Journal of Official Statistics, Vol. 28, No. 3, 2012, pp. 373–394

Improving RDD Cell Phone Samples. Evaluation of Different Pre-call Validation Methods

Tanja Kunz[1] *and Marek Fuchs*[1]

Using cell phones in surveys poses new challenges to survey researchers. Amongst others, high proportions of nonworking numbers in randomly generated cell phone samples are reflected in increasing survey costs and extended fieldwork periods. In addition, ambiguous voicemail and operator messages which do not clearly indicate the status of a number negatively affect the proportion of numbers of unknown eligibility as well as the reliability of response rates. The simulated experiment reported in this article aimed at increasing the proportion of working cell phone numbers in the sample by screening out technically invalid or temporarily inactive numbers prior to fieldwork. Two methods were tested: number validation and text messaging. Findings from the application of different screening rules varying in strictness show that pre-call validation increases working number rates whereby survey costs and numbers of unknown eligibility can be reduced. Strict screening seems to be more efficient, however, at the expense of higher screening error.

Key words: HLR lookup; text message (sms); eligibility; nonworking numbers; survey costs.

1. Introduction and Background

In recent years, cell phone penetration rates in European countries have been on the rise: on average, 91 percent of the European households are currently equipped with at least one cell phone according to Eurobarometer data 2011 (European Commission 2011). Whereas Nordic countries like Sweden and Finland rank among the European front-runners with cell phone penetration rates of up to 97 percent, Germany ranks in the lower third with 86 percent. The changing cell phone usage patterns are also reflected in an increasing share of cell phone-only households who have abandoned their landlines in favor of exclusive cell phone usage (Busse and Fuchs 2012). Unlike general cell phone penetration rates, cell phone-only household rates differ to a greater extent across the 27 European (EU-27) countries. The countries with the highest share of cell phone-only households are Finland and the Czech Republic with 78 and 80 percent, respectively. Compared to an EU-27 average of 34 percent, Germany again ranks in the lower third with 8 percent cell phone-only households (European Commission 2011). The upward tendency concerning the share of cell phone-only households is not restricted to the EU. In the United States for

[1] Darmstadt University of Technology, Institute of Sociology, 64283 Darmstadt, Germany. Emails: kunz@ifs.tu-darmstadt.de and fuchs@ifs.tu-darmstadt.de
Acknowledgment: Support from Deutsche Forschungsgemeinschaft to the second author (#FU 389/16-2) is gratefully acknowledged.

instance, wireless substitution in households is about 30 percent at present and increasing further (Blumberg and Luke 2011).

Considering the rising number of cell phone-only households all over the world, traditional random digit dialing (RDD) landline surveys have been challenged considerably. Since cell phone-only individuals differ in certain sociodemographic characteristics from people who are also reachable via landline, traditional fixed-line surveys are prone to coverage biases due to the underrepresentation of specific groups of individuals (Blumberg and Luke 2009; Busse and Fuchs 2012; Peytchev et al. 2010). In Europe and the United States, cell phone-only individuals tend to be younger and male, have a higher probability of being unmarried, and tend to have lower incomes (Blumberg and Luke 2011; Graeske and Kunz 2009; Hu et al. 2011; Peytchev et al. 2010; Zuwallack 2009). Considering these differences between persons with and without landlines, the need to implement dual-frame surveys is meanwhile seen as unavoidable to prevent potential coverage biases in telephone surveys (Gabler and Ayhan 2007; Hu et al. 2011; Keeter et al. 2010; Kennedy 2007). However, the need for integrating cell phone samples into telephone samples creates new challenges for survey researchers.

1.1. Challenges Concerning Cell Phone Samples

As cell phone number directories are not available in most western countries, a sampling frame of listed cell phone numbers and randomly generated numbers is commonly used in Germany (Gabler and Häder 2009). Despite a continuous optimization of the underlying sampling frame, a remaining high share of nonworking numbers such as technically invalid numbers which are not yet assigned to a subscriber, or – although being technically valid – numbers which are never answered by a subscriber in each random sample from this frame makes data collection more difficult (Häder et al. 2009). Due to a high proportion of nonworking cell phone numbers in the sample, more time and effort is required on average for contacting a person and completing an interview. This inefficiency in fieldwork results in an extension of data collection periods, lower interviewer motivation, and higher total survey costs (AAPOR 2010; Buskirk et al. 2009; Keeter et al. 2008). The use of predictive dialers could potentially lessen this problem by automatically recognizing phone calls answered by a person, as well as phone calls with no answers or busy signals, and voicemail or operator messages (Jesurum 1990). Accordingly, technically invalid numbers could be screened out without the need for manual 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 other countries, guidelines for conducting survey research via cell phones impose certain limitations concerning the use of predictive dialers. For example, according to the Communications Act 2003 in the UK (OFCOM 2010) and according to the guidelines for telephone surveys in Germany (ADM 2008), there are restrictions, amongst others, regarding "abandoned" or "silent calls" which occur when no call agent is available although the call recipient already answered the phone. Even though in many European countries the use of predictive dialing in survey research is not prohibited in general, the efficiency of using these devices in cell phone samples remains doubtful. In particular, heterogeneous operator messages pose difficulties for predictive dialers in distinguishing reliably between technically valid and invalid cell phone numbers (AAPOR 2008; Callegaro et al. 2007; Häder et al. 2009).

Besides a high proportion of technically invalid numbers, RDD cell phone samples also imply a high share of numbers of unknown eligibility (AAPOR 2010; Häder et al. 2009). Many cases of unknown eligibility limit the informative value of final disposition codes and – in conclusion – the reliability of response rates derived from these codes. Compared to landline numbers, interviewers encounter more pronounced problems confirming the eligibility of cell phone numbers mainly because of ambiguous voicemail and operator messages (AAPOR 2010; Callegaro et al. 2007). Especially in cases where cell phone subscribers use their phones merely sporadically and thus are hard to reach, final disposition codes remain unclear despite several call attempts. Thus, a reliable calculation of response rates is threatened. Furthermore, ambiguous operator messages indicating that the subscriber is temporarily not available or may have the phone turned off do not provide reliable information on the technical validity of cell phone numbers (Callegaro et al. 2007; Häder et al. 2009). Under these circumstances, more call attempts are needed to ascertain the status of cell phone numbers which are never answered by a person but nonetheless seem to be technically valid numbers.

Altogether, high proportions of technically invalid cell phone numbers and numbers of unknown eligibility in cell phone samples necessitate more call attempts, lengthen the field period, and increase survey costs. Thus, an appropriate method for predicting the validity of cell phone numbers is needed. For this reason, Buskirk et al. (2009) created a call design developed especially for the United States based on provider-specific internet and text messaging services. Additional information about the status of cell phone numbers was gathered without contacting the cell phone owners by conducting a network lookup via the provider's website and by using text messaging services that provide information about the working status of the cell phone number after the number is entered on the web interface prior to further prompts for sending a text message. Although this was a reasonable method to reduce the effort required for determining the status of cell phone numbers, their validation method was restricted to cell phone providers which offer such validation services. Therefore, a complete coverage of the cell phone frame cannot be ensured as long as the survey is restricted to certain providers in the United States. Steeh et al. (2007) used text messages in order to obtain additional information about the status of cell phone numbers. Using text messaging return codes, they detected ineligible numbers at least as efficiently as by calling them. Furthermore, fewer cases remained undetermined after sending text messages. However, because of the legal restrictions of sending bulk text messages in the U.S. it is difficult to send unsolicited messages to cell phone subscribers. In Germany as well as in other European countries, such restrictions regarding provider-specific services or legal requirements do not exist insofar as sending bulk text messages is exclusively applied for research purposes, and certain requirements are met (ESOMAR 2010). Initial attempts to use text messaging delivery reports to determine the working status of cell phone numbers in Germany were sobering. A comparison between results of actual call attempts and text messaging return codes showed inconsistent and unreliable results (Schneiderat and Schlinzig 2009). By contrast, Struminskaya and colleagues (2011) used Home Location Register (HLR) lookups to detect technically invalid numbers and numbers of unknown eligibility. Their results provided initial evidence of the applicability of this method.

1.2. Pre-call Validation Method: Number Validation and Text Messaging

There are two different technical possibilities in Germany (and most European countries) for pre-call validation of cell phone numbers which we tested in the simulated experiment reported in this article: number validation using a Home Location Register (HLR) lookup and validation via detailed delivery reports received after sending bulk text messages.

Cell phone *number validation services* provide verification in real time by performing a HLR lookup (Noldus 2006). The HLR lookup indicates the current status of the cell phone number by providing a validation code. This code specifies whether a given cell phone number is “active and in use”, “invalid”, or “not logged in for a longer time”. The latter code is provided after the corresponding cell phone has been switched off for about an hour or longer. The method is inexpensive (€0.04 per check), and validation codes are available without contacting the cell phone subscriber. Furthermore, the validation codes are returned within a few seconds of the query. Thus, multiple checks can be carried out in a relatively short time at moderate costs. However, by conducting a HLR lookup the number validation method provides only a snapshot.

Text messaging services can also be applied for pre-call validation by sending bulk messages. Sending text messages offers additional information by means of detailed delivery status codes, and reason codes in case of failed delivery. These codes are available from providers of bulk text messaging services. If delivery succeeds, it can be assumed that the number is technically valid. If delivery fails, several reasons for failure may exist. The most interesting reason codes regarding pre-call validation are the additional information of “subscriber is unknown” or “subscriber is temporarily not available”. Based on this additional information, the respective case could be excluded from the sample. Though more expensive (about €0.09 per text message), sending text messages at first glance has multiple advantages compared to number validation via HLR lookups. Compared to number validation, using text messaging return codes has the advantage of seemingly more reliable validation information due to a longer delivery attempt. If the message cannot be delivered initially the message will be queued for retrying at a later point in time for up to 48 hours. Furthermore, sending an advance text message can also be used for prenotification of respondents. 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) did not find any evidence that sending text messages has an effect on response rates. Steeh et al. (2007) showed that using text messages for prenotification tended to have a positive effect on several measures of survey participation such as cooperation rates and response rates. However, these differences were not statistically significant. In a German study, positive effects of prenotification on response rates could be shown (Häder and Schneiderat 2009). Thus, the effectiveness of sending text messages on improving cooperation rates can be an advantageous side effect of this pre-call validation method.

2. Research Question

The study aimed to maximize the proportion of working cell phone numbers in RDD cell phone samples (1) by eliminating technically invalid cell phone numbers which are not assigned to a subscriber, and (2) by decreasing the share of numbers of unknown eligibility which remain unanswered by a person and have ambiguous voicemail or operator messages through an appropriate pre-call validation method. As a result, a reduction of the interviewers' workload and survey costs was expected and the reliability of the response rate calculation was expected to improve. To decrease time and cost efforts without compromising data quality, the application of both pre-call validation methods had to be examined carefully.

Sending text messages is a convenient and frequently used mode of communication. A prompt and reliable delivery of text messages is ensured by text message services. Considering the advantages of sending text messages, such as multiple delivery attempts and detailed delivery return codes, it can be assumed that applying this procedure yields the most accurate and reliable information about the working status of a cell phone number.

Number validation services are less common and the application of information obtained from HLR lookups developed only recently. At present, little is known about the validity of the returned number validation codes. In addition, number validation provides information concerning the status of a cell phone solely at a single point in time. For example, temporary technical reasons might harm the value of number validation results. Therefore, the validity of this procedure needs to be checked carefully.

In detail, the present article aims at answering the following questions: Regarding *interviewer workload* and *survey costs* we examined by means of pre-call validation (1) whether contact rates and interview rates could be increased, and (2) whether mean number of call attempts and mean overall call duration per completed interview could be reduced. With respect to *data quality* we considered (3) whether the share of numbers of unknown eligibility could be reduced by use of pre-call validation, and consequently, whether response rates could be calculated more reliably. Also, we asked (4) whether the two validation methods falsely screened out cell phone numbers although they were working (false negatives). In a final step relating to the *predictive efficiency* of the respective pre-call validation methods, (5) we examined each validation method concerning the precision with which working and nonworking cell phone numbers were classified correctly.

3. Method

Within the scope of the "Experimental Mobile Phone Panel" in Germany (supported by Deutsche Forschungsgemeinschaft), recruitment interviews were conducted from September 2010 to January 2011 at Darmstadt University of Technology (Germany). An initial sample of 24,999 cell phone numbers drawn from a sampling frame of randomly generated cell phone numbers (Gabler and Häder 2009) was provided by GESIS Mannheim (Germany). This method generates a cell phone frame using the following methodology: All possible banks of 10,000 cell phone numbers are assessed for every cell phone carrier access code. Within banks with at least one listed cell phone number (according to directories and internet research), all potential numbers are generated. Number banks reserved for technical services are excluded from the sampling frame. From the remaining frame of generated numbers, a random sample of cell phone numbers stratified by provider access code is drawn. This sampling design enables equal inclusion probabilities for all working cell phone numbers in Germany. The varying density of assigned numbers which is quite in some banks low is taken into account. Although more efficient than classical RDD, the gross samples drawn from this frame still contain high proportions of technically invalid numbers and numbers of unknown eligibility (Häder et al. 2009).

Results presented in this article are based on a random subsample of 3,275 cell phone numbers from the initial sample of 24,999. For this subsample final disposition codes from the field as well as pre-call validation results from both validation methods were available. Resulting from an application of up to 15 call attempts during the fieldwork period of the study, final disposition codes were generated according to AAPOR Standard Definitions (AAPOR 2011) and were used as reliable information about the status of each cell phone number. Originally developed for RDD landline telephone surveys, these Standard Definitions can also be applied to RDD cell phone surveys. Only slight modifications were necessary regarding voicemail and operator messages indicating that the cell phone subscriber is temporarily not available or that the cell phone is turned off. Numbers with these new additional final disposition codes were classified as cases of unknown eligibility.

Referring to the subsample of 3,275 cell phone numbers, response rates calculated on the basis of AAPOR Standard Definitions RR1 (AAPOR 2011) amounted to 11 percent without sending a text message, and increased to 13 percent when sending a text message as initial contact (difference not statistically significant). The advance text message contained the following text: "Darmstadt University of Technology will soon invite you to a survey via cell phone and would like to thank you for your participation in advance".

Regarding pre-call validation services, three different providers for text messaging and four providers for number validation were examined in a pilot study with a sample of 1,500 cell phone numbers. We selected one provider each on the basis of (1) the price, (2) the level of detail of the delivery reports or reason codes in case of delivery failure, (3) the reliability of the provider, and (4) the usability of the provider's service.

As described above, two *pre-call validation methods* were applied to screen out nonworking numbers: *number validation* and *text messaging*. Also a *combination* of both methods was tested. By combining number validation with sending a text message, it was to be examined whether additional costs and efforts were justified with respect to an improved efficiency compared to a separate application of each validation method. Thus, all 3,275 cell phone numbers were randomly assigned to one of three validation conditions implementing the two pre-call validation methods, and a combination of both methods in a between-subjects design (see Table 1).

In a first step, after conducting pre-call validation based on number validation and/or text messaging, all cases of the three validation conditions ($u = 3{,}275$) were fielded and final disposition codes were assigned according to AAPOR standards (see above). In a second step, we simulated the resulting gross sample as if we had applied screening rules prior to fieldwork using the following approach. In order to assess the effects of the pre-call validation methods, we restricted the gross sample to those cases which would not have been excluded if we had applied screening rules prior to fieldwork. We then compared the final disposition codes resulting from fieldwork in the full sample ('no screening' condition) to the restricted sample (after the application of 'loose screening' and 'strict screening' rules).

Within each of the three validation conditions, two classes of *screening rules* that varied in strictness were employed resulting in six different simulated screening conditions in total. By applying *loose* screening rules, numbers were screened out that were definitely technically invalid according to the pre-call validation return codes. Numbers that were not explicitly confirmed as technically invalid by the pre-call validation method were treated as working cell phone numbers and remained in the sample. By using loose screening rules, a cell phone number was screened out as nonworking

- if the number was unknown according to number validation (condition 1a),
- if text message delivery failed due to an unknown subscriber (condition 2a),
- if text message delivery failed due to an unknown subscriber *or* if the number was unknown according to number validation (condition 3a).

In terms of *strict* screening rules, numbers were not only screened out if they were definitely technically invalid but also if they were temporarily not available according to the pre-call validation method. Thus, numbers that were not explicitly confirmed as technically invalid and as temporarily not available were treated as working cell phone numbers and remained in the sample. According to text messaging return codes, a cell phone subscriber is temporarily absent if the message cannot be delivered within 48 hours, whereas the number validation return code indicating that the number is temporarily not logged in is sent after the cell phone is switched off for about an hour. Therefore, a cell phone number was screened out according to strict screening rules as nonworking

- if the number was unknown *or* temporarily not logged in according to number validation (condition 1b),
- if text message delivery failed due to an unknown *or* temporarily absent subscriber (condition 2b),
- if text message delivery failed due to an unknown *or* temporarily absent subscriber *or* the number was unknown *or* temporarily not logged in according to number validation (condition 3b).

Simulating the application of loose and strict screening rules resulted in nested subsets of dependent samples. Accordingly, calculations comparing the full sample with the restricted samples resulting from the application of the loose and strict screening rules, as well as calculations comparing the loose and strict samples (comparing columns in Table 1) were based on chi-squared goodness-of-fit tests for dependent samples. Calculations comparing the impact of the two pre-call validation methods of number validation and text messaging, as well as the impact of a combination of both methods (comparing rows in Table 1) were based on chi-squared tests for independent samples.

4. Results

According to the various screening rules displayed in Table 1, percentages of screened out cell phone numbers varied considerably, ranging from 42 percent in the loose number

Table 1. Pre-call validation conditions and the respective screening rules

Validation condition	No screening[a]	Applied screening rules: (a) Loose[b]	(b) Strict[b]
(1) **Number validation**	-	number unknown	number unknown *or* not logged in
screened out (%)	0	42	64
sample size	1,025	599	369
(2) **Text messaging**	-	subscriber unknown	subscriber unknown *or* absent
screened out (%)	0	50	70
sample size	1,125	557	343
(3) **Combination**	-	number unknown *or* subscriber unknown	number unknown *or* not logged in *or* subscriber unknown *or* absent
screened out (%)	0	54	74
sample size	1,125	523	294
total sample size	3,275		

Note.
[a] For the 'no screening' conditions both sets of information were available: final disposition codes and pre-call validation return codes.
[b] Loose and strict screening rules were applied to the cases of the respective experimental condition resulting in differing sample sizes in the six simulated screening conditions.

validation condition (1a) up to 74 percent in the strict combined condition (3b). Compared to text messaging, the application of number validation seemed to be a less strict method of excluding cell phone numbers with the smallest percentages of screened out numbers of 42 percent in the loose (1a), and 64 percent in the strict screening condition (1b). Strikingly, there were only small differences between the loose text messaging (2a) and loose combined condition (3a), with 50 versus 54 percent of screened out numbers. The same applied to the strict screening conditions 2b and 3b with 70 versus 74 percent of screened out numbers.

As expected, irrespective of the pre-call validation method more cell phone numbers were excluded by applying the strict screening rules compared to the loose screening rules. Percentages of screened out numbers ranged from 64 to 74 percent in the strict screening conditions (1b, 2b, and 3b) compared to 42 to 54 percent in the loose screening conditions (1a, 2a, and 3a). By additionally excluding numbers that were temporarily not available in the strict screening conditions, percentages of screened out cell phone numbers were about 20 percentage points higher on average compared to the respective loose screening conditions.

4.1. Workload and Cost Measures

The present experiment aimed at increasing the proportion of working cell phone numbers in the fielded sample by ideally screening out all – according to the respective screening rule – nonworking cell phone numbers prior to fieldwork. This was expected to decrease interviewer workload and survey costs. Because the total quantity of fielded numbers declined to varying degrees when different screening rules were applied, the percentage basis of calculations changed in each of the six simulated screening conditions. Thus, changes in the following outcome rates (working number rate, contact and interview rate) did not result from an increased number of completes but from a reduced quantity of fielded cell phone numbers according to the respective pre-call validation condition and screening rule. Due to the random assignment of the full sample to one of the three validation conditions, differences in working number rates, contact rates, and interview rates were due to chance only. No significant differences occurred according to a chi-squared test of independence.

Working number rates were defined as the proportion of working cell phone numbers among all fielded numbers. As shown in Table 2, number validation and text messaging as well as the combination of both methods were appropriate pre-call validation methods to increase the percentage of working numbers in the field irrespective of the strictness of the screening rules. Working number rates in the loose and strict screening conditions ranged from 91 to 98 percent according to the particular validation condition, and were all significantly higher than in the respective 'no screening' condition with working number rates of 53 to 56 percent (all comparisons $p < .001$). With the exception of the number validation method, strict screening conditions achieved significantly higher working number rates compared to the respective loose screening condition (text messaging conditions 2a and 2b: $p < .001$; combined conditions 3a and 3b: $p < .001$). Especially text messaging in the strict screening condition (2b) as well as the loose (3a) and strict (3b) combined condition yielded particularly high working number rates of 95 percent or higher, which differed significantly from number validation (1a, 1b) and loose text messaging (2a) (all comparisons $p < .05$ or better).

For the purpose of this project, *contact rates* indicated the proportion of cell phone numbers where an individual was reached by an interviewer among all fielded numbers. As we were primarily interested in the technical validity of cell phone numbers, the contact rates denoted a direct contact with a person, whether this person was eligible or not. Thus, contact rates reported here differ from contact rates defined in the AAPOR Standard Definitions (AAPOR 2011). Contact rates were significantly increased by means of pre-call validation compared to the respective 'no screening' condition (all comparisons $p < .001$). In the loose screening conditions, contact rates ranged from 44 to 55 percent. Even higher contact rates were achieved in the strict screening conditions with 68 to 84 percent. By applying strict screening rules, contact rates were increased by 24 percentage points in the number validation condition (1b), 26 percentage points in the text messaging condition (2b), and 29 percentage points in the combined condition (3b) compared to the respective loose condition (all comparisons $p < .001$). However, there is no significant difference between contact rates achieved in the strict text messaging condition (2b) and the strict combined condition applying text messaging and number validation (3b). The same is true for the comparison of the loose conditions 2a and 3a (not significant). Nevertheless, irrespective of the strictness of the screening rules, text messaging and a combination of number validation and text messaging yielded significantly higher contact rates compared to the sole use of number validation ($p < .001$ for all comparisons except

Table 2. Workload and cost measures depending on the respective validation condition and screening rule.

			Applied screening rules	
	Validation condition	*No screening*	(a) *Loose*	(b) *Strict*
Working number rate (%)	(1) number validation	56	91$^{*, (2b), (3a), 3b}$	91$^{*, (2b), (3a), 3b}$
	(2) text messaging	54	91$^{*, (3a), 3b}$	97$^{*, \dagger, (1a), (1b)}$
	(3) combination	53	95$^{*, (1a), (1b), (2a)}$	98$^{*, \dagger, 1a, 1b, 2a}$
Contact rate (%)	(1) number validation	26	44$^{*, (2a), 2b, 3a, 3b}$	68$^{*, \dagger, 2a, 2b, 3a, 3b}$
	(2) text messaging	28	54$^{*, (1a), 1b, 3b}$	80$^{*, \dagger, 1a, 1b, 3a}$
	(3) combination	28	55$^{*, 1a, 1b, 2b}$	84$^{*, \dagger, 1a, 1b, 2a}$
Interview rate (%)	(1) number validation	6	11$^{*, 2b, 3b}$	17$^{*, \dagger}$
	(2) text messaging	7	14$^{*, (3b)}$	22$^{*, \dagger, 1a, (3a)}$
	(3) combination	7	14$^{*, (2b)}$	22$^{*, \dagger, 1a, (2a)}$
Call attempts[a]	(1) number validation	106	87	43
	(2) text messaging	90	62	33
	(3) combination	100	68	32
Overall call duration (minutes)[a]	(1) number validation	71	65	50
	(2) text messaging	63	53	43
	(3) combination	70	58	44

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.

Calculations were based on chi-squared goodness-of-fit tests for dependent samples:

* $p < .001$ compared to the respective 'no screening' condition,

† $p < .001$ compared to the respective loose screening condition.

Calculations were based on chi-squared tests for independent samples:

^{1a} $p < .001$ compared to condition 1a, $^{(1a)}$ $p < .05$ compared to condition 1a,

^{1b} $p < .001$ compared to condition 1b, $^{(1b)}$ $p < .05$ compared to condition 1b,

^{2a} $p < .001$ compared to condition 2a, $^{(2a)}$ $p < .05$ compared to condition 2a,

^{2b} $p < .001$ compared to condition 2b, $^{(2b)}$ $p < .05$ compared to condition 2b,

^{3a} $p < .001$ compared to condition 3a, $^{(3a)}$ $p < .05$ compared to condition 3a,

^{3b} $p < .001$ compared to condition 3b, $^{(3b)}$ $p < .05$ compared to condition 3b.

condition 2a compared to condition 1a with $p < .05$). Accordingly, using text messaging return codes seemed to be a more effective pre-call validation method yielding higher contact rates than the number validation method.

Higher contact rates were reflected in significantly higher *interview rates* in all six simulated screening conditions compared to the respective 'no screening' conditions (all comparisons $p < .001$). It was possible to increase the proportion of completed interviews among all fielded numbers to up to 22 percent in the strict combined condition (3b). Without implementing any pre-call validation method, interview rates were as low as six or seven percent in the 'no screening' conditions. Irrespective of the pre-call validation method, interview rates significantly increased in the strict screening conditions compared to the respective loose screening conditions (all comparisons $p < .001$): The increase amounted to six percentage points in the strict number validation condition (1b), and eight percentage points in the strict text messaging (2b) and strict combined condition (3b), respectively. Interestingly, in the text messaging condition and combined condition higher interview rates were obtained than in the number validation condition (three percentage points for the loose condition and five percentage points for the strict condition). Again, there was no significant difference in the interview rates between the text messaging condition and combined condition, neither regarding the loose screening conditions 2a and 3a, nor with respect to the strict screening conditions 2b and 3b.

Mean number of call attempts and *mean overall call duration per completed interview* (calculated by dividing the sum of all call attempts and the added overall call duration by the total number of completed interviews) as further indicators of the interviewers' workload were also described in Table 2. As expected, mean number of call attempts and mean overall call duration for each completed interview tended to be lower in the loose and strict screening conditions compared to the corresponding 'no screening' conditions (no statistical testing possible, see note (a) below Table 2). Especially in the strict screening conditions, huge reductions of the mean number of call attempts were achieved compared to the respective 'no screening' conditions with a reduction of 59 percent (from 106 to 43 call attempts) in condition 1b, a reduction of 63 percent (from 90 to 33 call attempts) in condition 2b, and a reduction of 68 percent (from 100 to 32 call attempts) in condition 3b. In the loose screening conditions, reductions amounted to 18 percent in condition 1a (from 106 to 87 call attempts), 31 percent in condition 2a (from 90 to 62 call attempts), and 32 percent in condition 3a (from 100 to 68 call attempts). Reductions of the mean overall call duration in the strict screening conditions were smaller, and ranged from 30 to 37 percent in the strict screening conditions, and from 8 to 17 percent in the loose screening conditions.

All indicators reported so far suggest that the application of pre-call validation methods drastically decreased the interviewers' workload. Additionally, screening out nonworking cell phone numbers was also reflected in *total survey costs*. In Figure 1, overall costs per completed interview and cost savings due to various pre-call validation methods are depicted. Cost calculations were based on direct labor costs for interviewers and supervisors (€11 per hour, respectively), telephone charges (€0.14 per minute) and additional costs for pre-call validation (€0.04 per number validation, €0.09 per text message). Costs in staff time for conducting pre-call validation and analyzing return codes were not included in the cost calculations, because the initial development of a systematic

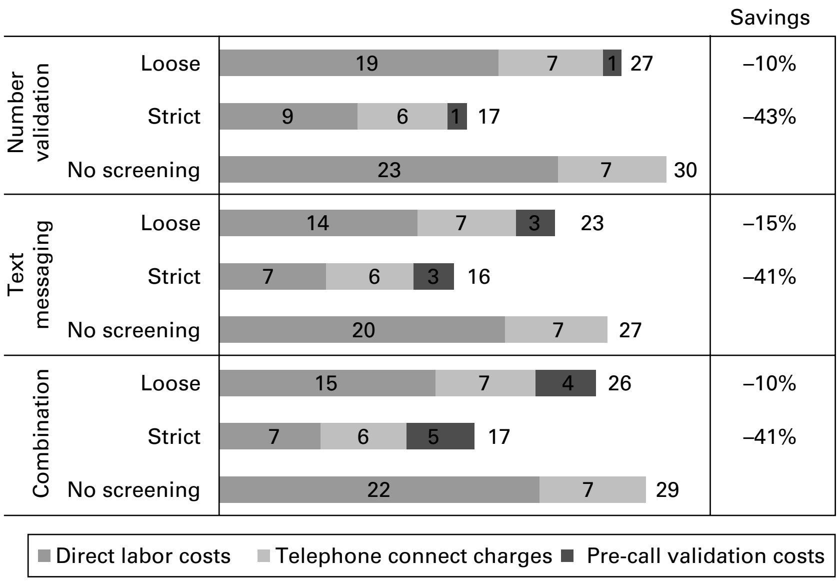


Fig. 1. Total survey costs and cost savings per completed interview due to pre-call validation. Significance tests were not reported because calculations of overall costs involved derived values based on the number of completed interviews, and not on the original elements of the sample.

use of 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 lower. Assuming that interviewers made on average 88 call attempts per hour, absolute costs ranged from €16 to €30 per completed interview. The largest cost savings were achieved in all strict screening conditions. Irrespective of whether conducting number validation (1b), sending text messages (2b), or applying a combination of both methods (3b), cost savings exceeded 40 percent. Cost reductions were also achieved in the loose screening conditions, albeit to a lesser extent. Compared to overall costs in the 'no screening' conditions ranging from €27 to €30 per complete, only slight savings were obtained by applying loose screening rules, amounting to 10 to 15 percent. By additionally excluding cell phone numbers that were temporarily not available, strict screening rules were thus more efficient with regard to time and cost savings than loose screening rules. Nevertheless, cost savings were achieved in each of the three validation conditions, irrespective of the strictness of screening rules. In particular, additional costs caused by the application of the pre-call validation methods were comparably low with €1 to €5 per completed interview, and did not exceed the cost savings achieved.

4.2. Data Quality Measures

Preliminary findings revealed that excluding a higher proportion of cell phone numbers following stricter screening rules resulted in a remarkable reduction of the interviewers' workload and considerable cost savings. However, when deciding on an adequate pre-call validation method, data quality needed to be taken into account as well.

Decreasing the proportion of numbers of unknown eligibility is a major concern in order to enable a reliable calculation of response rates. As depicted in Figure 2, *unknown*

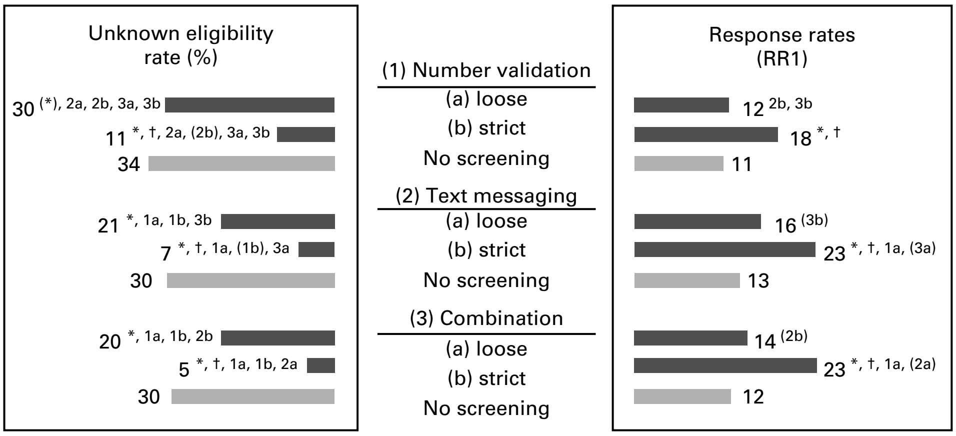


Fig. 2. Unknown eligibility rates and response rates depending on the respective validation condition and screening rule.
Calculations based on chi-squared goodness-of-fit tests for dependent samples:
* *p* < .001 *compared to the respective 'no screening' condition,*
$^{(*)}$ *p* < .05 *compared to the respective 'no screening' condition;*
† *p* < .001 *compared to the respective loose screening condition.*
Calculations based on chi-squared tests for independent samples:
^{1a} *p* < .001 *compared to condition 1a,* $^{(1a)}$ *p* < .05 *compared to condition 1a,*
^{1b} *p* < .001 *compared to condition 1b,* $^{(1b)}$ *p* < .05 *compared to condition 1b,*
^{2a} *p* < .001 *compared to condition 2a,* $^{(2a)}$ *p* < .05 *compared to condition 2a,*
^{2b} *p* < .001 *compared to condition 2b,* $^{(2b)}$ *p* < .05 *compared to condition 2b,*
^{3a} *p* < .001 *compared to condition 3a,* $^{(3a)}$ *p* < .05 *compared to condition 3a,*
^{3b} *p* < .001 *compared to condition 3b,* $^{(3b)}$ *p* < .05 *compared to condition 3b.*

eligibility rates defined as the proportion of cases of unknown eligibility among all cases of the fielded sample were significantly reduced due to pre-call validation compared to the corresponding 'no screening' condition (all comparisons $p < .05$ or better). Due to the random assignment of the full sample to one of the three validation conditions, differences in unknown eligibility rates occurred by chance only. When applying strict screening rules, unknown eligibility rates were drastically reduced to 5 to 11 percent, whereas the rates in the corresponding 'no screening' conditions reached 30 percent or more. In the loose screening conditions, noticeable reductions could only be achieved in the text messaging condition (2a) and in the combined condition (3a) with unknown eligibility rates of 21 and 20 percent, respectively. Strict screening rules achieved a significantly more pronounced decrease of unknown eligibility rates compared to the loose screening rules, by 19 percentage points in the number validation condition, by 14 percentage points in the text messaging condition, and by 15 percentage points in the combined condition (all comparisons $p < .001$).

Furthermore, by increasing the proportion of numbers with unambiguous final disposition codes, response rates could be calculated more reliably. Additionally, unknown eligibility rates affected the magnitude of response rates. In particular, by applying strict screening rules response rates (RR1) according to AAPOR standards (AAPOR 2011) were significantly higher compared to the 'no screening' conditions (all comparisons $p < .001$). It was possible to achieve noticeable increases in the strict text messaging condition (2b) and strict combined condition (3b) with response rates of 23 percent, respectively. Loose screening conditions did not yield significant increases in response rates compared to the respective 'no screening' conditions.

Considering the preliminary effects of the various pre-call validation methods and screening rules, excluding nonworking cell phone numbers from the sample prior to fieldwork seemed to be efficient in reducing the interviewer workload, survey costs, and unknown eligibility rates. However, it was suspected that the more cell phone numbers were excluded by pre-call validation, the higher the risk of false negative cases that would be mistakenly screened out although they were working cell phone numbers. Especially strict screening rules that excluded not only technically invalid cell phone numbers but also numbers that were temporarily not logged in (number validation) or absent (text messaging) hold the potential to induce bias in the sample. These numbers were excluded under the assumption that many of them were in fact no longer in use, and that calls to these numbers would never connect to any potential respondent. However, it is well possible that several of these numbers were actually still in use while the associated cell phone might temporarily not have been turned on. If this was the case, there is a tendency to systematically exclude subgroups of the population who use their cell phones less frequently, and therefore turn their cell phones off more often, similar to the systematic exclusion of answering machine owners in landline samples (Oldendick 1993). In order to assess the potential loss in accuracy due to this potential screening error, we first analyzed the proportion of falsely excluded cases and then estimated sociodemographic variables with and without wrongly screened out cases.

In general, to avoid biases due to many false negative cases resulting from pre-call validation *false negative rates* should be kept to a minimum. In order to determine false negative cases, we relied on the following final disposition codes from the field. Applying AAPOR Standard Definitions (AAPOR 2011), we distinguished cases which resulted in an interview, eligible cases with no interview (including refusals, noncontact, and other noninterview cases), cases of unknown eligibility, and ineligible cases. In terms of the AAPOR Standard Definitions, cases that are ineligible by definition (e.g., nonresidential numbers) are typically excluded from calculation of outcome rates. In contrast to this, ineligible cases were not excluded in our simulation study as long as they were working numbers like businesses or government offices. Thus, cell phone numbers classified as ineligible by AAPOR Standard Definitions were considered as false negative cases because by means of pre-call validation solely nonworking numbers were to be excluded irrespective of the ineligibility of the number. For the purpose of this study, we defined the false negative rate as the proportion of false negative cases among all cases classified as working by final disposition codes from the field. Table 3 depicts false negative rates for the various validation conditions and screening rules.

Overall, loose as well as strict screening rules screened out numbers that seemed to be working cell phone numbers based on the information from fieldwork. In all strict screening conditions in particular, false negative rates were tremendously high, with up to 52 percent in the strict combined condition considering number validation and text messaging return codes. The proportions of false negative cases in the strict screening conditions were all significantly higher compared to the corresponding loose screening condition (all comparisons $p < .001$). However, taking detailed final disposition codes

Table 3. False negative cell phone numbers, final disposition codes of false negative numbers, and screening bias depending on the respective validation condition and screening rule

		Final disposition codes (AAPOR) (%)				Screening bias	
Validation condition + screening rule	False negative rate %	*interview*	*eligible, non-interview*	*unknown eligibility, non-interview*	*not eligible*	*age*	*gender*
(1) **Number validation**							
(a) loose	**4** [1b, 2a, 2b, 3a, 3b]	-	0.5	3	-	0.6	0.0
(b) strict	**41** [1a, 2a, 3a, 3b]	0.7	2	38	-	0.4	0.9
(2) **Text messaging**							
(a) loose	**17** [1a, 1b, 2b, 3b]	0.7	1	15	0.3	0.6	0.3
(b) strict	**46** [1a, 2a, 3a, (3b)]	1	4	40	0.5	0.4	0.6
(3) **Combination**							
(a) loose	**17** [1a, 1b, 2b, 3b]	0.5	3	14	-	0.5	1.3
(b) strict	**52** [1a, 1b, 2a, (2b), 3a]	2	8	42	0.2	1.0	2.9

Notes: Calculations were based on chi-squared tests for independent samples; direct comparison of 1a with 1b, 2a with 2b, and 3a with 3b were based on chi-squared goodness-of-fit tests for dependent samples:
[1a] $p < .001$ compared to condition 1a, [(1a)] $p < .05$ compared to condition 1a,
[1b] $p < .001$ compared to condition 1b, [(1b)] $p < .05$ compared to condition 1b,
[2a] $p < .001$ compared to condition 2a, [(2a)] $p < .05$ compared to condition 2a,
[2b] $p < .001$ compared to condition 2b, [(2b)] $p < .05$ compared to condition 2b,
[3a] $p < .001$ compared to condition 3a, [(3a)] $p < .05$ compared to condition 3a,
[3b] $p < .001$ compared to condition 3b, [(3b)] $p < .05$ compared to condition 3b.

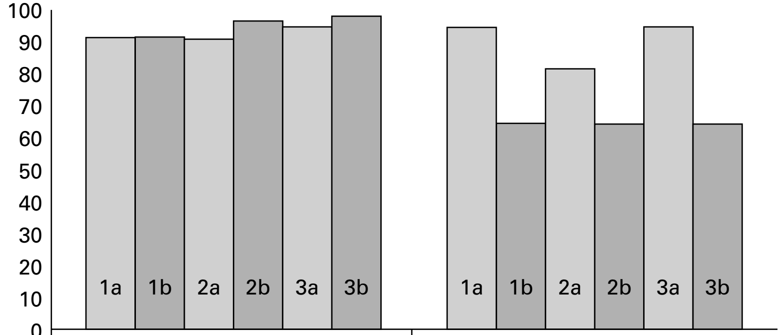
into account (see columns two to six in Table 3) these high percentages of false negative cases seemed less severe. In the strict combined condition with the highest false negative rate of 52 percent, only a small proportion of all mistakenly screened out numbers resulted in an interview (representing two percentage points of 52 percent false negative cases). In fact, irrespective of the validation condition, numbers of unknown eligibility made up the largest part of the false negatives with up to 42 percentage points of 52 percent false negative cases in the strict combined condition.

We further calculated screening biases for age and gender based on the respondents' answers provided in the interview. Additionally, based on the "Basic Question Procedure" (BQP) as proposed by Kersten and Bethlehem (1984) nonrespondents who refused to complete the full questionnaire were asked for age and gender, their highest level of formal education, and their current employment status. If nonrespondents refused to answer these four questions, interviewers were advised to estimate age and gender. Previous findings showed that in most cases interviewers can reliably estimate the refusals' gender and age (Lavrakas 2010). Based on the answers to sociodemographic questions (for respondents) and estimated information (for refusals), we compared the response distribution for the sample that contained the false negative cases to the sample without these cases of respondents and refusals. Accordingly, we were able to assess the impact of falsely excluded numbers on sample composition. Screening biases were derived by dividing the difference of two percentages by the percentage provided by the full sample (including false negatives). Results (rightmost two columns in Table 3) indicated small but insignificant biases due to false negatives. Thus, sample composition concerning age and gender in the loose and strict screening conditions (excluding false negative cases) did not differ significantly from the sample composition in the 'no screening' conditions. No screening biases were calculated for education and employment status of nonrespondents, because interviewers were not able to estimate these indicators. Also, the total number of nonrespondents who answered the four basic questions was too small to calculate reliable screening biases for education and employment status. Further analysis of the screening bias was restricted by the fact that we had no further information on sociodemographic characteristics or substantive variables for screened out cases.

4.3. Predictive Efficiency Measures

To reliably evaluate the pre-call validation methods compared in this paper we need to determine each validation method's probability for correctly predicting the status of cell phone numbers. Prediction accuracy and thus the reliability of a pre-call validation method can be expressed in terms of positive and negative predictive values. The positive and negative predictive values do not add up to 100 percent because calculations are based on different groups of units: the positive predictive value refers to all positively classified cases, whereas the negative predictive value relates to all negatively classified cases.

The *positive predictive value* (PPV) describes the proportion of truly working numbers among all numbers classified as working by pre-call validation method. As a diagnostic measure the PPV indicates the probability that a cell phone number validated as working is in fact a working number, and thus, describes the degree of accuracy of the respective pre-call validation method in detecting working numbers. A small PPV indicates that many of



	PPV	NPV
(1a) number validation - loose	91 $^{(2b), (3a), 3b}$	95 $^{1b, 2a, 2b, 3a, 3b}$
(1b) number validation - strict	91 $^{(2b), (3a), 3b}$	65 $^{1a, 2a, 3a}$
(2a) text message - loose	91 $^{2b, (3a), 3b}$	81 $^{1a, 1b, 2b, 3b}$
(2b) text message - strict	97 $^{(1a), (1b), 2a}$	64 $^{1a, 2a, 3a}$
(3a) combination - loose	95 $^{(1a), (1b), (2a), 3b}$	83 $^{1a, 1b, 2b, 3b}$
(3b) combination - strict	98 $^{1a, 1b, 2a, 3a}$	63 $^{1a, 2a, 3a}$

Fig. 3. Positive and negative predictive value (percent) by pre-call validation method. Calculations were based on chi-squared tests for independent samples; direct comparisons of 1a with 1b, 2a with 2b, and 3a with 3b were based on chi-squared goodness-of-fit tests for dependent samples:
^{1a} $p < .001$ compared to condition 1a, $^{(1a)}$ $p < .05$ compared to condition 1a,
^{1b} $p < .001$ compared to condition 1b, $^{(1b)}$ $p < .05$ compared to condition 1b,
^{2a} $p < .001$ compared to condition 2a, $^{(2a)}$ $p < .05$ compared to condition 2a,
^{2b} $p < .001$ compared to condition 2b, $^{(2b)}$ $p < .05$ compared to condition 2b,
^{3a} $p < .001$ compared to condition 3a, $^{(3a)}$ $p < .05$ compared to condition 3a,
^{3b} $p < .001$ compared to condition 3b, $^{(3b)}$ $p < .05$ compared to condition 3b.

the cell phone numbers which are deemed working are false positive cell phone numbers which affect survey costs negatively due to many numbers that have to be dialed for nothing. As described in Figure 3, there were only small differences in the PPV of loose and strict screening, and all three validation conditions applying number validation, text messaging, and a combination of both methods yielded comparably high PPVs ranging from 91 to 98 percent. Nevertheless, PPVs were significantly higher in the strict text messaging condition (2b) with a PPV of 97 percent compared to the respective loose condition (2a) ($p < .001$) with a PPV of 91 percent. Also, the strict combined condition (3b) yielded a significantly higher PPV of 98 percent compared to the corresponding loose condition (3a) with a PPV of 91 percent ($p < .001$). However, there is no significant difference between PPVs achieved by applying loose (1a) and strict number validation (1b).

The *negative predictive value* (NPV) denotes the proportion of truly nonworking numbers among all numbers classified as nonworking by the validation method. The NPV indicates the probability that a cell phone number determined as nonworking is actually nonworking. Therefore, a small NPV refers to a high proportion of false negative numbers. To avoid screening biases due to many false negative cases the NPV of the pre-call validation method should be maximized. As expected, the strict screening rules in condition 1b, 2b, and 3b yielded a significantly lower NPV than the respective loose screening conditions (all comparisons $p < .001$). The highest NPV of 95 percent was obtained by number validation in the loose screening condition (1a), followed by the loose combined condition (3a) with 83 percent, and the loose text messaging condition (2a) with a NPV of 81 percent. NPVs in the strict screening conditions ranged from 63 to 65 percent. While there were no significant differences between the strict conditions, each strict condition yielded significantly lower values compared to all loose screening conditions ($p < .001$ for all comparisons). Thus, applying strict screening rules yielded a smaller predictive efficiency regarding the detection of nonworking cell phone numbers irrespective of the applied pre-call validation method.

5. Discussion

The priority objective of this study was an efficiency examination of the use of number validation via Home Location Register (HLR) lookups and/or text messaging return codes to detect and screen out nonworking cell phone numbers prior to fieldwork. For each of the three validation conditions – implementing (1) number validation, (2) text messaging, or (3) a combination of both methods – different screening rules were applied which differed in strictness: loose screening implied the exclusion of technically invalid cell phone numbers, strict screening excluded technically invalid and temporarily inactive numbers. Several criteria were taken into account to evaluate the efficiency of the pre-call validation methods regarding costs and data quality.

In particular, based on a simulated experiment it has been shown that number validation and text messaging are both appropriate methods to increase the percentage of working numbers in the field. By excluding nonworking cell phone numbers without any call attempt, the interviewers' workload was decreased. On average they conducted fewer call attempts and needed a shorter overall call duration per completed interview.

Consequently, total survey costs could be decreased. In addition, pre-call validation resulted in a reduction of numbers of unknown eligibility. This resulted in a cost reduction, too, since many of these numbers had no informative final disposition code even after standard fieldwork procedures (up to 15 call attempts) due to ambiguous provider messages or permanent "ring-no-answer". Furthermore, response rates could be calculated more reliably after increasing the proportion of numbers with unambiguous final disposition codes.

Predominantly, differences between the three validation conditions (number validation, text messaging, and a combination of both methods) tended to be rather small with the exception of number validation in the loose screening condition (1a). Using this pre-call validation method, a high working number rate of 91 percent was achieved. However, loose number validation showed small improvements compared to the 'no screening' condition, especially regarding the reduction of the unknown eligibility rate. Thus, number validation predominantly screened out technically invalid cell phone numbers at the expense of a fairly small false negative rate of only 4 percent. Since technically invalid cell phone numbers can mostly be identified during fieldwork by a single call attempt, number validation in the loose screening condition achieved only minor savings in terms of costs and effort. In the strict screening condition, number validation also fell behind text messaging and the combination of both methods, albeit to a lesser extent. Nevertheless, the text messaging method yielded somewhat larger cost savings. The decision whether the application of number validation or text messaging should be the preferred pre-call validation method is also a matter of whether contacting the respondent prior to fieldwork using a short text message is desired or not.

Irrespective of loose or strict screening rules, additional cost and effort for implementing a combination of number validation and text messaging was not justified compared to the sole use of text messaging. In particular, regarding the increase of contact and interview rates, and regarding the reduction of unknown eligibility rates, the combination of both methods did not achieve pre-call validation results superior to the sole use of text messaging.

Screening out technically invalid numbers seems to work reliably using both number validation and text messaging return codes. However, gains in terms of cost and effort are limited since technically invalid numbers can easily be detected in the first call attempt during fieldwork. More effective in terms of survey costs is the reduction of numbers that are no longer in use or have not yet been assigned to a subscriber even though they are technically valid numbers. These numbers cannot be detected during fieldwork and cause a lot of useless call attempts since interviewers typically get a 'ring-no-answer' or an ambiguous provider message when calling such numbers. Thus they remain numbers of unknown eligibility. Number validation and text messaging return codes using loose screening rules do not provide a reliable method of identifying such numbers prior to fieldwork, since the group of numbers that is temporarily not logged in or absent consists not only of technically valid numbers that are no longer in use but also of numbers associated to cell phones that are turned off during the night, during the day, over the weekend or other periods. Thus, when screening out these numbers using strict screening rules, one has to weigh gains in terms of costs and effort against any potential nonignorable screening bias resulting in a distorted sample composition. Since the available data on bias concerning age and gender are limited in scope, future validation studies should try to assess the magnitude and direction of this bias with respect to sociodemographic and substantive estimates.

At this stage, little is known about the allocation and reallocation of cell phone numbers over time. Thus, pre-call validation can also be used to elicit the validity of randomly generated cell phone numbers over time. In particular, pre-call validation could be used to determine the time span from number termination and reassignment to a new subscriber. Using this information, sampling frames of randomly generated cell phone numbers can be optimized further, since this time span may hint at a proper refreshment interval for sampling frames.

Furthermore, number validation and text messaging as pre-call validation methods have to be tested in other countries. For most European countries similar service providers are available offering number validation services as well as text messaging services. Implementation and return codes are almost identical. However, the situation in the U.S. 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, number validation using HLR lookups is problematic in the U.S. because of an insufficient coverage in the GSM mobile network. Therefore, further research on appropriate methods for pre-call validation of cell phone samples is needed in the U.S. as well as in other European countries. For now, both pre-call validation methods tested in this study are promising procedures to ease the implementation of cell phone surveys.

6. 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. Deerfield, IL: AAPOR.

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. Deerfield, IL: AAPOR.

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

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

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

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

Brick, J.M., Brick, P.D., Dipko, S., Presser, S., Tucker, C., and 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.

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

Busse, B. and Fuchs, M. (2012). The Components of Landline Telephone Survey Coverage Bias. The Relative Importance of No-Phone and Mobile-Only Populations. Quality & Quantity, 46, 1209–1225.

Callegaro, M., Steeh, C., Buskirk, T.D., Vehovar, V., Kuusela, V., and 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 A, 170, 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:10.4232/1.10,128.

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

Gabler, S. and Ayhan, Ö. (2007). Gewichtung bei Erhebungen im Festnetz und über Mobilfunk. Ein Dual Frame Ansatz. ZUMA-Nachrichten Spezial, 13, 39–46.

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

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

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

Häder, S. and 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., and Frankel, M.R. (2011). Improving Public Health Surveillance Using a Dual-Frame Survey of Landline and Cell Phone Numbers. American Journal of Epidemiology, 173, 703–711.

Jesurum, R. (1990). Data Gathering Made Easy. Using Predictive Dialing With CATI Software. Paper presented at the 133rd ESOMAR Seminar "How to Cope With Data Overload: The Impact of New Technology on Marketing Information", Paris, France.

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

Keeter, S., Dimock, M., Kennedy, C., Best, J., and 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, 750–771.

Kersten, H.M.P. and Bethlehem, J.G. (1984). Exploring and Reducing the Nonresponse Bias By Asking the Basic Question. The Statistical Journal of the United Nations Commission for Europe, 2, 369–380.

Lavrakas, P.J. (2010). Telephone Surveys. Handbook of Survey Research, P.V. Marsden and J.D. Wright (eds). Bingley, UK: Emerald Group Publishing Limited, 471–498.

Noldus, R. (2006). CAMEL: Intelligent Networks for the GSM, GPRS and UMTS Network. Chichester: John Wiley & Sons, Ltd.

OFCOM (2010). Tackling Abandoned and Silent Calls. Consultation. Retrieved from http://stakeholders.ofcom.org.uk/binaries/consultations/silentcalls/summary/condoc.pdf

Oldendick, R.W. (1993). The Effect of Answering Machines on the Representativeness of Samples in Telephone Surveys. Journal of Official Statistics, 9, 663–672.

Peytchev, A., Carley-Baxter, L.R., and 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, 287–302.

Schneiderat, G. and Schlinzig, T. (2009). Teilnahmebereitschaft und Teilnahmeverhalten bei Telefonumfragen in der Allgemeinbevölkerung über das Mobilfunknetz. In Telefonbefragungen über das Mobilfunknetz. Konzept, Design und Umsetzung einer Strategie zur Datenerhebung, M. Häder and S. Häder (eds). Wiesbaden: VS Verlag für Sozialwissenschaften, 83–97.

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

Struminskaya, B., Kaczmirek, L., Schaurer, I., Bandilla, W., Gabler, S., and 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, 388–406.

Received September 2011
Revised March 2012