents.
  - For example, if 4 people are eligible in the household, one individual is randomly selected from the following.
    - (1) The youngest; (2) **The one who is 20 to 39 years old**; (3) The second oldest; (4) The oldest; and (5) The second youngest.
- A total of 65,732 complete cases are interviewed for this study, with a response rate (RR5) of 18.69%.
- Post-stratification weighting is applied using gender, age, and residential area at the county level as weighting variables.

# Results of Within-household Selection

No. of people in household	No. of household (A)	% of population (B)	% of sample (C)	Prob. of adding 20-39 yrs in each HH
1	2,747,386	32.1	<b>19.0</b>	
2	1,688,064	19.7	18.6	1/3
3	1,555,706	18.2	18.1	1/4
4	1,326,372	15.5	<b>20.7</b>	1/5
5	641,974	7.5	<b>11.5</b>	1/6
6	310,301	3.6	<b>6.3</b>	1/7
7	133,452	1.6	2.7	1/8
8	67,096	0.8	1.4	1/9
9	27,543	0.3	0.6	1/10
10	32,294	0.4	0.7	1/11
11 or more	30,647	0.4	0.6	1/12
Total	8,560,833	100.0	100.0	

- Among successful interviews, fewer single households while more households with 4-6 people than those in the population.
- E is the increased % of respondents aged 20 to 39:
  - $E = \sum_1^{11} B_i \times D_i - M = 15.4\%$
  - Where M=a selected household without individual aged 20-39 /total number of household=0.9%

# Age Distribution after the Adjustment

Age	Population		Sample distribution before weighting		Sample distribution after weighing	
	No.	%	n	%	n	%
Total	23,492,074	100.00	65,732	100.00	65,732	100.00
9 years or under	2,029,973	8.64	3,345	5.09	5,674	8.63
10-19 years old	2,657,642	11.31	5,316	8.09	7,442	11.32
20-29 years old	3,190,390	13.58	5,468	8.32	8,927	13.58
30-39 years old	3,926,998	16.72	7,372	11.22	10,988	16.72
40-49 years old	3,620,156	15.41	9,132	13.89	10,129	15.41
50-59 years old	3,608,663	15.36	13,185	20.06	10,097	15.36
60-69 years old	2,496,687	10.63	12,395	18.86	6,986	10.63
70 years or older	1,961,565	8.35	9,519	14.48	5,489	8.35

Compared to the 2014 results:

+0.8 %

+ 2.3%



# Cell-only Sample from An On-line Panel

- Respondents aged 15 or older were randomly selected from an on-line panel called *EZchoice*. Data are collected using a software application on cell-phones.
- Using the same questionnaire as in CATI, a total of 4,005 surveys were completed from Aug. 19 to Oct. 4, 2016. Among them, 2,532 respondents were identified as cell-only.
- Raking is applied for data weighting, using gender, age, and residential area at the county level as weighting variables.

\*Data for cell-only population is provided by Statinc, an affiliated company of Dataa.

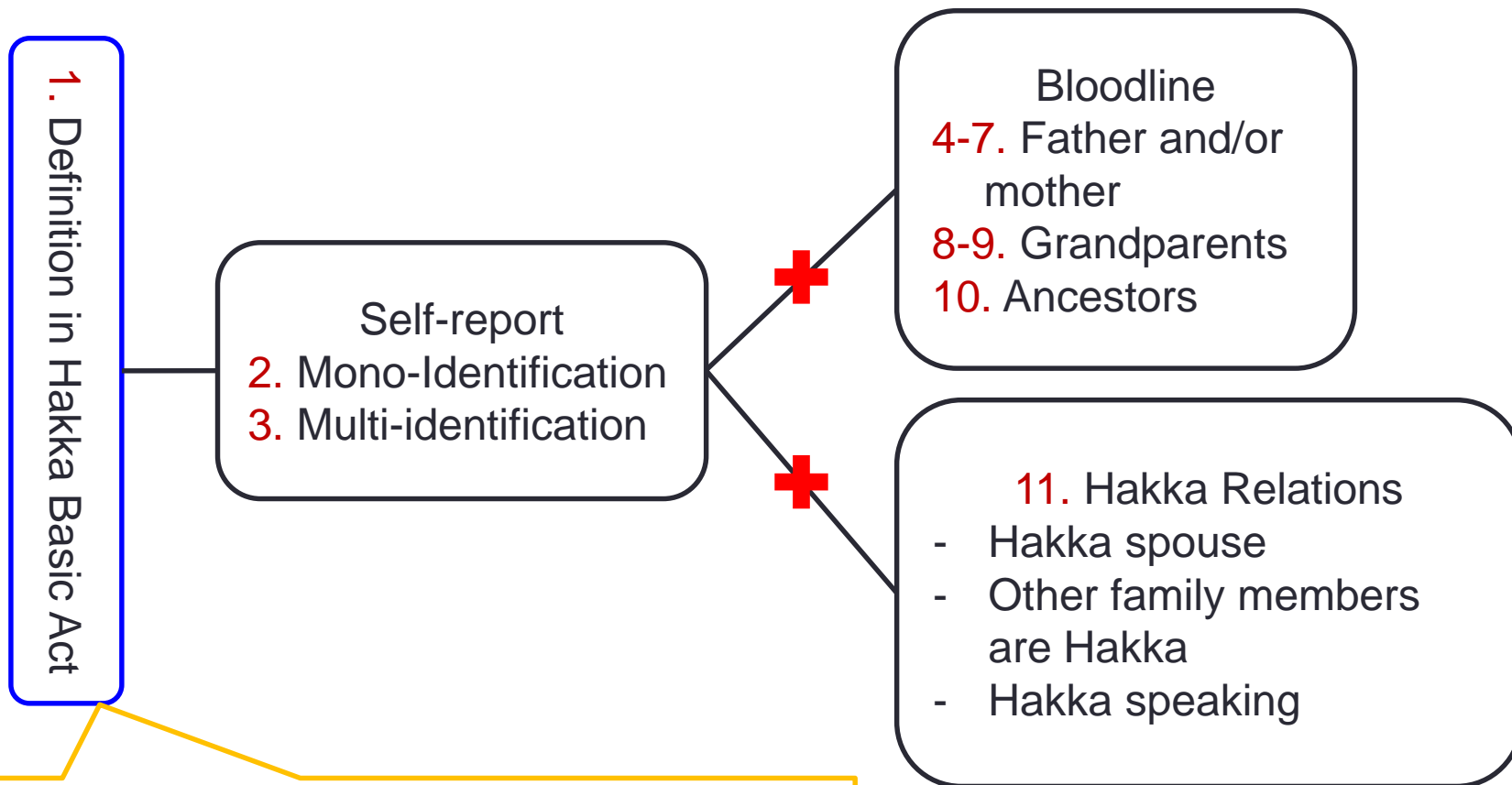
# Descriptive of Cell-only Sample

Variable	%	Variable	%
Gender		Residential area	
Male	45.2	North I	33.9
Female	54.8	North II	16.8
Age		Middle	18.8
15-19	10.9	South I	12.4
20-29	25.2	South II	14.4
30-39	28.2	East	3.6
40-49	17.8		
50-59	9.3		
60-69	7.0		
70 or older	1.6		

n=2,532, before raking.



# Definition conforming to Hakka Basic Act



*“Hakka refers to those who are biologically related or have certain relations with Hakka, and are self-identified as Hakka.”*

# CATI and Cell-only among 20-39 years old

- The cell-only Hakka aged 20-39 have lower proportions of mono-identification and having Hakka ancestors, while the proportion of having Hakka relations is higher, than the CATI Hakka in the same age group.

Definitions	CATI (A)	Cell-only (B)	Difference (B-A)	Sampling error
1. Hakka Basic Act	19.3	19.7	0.4	2.2
2. Multi-identification	20.7	21.6	0.9	2.3
3. Mono-identification	15.8	11.1	<b>-4.7*</b>	1.7
4. Father's bloodline	16.2	15.1	-1.1	2.0
5. Mother's bloodline	15.5	15.3	-0.2	2.0
6. Father and mother	9.7	10.9	1.2	1.7
7. Father or mother	22	19.5	<b>-2.5*</b>	2.2
8. Grand parents on father's side	18.4	19	0.6	2.2
9. Grand parents on mother's side	17.4	19.2	1.8	2.2
10. Ancestors	26.8	17.6	<b>-9.2*</b>	2.1
11. Hakka relations	29.5	39.6	<b>10.1*</b>	2.7

# Cell-only Hakka as a whole

- The followings are found for the cell-only Hakka (n=1,352), when compared to CATI Hakka (n=19,912):
  - A higher proportion of females.
  - Higher proportions of respondents aged 39 or younger; and the proportion of 40 to 49 years old is the lowest.
  - Fewer respondents with a degree of primary school but more with a degree of senior and junior high schools.
  - More respondents working as professional or technical.
  - Higher proportion of them residing in non-Hakka concentrated townships/boroughs.

# CONCLUDING REMARKS

---

# Findings

- Probability adjustment of within-household selection has obtained some success although not satisfactory.
  - The increased percentage of the respondents aged 20-39 is 15.4%.
  - The resulted share of respondents aged 30-39 increased 2.8% than the 2014 National Hakka Survey, but the proportion is still 5.5% lower than that in the population.
  - There is no significant improvement for those aged 20-29 when compared to the 2014 survey.
- Cell-only sample includes higher proportions of respondents aged 20-39.
  - Distribution of Hakka identification differed from the CATI result.
  - For the purpose of the National Hakka Survey, web/app or alternative modes can be considered as a supplement to household landlines.

# Practice/Research in the Future

- Different adjustment for selecting respondents in different age groups.
- Integration of cell-only data into CATI data?
  - Dual-frame sampling for telephone surveys to improve coverage rate.
- Use of web/mobile applications for data collection.
  - Issues of coverage, sampling error, data quality, etc.