Pre-Call Validation of RDD Cell Phone Numbers
A Field Experiment

Tanja Kunz[1], Marek Fuchs[1]
[1] Institute of Sociology, Darmstadt University of Technology
Residenzschloss, 64283 Darmstadt (Germany)

Abstract

The field experiment (n = 24,999) reported in this paper was designed to decrease survey costs and the interviewers' workload by screening out technically invalid numbers and numbers of unknown eligibility prior to fieldwork of a telephone survey in the cell phone frame. In addition, effects of pre-call validation on data quality were examined. Two methods were tested: number validation and text messaging. Furthermore, several screening conditions of different strictness were examined. Results indicate that both number validation and text messaging are effective methods to increase the percentage of working numbers in the field. Contact and interview rates can also be increased. A reduction of the proportion of numbers of unknown eligibility due to pre-call validation results in an increase of response rates. Altogether, high percentages of screened out numbers achieve considerable cost savings. However, pre-call validation of cell phone numbers comes at the risk of screening out valid cell phone numbers which potentially causes biases.

Key Words: cell phone survey, RDD, response rates, survey cost, HLR lookup, text messages

1. Introduction and background

The increase of cell phone penetration rates and the rising share of cell phone-onlys who can exclusively be contacted by cell phone pose new challenges to survey researchers. In Germany, the proportion of households with cell phones has reached 86 percent and the percentage of cell phone-onlys amounts to 8 percent according to Eurobarometer data 2011 (European Commission, 2011). Regarding other European countries, the share of households with cell phones is 91 percent on average, and the proportion of people who rely solely on cell phones ranges from 2 percent in Sweden up to 75 percent in Finland or even 80 percent in the Czech Republic. The EU-27 average of cell phone-only households is 34 percent (European Commission, 2011). The upward tendency concerning the share of cell phone-onlys is not restricted to the EU. In the United States for instance, wireless substitution in households is about 30 percent at present and further increasing (Blumberg & Luke, 2011). Since cell phone-onlys are significantly different with respect to several sociodemographic characteristics from those reachable on a landline phone, their rising share may pose a problem to telephone surveys. They are typically younger, predominantly male, less educated, and less affluent (Blumberg & Luke, 2011; Graeske & Kunz, 2009; Hu, Balluz, Battaglia, & Frankel, 2011; Peytchev, Carley-Baxter, & Black, 2010; Zuwallack, 2009). As long as the percentage of the cell phone-only population in the

total population is relatively small, the potential coverage bias in landline surveys can be ignored. However, for certain subgroups with larger proportions of cell phone-onlys the potential risk of a coverage bias is considerably higher (Blumberg & Luke, 2009; Busse & Fuchs, 2011; Peytchev et al., 2010). Therefore, the increasing percentage and distinctive demographics of cell phone-onlys led to the application of dual-frame samples, i.e., a combination of cell phone samples with conventional landline surveys (Gabler & Ayhan, 2007; Hu et al., 2011; Keeter, Christian, & Dimock, 2010; Kennedy, 2007).

Previous studies on cell phone number usage in telephone surveys discussed the importance of integrating cell phone numbers into telephone samples, and the challenges associated (AAPOR, 2008, 2010; Brick, Edwards, & Lee, 2007; M. Häder & Häder, 2009; Lavrakas, Shuttles, Steeh, & Fienberg, 2007; Zuwallack, 2009). By contrast, little research has been conducted to improve the efficiency of cell phone samples and to reduce the high proportion of technically invalid numbers. Because cell phone numbers are mostly not listed in Germany as well as in most western countries, the usage of some kind of random digit dial procedure is the only way of conducting cell phone surveys. In Germany, a well established sampling method is the Gabler-Häder design (Gabler & Häder, 2009) for creating a sampling frame by generating cell phone numbers at random. However, despite continuous optimization of the sampling frame each random sample from this frame still comprises high proportions of nonworking numbers and numbers of unknown eligibility (S. Häder, Häder, Graeske, Kunz, & Schneiderat, 2009). Dialing technically invalid cell phone numbers results in additional interviewer time, in decreasing interviewer motivation, and in increasing survey costs (AAPOR, 2010; Buskirk, Callegaro, & Rao, 2009; Keeter, Dimock, Kennedy, Best, & Horrigan, 2008). In addition, the high proportion of numbers of unknown eligibility mainly due to ambiguous voicemail and operator messages increases the number of call attempts needed to gain certainty about the status of such cases (AAPOR, 2010; Callegaro et al., 2007). Both factors result in an inefficient data collection and extended fieldwork period, and ultimately increase total survey costs. Also, if the status of cell phone numbers remains uncertain despite the maximum number of call attempts, this will result in an unreliable calculation of response rates. The use of predictive dialers could potentially solve the problem of high proportions of invalid numbers by recognizing automatically live persons, no answers or busy signals, and voicemail or operator messages. Technically invalid numbers could be screened out easily without the need for dialing. However, in the U.S. calling cell phone numbers via an automated or predictive dialer without prior consent of the called cell phone subscriber is prohibited by the FCC (AAPOR, 2010). In several European countries, the use of predictive dialing in survey research is permitted in accordance with certain limitations (ADM, 2008; OFCOM, 2010), however, the efficiency of the application for cell phone samples remains doubtful. In particular, high proportions of heterogeneous operator messages in cell phone samples (AAPOR, 2008; Callegaro et al., 2007; S. Häder et al., 2009) make it difficult for automatic answer detection technologies to reliably distinguish technically valid and invalid cell phone numbers.

The integration of cell phone numbers in telephone samples still involves high costs and effort what restrain not least some researchers from conducting cell phone interviews. Therefore, the present experiment was designed to develop and test a standardized approach which can be used to validate random cell phone numbers prior to fieldwork. Thereby, cost and time efforts should be decreased without affecting data quality negatively. We explored two options of pre-call validation of cell phone numbers: simple number validation and validation via detailed delivery reports received after sending text messages. *Cell phone number validation services* provide verification in real time by performing a Home Location Register (HLR) lookup resulting in a validation code which indicates the current status of a cell phone number. This provides information on whether a given cell phone number is "active and currently available", "invalid" or "not logged in for a longer time". The method is inexpensive ($0.05 per check) and fast, no message has to be sent to the subscriber and most queries are returned within a few seconds. So, multiple checks can be done in a relatively short time at moderate costs. However, this procedure provides only a snapshot in time and thus, its reliability has to be questioned. *Text messaging services* can also be applied for pre-call validation by sending bulk messages to cell phones. Though being more expensive ($0.12 per text message), sending text messages has at first glance multiple advantages compared to a simple number validation. Sending text messages offers additional information in terms of detailed delivery status codes, and reason codes in case of failed delivery. If delivery failed, the distinction between "subscriber is unknown" or "subscriber is temporarily not available" can be used to decide on whether a number should be fielded or not. Also, it is not only a snapshot in time because delivery-attempts continue up to 48 hours before a status code is generated. In addition, sending text messages can be used both for number verification and pre-notification of the survey request. In compliance with certain requirements, sending bulk text messages is permitted in Europe as far as it is exclusively applied for research purposes (ESOMAR, 2010). Previous studies considering the effectiveness of advance text messages in cell phone surveys reported mixed results concerning an improvement of the respondents' participation. Brick and colleagues (2007) found no effect on an increase of response rates by sending text messages. Steeh, Buskirk, and Callegaro (2007) showed that using text messages for pre-notification tended to have a positive effect on several measures of survey participation like cooperation and response rates. However, these differences were statistically not significant. In a German study, positive effects of pre-notification on response rates could be shown (S. Häder & Schneiderat, 2009). Thus, the effectiveness of sending text messages in improving cooperation rate can be a positive side effect.

2. Research Questions

The study had two main objectives concerning survey costs and data quality. First, survey costs and effort should be decreased by means of pre-call validation. Dialing invalid numbers is extremely time-consuming, inefficient, and decreases interviewer motivation. By screening out technically invalid numbers using an appropriate pre-call validation method prior to data collection, contact rates and interview rates should be increased. Furthermore, with an increase of the percentage of working cell phone numbers in the field, the interviewers' workload in terms of mean number of call attempts and mean overall call duration needed for completing an interview should be reduced. In total, overall survey costs should be decreased. Second, the reduction of survey costs should not be at the expense of reduced data quality. Quite the opposite, excluding technically invalid numbers by means of pre-call validation should reduce

the proportion of numbers of unknown eligibility. Hence, response rates should be calculated more reliably because the number of cases with unambiguous final disposition codes will increase. However, the reliability of the respective validation method had to be examined carefully. The increase of the proportion of valid cell phone numbers should not be achieved at the expense of a large proportion of screened out numbers that are in fact working numbers (false negative numbers). A high percentage of false negative numbers has the potential to induce a bias in estimates in case certain groups of respondents are prone to higher risks of wrongful exclusion.

In detail, the present paper aimed at answering the following questions: Relating to survey costs we examined (1) whether contact and interview rates can be increased, and (2) whether mean number of call attempts and mean overall call duration needed for a completed interview can be reduced by means of pre-call validation. Regarding data quality we determined (3) whether response rates can be calculated more reliably by decreasing the share of numbers of unknown eligibility, and (4) how the proportion of false negative numbers defined as numbers which are falsely screened out although being valid cell phone numbers is affected by various screening methods.

3. Method

In order to identify an appropriate method for pre-call validation of cell phone numbers, a field-experimental study was designed. Within the scope of the "Experimental Mobile Phone Panel" in Germany, recruitment interviews were conducted from September 2010 to January 2011. The sample of 24,999 cell phone numbers was based on the Gabler-Häder design (Gabler & Häder, 2009). In contrast to classical random digit dialing (RDD), this method builds a sampling frame of randomly generated cell phone numbers. According to the Gabler-Häder design, all possible banks of 100,000 numbers are assessed for every cell phone carrier access code. Within number banks with at least one listed cell phone number (according to directories and internet research), all theoretically possible numbers are generated. From this frame of randomly generated numbers, a systematic random sample is drawn. This sampling method enables equal inclusion probabilities but is at the same time more efficient than RDD. Due to a differential and in some banks quite low density of assigned numbers, classical RDD is very inefficient in Germany. Although the Gabler-Häder design is more efficient than classical RDD, the key problem remains: the lower the density of working numbers within a bank, the higher the share of invalid numbers in the sample.

In our experiment we started with an initial sample of 24,999 cell phone numbers. All numbers were randomly assigned to either an experimental group (EG) or a control group (CG) in a between-subjects design. We differentiated checking by number validation, checking by sending text messages or by applying a combination of both methods. In a pilot study, we examined three different providers for text messaging and four providers for number validation. We selected one provider each (1) on the basis of the price, (2) the detailedness of the delivery reports or reason codes in case of delivery failure, (3) on the basis of the reliability of the provider, and (4) the usability of the provider's service.

In total, six experimental groups with distinct screening conditions were distinguished (see Table 1). A cell phone number was screened out,

- if the number was unknown by a one-time number validation check (EG 1),
- if the number was unknown by a two-time number validation check with a time lag of four days between first and second check (EG 2),
- if text message delivery failed due to an unknown subscriber (EG 3a),
- if text message delivery failed due to an unknown *or* temporarily absent subscriber (EG 3b),
- if text message delivery failed due to an unknown subscriber *and* if the number was unknown by a one-time number validation check (EG 4a),
- if text message delivery failed due to an unknown *or* temporarily absent subscriber *and* the number was unknown *or* temporarily not logged in by a one-time number validation check (EG 4b).

Pre-call validation information by means of a single number validation check in EG 1 was limited to a single point in time. This apparent disadvantage regarding reliability should be tested in EG 2 by conducting a second check four days after the first check. Thus, a cell phone number was only excluded if the number was unknown according to the first and second check. By combining number validation and text messaging in EG 4a and EG 4b, it was examined whether additional cost and effort were justified with respect to an improved efficiency of a combined pre-call validation process compared to applying only one of the two methods.

Table 1: Description of the Experimental Design

	validation method	screening condition	initial sample (n)	screened out (%)	final sample (n)
CG 0	-	-	6,200	-	6,200
EG 1	one-time check	number unknown	2,250	42	1,295
CG 1	one-time check	-	1,025	-	1,025
EG 2	two-time check	number unknown	2,250	42	1,307
CG 2	two-time check	-	1,025	-	1,025
EG 3a	text messaging	subscriber unknown	2,500	50	1,238
EG 3b	text messaging	subscriber unknown *or* absent	2,500	70	753
CG 3	text messaging	-	1,125	-	1,125
EG 4a	one-time check & text messaging	number unknown *and* subscriber unknown	2,500	42	1,440
EG 4b	one-time check & text messaging	number unknown *or* not logged in *and* subscriber unknown *or* absent	2,499	64	911
CG 4	one-time check & text messaging	-	1,125	-	1,125
Total			24,999		17,444

For each experimental group a corresponding control group was set up. Cell phone numbers in the control groups were fielded in parallel to the numbers of the respective experimental groups. By doing this, potential unobserved confounding factors due to seasonal effects, interviewers' motivation, and composition of the sample in the field should be controlled. Additionally, in each control group the same pre-call validation method was applied as in the respective experimental condition (with the exception of CG 0 where no validation method was applied for cost reasons). However, none of the numbers were screened out in the control conditions 1, 2, 3, and 4 despite available pre-call validation return codes. This approach enabled a detailed evaluation of the applied validation method and the respective screening condition by comparing field information in terms of final disposition codes as reliable information about the working status of the number with the estimated status based on the validation return codes. After screening, a final sample of 17,444 cell phone numbers was fielded successively in five packages of equal size over the entire field period. In the control conditions without text messaging, the response rate (AAPOR RR1) was 11 percent. In the control conditions using text messages to determine the status of a cell phone number and to announce the survey request the response rate increased to 13 percent ($\chi^2 = 2.41$, $p = .12$).

4. Results

According to the different screening conditions displayed in Table 1 (second to last column), the percentage of excluded numbers varied considerably ranging from 42 percent in EG 1, EG 2 and EG 4a up to 64 percent in EG 4b, or even 70 percent in EG 3b. Therefore, the application of number validation seemed to be a less strict strategy of excluding cell phone numbers than excluding cell phone numbers based on text messaging return codes.

Considering the *working number rate* in terms of the proportion of valid cell phone numbers among all numbers after screening (first column of Table 2), the efficiency of both number validation and text messaging became obvious. The percentage of nonworking numbers dropped from almost 50 percent without pre-call validation to about 10 percent or even less after screening due to number validation and/or text messaging return codes. In every experimental group a significantly higher working number rate was achieved compared to the respective control group (all comparisons *p* < .001). Remarkably, EG 3b differed significantly from all other experimental groups with a working number rate of 98 percent (all comparisons *p* < .001). Therefore, more strict screening, and thus, higher percentages of screened out numbers resulted in a higher proportion of technically valid cell phone numbers in the field.

Table 2: Effects of Pre-Call Validation Regarding Workload and Cost Measures

	working number rate (%)	contact rate (%)	interview rate (%)	call attempts[a]	overall call duration (minutes)[a]
CG 0	53	28	6	102	70
EG 1	92[*, (2), 3b, (4a)]	49[*, 3b, 4b]	10[*, (3a), 3b, 4b]	89	67
CG 1	56	26	6	106	71
EG 2	89[*, (1), 3b]	49[*, 3b, 4b]	10[*, (3a), 3b, 4b]	87	68
CG 2	51	26	4	141	92
EG 3a	91[*, 3b, (4a)]	52[*, 3b, 4b]	14[*, (1), (2), (3b), (4a), (4b)]	62	53
EG 3b	98[*, 1, 2, 3a, 4a, 4b]	80[*,1, 2, 3a, 4a, 4b]	20[*, 1, 2, (3a), 4a]	34	46
CG 3	54	28	7	90	63
EG 4a	88[*, (1), (3a), 3b]	49[*, 3b, 4b]	11[*, (3a), 3b, 4b]	75	60
EG 4b	90[*, 3b]	68[*,1, 2, 3a, 3b, 4a]	18[*, 1, 2, (3a), 4a]	41	46
CG 4	53	28	7	100	70
Total	68	39	9	84	64

Note. [a] Data on call attempts and overall call duration referred to mean per completed interview; no significance tests were calculated for these two indicators because calculations involved derived values based on the number of completed interviews, and not on the original elements of the sample. Thus, reliable variance estimations were not available.
Calculations were based on chi-squared goodness-of-fit tests for dependent samples:
[*] $p < .001$ compared to the respective CG.
Calculations were based on chi-squared tests of independence to calculate pairwise comparisons of all six experimental groups:

[1] $p < .001$ compared to EG 1	[(1)] $p < .05$ compared to EG 1
[2] $p < .001$ compared to EG 2	[(2)] $p < .05$ compared to EG 2
[3a] $p < .001$ compared to EG 3a	[(3a)] $p < .05$ compared to EG 3a
[3b] $p < .001$ compared to EG 3b	[(3b)] $p < .05$ compared to EG 3b
[4a] $p < .001$ compared to EG 4a	[(4a)] $p < .05$ compared to EG 4a
[4b] $p < .001$ compared to EG 4b	[(4b)] $p < .05$ compared to EG 4b.

4.1. Survey costs

Contact and interview rates
As shown in Table 2, *contact rates* were significantly higher in the experimental conditions compared to the corresponding control group (all comparisons $p < .001$). Whereas contact rates did not exceed 28 percent in any of the control groups, contact rates in the experimental conditions ranged from 49 to 80 percent. Especially EG 3b and EG 4b differed significantly from all other experimental conditions with contact rates of 80 and 68, respectively (all comparisons $p < .001$). As we were interested in the technical validity of cell phone numbers, the contact rate denoted a direct contact with a person, whether this person was eligible or not. Therefore, this approach is different from the AAPOR calculation of contact rates which includes only eligible persons. Referring to the *interview rates*, the percentage of completed interviews with eligible respondents was significantly higher in the experimental groups compared to the respective control group (all comparisons $p < .001$). In this respect, EG 3b and 4b were again particularly noticeable. In these two experimental conditions, interview

rates were the highest with a significant difference to all other experimental conditions (all comparisons $p < .05$ or lower). The proportions of cell phone numbers leading to a completed interview were comparable high in both conditions with about 20 and 18 percent, respectively ($\chi^2 = 1.35$, $p = .22$).

Mean number of call attempts and mean overall call duration per interview
As expected, *mean number of call attempts* and *mean overall call duration* (in minutes) needed for completing an interview (calculated by dividing the sum of all call attempts and the summed up overall call duration by the total number of completed interviews) tended to be lower in the experimental groups than in the corresponding control group (Table 2). Hence, less interviewer effort was needed to complete an interview after pre-call validation. Especially the strict text messaging screening condition (EG 3b), and the combined strict text messaging and strict number validation screening condition (EG 4b) seemed to be most effective in decreasing the interviewers' time and effort needed for conducting an interview.

Total survey costs
Contact and interview rates as well as mean number of call attempts and mean overall call duration influenced not only the interviewers' workload but also *overall survey costs* for conducting an interview via cell phone. Results concerning overall costs per completed interview, and cost savings due to different pre-call validation methods were reported in Figure 1.
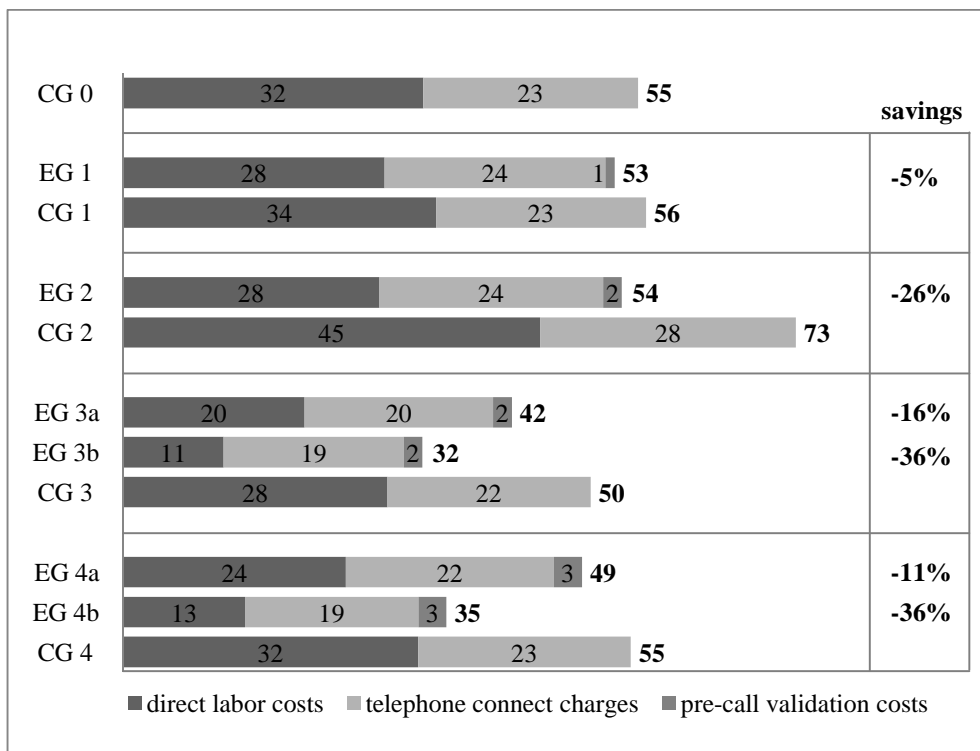

Figure 1: Costs per completed interview (in US$) and cost savings (in %) due to pre-call validation. Significance tests were not reported because calculations of overall costs involved derived values based on the number of completed interviews, and not on the original elements of the sample. Thus, reliable variance estimations were not available.

Cost calculations were based on direct labor costs for interviewers and supervisors ($16 per hour, respectively). Assuming that interviewers made on average 88 call attempts per hour and supervisors spent three-fourths of their time for direct supervision of the interviewers' work, total direct labor costs varied between $11 and $45 per completed interview depending on the experimental or control group. Telephone charges were included in the calculations with $0.44 per minute. Additional costs for pre-call validation were included, based on $0.05 per number check and $0.12 per text message. Additional costs in staff time for conducting pre-call validation and analyzing return codes were not included because the initial development of a systematic use of the pre-call validation methods was quite time-consuming while it can be assumed that labor costs for conducting pre-call validation in a production environment will be distinctly less.

Altogether, cost savings could be achieved in every experimental condition, and additional costs for pre-call validation were comparably low ranging from $1 to $3 per completed interview. Thus, cost savings due to pre-call validation exceeded the incurred extra charges. The largest cost savings could be achieved by screening out cell phone numbers with an unknown or temporarily absent subscriber (EG 3b) as well as by screening out numbers that were unknown or not logged in for a longer time in addition to unknown or absent subscribers (EG 4b). By contrast, excluding numbers that were unknown after a one-time number validation check (EG 1) or two-time number validation check (EG 2) yielded relatively small savings (leaving aside the conspicuously high costs in the CG 2 which were presumably due to a low interview rate in this control group for which no adequate explanation could be found despite thorough examination of the relevant factors).

4.2. Data quality

So far, strict screening conditions, and thus, a higher percentage of screened out cell phone numbers resulted in a substantial reduction of the interviewers' workload and considerable cost savings. However, when deciding on a proper pre-call validation method data quality needs to be taken into account as well.

Reduction of cases of unknown eligibility
Cases of unknown eligibility mainly occur when no information than just a voicemail and/or operator message or dialing tone is available to determine the final disposition code for a telephone number. *Unknown eligibility rates* of the experimental and control conditions were compared to examine the efficiency of pre-call validation in reducing unknown eligible numbers (results are reported in Figure 2).

Each pre-call validation method led to a significant reduction of the proportion of unknown eligible numbers compared to the respective control group (all comparisons $p < .05$ or lower). However, unknown eligibility rates varied drastically depending on the respective screening condition. In particular, the use of text messaging return codes in the strict screening condition of EG 3b and a combined screening rule according to EG 4b were most helpful in reducing numbers of unknown eligibility to 7 and 10 percent, respectively. By contrast, unknown eligibility rates amounted to about 30 percent in the corresponding control groups. The sole use of number validation in EG 1 and EG 2, and a combination of number validation and text messaging delivery reports in EG 4a also resulted in a reduction of the percentage of numbers of unknown

eligibility. However, unknown eligibility rates remained on a significantly higher level compared to experimental groups that were based on text messaging return codes like EG 3a, 3b, and 4b (all comparisons $p < .001$). With a proportion of screened out numbers of 42 percent in EG 1, EG 2, and EG 4a (see Table 1), the excluded cell phone numbers seemed to be mainly sample numbers classified as nonworking numbers and not as numbers of unknown eligibility. Therefore, excluding numbers that were unknown according to number validation return codes predominantly reduced the interviewers' workload but not numbers with an ambiguous status.
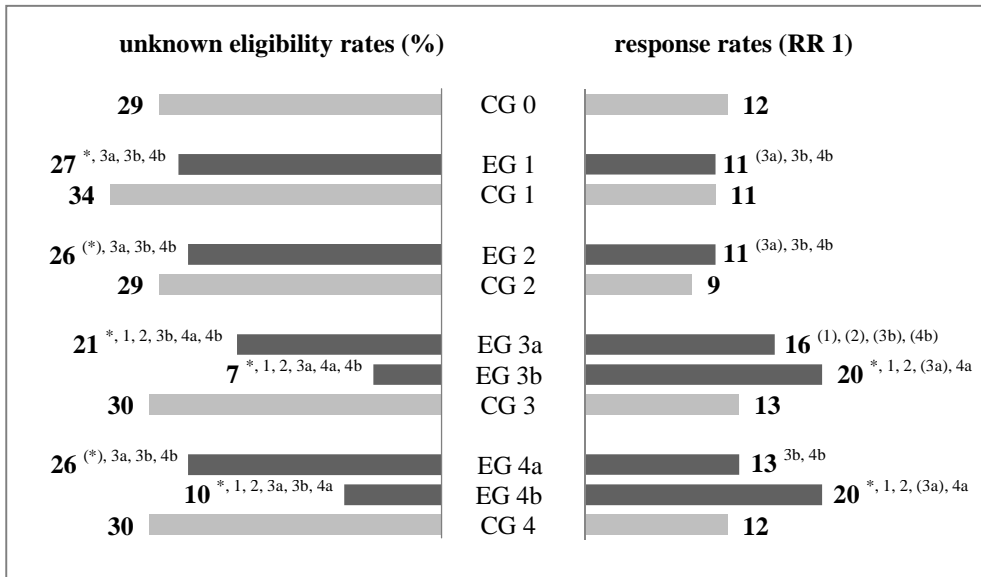

Figure 2: Unknown eligibility rates and response rates due to pre-call validation. Calculations were based on chi-squared goodness-of-fit tests for dependent samples: [*] $p < .001$ compared to the respective CG, [(*)] $p < .05$ compared to the respective CG. Calculations were based on chi-squared tests of independence to calculate pairwise comparisons of all six experimental groups:

[1] $p < .001$ compared to EG 1   [(1)] $p < .05$ compared to EG 1
[2] $p < .001$ compared to EG 2   [(2)] $p < .05$ compared to EG 2
[3a] $p < .001$ compared to EG 3a   [(3a)] $p < .05$ compared to EG 3a
[3b] $p < .001$ compared to EG 3b   [(3b)] $p < .05$ compared to EG 3b
[4a] $p < .001$ compared to EG 4a   [(4a)] $p < .05$ compared to EG 4a
[4b] $p < .001$ compared to EG 4b   [(4b)] $p < .05$ compared to EG 4b

Response rates can be improved among others by increasing the number of call attempts, by refusal conversion, by extending the field period, or by improving sample quality. The reduction of the proportion of numbers of unknown eligibility prior to fieldwork also affects response rate calculations. The higher unknown eligibility rates the less reliable calculations of response rates. Additionally, the proportion of numbers of unknown eligibility affects the magnitude of response rates negatively. Thus, response rates could be calculated more reliably and could be increased by reducing the amount of cases with uncertain status through pre-call validation (see Figure 2). EG 3b and 4b yielded the lowest unknown eligibility rates. Accordingly, these experimental groups achieved the highest response rates of 20 percent, respectively. Both groups differed significantly from the corresponding control condition (EG 3b: χ^2 = 11.86, $p < .001$; EG 4b: χ^2 = 14.46, $p < .001$), and from all other experimental groups

(all comparisons $p = < .05$ or lower). Pre-call validation in EG 1, 2, 3a and 4a achieved no significant increase of response rates compared to the respective control condition (all comparisons $p = $ n.s.).

Proportion of false negative numbers

Preliminary findings indicated a drastic decrease of the interviewers' workload and survey costs due to an increase of the percentage of technically valid cell phone numbers in the field. Furthermore, we found evidence for improved calculations of response rates as well as increased response rates due to reduced unknown eligibility rates as a result of pre-call validation. However, in order to examine the effects of the various strategies of pre-call validation, *false negative rates* defined as the proportion of false negative cases among all cases classified as valid by final disposition had to be taken into account. False negative cases are cell phone numbers which were mistakenly screened out in consequence of the experimental screening conditions. Given our experimental design, we were able to estimate the proportion of false negatives since we had number validation and text messaging return codes from pre-call validation as well as final disposition codes from the field for all cases in the control conditions. Thus, false negative rates were estimated by simulating the screening conditions of the experimental conditions on the basis of the respective control conditions.

As shown in Table 3, there were small false negative rates in EG 1, 2, and 4a (4 percent or less). By contrast, EG 3b and 4b which were the experimental conditions with the strictest screening conditions and with the highest working number rates also had the highest percentages of false negative numbers with 46 and 40 percent, respectively.

Table 3: False Negative Rates and Respective Final Disposition Codes

condition[a]	false negative rate %	final disposition codes (AAPOR) %			
		interview	eligible, non-interview	unknown eligibility, non-interview	not eligible
EG 1	4 [3a, 3b, 4b]	-	0.5	3	-
EG 2	2 [3a, 3b, 4b]	-	0.4	2	-
EG 3a	17 [1, 2, 3b, 4a, 4b]	0.7	1	15	0.3
EG 3b	46 [1, 2, 3a, 4a, (4b)]	1	4	40	0.5
EG 4a	4 [3a, 3b, ´4b]	0.3	1	2	-
EG 4b	40 [1, 2, 3a, (3b), 4a]	0.7	4	35	0.2

Note. [a] Estimation of false negative rates due to different pre-call validation methods were calculated on the basis of the respective control conditions.

Calculations were based on chi-squared tests of independence for independent samples; direct comparison of 3a with 3b, and 4a with 4b were based on chi-squared goodness-of-fit tests for dependent samples:

[1] $p < .001$ compared to EG 1 [(1)] $p < .05$ compared to EG 1
[2] $p < .001$ compared to EG 2 [(2)] $p < .05$ compared to EG 2
[3a] $p < .001$ compared to EG 3a [(3a)] $p < .05$ compared to EG 3a
[3b] $p < .001$ compared to EG 3b [(3b)] $p < .05$ compared to EG 3b
[4a] $p < .001$ compared to EG 4a [(4a)] $p < .05$ compared to EG 4a
[4b] $p < .001$ compared to EG 4b [(4b)] $p < .05$ compared to EG 4b.

False negative cases were classified by final disposition codes according to the AAPOR Standard Definitions (AAPOR, 2011): cases which resulted in an interview, eligible cases without an interview (including refusals, non-contacts, and others), cases of unknown eligibility, and not eligible cases. Not eligible but technically working cases like non-residential cell phone numbers were also considered as false negative cases because only those numbers which were invalid in a technical sense should be screened out by pre-call validation. Despite high false negative rates in some experimental conditions, proportions of potentially screened out numbers that had led to an interview were relatively small in all conditions (Table 3). In EG 3b for instance, only a small proportion of all mistakenly screened out numbers resulted in an interview with 1 percentage point of 46 percent false negative numbers. By contrast, cell phone numbers of unknown eligibility made up the largest part of the false negative cases. Thus, number validation and especially text messaging in combination with strict screening rules seemed to be particularly helpful in excluding numbers of unknown eligibility. Even though the percentage of valid but mistakenly screened out numbers that could potentially result in an interview was relatively small for all pre-call validation methods tested, we calculated screening biases for age and gender. Calculations were based on data given by the interviewees, and in addition on interviewer estimations of age and gender for refusals. Results (not in the table) based on chi-squared goodness-of-fit tests for dependent samples indicated small and insignificant biases due to false negatives.

Summary and Discussion

The present experiment was designed to develop and test an appropriate pre-call validation method for the improvement of cell phone samples. Different validation methods using number validation by means of a HLR lookup and/or text messaging were examined concerning their efficiency in excluding technically invalid cell phone numbers reliably. In summary, findings indicated a drastic increase of the percentage of working cell phone numbers in the field by means of pre-call validation. Additionally, a significant increase of contact and interview rates, and an – at least in part – considerable decrease of the interviewers' workload regarding mean number of call attempts and mean overall call duration per completed interview could be achieved. Consequently, total survey costs decreased – especially when using more strict screening rules. Therefore, both the use of number validation and text messaging return codes were appropriate methods to decrease survey costs. Furthermore, proportions of unknown eligible numbers could be reduced. Thus, response rates could be calculated more reliably after increasing the percentage of numbers with unambiguous final disposition codes. These results are supported by initial findings of Struminskaya and colleagues (2011) who also used HLR lookups to gain more certainty about numbers of unknown eligibility. However, relying merely on text messaging return codes seemed to be, at a first glance, more efficient. Excluding cell phone numbers on the basis of text message delivery reports resulted in higher proportions of screened out numbers than the sole use of number validation return codes. That was even true for rather loose screening conditions which merely excluded cell phone numbers that were definitely invalid according to pre-call validation results. Furthermore, using text messaging return codes in combination with strict screening rules (by excluding not only unknown cell phone numbers but also numbers of temporarily not available subscribers), the proportion of excluded cell phone numbers even increased to 70 percent. However, screening out up to 70 percent of the numbers of the initial cell phone sample could potentially harm data quality and should

therefore not be done carelessly. Of course, cost savings due to more strict pre-call validation methods are substantial and make such methods appealing to survey researchers. However, lower survey costs should not come at the expense of data quality. The extent of screened out cell phone numbers prior to data collection was directly related to the amount of mistakenly screened out cell phone numbers. When applying more strict screening rules as used in EG 3b or EG 4b, false negative rates amounted to over 40 percent. In the present study, this comparable high proportion of false negative numbers had no negative consequences on sample composition. Even though we found only minor non-significant biases due to quite high false negative rates, special precautions should be applied before generalizing our findings. In particular, further research is needed regarding the effects of screening rules on sample composition. Findings concerning potential screening biases have to be replicated before the application of text messaging pre-call validation is unconditionally suggested. At this point in time, the usage of return codes of simple number validation is an appropriate pre-call validation method for identifying invalid numbers to a satisfying extent without threatening data quality due to potential screening biases through high false negative rates.

Furthermore, simple number validation and text messaging as pre-call validation methods have to be tested in other countries. For most European countries, service providers offering HLR lookups and bulk text messaging are available. Implementation and received return codes are almost identical. However, the situation in the USA is quite different. No text messages may be sent to cell phone subscribers without prior consent in order to comply with the Telephone Consumer Protection Act (TCPA) legislation (AAPOR, 2010). In addition, HLR lookups are limited in scope in the U.S. because of an insufficient coverage in the GSM mobile network. Initial attempts for gathering additional information about the status of cell phone numbers, and using them for an improvement of working number rates in cell phone samples were made by Buskirk, Callegaro and Rao (2009). They developed a call design based on provider-specific internet services in the United States. Similar to our method of simple number validation they gained auxiliary information without the need of contacting the cell phone owner. However in contrast to Europe, these services are restricted to certain providers. Thus, their findings were confined to cell phone numbers of only three U.S. providers. By contrast, using HLR lookups for number validation has the great advantage that validation information about the working status of cell phone numbers can be gathered from a single source irrespective of the cell phone provider. This is true for large parts of Europe, and at least for some regions in the United States where GSM mobile network coverage is given.

References

American Association for Public Opinion Research (AAPOR) (2008). *Guidelines and Considerations for Survey Researchers When Planning and Conducting RDD and Other Telephone Surveys in the U.S. With Respondents Reached via Cell Phone Numbers*. Ann Arbor, MI.

American Association for Public Opinion Research (AAPOR) (2010). *New Considerations for Survey Researchers When Planning and Conducting RDD Telephone Surveys in the U.S. With Repsondents Reached via Cell Phone Numbers*. Ann Arbor, MI: AAPOR.

American Association for Public Opinion Research (AAPOR) (2011). *Standard Definitions. Final Dispositions of Case Codes and Outcome Rates for Surveys*. Ann Arbor, MI: AAPOR.

ADM (2008). Richtlinie für telefonische Befragungen. Retrieved from http://www.adm-ev.de/fileadmin/user_upload/PDFS/R04_D.pdf

Blumberg, S. J., & Luke, J. V. (2009). Reevaluating the Need for Concern Regarding Noncoverage Bias in Landline Surveys. *American Journal of Public Health, 99*(10), 1806-1810.

Blumberg, S. J., & Luke, J. V. (2011). *Wireless Substitution: Early Release of Estimates From the National Health Interview Survey, July-December 2010*: National Center for Health Statistics.

Brick, J. M., 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*(1), 23-39.

Brick, J. M., Edwards, W. S., & Lee, S. (2007). Sampling Telephone Numbers and Adults, Interview Length, and Weighting in The California Health Interview Survey Cell Phone Pilot Study. *Public Opinion Quarterly, 71*(5), 793-813.

Buskirk, T. D., Callegaro, M., & Rao, K. (2009). "N the Network?" : Using Internet Resources for Predicting Cell Phone Number Status. *Social Science Computer Review, 28*(3), 271-286.

Busse, B., & Fuchs, M. (2011). The Components of Landline Telephone Survey Coverage Bias. The Relative Importance of No-Phone and Mobile-Only Populations. *Quality & Quantity*, 1-17.

Callegaro, M., Steeh, C., Buskirk, T. D., Vehovar, V., Kuusela, V., & Piekarski, L. (2007). Fitting Disposition Codes to Mobile Phone Surveys: Experiences From Studies in Finland, Slovenia and the USA. *Journal of the Royal Statistical Society, 170*(3), 647-670.

European Commission (2011). Eurobarometer 75.1. February-March 2011. TNS OPINION & SOCIAL, Brussels [Producer]; GESIS, Cologne [Publisher]: ZA5479, data set version 2.0.0, doi:10.4232/1.10128.

ESOMAR (2010). Conducting Survey Research Via Mobile Phone. Retrieved from http://www.esomar.org/uploads/professional_standards/guidelines/ESOMAR_Codes&Guidelines_Conducting-survey-research-via-mobile-phone.pdf

Gabler, S., & Ayhan, Ö. (2007). Gewichtung bei der Erhebung im Festnetz und über Mobilfunk. Ein Dual Frame Ansatz. *ZUMA-Nachrichten Spezial*(13), 39-46.

Gabler, S., & Häder, S. (2009). Die Kombination von Mobilfunk- und Festnetzstichproben in Deutschland. In M. Weichbold, J. Bacher & C. Wolf (Eds.), *Umfrageforschung. Herausforderungen und Grenzen* (pp. 239-252): VS Verlag für Sozialwissenschaften.

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

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

Häder, S., Häder, M., Graeske, J., Kunz, T., & Schneiderat, G. (2009). Realisierung der Stichprobe. In M. Häder & S. Häder (Eds.), *Telefonbefragungen über das Mobilfunknetz. Konzept, Design und Umsetzung einer Strategie zur Datenerhebung* (pp. 71-82). Wiesbaden: VS Verlag für Sozialwissenschaften.

Häder, S., & Schneiderat, G. (2009). Antwortbereitschaft bei Mobilfunkumfragen. *Sozialwissenschaftlicher Fachinformationsdienst soFid (Methoden und Instrumente der Sozialwissenschaften), 2009/2*, 41-50.

Hu, S. S., Balluz, L., Battaglia, M. P., & Frankel, M. R. (2011). Improving Public Health Surveillance Using a Dual-Frame Survey of Landline and Cell Phone Numbers *American Journal of Epidemiology, 173*(6), 703-711.

Keeter, S., Christian, L., & Dimock, M. (2010). *The Growing Gap between Landline and Dual Frame Election Polls. Republican Vote Share Bigger in Landline-Only Surveys*: Pew Research Center for the People & the Press.

Keeter, S., Dimock, M., Kennedy, C., Best, J., & Horrigan, J. (2008). *Cost and Benefits of Full Dual Frame Sample Telephone Survey Design*. Paper presented at the 63th Annual conference of the American Association for Public Opinion Research, New Orleans, LA.

Kennedy, C. (2007). Evaluating the effects of screening for telephone service in dual frame RDD surveys. *Public Opinion Quarterly, 71*(5), 750-771.

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*(5), 840-854.

OFCOM (2010). Tackling abandoned and silent calls. Consultation. Retrieved from http://stakeholders.ofcom.org.uk/binaries/consultations/silentcalls/summary/condoc.pdf

Peytchev, A., Carley-Baxter, L. R., & Black, M. C. (2010). Coverage Bias in Variances, Associations, and Total Error From Exclusion of the Cell Phone-Only Population in the United States. *Social Science Computer Review, 28*(3), 287-302.

Steeh, C., Buskirk, T. D., & Callegaro, M. (2007). Using Text Messages in U.S. Mobile Phone Surveys. *Field Methods, 19*(1), 59-75.

Struminskaya, B., Kaczmirek, L., Schaurer, I., Bandilla, W., Gabler, S., & Häder, S. (2011). Identifying Non-working Numbers in Cell Phone RDD Samples via HLR-Lookup Technology. *Survey Practice*, August: www.surveypractice.org.

Zuwallack, R. (2009). Piloting Data Collection via Cell Phones: Results, Experiences, and Lessons Learned. *Field Methods, 21*(4), 388-406.