Constructing Hard to Survey Index in the Korean Labor Force Survey

Youngshil Park (Statistics Korea)

Sunwoong Kim(Dongguk University)

Okhee Choi (Statistics Koera)

Contents

Background

Previous Studies

Alternatives for Sample Surveys

Results

Conclusion

Background

Background

- Nonresponse rates in national household surveys have increased over the past years

Nonresponse rate in Korean Labor Force Survey: 2008.1~2011.4

Background (Cont.)

- Decline of response rate is associated with increase of survey error and cost

- The appropriate field strategies are needed to improve response rate

- Hard to Count (HTC) Score was used to increase cooperation or to improve estimation value
  - Census Bureau, US (2010)
  - Office for National Statistics, UK (2001)

Background (Cont.)

- Definition: HTC Score / Index

  ➡ "How difficult an area is expected to be to enumerate in the census"

- It is expected that underenumeration in the census will be higher in areas characterised by particular social, economic and demographic characteristics.

  - Ex) multi-occupancy

Background (Cont.)

- In Korea, this kind of indicator would be useful to effectively manage enumeration districts (ED) that are difficult to survey

ex)

Previous Studies

Previous Studies

Office for National Statistics, UK

- ✓ 2001 Census : Census Coverage Survey

Census Bureau, US

- ✓ 2010 Census : Integrated Communication Program

Previous Studies : US

○ Purpose

- Identifying hard to count areas
- Identifying areas with potentially low response rates
- Identifying areas where special attention may be needed for:
  - ○ Questionnaire Assistance Centers
  - ○ Distribution of Be Counted Forms in languages other than English

○ Level of Analysis : Track

Previous Studies : US (Cont.)

○ Variables

- Housing, demographic, and socioeconomic variables that are correlated to mail nonresponse

- Guided by extensive research conducted by the U.S. Census Bureau, and others to measure census coverage.

Previous Studies : US (Cont.)

○ Variables

| No.	Description	
1	% renter occupied units	Housing
2	% vacant units	
3	% non-single family attached	
4	% units with >1.5 person per room	
5	% occupied units with no telephone service	
6	% people below poverty level	Demographic
7	% households with public assistance income	
8	% not high school graduate (ages +25)	
9	% people unemployed	
10	% households that are not husband/wife family	
11	% occupied units where householder moved into unit in 1999-2000	
12	% linguistically isolated households	

Previous Studies : US (Cont.)

○ Calculation of HTC Score

- The value of each individual variable is sorted across geographic areas from high to low

- Scores (0 to 11) are assigned to each variable for each tract

- The scores assigned to each of the 12 variables for a tract are summed to form a composite HTC scores for the tract
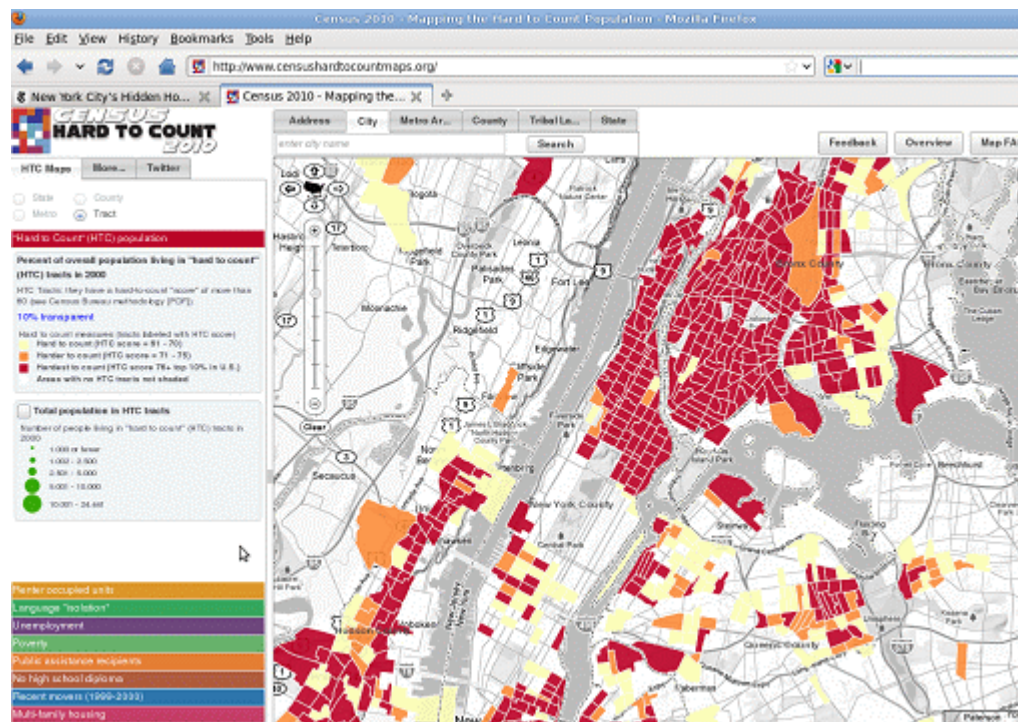
Previous Studies : US (Cont.)

○ Validity

- Has been proven by testing against empirical measures of mail return rates in the 1990 and 2000 census .

Figure 1

Comparison of Trends in 1990 and 2000 Mail Return Rates by Hard-To-Count Scores
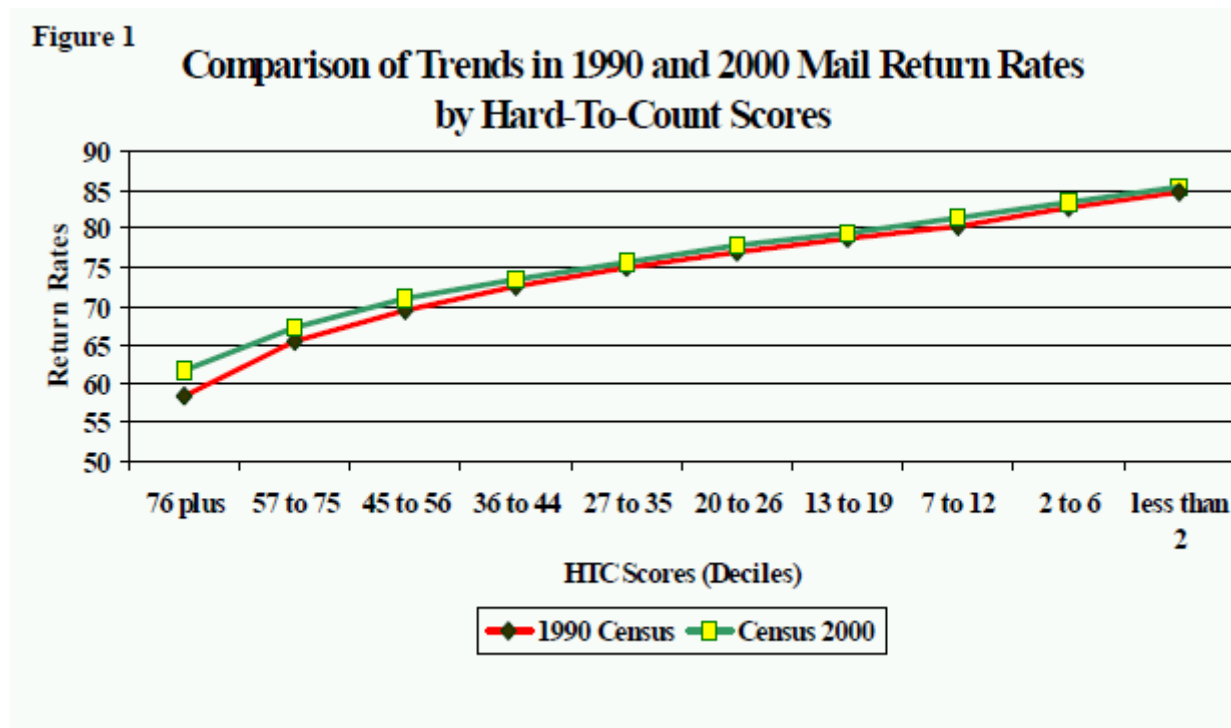
Previous Studies : US (Cont.)

○ Usage

- linking with spatial map data files

Previous Studies : UK

- Purpose

  - Providing a stratification tool for the first stage of the Census Coverage Survey Design, to assign postcodes into groups which should have a similar underenumeration pattern

- Level of Analysis : Enumeration Districts

Previous Studies : UK (Cont.)

○ Variables

- Proposed variables that contribute to under enumeration

- Exploring of the link between the proposed variables and the coverage levels in the 2001 census.

Previous Studies : UK (Cont.)

○ Variables

No.	Description
1	% unemployed persons
2	% persons whose country of birth is non English speaking
3	% households in multiply-occupied buildings
4	% households which were privately rented
5	% imputed households

Previous Studies : UK (Cont.)

- ○ Calculation of HTC score

  - Sum of the proportions of the variables

  $$HtC_{score} = \frac{multiocc\ HHs}{total\ HHs} + \frac{imputed\ HHs}{total\ HHs} + \frac{priv.\ rent\ HHs}{total\ HHs} + \frac{unemployed}{total\ pers} + \frac{CoB\ is\ non\ english\ speaking\ persons}{total\ pers}$$

  - The EDs are ordered by the HTC Score and split into a 40% 40% 20% distribution at the national level.

  - Each group is assigned an index value from 1 (easiest to count) to 3 (hardest to count) , with the top 20% being the EDs with the highest hard to count score

Alternatives for Sample Surveys

Alternatives for Sample Surveys

Data

✓ Korean Labor Force Survey: 2010 (May and June)

  ✓ Focusing on the specific metropolitan area (Kwangju)
  ✓ Some data (ex. migration rate, cooperation rate etc) was supplemented by interviewers

Alternatives for Sample Surveys (Cont.)

Data

✓ Level of Analysis: ED

✓ Average Number of Households in ED: 17

✓ Limitation
  ✓ Data is subjected to sampling error
  ✓ ED information was made by using only response household
    ✓ Nonresponse household information was not collected

Alternatives for Sample Surveys (Cont.)

Analysis

✓ Defining the underlying constructs for measuring how difficult to conduct sample survey in a certain sampling unit

➡️ **Hard to Survey (HTS)**

✓ Selecting the variables and Computing the HTS scores

    ✓ Test: methods of UK and US

✓ Examining the performance of the score by analyzing correlation with future nonresponse rate

Results

Results

- Selection of Variables

  - ✓ Literature review about nonresponse
  - ✓ Expert review

  ⬇

  - ✓ Choosing six variables which are expected to be correlated with nonresponse

  ⬇

  - ✓ Exploring the link between the proposed variables and the nonresponse rate, based on correlation and multiple regression analysis

Results (Cont.)

o Proposed Variables

Variables	Description
Nonresponse Rate	% non-interviewed household of total eligible household
Children	% households having children under 15 years old
Single	% single person households
Size	% housing which of size is top 25%
Moved	% households moved within January to May in 2010
Rented	% households which are rented
Unemployed	% people unemployed

Results (Cont.)

○ Correlation Analysis

	1)	2)	3)	4)	5)	6)	7)
1) NR (May)	-						
2) Children	0.324***						
3) Single	-0.337***	-0.507***					
4) Size	0.180 ∮	0.231*	-0.370***				
5) Moved	0.176 ∮	-0.011	0.023	-0.152			
6) Rented	0.089	-0.070	0.092	-0.295**	0.297**		
7) Unemployed	0.233*	0.101	0.158	-0.015	-0.197*	0.048	-

∮ P<0.1 , * P< 0.05, **<0.01 ***<0.001

Results (Cont.)

○ Multiple Regression Analysis

	coeff.	s.e	t	P-value
constant	0.059	0.026	2.282	0.025
Child	0.052	0.041	1.254	0.213
Single	**-0.175**	**0.062**	**-2.813**	**0.006**
Size	0.022	0.020	1.094	0.276
Moved	**0.234**	**0.088**	**2.659**	**0.009**
Rented	0.025	0.033	0.755	0.452
Unemp	**0.443**	**0.127**	**3.495**	**0.001**
F	6.642			
R-squared	0.281			

Dependent variable: Nonresponse Rate in May

Results (Cont.)

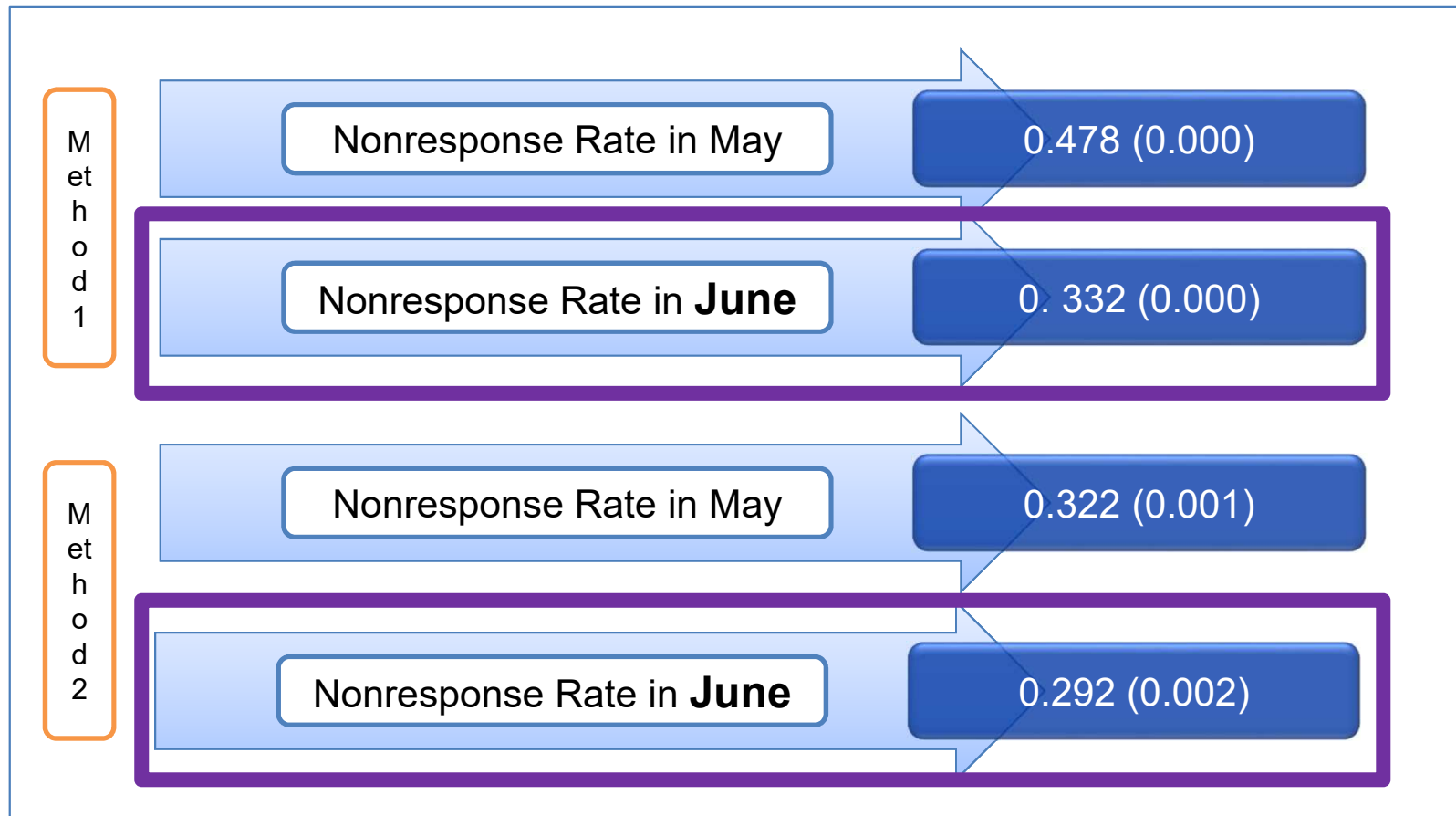- Calculation of Score

Method 1

✓ HTS Score = non-single + moved + unemployed

Method 2

✓ Assign 0 to 11 scores for each variables
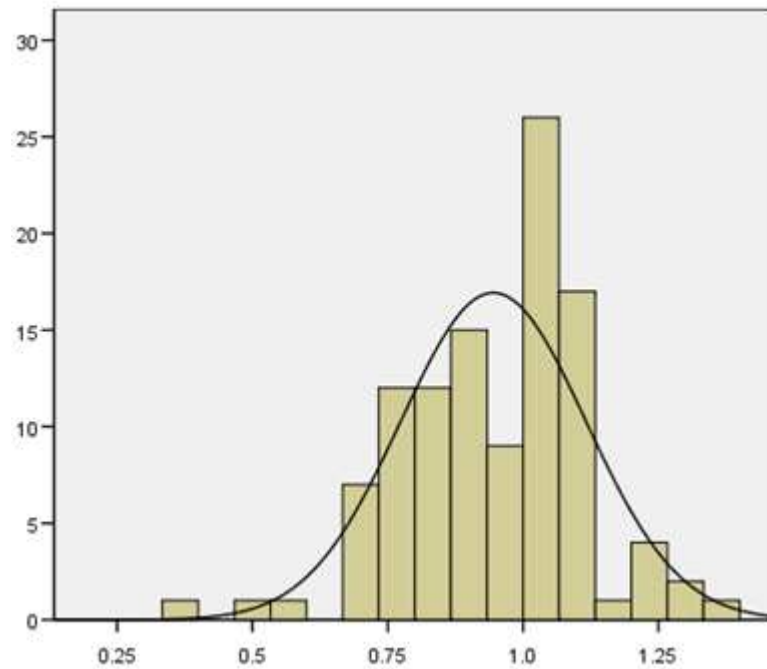✓ Sum of the three variables of score

Results (Cont.)

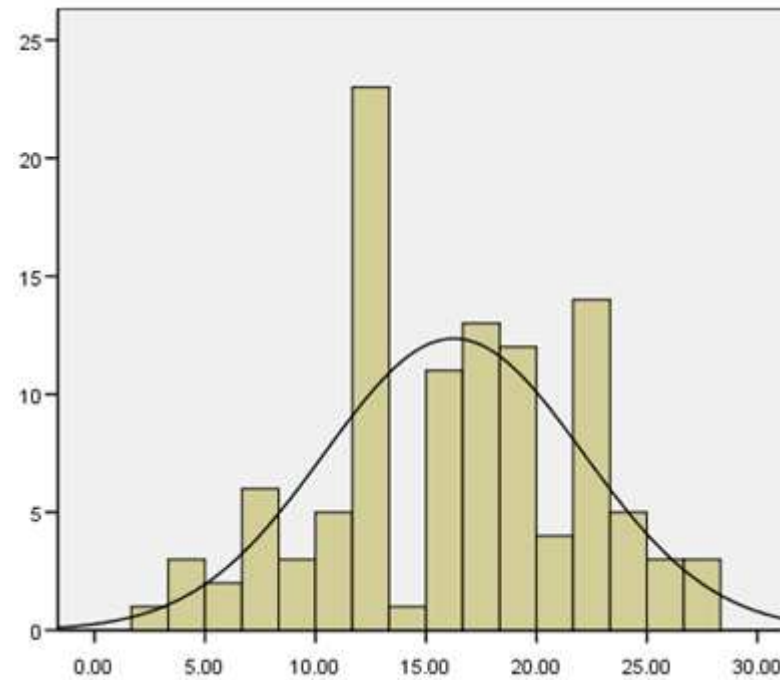○ Correlation HTS Score with Nonresponse Rate

Method 1		
	Nonresponse Rate in May	0.478 (0.000)
	Nonresponse Rate in **June**	0. 332 (0.000)

Method 2		
	Nonresponse Rate in May	0.322 (0.001)
	Nonresponse Rate in **June**	0.292 (0.002)

Results (Cont.)

- Distribution of HTS Score

Method1
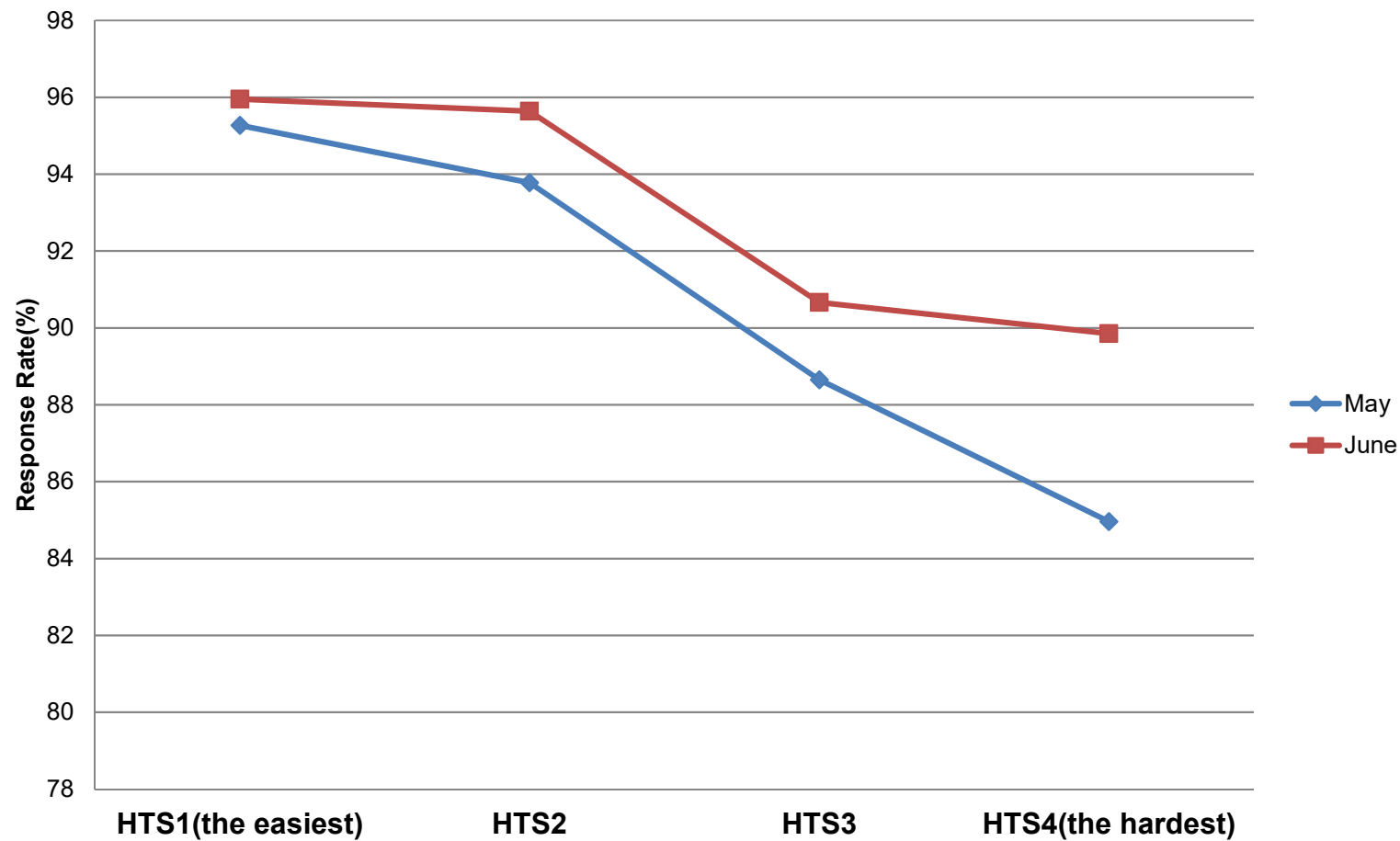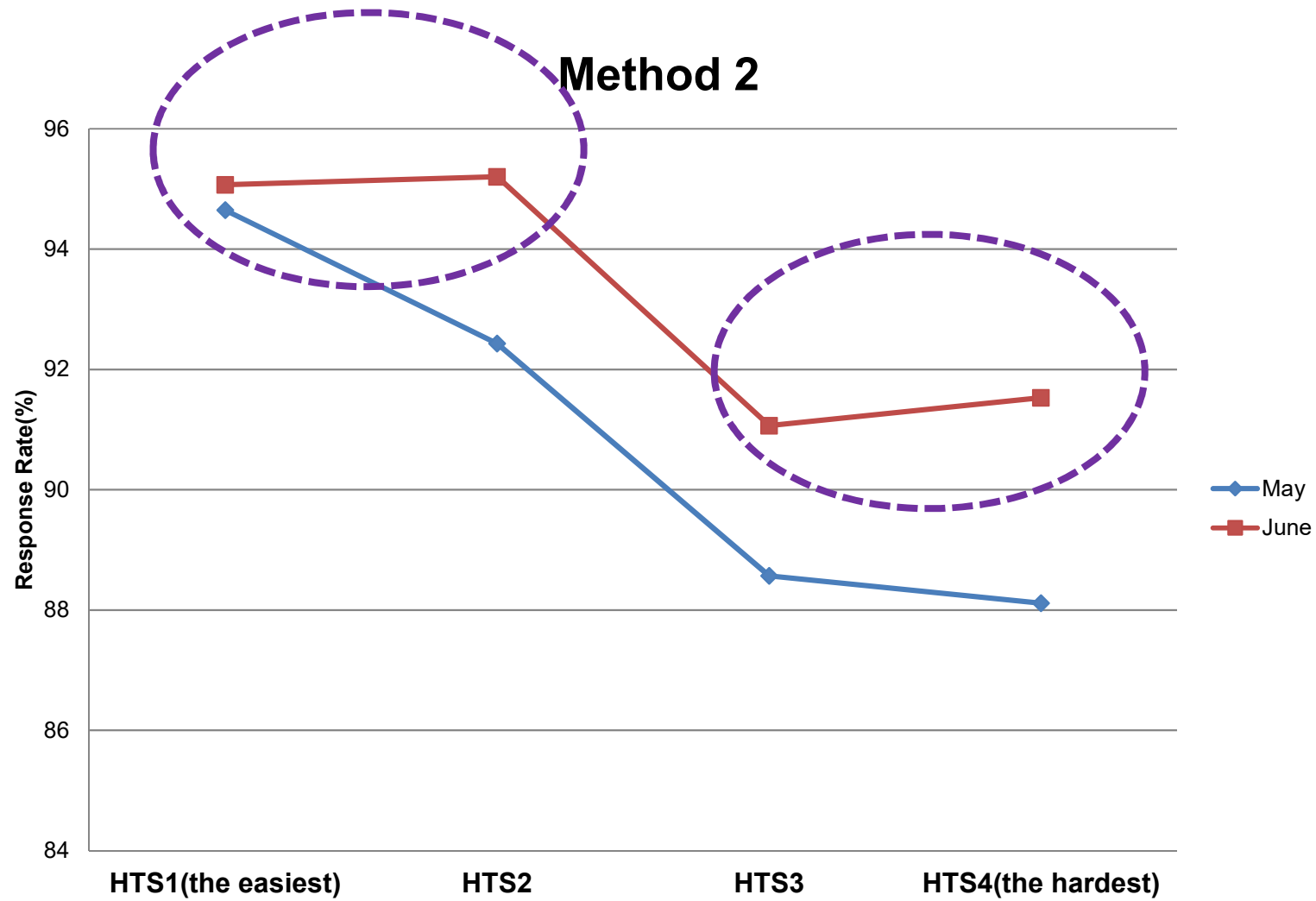
Method2

Results (Cont.)

Method 1

Conclusion

Conclusion

Summary

✓ Selected variables for HTS Score

    ✓ Single
    ✓ Moved
    ✓ Unemployed

✓ Calculation

    ✓ Method 1 preferred

Conclusion (Cont.)

o Comparison of Variables

variables	US	UK	KOREA
% vacant units	0		
% non-single family attached	0	0	
% renter occupied units	0	0	
% units with >1.5 person per room	0		
% hh that are not husband/wife family	0		
% occupied units with no telephone service	0		
% not high school graduate (ages +25)	0		
% people below poverty level	0		
% hh with public assistance income	0		
% people unemployed	0	0	0
% linguistically isolated households	0		
% occupied units where householder moved into unit in 1999-2000	0		0
% persons whose country of birth is non-English speaking		0	
% imputed households		0	
% single person household			0

Conclusion (Cont.)

Implication

✓ Used linking for the map and easily identified which areas are more difficult to survey than other areas

✓ Interviewers or other resources could be effectively assigned based on the HTS Score

Future study

✓ Tests on a national level using more variables would be useful

References

- ONC(SC), 2000, 2001 Hard to Count Index

- ONC(SC), 2001, Transformation of the Hard to Count Variables

- Census Bureau, Tract Level Planning Database With Census 2000 Data

Thank you

Questions: youngshil@korea.kr