gesis
Leibniz-Institut
für Sozialwissenschaften

Working *Papers* 2011|13

Combining landline and mobile phone samples

A dual frame approach

Mario Callegaro, Oztas Ayhan, Siegfried Gabler, Sabine Haeder, Ana Villar

GESIS-Working Papers 2011|13

Combining landline and mobile phone samples

A dual frame approach

Mario Callegaro, Oztas Ayhan, Siegfried Gabler, Sabine Haeder, Ana Villar

GESIS – Leibniz-Institut für Sozialwissenschaften 2011

GESIS-Working Papers

GESIS – Leibniz-Institut für Sozialwissenschaften
Postfach 12 21 55
68072 Mannheim
Telefon: (0621) 1246 - 282
Telefax: (0621) 1246 - 100
E-Mail: sabine.haeder@gesis.org

ISSN: 1869-0483 (Print)
ISSN: 1869-0491 (Online)
Herausgeber,
Druck und Vertrieb: GESIS - Leibniz-Institut für Sozialwissenschaften
Unter Sachsenhausen 6-8, 50667 Köln

Contents

page

Abstract 7

1 Introduction 8

2 Dual frame approaches 12

3 Dual frame approach and optimal allocation 14

4 Dual frame application 19

5 Conclusions 27

6 References 28

Appendix 32

Abstract

More and more households abandon their landline phones and rely solely on cell phones. This implies a challenge for survey researchers: since the cell phone only households are not included in the frames for landline telephone surveys, samples based on these frames are in danger to be seriously biased due to undercoverage, if respondents who do not have a landline are systematically different from respondents who have a landline. Thus, strategies for combining samples from different frames need to be developed. In this paper we give theoretical foundations for a dual frame approach to sampling, explain how samples can be optimally allocated from these two frames, and describe an empirical application of a survey conducted in Germany that used a dual frame approach.

1 Introduction

Phone surveys since their beginning until 2007 used mainly two strategies to build telephone sampling frames: lists of landline phone numbers, and random digit dialling (RDD). The landline sampling frame provided good coverage for telephone surveys, given that 95% of households in countries like the United States or Germany had landline phones and hence had the potential of being in the frame, and a non-zero probability of being part of the sample. The remaining 5% of households were unreachable by phone and thus had a zero chance of being part of the sample. In spite of known differences between households with landlines and households with no phone service, the low percentage of the latter kept bias reasonably small for many survey estimates (Tucker & Lepkowski, 2008). However, a trend is clear in almost all industrialized countries: more and more people are "cutting the cord" becoming cell phone-only households and more people, especially younger individuals who leave their parents' homes and move into a new household, do not even install a landline phone thus relying only on their cell phone for telephone service. For example in 2007 in the European Union member states, 24% of households could be reached only via mobile phone (TNS Opinion & Social, 2008). Table 1 shows more recent data from the European Social Survey (ESS) which was fielded in 2008/2009[1]. Especially in new European member states such as Bulgaria and Latvia it is obvious that telephone samples including only landline phone households would lead to large undercoverage rates.

[1] http://ess.nsd.uib.no/ess/round4/download.html

Table 1: Percent of households with no landlines and individuals with mobile phones, results from the European Social Survey Round 4, weighted data, countries in alphabetical order

Country	**No landline phone in household (in %)**	**Personally have mobile phone (in %)**
Belgium	23.1	89.4
Bulgaria	43.5	72.2
Croatia	12.5	85.2
Cyprus	10.5	85.2
Czech Republic	73.3	92.0
Denmark	20.1	92.5
Estonia	42.6	90.2
Finland	65.6	95.2
France	7.6	82.6
Germany	6.1	84.9
Greece	15.8	90.8
Hungary	46.2	79.5
Israel	12.3	89.9
Latvia	56.5	82.0
Netherlands	12.3	91.3
Norway	28.1	96.6
Poland	34.1	80.2
Portugal	35.0	81.7
Romania	59.4	77.4
Russian Federation	48.3	71.7
Slovakia	45.2	85.6
Slovenia	10.2	87.9
Spain	24.7	82.0
Sweden	8.9	92.0
Switzerland	10.2	88.6
Turkey	38.3	66.7
Ukraine	33.2	77.1
United Kingdom	11.7	85.3

In North America, Canada shows a slower trend with only 8% of CPO-HH as of December 2008 (Statistics Canada, 2009). On the other hand, the United States is following the European trend (Blumberg & Luke, 2009). In the first half of 2009, 22.7 percent of all households were cell phone-only households. More importantly, differences between cell phone-only households and non cell phone-only households can be found in sociodemographic variables such as age, marital status, or household size or poverty level (cites).

With the increasing number of households becoming cell phone only, the coverage error of traditional landline RDD is increasing, thus forcing researchers to consider augmenting landline samples with cell phone samples (AAPOR cell phone task force, 2010; Callegaro & Poggio, 2004; Kuusela & Simpanen, 2002; Lavrakas, Shuttles, Steeh, & Fienberg, 2007; Taylor, 2003). If telephone surveys are conducted using a landline frame only, the coverage error could introduce possible bias in the survey estimates thus becoming non-ignorable. Enough evidence has been found of coverage bias for landline-only sampling frames. For instance, in the above cited estimates (Blumberg & Luke, 2009) wireless-only adults (keeping everything else constant)

- were more often binge drinkers (35.3%) than adults living in landline households (19.3%)
- were more often reporting an excellent or very good health status
- were more likely to have no health insurance coverage than adults with landline phone.

Mokrzycki, Keeter and Kennedy (2009) showed remarkable differences in presidential preferences between adults in cell phone-only households and adults in landline households. With the help of the 2008 Exit Poll, a clear preference for candidate Obama by the wireless-only electors – especially among the younger adults - was documented. This has of course implications for post-stratification weighting adjustments to age because age groups are obviously not homogeneous with respect to phone ownership. Thus it may become more and more difficult to compensate non-coverage bias through weighting in pure, traditional landline telephone surveys.

The majority of studies attempting to measure coverage error use a face to face survey with very high response rate and compare the answers of cell phone-only households to households with a landline phone. Delnevo, Gundersen and Hagman (2008) show evidence of the increasing coverage bias. Analyzing five consecutive years of the landline telephone interview Behavioral Risk Factor Surveillance System (BRFSS), the authors found a declining prevalence of alcohol drinking and smoking among young adults. Comparing the data with use area probability samples, such as the National Health Interview Survey and the National Survey of Drug Use and Health, they attribute this decline to coverage error due to wireless substitution instead of a "real" decline of these behaviours in the U.S. young adult population. Confirmation to the Delnevo and colleagues findings comes from Link et al. (2007). Using a parallel dual frame survey (cell and landline) in three states of the BRFSS the authors show how respondents on the cell phone frame have significantly different health practices and behaviours than respondents reached on the landline frame.

A major line of research established by the Pew Center for the People and the Press started polling landline and cell phone-only respondents in the United States. Keeter et al. (2007) summarized the results of four dual frame telephone surveys conducted in 2006, finding minimal impact of including a sample of CPO respondents on general population estimates of cell phone use, media consumption, political attitudes and technology consumption. The authors, however, found that for young adults (aged 18-25 years) estimates of certain variables would have been biased if a cell phone-only sample was not included in the design. A follow-up study analyzed four dual frame surveys conducted between October 2007 and February 2008 (Keeter, Dimock, Kennedy, Best, & Horrigan, 2008). Although the findings were in line with the previous study on some variables, cell phone-only respondents differed from landline respondents in certain survey estimates (e.g. technology use). The authors conclude stating that "The decline in the number of young respondents – a casualty of the cell-only phenomenon – is one of the most important problems facing the telephone survey community today" (p. 14). In comparison to the previous study Keeter et al. (2008) point out that another group where differences between CPO-HH and landline HH are becoming of a certain size are Hispanics. In fact both young adults and Hispanics were approximately twice as likely to be reached on a cell phone than on a landline phone. A specific study on the topic has been reported by Dutwin, Kennedy, Keeter & Kulp (2008). Other studies comparing cell phone-only households and landline households are presented by Keeter (2008), Witt, Best & Rainie (2008), the Pew Research Center (2008) and Brick et al. (2006; 2007).

In the following we want to concentrate on the results of the German dual frame telephone survey CELLA1 (Haeder & Haeder, 2009). CELLA1 was developed with funding from the German Research Foundation in a joint project of the University of Technology in Dresden and GESIS in Mannheim. The project aims to develop appropriate ways to combine landline and cell phone frames. For this, in the first phase of the project we conducted two surveys – one on fixed line phones and one on mobile phones with around 1,000 interviews each.

In this paper we want to describe the model we have applied for the combination of both surveys. First, the theoretical model for the dual frame approach will be explained. Then, a general formula for the optimal allocation for landline and cell phone samples under cost restrictions is developed. After this, we introduce the CELLA1 study and show results of the landline and cell phone subsamples as well as of the combined sample. It will become clear that the use of dual frame approaches is workable and worth the effort.

2 Dual frame approaches

Dual frame surveys have been a concern of survey statisticians for many years. When single frames are not sufficient to cover the target population, dual frames are necessary. This issue was first confronted by Stephan (1936), followed by Goodman (1949; 1952), for the estimation of *k*–classes (lists) in the population. Hartley (1962; 1974) and Cochran (1967) have used this concept in today's context. Hartley (1962) has proposed the unbiased estimator of Y,

$$\hat{Y}_H = f_A^{-1}\ y_a + p\ f_A^{-1}\ y_{ab}^{'} + (1-p) f_B^{-1} y_{ab}^{''} + f_B^{-1} y_b$$

where *p* is a constant to be chosen to minimize the sampling variance of the estimator [$\mathrm{var}(\hat{Y}_H)$]. Here, $f_A^{-1} = (n_A / N_A)^{-1} = (N_A / n_A)$ and $f_B^{-1} = (N_B / n_B)$ refers to the expansion factors for two respective frames, and y_a, $y_{ab}^{'}$, $y_{ab}^{''}$, y_b refers to sample totals of the domains.

Cochran (1967) has also proposed estimators for dual and triple frame surveys. On the other hand, Lund (1968) has proposed the following alternative estimator to improve on the Hartley's estimator.

$$\hat{Y}_L = f_A^{-1}\ y_a + (p\ f_A^{-1}\ n_{ab}^{'} + (1-p) f_B^{-1}\ n_{ab}^{''}) \bar{y}_{ab} + f_B^{-1} y_b$$

Here, $n_{ab}^{'}$ *and* $n_{ab}^{''}$ refers to number of sample observations in respective domains and $\bar{y}_{ab} = y_{ab} / n_{ab}$ is the sample mean of variable Y in the intersection domain. This estimator is also unbiased and $\mathrm{var}(\hat{Y}_L) \le \mathrm{var}(\hat{Y}_H)$ for any value of *p*. Lund also proposed that *p* be chosen to minimize $vay(\hat{Y}_L)$. On the other hand, Fuller and Burmeister (1972) had proposed the following estimator,

$$\hat{Y}_{FB} = \left(N_A - \hat{N}_{ab}\right) \bar{y}_a + \hat{N}_{ab}\ \bar{y}_{ab} + \left(N_B - \hat{N}_{ab}\right) \bar{y}_b$$

where, $N_A = N_a + N_{ab}$ and $N_B = N_b + N_{ab}$ and also ignoring the finite population corrections, $\hat{N}_{ab}$ is taken as the smallest root of the following quadratic equation,

$$\left(n_A + n_B\right) x^2 - \left(n_A\ N_B + n_B\ N_A + n_{ab}^{'}\ N_A + n_{ab}^{''}\ N_B\right) x + n_{ab}\ N_A\ N_B = 0$$

They have shown that, asymptotically $\hat{Y}_{FB}$ is unbiased, and $\mathrm{var}(\hat{Y}_{FB}) \le \mathrm{var}(\hat{Y}_L)$ for any value of *p*.

Bankier (1986) has proposed the use of the Horvitz-Thompson estimator of the simple form. $\hat{Y}_S = f_A^{-1}\, y_a + f_B^{-1} y_b + (f_A + f_B)^{-1}\, y_{ab}$

This estimator also requires that the duplicate units have to be identified. Skinner (1991) has proposed the following raking ratio (RR) estimators for multiple frame surveys. As $r \to \infty$, $\hat{Y}_{RR}^{(r)} \to \hat{Y}_{RR}^{(\infty)}$, where $\hat{Y}_{RR}^{(\infty)} = \left(N_A - \widetilde{N}_{ab}\right)\bar{y}_a + \widetilde{N}_{ab}\bar{y}_{ab} + \left(N_B - \widetilde{N}_{ab}\right)\bar{y}_b$

and $\widetilde{N}_{ab}$ is the smallest root of the following quadratic equation

$$n_{ab}\, x^2 - \left[\, n_{ab}\, (N_A + N_B) + \left(f_A^{-1} + f_B^{-1}\right) n_a\; n_b \,\right] x + n_{ab}\; N_A\; N_B = 0$$

Further developments on dual frame survey methodologies and related weighting issues are extensively covered by Skinner and Rao (1996), Lohr and Rao (2000), and Lohr and Rao (2006).

Gabler and Haeder (2002) have provided detailed information on the landline telephone system in Germany. The paper illustrates the number of possible digits which are used for area codes as well as the suffix (*central office codes + local numbers*). This certainly will provide essential information for one of the basic data sources of the dual frame for German telephone systems.

Brick *et.al.* (2006) proposed an alternative estimator where the non-response adjusted weights for households with landlines only is y_a, and for households with cell phones only is y_b. A composite estimator is $y_{comp} = y_a + y_b + y_\lambda$ where the overlap population is estimated by $y_\lambda = \lambda\, y'_{ab} + (1-\lambda)\, y''_{ab}$ with $0 < \lambda < 1$, and y'_{ab} and y''_{ab} are the non-response adjusted estimators of households with both cell and landlines from frame A and frame B, respectively. We refer to these weights, with $\lambda = 0.5$, as the "simple composite" weights.

Recently Gabler and Ayhan (2007) have proposed a weighting procedure for combining cell phone frames with the landline telephone frames which is explained in more detail in Section 3.

3 Dual frame approach and optimal allocation

By definition a landline sample and a cell phone sample overlap for respondents in households with both landline and cell phones. As extensively debated in the AAPOR cell phone task force document (2010), two different approaches have been used to handle the overlap: the screening approach and the overlap approach.

- In the screening approach the interviewers terminate interviews with cell phone respondents who have at least one landline phone in the household and thus could potentially be reached through the landline sample frame. These individuals will be eligible only when sampled from the landline frame.
- In the overlap approach the interview is conducted regardless of the frame from which the respondent is selected (landline or cell) and their phone ownership status, and information about each respondent phone status is collected from both frames.

At this current stage of knowledge there is no general consensus on what method is preferable (AAPOR cell phone task force 2010, p.6). We decided to use the second approach since the cell phone only population in Germany - where the survey was conducted - is still so small that screening for cell phone-only respondents was deemed too time and money consuming.

The resulting sample included 8.1% cell phone-only households, 6.5% landline-only households, and 85.4% households where both landline and cell phones were present.

To combine these samples from both frames we computed the inclusion probability of each household so as to account for multiplicity. To compute inclusion probabilities we used the Gabler and Ayhan's (2007) definition of the following relevant parameters:

Landline		**Cell phone**	
M^L	frame size of numbers	M^C	frame size of numbers
m^F	sample size of numbers	m^C	sample size of numbers
k_i^L	count of landline numbers allowing access to target person i	k_i^C	count of cell phone numbers allowing access to target person i
z_i	size of household to which target person i belongs		

To simplify the formulae we made the following fundamental assumption:

> *The probability that two (not necessarily distinct) members of the same household are selected into the sample through different ways is negligible. The probability that two (not necessarily distinct) members of the same household are selected into the sample through different ways is negligible.*

Then, the inclusion probability of a target person via the landline frame is

$$\pi_i^L = \sum_{j=0}^{k_i^L} \left(1 - \left(1 - \frac{1}{z_i} \right)^j \right) \frac{\binom{k_i^L}{j}\binom{M^L - k_i^L}{m^L - j}}{\binom{M^L}{m^L}} \approx k_i^L \frac{m^L}{M^L} \cdot \frac{1}{z_i} \, .$$

Quite analogously, the inclusion probability of a target person via the cell phone frame is

$$\pi_i^C = \sum_{j=1}^{k_i^C} \frac{\binom{k_i^C}{j}\binom{M^C - k_i^C}{m^C - j}}{\binom{M^C}{m^C}} \approx k_i^C \frac{m^C}{M^C} \, .$$

In consideration of the fundamental assumption above we deduce

$$\pi_i^{L \cap C} = \pi_i^L \pi_i^C \approx 0$$

and it follows for the inclusion probability of the target person *i*

$$\pi_i \approx k_i^L \frac{m^L}{M^L} \cdot \frac{1}{z_i} + k_i^C \frac{m^C}{M^C} .$$

From a practical viewpoint, to combine the samples, researchers need to include items in the questionnaire for telephone group classification (AAPOR 2010, p. 68). These questions determine whether a respondent selected through one frame could have been selected in the other frame. However, the true value for k_i^L is often not known. In Germany, one possibility to estimate k_i^L is to distinguish between households with ISDN (Integrated Services Digital Network) and households with analog modem, and to assign a fixed value, say k_i^L = 2.5 for a target person living in an ISDN-household and k_i^L = 1 for a target person living in a household with analog modem. These ISDN values were based on data from a previous study where f respondents had been asked about the number of landlines in their household[2], and the mean was 2.5. However, in a recent study conducted at GESIS in January 2010 we found the following distribution of numbers belonging to ISDN that leads to a mean of 2.8:

Table 2: Percent of respondents with landline numbers by ISDN access (Data Source: Mobilpanel survey, GESIS 2010)

Landline numbers	Respondents with ISDN
2	76
3	51
4	5
5	7
6 and more	9
	148

Unfortunately, a lot of the generated numbers are non-working numbers. We define U^L as the set of all persons having a landline phone. Quite analogously we define U^C as the set of all persons having a cell phone. It should be clear that the size N^L of U^L is much smaller than M^L and that the size N^C of U^C is much smaller than M^C. Using $\pi_i^L \approx k_i^L \frac{m^L}{M^L} \cdot \frac{1}{z_i}$, $i = 1,\dots,N^L$, and $\pi_i^C \approx k_i^C \frac{m^C}{M^C}$, $i = 1,\dots,N^C$, the usual estimation strategy for the dual frame

$$\hat{t}_y = \left(\sum_{i \in s^{L\cap\bar{C}}} \frac{y_i}{\pi_i^L} + \sum_{i \in s^{L\cap C}} \lambda_i \frac{y_i}{\pi_i^L} \right) + \left(\sum_{i \in s^{L\cap C}} (1-\lambda_i) \frac{y_i}{\pi_i^C} + \sum_{i \in s^{\bar{L}\cap C}} \frac{y_i}{\pi_i^C} \right)$$

where $\left\{ s^{L\cap\bar{C}}, s^{L\cap C}, s^{\bar{L}\cap C} \right\}$ is the partition of the sampled persons in $\left\{ U^{L\cap\bar{C}}, U^{L\cap C}, U^{\bar{L}\cap C} \right\}$. $U^{L\cap\bar{C}}$ denotes the landline-only individuals, $U^{\bar{L}\cap C}$ the set of cell phone-only individuals and $U^{L\cap C}$ the set of all person having both kinds of phones. It is obvious that the size of sampled persons is not fixed but random,

2 We need to correct the different inclusion probabilities one has because his fixed line phone is in our frame with multiple numbers. We also made an experiment at the University of Dresden were we asked people about the number of their fixed line phones and then looked at their contracts if this information was correct. We found that people are not able to give valid information on the number of active telephone numbers they have. That's why we use the fixed value as approximation.

depending on how many non-working numbers are selected. n^L denotes the sample size of all persons in $s^L = s^{L\cap\bar{C}} \cup s^{L\cap C}$. n^C denotes the sample size of all persons in $s^C = s^{\bar{L}\cap C} \cup s^{L\cap C}$. The set of all sampled person is $s = s^L \cup s^C$.

For $\lambda_i = \dfrac{\pi_i^L}{\pi_i^L + \pi_i^C}$ we get the Horvitz-Thompson estimator $\sum_{i\in s} \dfrac{y_i}{\pi_i}$ for the total of the y-values and are in the situation of a single frame approach (Bankier, 1986; Lohr, 2009; Lohr & Rao, 2006). In Brick et al. (2006) $\lambda_i = 0.5$ or $\lambda_i = 0.42$ are used.

A well known possibility to adjust the Horvitz-Thompson estimator because of nonresponse is to use the generalized regression estimator $\sum_{i\in r} g_i \dfrac{y_i}{\pi_i}$, where $r \subset s$ is the subset of respondents and

$$g_i = \frac{1}{q_i}\left(1 + c_i\left(\sum_{k=1}^{N}\mathbf{x}_k - \sum_{k\in r}\frac{1}{\pi_k q_k}\mathbf{x}_k\right)'\left(\sum_{k\in r}\frac{c_k}{\pi_k q_k}\mathbf{x}_k\mathbf{x}_k{}'\right)^{-1}\right)\mathbf{x}_i .$$

$\mathbf{x}_k$ denotes a vector of auxiliary variables for unit k, q_k is the response probability of unit k and c_k a fixed positive number which is often 1, but may be chosen differently for units in the sets $U^{L\cap\bar{C}}, U^{L\cap C}, U^{\bar{L}\cap C}$.

To find a solution for optimal allocation of the sample size to landline and cell phone survey under cost control we define

$$\theta_i^L = \begin{cases} y_i & \text{for } i \in U^{L\cap\bar{C}} \\ \lambda_i y_i & \text{for } i \in U^{L\cap C} \end{cases}$$

$$\theta_i^C = \begin{cases} y_i & \text{for } i \in U^{\bar{L}\cap C} \\ (1-\lambda_i) y_i & \text{for } i \in U^{L\cap C} \end{cases}$$

and have to minimize $\operatorname{var}\left(\hat{t}_y\right)$ under the cost boundary $m^L p^L + m^C p^C \le g$ where p^L is the average cost arising for a number in the landline part of the survey and p^C is the average cost arising for a number in the cell phone part of the survey and m^L and m^C are the corresponding sample sizes. g denotes the whole cost boundary. It is obvious that the true cost for a selected number varies from "low" for a non-working number to "higher" for a number where an interview can be conducted.

Taking the derivative of

$$\operatorname{var}\left(\hat{t}_y\right)-\mu\left(m^L p^L+m^C p^C\right)$$

with respect to m^L and m^C approximately results

$$m^L=g\cdot\frac{\sqrt{\frac{\gamma}{p^L}}}{\sqrt{\gamma p^L}+\sqrt{p^C}}=g\cdot\frac{\sqrt{\gamma}}{\sqrt{\gamma}p^L+\sqrt{p^L p^C}}$$

$$m^C=g\cdot\frac{\sqrt{\frac{1}{p^C}}}{\sqrt{\gamma p^L}+\sqrt{p^C}}=g\cdot\frac{1}{\sqrt{\gamma p^L p^C}+p^C}$$

where

$$\gamma=\frac{V^L_{\psi^L}}{V^C_{\psi^C}}=\frac{\sum_{i\in U^L}\psi_i^L\left(\frac{\theta_i^L}{\psi_i^L}-t_\theta^L\right)^2}{\sum_{i\in U^F}\psi_i^C\left(\frac{\theta_i^C}{\psi_i^C}-t_\theta^C\right)^2}$$

$\psi_i^L=\frac{\pi_i^L}{m^L}=\frac{1}{M^L}\cdot\frac{k_i^L}{z_i}$ for $i\in U^L$ and $\psi_i^C=\frac{\pi_i^C}{m^C}=\frac{k_i^C}{M^C}$ for $i\in U^C$

$$t_z^L=\sum_{i\in U^L}\theta_i^L\ ;\ t_z^C=\sum_{i\in U^C}\theta_i^C\ .$$

Thus $\frac{m^L}{m^C}=\sqrt{\gamma\cdot\frac{p^C}{p^L}}$.

4 Dual frame application

Project outline

To shed light on the problem of how to conduct cell phone surveys we got funding for a research project from the German Research Foundation. The data in the following section were collected in the framework of the CELLA project (*CEL*l phone and *LA*ndline phone surveys). Cooperation partners in CELLA are the Technical University of Dresden, and GESIS – Leibniz Institute for the Social Sciences.

The field period lasted from October 2007 to April 2008. The rather long period aimed at making sure that, where necessary, numbers were contacted up to 15 times and in different weeks in order to determine whether the number was a working or a nonworking number. Altogether 2,171 interviews were conducted by the telephone lab of the Technical University of Dresden, 1,099 via cell phone and 1,072 via landline phone.

For the selection of landline phone numbers we used the in Germany well known "Gabler-Haeder-Design", which is a list assisted Random Digit Dialling (see Gabler & Haeder 2002). The landline phone frame consisted of M^L =125,314,800 numbers. The frame for the selection of cell phone numbers was developed by BIK- ASCHPURWIS+BEHRENS GmbH, Germany. It contains all theoretically possible numbers of the different providers. At the time of sampling for CELLA these were M^C =178,050,000 numbers.

In Germany, as in many other European countries, cell phone numbers are issued nationwide with their specific suffix. For this reason, unlike the U.S., there is no way to infer an approximate location of the respondent based on their cell phone numbers, thus preventing local or geographic specific cell phone surveys (Callegaro & Poggio, 2004).

Disposition codes from the survey are shown in Tables 3 and 4:

Table 3: Disposition codes for the landline phone survey

Not eligible	
Fax line	667
Non-working number: computer voice	7947
Non-working number: non-contact	2405
Nonresidence	704
Other	145
Unknown eligibility, non-interview	
Voice mail – don't know if household	1135
Eligible, non interview	3811
Refusal	2327
Break off	17
Physically or mentally unable/incompetent	14
Other	377
Interview	
Complete	1009
Partial	67
Total phone numbers used	16814

We selected about 17,000 phone numbers for the gross sample. Nearly a half of them were clearly non-working numbers which is only a problem if you do not have access to an automatic dialler. Our final AAPOR (2009) response rate 3 (RR3) was 24.7% using the proportional allocation method to estimate *e* (Smith, 2009).

In Table 4 the disposition codes for the cell phone survey are shown. The sample was split in two: half of the sample was randomly assigned to receive an SMS (Short Message Service) with a prenotification of the survey. The other half of the sample received no prenotification of any type.

Table 4: Disposition codes of the cell phone survey.

	No SMS	Advanced SMS
Not eligible		
Fax line	23	40
Non-working number: computer voice	5,680	7,905
Non-working number: non-contact	716	994
Nonresidence	44	112
Other	150	174
Unknown eligibility, non-interview		
Voice mail – don't know if eligible	2,140	3,010
Eligible, non interview	**1,232**	**1,825**
Refusal	717	979
Break off	39	14
Physically or mentally unable/incompetent	7	10
Other	36	47
Interview		
Complete	415	747
Partial	18	28
Total phone numbers used	9,985	14,060

Tables 3 and 4 show that a number of challenges were met. The sampling frame currently contains about 180 million numbers - and, obviously, many of them are nonworking. Given that there are no official lists with working numbers, all theoretically potentially working numbers need to be generated. Voice mails are problematic because there is oftentimes no definite hint of whether the phone is used for personal calls or not. From our point of view, it seems very likely that they are nonworking numbers given the long survey period.

A second computation of AAPOR RR3 was obtained, where *e* was modified using an estimation that 90% of the voice mail messages of unknown eligibility were in fact not eligible cases. This yields the following AAPOR response rates:

RR3 no SMS sent (standard *e*) = 26.5%
RR3 advanced SMS (standard *e*) = 33.0%

RR3 no SMS sent (modified *e*) = 32.2%
RR3 advanced SMS (modified *e*) = 40.1%

The advanced SMS had a positive effect on response rates. In addition, response rates for the cell phone survey was higher than the landline survey, especially with the SMS notification.

The efficiency of the frames is still low. Only about 5% of the dialled numbers of the gross sample lead to an interview.

Calculation of the weights

As explained in section 2, the inclusion probability of a person *i* can be calculated as follows:

$$\pi_i \approx \pi_i^L + \pi_i^C \approx k_i^L \frac{m^L}{M^L} \cdot \frac{1}{z_i} + k_i^C \frac{m^C}{M^C}.$$

Now we will describe how we fixed the parameters of the model.

It has already been mentioned that M^F is the quantity of landline phone numbers in the frame. In our case M^L = 125,314,800. m^L is the quantity of landline phone numbers in the sample (the gross sample size), here m^L = 16,815.

M^C denotes the quantity of cell phone numbers in the frame, and in our case M^C = 178,050,000. m^C is the quantity of cell phone numbers in the sample, and in our case m^C = 24,045.

A decision must be made regarding the measurement of k_i^C, the quantity of cell phone numbers allowing access to person *i*. Firstly, we can assume that cell phones are used predominantly as personal equipment. That is of interest because otherwise we would have to select one person out of those persons who usually use the cell phone together (see AAPOR 2010, p. 66). But only a minority of cell phone users in Germany shares them with other persons, a proportion we can neglect if we consider that cell phone use is likely to become even more common in the future.

Table 5: Cell phone – Sharing with others?

	Landline sample		Cell phone sample	
	n	%	*n*	%
Exclusive use by myself	790	91.1	1051	90.6
Other people use it from time to time	57	6.0	96	8.3
Share it with others	20	2.3	13	1.1
Total	867	100.0	1160	100.0

In our survey we asked on how many cell phone numbers the target person is reachable on. The answers are shown in Table 6.

Table 6: Quantity of mobile phone numbers k_i^C

	Landline sample		Cell phone sample	
	N	%	*n*	%
One	789	90.90	948	81.6
Two	63	7.26	167	14.4
Three	11	1.27	31	2.7
Four	2	0.23	4	0.3
Five and more	1	0.12	7	0.6
Don't know/ no answer	2	0.23	5	0.5
Total	868	100	1162	100

We use these values as variable k_i^C in the dual frame model. Furthermore, k_i^L, the quantity of landline phone numbers - allowing access to the household to which person *i* belongs – has to be fixed. The difficulty is that people often do not know exactly how many landline phone numbers they have (Meier, 2007). As stated earlier in this paper here we have found a rule of thumb which is reasonable for German telephone equipment: We fix that an analog telephone connection has one working number while a digital connection has 2.5 working numbers.

Because for the landline sample the respondent was selected with the modified birthday method, the inclusion probabilities for the household members vary. They depend on z_i, the size of the household (belonging to the target population) to which person *i* belongs. This information has to be asked in the interview.

With these model parameters we computed the weights for the combination of the landline and the cell phone survey – for interviewees selected from the landline phone frame and interviewees selected from the cell phone frame. We found the following distribution of the weights (see figure 1a and 1b). Obviously, larger weights occur only for persons selected from the landline phone frame when they life in households where many members belong to the target population.

Figure 1a: Histogram of the weights for the Landline phone respondents

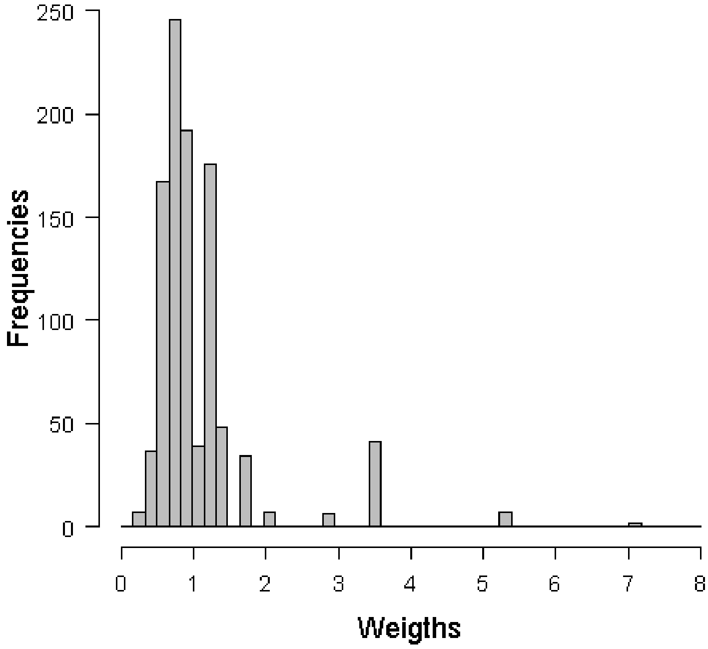
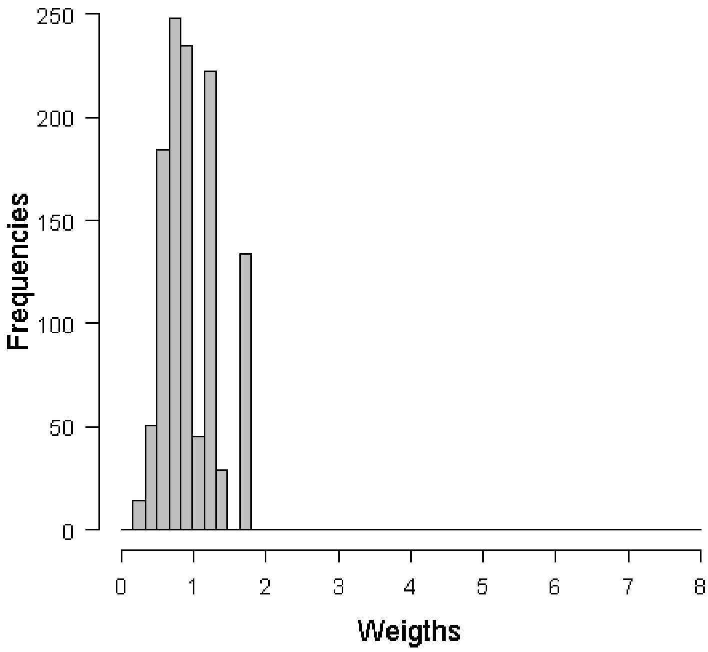


Figure 1b: Histogram of the weights for the Cell phone respondents



Application

In this paper we want to examine how the application of the above described weights changes the distributions of survey variables such as

- impact on mobile phone nets on quality of live
- interest in development of mobile phone nets
- having mobile-only friends.[3]

This way we want to further evaluate our approach for the combination of landline and cell phone samples.

In our survey persons contacted via landline phone as well as persons contacted via cell phone were asked: How do you evaluate the impact of the development of cell phones on the quality of life? In Table 7 it can be seen that distributions of landline and cell phone populations differ considerably. The cell phone community assesses the effect of the further development of cell phone nets more positive than the landline population. As it was to expect the combination of both samples leads to an estimation that evens out the extreme opinions.

Table 7: Impact of the development of cell phones on quality of life[4]

	Landline		Cell phone		Combined	
1 positively	112	11.8	146	15.9	273	14.7
2	267	28.2	276	30.1	498	26.8
3	331	35.0	308	33.6	658	35.4
4	150	15.9	114	12.4	273	14.7
5 negatively	86	9.1	73	7.9	155	8.4
Total	946	100.0	916	100.0	1858	100

The same trend can be observed for the indicator whether or not someone is interested in the development of cell phones (see Table 8). Clearly, the cell phone population is much more interested in this topic. Again, the combined population estimator lies in the middle of both subpopulations.

3 Graeske and Kunz (2009) have already shown that the weighted sample distributions of demographic variables such as age, gender and household size are more similar to the population distributions than the unweighted sample distributions.

4 Only valid answers

Table 8: Interest in the development of cell phones[5]

	Landline		Cell phone		Combined	
1 very interested	53	5.4	96	10.3	144	7.5
2	130	13.2	157	16.9	279	14.5
3	274	27.9	277	29.6	542	28.3
4	228	23.1	189	20.3	406	21.2
5 not at all interested	298	30.3	214	22.9	548	28.6
Total	984	100.0	934	100.0	1918	100.0

The same is true for the question if the interviewee has relatives or friends who are CPOs (Table 9). The landline phone population obviously knows less people who are not reachable via landline phones.

Table 9: Relatives, friends – cell phone-onlys[6]

	Landline		Cell phone		Combined	
Yes	361	36.8	534	47.4	877	41.9
No	619	63.2	593	52.6	1216	58.1
Total	979	100.0	1126	100.0	2092	100.0

Altogether it can be seen that at least for variables which deal with cell phone matters it is absolutely necessary to conduct interviews from cell phone and landline phone users and to combine these samples. If CPOs are missing in the survey – as it is still frequently the case in Germany nowadays - estimates are considerably biased due to undercoverage. This is of course even more true for countries with a higher proportion of CPOs, such as south Eastern European states.

Moreover, it seems to be the case that people who are interviewed on a cell phone are more open-minded for questions dealing with cell phone related attitudes. This result supports our demand to include people interviewed on cell phones in social surveys - even if this is more money consuming.

5 Only valid answers

6 Only valid answers

5 Conclusions

The goal of this study was to show that the dual frame approach for combining landline and cell phone samples is practical and applicable - at least in Germany. In the CELLA1 project, two parallel surveys were conducted, one via landline phones and one via cell phones. The sample sizes were about 1,000 in each survey, to secure enough statistical power among cell phone only households. However, under cost restrictions and with a different scope of the survey a different relation of the sample sizes may be more efficient. A formula for the optimal allocation is given in Section 3.

The inclusion probabilities could be computed in a relatively easy way: The sampling fractions need to be known as well as certain information on the respondents (number of landline and cell phone numbers, number of eligible individuals in the household) had to be gathered in the interviews.

The theoretical foundations for the dual frame approach and the computation of the weights are given in sections 2 and 3.

The variation of the weights for CELLA1 is not too large. The design effect due to differing inclusion probabilities (Ganninger, 2010; Kish, 1965) is 1.3, i.e. 1,300 respondents were needed to get the same precision of an estimate as in a simple random sample of size 1,000.

For different variables of interest it could be shown that weighting matters in this respect and that weighted estimators seem to be more meaningful. In so far cell phone-only households should not be left out in further telephone surveys - otherwise estimators could be seriously biased - especially in countries where the proportion of cell phone only households is considerable.

A surprising aspect was that the response rate for the cell phone survey was higher than for the landline survey, especially with the SMS notification. This finding is in contrast to the American results from the AAPOR task force report (2010) were it is stated that non response is greater or at best the same in cell surveys. In order to further validate this effect two subsequent surveys will be conducted by the authors at GESIS - again one via landline and one via cell phone, with sample sizes 1,500 each. The fieldwork will start in July 2010 and the scope of the study is again response behavior (with psychological background variables), weighting and mode effects.

6 References

AAPOR Cell Phone Task Force (2010) New Considerations for Survey Researchers when Planning and Conducting RDD Telephone Surveys in the U.S. With Respondents Reached via Cell Phone Numbers.

AAPOR Cell Phone Task Force (2009) American Association for Public Opinion Research. *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (6th ed.). Deerfield, IL: AAPOR.

Bankier, M. D. (1986) Estimators based on several stratified samples with applications to multiple frame surveys. *Journal of the American Statistical Association*, **81**, 1074-1079.

Blumberg, S. J., & Luke, J. V. (2009, May 2010). Wireless substitution: Early release of estimates from the National Health Interview Survey, January - June 2009. (Available from http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless200912.pdf)

Brick, M. J., Dipko, S., Presser, S., Tucker, C., & Yuan, Y. (2006) Nonresponse bias in dual frame sample of cell and landline numbers. *Public Opinion Quarterly*, **70**, 780-793.

Brick, M. J., Brick, P. D., Dipko, S., Presser, S., Tucker, C., & Yuan, Y. (2007) Cell phone survey feasibility in the U.S.: Sampling and calling cell numbers versus landline numbers. *Public Opinion Quarterly*, **71**, 23-39.

Callegaro, M., & Poggio, T. (2004) Espansione della telefonia mobile ed errore di copertura nelle inchieste telefoniche [Mobile telephone growth and coverage error in telephone surveys- English version at: http://eprints.biblio.unitn.it/archive/00000680]. *Polis*, **18**, 477-506.

Cochran, R. S. (1967) The estimation of domain sizes when sampling frames are interlocking. In American Statistical Association (ed.), *Proceedings of the Social Statistics Section* (pp. 332-335). Washington DC: AMSTAT.

Delnevo, C. D., Gundersen, D. A., & Hagman, B. T. (2008) Declining estimated prevalence of alcohol drinking and smoking among young adults nationally: Artifacts of sample undercoverage? *American Journal of Epidemiology*, **167**, 15-19.

Dutwin, D., Kennedy, C., Keeter, S., & Kulp, D. (2008) Dual frame (landline and cell RDD) estimation in a National Survey of Latinos. 63rd Annual conference of the American Association for Public Opinion Research. New Orleans, LA.

Fuller, W. A., & Burmeister, L. F. (1972) Estimators for samples selected from two overlapping frames. In American Statistical Association (ed.), *Proceedings of the Social Statistics Section* (pp. 245-249). Washington DC: AMSTAT.

Gabler, S., & Ayhan, Ö. (2007) Gewichtung bei Erhebungen in Festnetz und über Mobilfunk: Ein Dual Frame Ansatz. In S. Gabler & S. Haeder (eds.), *Mobilfunktelefonie - Eine Herausforderung für die Umfrageforschung* (pp. 39-45). Mannheim: GESIS-ZUMA.

Gabler, S., & Haeder, S. (2002) Idiosyncrasies in telephone sampling–The case of Germany. *International Journal of Public Opinion Research*, **14**, 338-345.

Ganninger, M. (2010) *Design effects: Model based versus design based approach. Thesis*. Mannheim: GESIS- Leibniz-Institute for the Social Sciences.

Goodman, L. A. (1949) On the estimation of number of classes in a population. *The Annals of Mathematical Statistics*, **20**, 572-579.

Goodman, L. A. (1952) On the analysis of samples from *k* lists. *The Annals of Mathematical Statistics*, **23**, 632-634.

Graeske, J., & Kunz, T. (2009) Stichprobenqualität der CELLA-Studie unter besonderer Berücksichtigung der Mobile-onlys. In M. Haeder & S. Haeder (eds.), *Telefonbefragungen über das Mobilfunknetz: Konzept, Design und Umsetzung einer Strategie zur Datenerhebung* (pp. 57-70). Wiesbaden: VS Verlag.

Haeder, M., & Haeder, S. (eds). (2009) *Telefonbefragungen über das Mobilfunknetz: Konzept, Design und Umsetzung einer Strategie zur Datenerhebung*. Wiesbaden: VS Verlag.

Hartley, H. O. (1962) Multiple frame surveys. In American Statistical Association (ed.), *Proceedings of the Social Statistics Section* (pp. 203-206). Washington DC: AMSTAT.

Hartley, H. O. (1974) Multiple frame methodology and selected applications. *Sankhya Series C*, **36**, 99-118.

International Telecommunication Union. (2010) *Measuring the information society Ver 1.01*. Geneve: ITU.

Keeter, S. (2008, June 2008). Research roundup: Latest findings on cell phones and polling. (Available from http://pewresearch.org/pubs/848/cell-only-methodology)

Keeter, S., Dimock, P., Kennedy, C., Best, J., & Horrigan, J. (2008, May 15-18) *Cost and benefits of full dual frame sample telephone survey design.* Paper presented at the 63rd Annual conference of the American Association for Public Opinion Research, New Orleans, LA.

Keeter, S., Kennedy, C., Clark, A., Tompson, T., & Mokrzyck, M. (2007) What's missing from national landline RDD surveys? The impact of the growing cell-only population. *Public Opinion Quarterly*, **71**, 772-792.

Kish, L. (1965) *Survey Sampling*. New York: Wiley.

Kuusela, V., & Simpanen, M. (2002, August, 25-28) *Effects of mobile phones on phone survey. Practices and results.* Paper presented at the International Conference on Improving Surveys, Copenhagen, Denmark.

Lavrakas, P. J., Shuttles, C. D., Steeh, C., & Fienberg, H. (2007) The state of surveying cell phone numbers in the United States. *Public Opinion Quarterly*, **71**, 840-854.

Link, M. W., Battaglia, M. P., Frankel, M. R., Osborn, L., & Mokdad, A. H. (2007) Reaching the U.S. cell phone generation: Comparison of cell phone survey results with an ongoing landline telephone survey. *Public Opinion Quarterly*, **71**, 814-839.

Lohr, S. (2009) Multiple-frame surveys. In D. Pfeffermann & C. R. Rao (eds.), *Handbook of statistics 29A Sample surveys: Design, methods and applications* (Vol. 29A). Amsterdam: Elsevier.

Lohr, S., & Rao, J. N. K. (2000) Inference from dual frame surveys. *Journal of the American Statistical Association,* **95**, 271-280.

Lohr, S., & Rao, J. N. K. (2006) Estimation in multiple frame surveys. *Journal of the American Statistical Association,* **101**, 1019-1030.

Lund, R. E. (1968) Estimators in multiple frame surveys. In American Statistical Association (ed.), *Proceedings of the Social Statistics Section* (pp. 282-288). Washington DC: AMSTAT.

Meier, G. (2007) Validierung eines Fragebogens zur Erfassung der Anzahl von Telefonnummern. In S. Gabler & S. Haeder (eds.), *Mobilfunktelefonie - Eine Herausforderung für die Umfrageforschung* (pp. 91-104). Mannheim: GESIS-ZUMA.

Mokrzycki, M., Keeter, S., & Kennedy, C. (2009) Cell-Phone-Only voters in the 2008 Exit Polls and implications for future noncoverage bias. *Public Opinion Quarterly,* **73**, 845-865.

Pew Research Center (2008) Ways of coping with a growing population segment. The impact of "Cell-Onlys" on public opinion polling. (Available from http://people-press.org/reports/pdf/391.pdf)

Skinner, C. J. (1991) On the efficiency of raking ratio estimation for multiple frame surveys. *Journal of the American Statistical Association,* **86**, 779-784.

Skinner, C. J., & Rao, J. N. K. (1996) Estimation in dual frame surveys with complex design. *Journal of the American Statistical Association,* **91**, 349-356.

Smith, T. W. (2009) A revised review of methods to estimate the status of cases with unknown eligibility.

Statistics Canada. (2009, December 2008). Residential Telephone Service Survey. (Available from http://www.statcan.gc.ca/daily-quotidien/090615/dq090615c-eng.htm)

Stephan, F. F. (1936) Practical problems of sampling procedure. *American Sociological Review,* **1**, 569-580.

Taylor, S. (2003) Telephone surveying for household social surveys: the good, the bad, and the ugly. *Survey Methodology Bulletin,* ***52***, 10-21.

Tucker, C. & Lepkowski, J. M. (2008). Telephone survey methods: Adapting to change. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. de Leeuw, L. Japec et al. (eds). *Advances in telephone survey methodology*. pp. 3-26. Hoboken, NJ.: Wiley.

TNS Opinion & Social. (2008) *E-Communication household survey. Special Eurobarometer 293 / Wave 68.2*. Brussels: European Commission.

Witt, E., Best, J., & Rainie, L. (2008) Internet access and use: Does cell phone interviewing make a difference? 63rd Annual conference of the American Association for Public Opinion Research. New Orleans, LA.

Appendix

Proof. Since the samples from landline frame and cell phone frame are independent we have

$$\begin{aligned}\operatorname{var}\left(\hat{t}_y\right) &= \operatorname{var}\left(\sum_{i\in s^{F\cap\bar{C}}} \frac{y_i}{\pi_i^F} + \sum_{i\in s^{F\cap C}} \frac{\lambda_i y_i}{\pi_i^F}\right) + \operatorname{var}\left(\sum_{i\in s^{F\cap C}} \frac{(1-\lambda_i) y_i}{\pi_i^C} + \sum_{i\in s^{\bar{F}\cap C}} \frac{y_i}{\pi_i^C}\right)\\ &= \operatorname{var}\left(\sum_{i\in s^F} \frac{\theta_i^F}{\pi_i^F}\right) + \operatorname{var}\left(\sum_{i\in s^C} \frac{\theta_i^C}{\pi_i^C}\right)\end{aligned}$$

As approximation we replace $\operatorname{var}\left(\sum_{i\in s^F} \frac{\theta_i^F}{\pi_i^F}\right)$ and $\operatorname{var}\left(\sum_{i\in s^C} \frac{\theta_i^C}{\pi_i^C}\right)$ by the corresponding formulae for sampling with replacement with selection probabilities $\psi_i^F = \frac{\pi_i^F}{m^F} = \frac{1}{M^F}\cdot\frac{k_i^F}{z_i}$ for $i\in U^F$ and $\psi_i^C = \frac{\pi_i^C}{m^C} = \frac{k_i^C}{M^C}$ for $i\in U^C$, respectively.

Then,

$$\begin{aligned}\operatorname{var}\left(\hat{t}_y\right) &\approx \operatorname{var}_{\psi^F}\left(\frac{1}{m^F}\sum_{i\in s^F} \frac{\theta_i^F}{\psi_i^F}\right) + \operatorname{var}_{\psi^C}\left(\frac{1}{m^C}\sum_{i\in s^C} \frac{\theta_i^C}{\psi_i^C}\right)\\ &= \frac{1}{m^F}\sum_{i\in U^F} \psi_i^F\left(\frac{\theta_i^F}{\psi_i^F} - t_\theta^F\right)^2 + \frac{1}{m^C}\sum_{i\in U^F} \psi_i^C\left(\frac{\theta_i^C}{\psi_i^C} - t_\theta^C\right)^2\\ &= \frac{1}{m^F}V_{\psi^F}^F + \frac{1}{m^C}V_{\psi^C}^C\end{aligned}$$

and we deduce

$$\begin{aligned}&\frac{\partial\left(\operatorname{var}\left(\hat{t}_y\right) - \lambda\left(m^F p^F + m^C p^C\right)\right)}{\partial m^F} = 0\\ &\frac{\partial\left(\operatorname{var}\left(\hat{t}_y\right) - \lambda\left(m^F p^F + m^C p^C\right)\right)}{\partial m^C} = 0\\ &m^F p^F + m^C p^C = g\end{aligned}$$

i.e.

$$-\frac{V_{\psi^F}^F}{m^{2F}}+\lambda p^F=0$$
$$-\frac{V_{\psi^C}^C}{m^{2C}}+\lambda p^C=0$$

or

$$\sqrt{\frac{1}{\lambda}}\sqrt{\frac{V_{\psi^F}^F}{p^F}}=m^F$$
$$\sqrt{\frac{1}{\lambda}}\sqrt{\frac{V_{\psi^C}^C}{p^C}}=m^C$$

From

$$m^F p^F+m^C p^C=g$$

follows

$$\sqrt{\frac{1}{\lambda}}\left(\sqrt{V_{\psi^F}^F p^F}+\sqrt{V_{\psi^C}^C p^C}\right)=m^F p^F+m^C p^C=g$$

i.e.

$$\sqrt{\frac{1}{\lambda}}=\frac{g}{\sqrt{V_{\psi^F}^F p^F}+\sqrt{V_{\psi^C}^C p^C}}$$

which implies

$$m^F=g\cdot\frac{\sqrt{\frac{\gamma}{p^F}}}{\sqrt{\gamma p^F}+\sqrt{p^C}}=g\cdot\frac{\sqrt{\gamma}}{\sqrt{\gamma}p^F+\sqrt{p^F p^C}}$$
$$m^C=g\cdot\frac{\sqrt{\frac{1}{p^C}}}{\sqrt{\gamma p^F}+\sqrt{p^C}}=g\cdot\frac{1}{\sqrt{\gamma p^F p^C}+p^C}$$

and

$$\frac{m^F}{m^C} = \sqrt{\gamma \cdot \frac{p^C}{p^F}} \quad .$$

Example. For $\gamma = 1, p^C = 4p^F$ we get $m^F = \frac{g}{3p^F}$ and $m^C = \frac{g}{6p^F}$. $m^F = 2m^C$, i.e. the sample size in the landline phone part of the survey is double to the cell phone part.

How to estimate γ from earlier surveys?

From

$$\sum_{i \in U^F} \psi_i^F \left(\frac{\theta_i^F}{\psi_i^F} - t_\theta^F \right)^2 \approx E^F \left[\frac{1}{n^F - 1} \sum_{i \in s^F} \left(\frac{\theta_i}{\psi_i} - \frac{1}{n^F} \sum_{j \in s^F} \frac{\theta_j}{\psi_j} \right)^2 \right]$$
$$\sum_{i \in U^F} \psi_i^C \left(\frac{\theta_i^C}{\psi_i^C} - t_\theta^C \right)^2 \approx E^C \left[\frac{1}{n^C - 1} \sum_{i \in s^C} \left(\frac{\theta_i}{\psi_i} - \frac{1}{n^C} \sum_{j \in s^C} \frac{\theta_j}{\psi_j} \right)^2 \right]$$

follows

$$\hat{\gamma} = \frac{s_x^{2F}}{s_x^{2C}} \quad .$$

For real data non-response adjusted quantities, $x_i^F = \frac{\theta_i^F}{\tilde{\psi}_i^F}$ and $x_i^C = \frac{\theta_i^C}{\tilde{\psi}_i^C}$ should be used.