Challenges to Telephone Survey Sampling: Frames and Weighting Strategies

Sun-Woong Kim[1)]

Abstract

The growing number of households with only cellular phones and no landline telephones, is decreasing the coverage of the landline frame. Dual frame sample designs are considered which draw phone numbers from the landline and the cellular number frame. Cellular phones may be used as personal or household devices. Weighting strategies appropriate to compensate for unequal probabilities in selecting dual frame samples are proposed. Also, some current issues are discussed.

Key Words: landline phones, cellular phones, noncoverage bias, cellular only households, dual frame design, weighting

1. Introduction

Telephone surveys with the lower cost of data collection compared to face to face surveys has been widely used in the United States since Mitofsky (1970) and Waksberg (1978) devised two-stage random digit dialing (RDD) sample designs. Also, the increase of the coverage rate of telephone households led to the wide spread of telephone surveys. In 1963, about 75 percent of the households in the country had their landline telephones at home and the coverage rate increased to 93 percent by 1886, as reported by Thornberry and Massey (1988).

But by the middle of 1980's many countries including Korea had low coverage rates under 70 percent that would suffer from significant bias if only the telephone households were interviewed (Trewin and Lee (1988)). Since the 1980's, telephone surveys has broadly used due to a rapid spread of landline telephones in the world.

At this time, as presented by Lepkowski (1988), there were concerns about the increase of unlisted telephone numbers and they stimulated the development of more efficient sampling techniques and reliable telephone sampling systems.

Since the early 1990's, the number of cellular (or cell) phone subscribers with advances in telecommunication technology has increased rapidly. In the end of 1990's most countries in Europe Union (EU) and some countries in Asia including Korea passed the cross-over level, which indicates the point that the number of cellular phone subscribers per 100 inhabitants is larger than the number of landlines per 100 inhabitants. Kim and Lepkowski (2002a, 2002b) investigated the effects of cellular phones in the world and concluded that

1)Assistant Professor, Department of Statistics, Dongguk University, e-mail:sunwk@dongguk.edu

there was a trend of decreased coverage of landline telephones due to an increase in celluar only households.

On the other hand, Kim and Lepkowski (2002a) asked a certain number of commercial survey research companies in the EU in order to know whether they were contacting cellular numbers with landline numbers for telephone surveys. Some of them reported that they started to call both numbers in 2000 or 2001, although it was not clear how they was combining the data obtained from the interviews with those numbers.

Since 2002, there have been a lot of presentations and discussions dealing with the issues on the introduction of cellular phone surveys in the American Association for Public Opinion Research (AAPOR), the American Statistical Association (ASA) and the Cell Phone Sampling Summit (CPSS) sponsored by the Nielsen Media Research.

In this paper, we first take a look at the noncoverage problems in recent years in the United States and Korea. Second, we consider dual frame sample designs which draw phone numbers from the landline and the cellular number frame. Third, we suggest some weighting strategies to compensate for unequal probabilities in selecting dual frame samples, which is one of the main issues in respect of survey sampling. Finally, we discuss some current issues.

2. Telephone Noncoverages

As Groves (1987) suggested, "coverage error is the forgotten child among the family of errors to which surveys are subject." In fact, a major historical concern about the telephone surveys has been noncoverage bias, while this error has been often ignored in telephone field work. With this spread of cellular phones, addressing noncoverage problems for landline telephones is necessary and specific treatments for the problems should be speculated.

Tucker and Brick (2005) presented the household telephone status shown in Table 2.1, based on the results from a supplement to the Current Population Survey (CPS) on telephone service in the U.S.

Table 2.1 Household Telephone Status in U.S.

Group	Coverage Rates
Landline and Cell	49.7
Landline only	36.8
Cell only	6.7
No phone	6.7

Table 2.1 shows that the coverage rate for households with landline phones is 86.5 percent and the one for households with cellular phones is 56.4 percent. This indicates that landline telephone noncoverage rate is higher than in the past and cellular phone nocoverage rate is still low. But if we consider the two telephone services together, the coverage rate

is 93.2 percent, which is quite level.

The data in Table 2.2 came from the Computer and Internet Use Survey (CIUS) conducted by the Korea National Statistical Office (2002).

Table 2.2 Household Telephone Status in Korea

Group	Coverage Rates
Landline and Cell	75.3
Landline only	16.3
Cell only	7.8
No phone	0.6

Table 2.2 shows that the total coverage rate for landline phones is 91.6 percent, while the one for cellular phones is 83.1 percent. Although the coverage for landline phones is comparatively high, the important finding in Table 2.2 is that if combining the two types of phones, the noncoverage rate for all households is only 0.6 percent. In this case almost complete coverage is achieved and hence the noncoverage bias that would result from the exclusion of nontelephone households may be disregarded.

As shown in Table 2.1 and Table 2.2, the household telephone statuses among the two countries are quite different. But it is noted that integrating data via the landline phone frame and cellular phone frame may be desirable in both countries.

3. Weighting for Dual Frame Designs

Like other data collection methods, the completeness of sampling frames is important in a survey conducted by telephone. This depends on the accuracy and maintenance of the frame. Here we do not deal with those issues on the frames because they are other current challenges in survey sampling.

Lepkowski and Groves (1986) suggested a dual frame, mixed mode survey design that the samples are selected from both an address frame and a telephone frame. This is an example of a dual frame design and exactly means a telephone survey supplemented with face-to-face interviews in nontelephone households to reduce noncoverage bias. Though this type of survey design can be used to avoid the noncoverage problem in a landline frame, we may prefer a different type of survey design, a mix of landline samples and cell samples, because of less costs. They also proposed a post-stratified estimator of the overall mean across some domains. This estimator includes a mixing parameter that is chosen to minimize the mean square error. A variety of sources are needed to find out the value of the estimator. We develop the following straightforward method.

To compensate for unequal probabilities when selecting samples in the dual frame, we need appropriate weighting schemes. But developing those schemes is not a simple matter. This comes from the following facts. There has been a primary concern that the cellular phone is a personal device unlikely the landline phone that. Also, it is clearly evident from

Tucker and Brick (2005) that the cellular phone is not used only as a personal device. For example, in 2004 there existed 33.1 percent of households having cellular phones used by multiple household members among the households with 3 cellular phones in the United States.

Since the types of phones depend on the selected household, the following concepts for the weights are considered:

(i) Household Level Weight
(ii) Person Level Weight in the household

We introduce the following notation to determine sample design weights in the dual frame.

C : cell phones
L : landline phones
CPD : cell phones used as a personal device
CHD : cell phones used as a household (HH) device
LPD : landline phones used as a personal device
LHD : landline phones used as a HH device
π_{C_i} : sampling rate for cell phone
π_{F_i} : sampling rate for landline phone
N_{CPD} : number of cell phones used as a personal device
N_{CHD} : number of cell phones used as a HH device
N_{LPD} : number of landline phones used as a personal device
N_{LHD} : number of landline phones used as a HH device
N_A : number of adults in the HH
N_{ACPD} : number of adults using cell phones as a personal device in the HH
N_{ACHD} : number of adults using cell phones as a HH device in the HH
N_{AOUC} : number of adults having overlapped uses for cell phones in the HH
N_{ALPD} : number of adults using landline phones as a personal device in the HH
N_{ALHD} : number of adults using landline phones as a HH device in the HH
N_{AOUL} : number of adults having overlapped uses for landline phones in the HH
$N_{AOUCPLH}$: number of adults having overlapped uses for both cell phone as a personal device and landline phone as a HH device
$N_{AOUCHLP}$: number of adults having overlapped uses for both cell phone as a HH device and landline phone as a personal device
$N_{AOUCHLH}$: number of adults having overlapped uses for both cell phone as a HH device and landline phone as a HH device

The weighting strategies for landline only households or cellular only households may be simpler than those for households having both. We suggest the specific schemes for the latter in Table 3.1.

<Table 3.1> Weighting strategies for households having both landlines and cells

Types of Phones	Household Level Weight	Person Level Weight
CPD+LPD	$\frac{1}{\pi_{C_i} N_{CPD_i} + \pi_{L_i} N_{LPD_i}}$	$\frac{N_A}{\pi_{C_i} N_{CPD_i} + \pi_{L_i} N_{LPD_i}}$
CPD+LHD	$\frac{1}{\pi_{C_i} N_{CPD_i} + \pi_{L_i} N_{LHD_i}}$	(1-1) (1-2) (1-3)
CHD+LPD	$\frac{1}{\pi_{C_i} N_{CHD_i} + \pi_{L_i} N_{LPD_i}}$	(2-1) (2-2) (2-3)
CHD+LHD	$\frac{1}{\pi_{C_i} N_{CHD_i} + \pi_{L_i} N_{LHD_i}}$	(3-1) (3-2) (3-3)

Note.(1-1): $\frac{N_{ACPD_i}}{\pi_{C_i} N_{CPD_i} + \pi_{L_i} N_{LHD_i}}$ if the adult has a cell only as a personal device

(1-2): $\frac{N_{ALHD_i}}{\pi_{C_i} N_{CPD_i} + \pi_{L_i} N_{LHD_i}}$ if the adult has a landline only as a HH device

(1-3): $\frac{N_{AOUCPLH_i}}{\pi_{C_i} N_{CPD_i} + \pi_{L_i} N_{LHD_i}}$ if the adult has both

For others the similar schemes are applied.

4. Current Issues

The distinction between household and personal use for cell phones is one that has not until now been made explicit in telephone surveys. The distinction is made in the weighting schemes only for the purposes of compensating nonequal probability selections of the households and persons. In order to calculate the weights we need to design specific questions in telephone interviews to find out the numbers concerning phones and adults in the household.

The proposed weighting strategies are still complicative in using in telephone surveys. There can be many different types of phones in the households. Therefore we need to develop the simpler strategies, which do not require a number of procedures and questions in interviews.

References

[1] Groves, R.M. (1987). Research on Survey Data Quality, *Public Opinion Quarterly,* 50th Anniversary Issue, S156-S172.

[2] Kim S.W. and Lepkowski, J.M.. (2002). Telephone household non-coverage and mobile telephones, paper presented at the *Annual Meeting of the American Association for Public Opinion Research*, St. Pete. Beach, FL.

[3] Kim S.W. and Lepkowski, J.M.. (2002). The changes and trends in telephone household coverage, *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 1851-1856.

[4] Korea National Statistical Office (2002). Computer and Internet Use Survey.

[5] Lepkowski, J.M. (1988). Telephone sampling methods in the United States, *Telephone Survey Methodology*, Groves, R.M., et al., John Wiley & Sons, New York, 73-98.

[6] Lepkowski, J.M. and Groves, R.M. (1986). A mean squared error model for dual frame, mixed mode survey design. *Journal of the American Statistical Association*, 930-937.

[7] Mitofsky, W. (1970). Sampling of telephone households, Unpublished CBS memorandum.

[8] Thornberry. O.W. and Massey, J.T. (1988). Trends in United States telephone coverage across time and subgroups, *Telephone Survey Methodology*, Groves, R.M., et al., John Wiley & Sons, New York, 25-49.

[9] Trewin M. and Lee, G. (1988). International comparisons of telephone coverage, *Telephone Survey Methodology*, Groves, R.M., et al., John Wiley & Sons, New York, 9-24.

[10] Tucker, C. and Brick, M. (2005). CPS estimates of cell-phone only population size and characteristics, presented at the *Cell Phone Sampling Summit II.* New York, New York.

[11] Waksberg, J. (1978). Sampling methods for random digit dialing, *Journal of the American Statistical Association*, 73, 40-46.